Exploring heterogeneity in diversification patterns across the tree of life using probabilistic models

Odile Maliet

Sorbonne Université

Ecole doctorale 227 : Sciences de la Nature et de l'Homme

Ecole Normale Supérieure / Institut de Biologie de l'Ecole Normale Supérieure

Exploring heterogeneity in diversification patterns across the tree of life using probabilistic models

Par Odile Maliet

Thèse de doctorat en écologie et évolution

Dirigée par

Hélène Morlon, Directrice de recherche CNRS à l'Institut de Biologie de l'École Normale Supérieure

Nicolas Loeuille, Professeur à l'Institut d'Écologie et des Sciences de l'Environnement de Paris, Sorbonne Université

Présentée et soutenue publiquement le 20 septembre 2018

Devant un jury composé de :

Hélène Morlon	Directrice de Recherche CNRS	Directrice de thèse
Nicolas Loeuille	Professeur Sorbonne Université	Directeur de thèse
Franck Jabot	Chargé de Recherche IRSTEA	Rapporteur
Rudolf Rohr	Maitre-assistant à l'Université de Fribourg	Rapporteur
Marianne Elias	Directrice de Recherche CNRS	
Michael Blum	Directeur de recherche CNRS	

Résumé

Dans cette thèse, je présente différentes approches pour quantifier et expliquer les variations dans le processus de diversification au sein de l'arbre du vivant. Toutes les approches présentées s'appuient sur des modèles probabilistes. Le premier chapitre s'intéresse à la forme générale des phylogénies, et propose une nouvelle mesure de la relation entre la richesse spécifique et la profondeur des clades au sein d'une phylogénie. Nous montrons que, dans les phylogénies empiriques, cette mesure s'écarte de la valeur attendue pour un processus de diversification homogène entre les lignées, possiblement à cause de la présence de variations du taux de diversification au sein des groupes étudiés. Dans le deuxième chapitre, je m'intéresse à une description à plus fine échelle de ces variations de taux, et présente une nouvelle méthode pour estimer des vitesses de diversification lignée-spécifiques dans les phylogénies empiriques. Contrairement aux méthodes existantes, qui considère que les taux de diversifications varient par des sauts rares et de grande amplitude, la nôtre propose une vision plus graduelle de l'évolution de la vitesse de diversification. Nous appliquons notre méthode à un jeu de données empirique et montrons que la variabilité du taux de diversification est aussi forte au sein des clades qu'entre les clades, ce qui s'accorde bien avec cette vision d'une évolution progressive des taux de diversification. Enfin, le troisième chapitre se concentre sur une des explications possibles à la présence de variabilité dans les vitesses d'accumulation d'espèces, en présentant un modèle individu-centré permettant d'étudier l'effet de différent types d'interactions écologiques sur le processus de diversification. Nous étudions les prédictions de ce modèle pour la diversité obtenue, ainsi que sur plusieurs mesures caractérisant la structure des réseaux d'interactions. Notre modèle génère des réseaux d'interactions écologiques réalistes, avec des réseaux plus modulaires et moins emboités dans les communautés antagonistes que dans les communautés mutualistes. La présence d'interactions antagonistes favorise la diversification de la communauté, du point de vue de la variabilité en traits comme de celui du nombre d'espèces, tandis que les interactions mutualistes entravent la création de diversité, du fait d'une forte sélection stabilisante.

Abstract

In this PhD thesis, I present different approaches based on probabilistic models for quantifying and explaining heterogeneity in the diversification process across the tree of life, all in the probabilistic modeling framework. In the first chapter, we focus on the general shape of phylogenetic trees, and propose a new metric for the quantification of the age-richness relationship of the subclades within a tree. The study of this metric in a dataset of empirical phylogenies shows that they diverge from the expectation under an homogeneous speciation model, possibly because of within-clade speciation rates variations. In the second chapter of the thesis, I focus on a finer scale description of diversification rates variations, and introduce a new method to estimate lineage-specific diversification rates within a phylogeny. Compared to previously existing methods that aim at identifying a few diversification rate shifts with large effect, ours propose a more gradual view of diversification rate evolution. We apply our approach to a dataset of empirical phylogenies, and show that Intra-clade variations accounts a large part of the rate variations in the whole dataset, suggesting suggest that models with many gradual changes may be more appropriate than models with few punctuated shifts for describing the evolution of diversification rates. Finally, the last chapter considers more directly one of the possible cause of variation in diversification rates, which is the presence of inter-species ecological interaction. We build an eco-evolutionary model for the emergence of mutualistic, antagonistic and neutral bipartite interaction communities, and study it prediction on species and trait diversity, as well as on several key network structure metrics. Our model generates realistic network structures, antagonistic communities being more modular and less nested than mutualistic ones. We find that antagonistic interactions foster both species and trait diversity, while mutualistic interactions generate strong stabilizing selection, with a negative impact on both diversity measures.

Table of Contents

Introduction

A striking pattern of present biodiversity is how unevenly it is distributed across space, the tree of life and ecological communities. Why are there so many species in the tropics? Why are some groups, like passerine birds or beetles, so much more diverse than their close relatives?

Explaining these variations in species richness is one of the key challenges for the study of macroevolution. Past differences in species diversification rates – the balance between speciation and extinction – are one of the major forces that shape today's biodiversity distribution. This thesis focus on different approaches, based on probabilistic models, that can be used to study past heterogeneity in diversification rates with the use of present day data.

I begin this introduction by presenting the general methods used to study diversification patterns, including phylogenetic tree shape characterization, statistical inference methods to estimate diversification rates from phylogenies, and models of cladogenesis below the species level. I then present extensions of the birth-death model that aim at incorporating heterogeneity in diversification rates, both through time and across lineages. Finally, I focus on one possible cause of rates variations, the presence of biotic interactions between species, and on how to describe their structure in a community.

1 Tools for studying diversification from reconstructed phylogenies.

1.1 Phylogenies of extant species.

All the work presented in this PhD thesis uses dated species phylogenies to learn about the diversification patterns. This approach complements the one using information from the fossil record, and enables to gain knowledge on the many clades for which the fossil record is too scarce to be used in diversification studies. With the rapidly increasing number of large scale dated phylogenies available, phylogenetic comparative methods have known a great development over the last decades.

Phylogenies represent the evolutionary relationships between *lineages*, with tips representing species and nodes representing speciation events. On a dated phylogeny, nodes depths are proportional to the amount of evolutionary time spent since the splitting event. The *complete* phylogeny contains information about all species in the clade, including those that are extinct at present. Yet the data one has to deal with is the *reconstructed* phylogeny, that contains only sampled extant species (Fig. 1). It is the data one usually has to deal with when working on phylogenetic comparative methods, as most phyogenies are reconstructed from molecular data obtained from extant species.

1.2 Phylogenetic tree shape. Phylogenies contain important information on the processes that contributed to shape them. The shape of a tree is composed of its *topology* – the particular branching pattern of the tree – and its *branch lengths*. The *labelled topology* refers to the topology with species identity at tips, while the *unlabelled topology* has no species at tips and refers only to the abstract tree shape.

Numbers aiming at summarizing information from the phylogenies are called summary statistics. They are useful to describe phylogenetic tree shapes without referring to a model, allowing to characterize empirical patterns without any prior expectation of what the underling mechanism would be. They can also be used to check the deviation of a tree from a model's expectations, as is done when performing posterior predictive checks to assess the goodness of fit of a particular model.

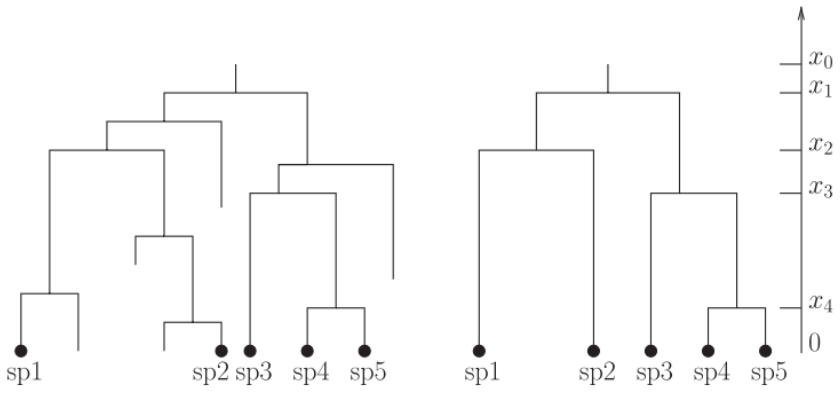


Figure 1: **A dated species phylogeny.** From Fig.1 in (Stadler, 2011). Left : The complete phylogeny, including all extinct species, with five sampled extant species (black dots). Rigth : The corresponding reconstructed phylogeny.

Examples of summary statistics of phylogenetic tree shape include simple descriptive numbers such as tip number, root age, number of cherries (nodes subtending two tips) within the tree. Others aim at summarizing complex information – *e.g.* tree imbalance – and necessitate more complex constructions.

Imbalance. The imbalance of a phylogeny is a measure of how evenly the species are distributed beween clades. In a balanced phylogeny, sister-clades tend to be of similar sizes (Fig. 2c), while they have very different size in an imbalanced tree (Fig. 2b). Imbalance in phylogenetic tree shape is of particular importance when looking for heterogeneity in diversification tempo across a tree, and does not require to have knowledge about the branch lengths. It can be quantified with the β statistic, as introduced by Aldous (1996). It is based on a statistical model in which a n-tipped tree is built by first assigning the n species independent positions on the $[0, 1]$ interval, according to a uniform distribution. The interval is then split according to a beta distribution with parameter $(\beta + 1, \beta + 1)$, with $\beta \in]-1, +\infty[$. The probability of i species being in the left subinterval, conditioned to the fact that there is at least one species per subinterval, is

$$p(i|n) = \frac{1}{a_n(\beta)} \frac{\Gamma(\beta + i + 1)\Gamma(\beta + n - i + 1)}{\Gamma(i + 1)\Gamma(n - i + 1)} \tag{1}$$

where $1 \leq i \leq n - 1$, $a_n(i)$ is a normalizing constant and Γ is the gamma function, defined for $x \in \mathbb{R}$ as $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$. The species are distributed accordingly within the two subclades of the phylogeny, and this is repeated within each subclade until subintervals contain only one species (Fig 2a). Even though the beta distribution is only defined for $\beta \in]-1, +\infty[$, $p(i|n)$ in equation (1) can be computed for $\beta \in [-2, +\infty[$, and the model can thus be extended to accommodate β values in $]-2, -1]$. For small β values, the intervals are split close to the edges and the resulting tree shapes tend to be very imbalanced. In particular, $\beta = -2$ always produces a *comb* (or *caterpillar*) tree (Fig. 2b). For $\beta = 0$, the position of the splits are uniformly distributed, and for large β values, splits happen close to the middle of the interval and we get a balanced, *bush* tree (Fig. 2c). Using equation (1), it is easy to compute the likelihood of this model for a given tree topology as the product of the probability of each split. β is computed as the Maximum Likelihood Estimate for this model.

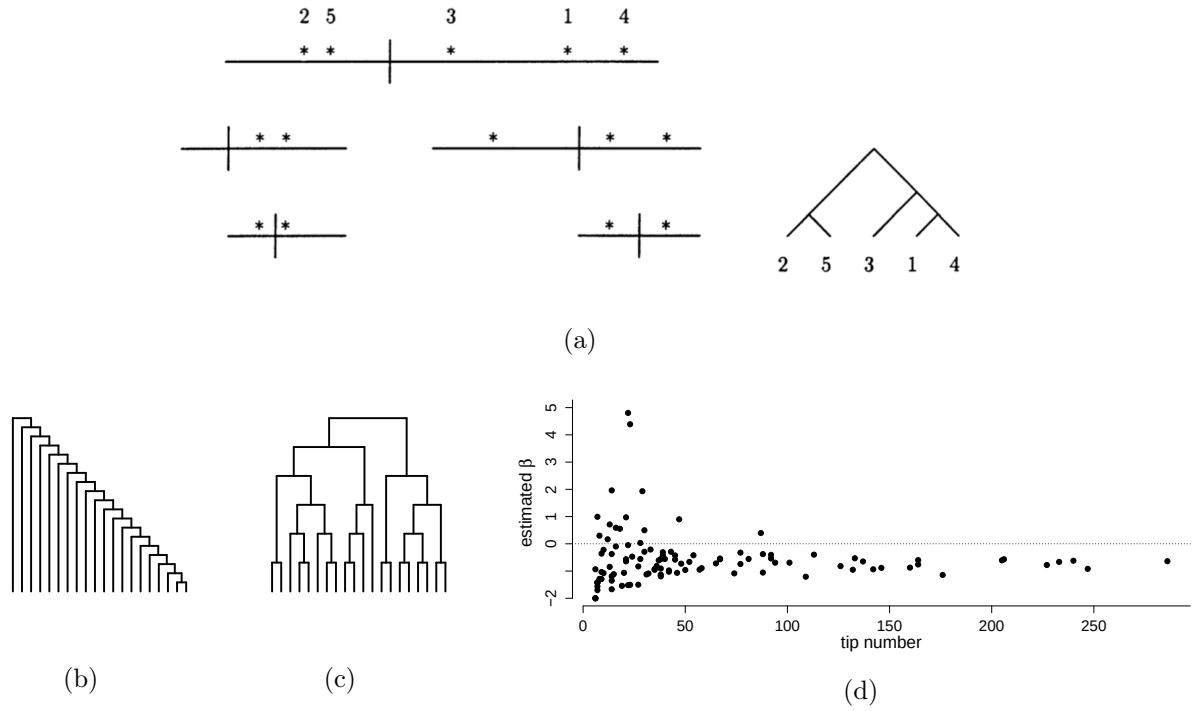


Figure 2: **The β statistic.** (a): The β-splitting model. Labels are uniformly distributed in the interval $[0, 1]$, which is then recursively split until all labels are in separate subintervals (see text). From Fig. 6 in Aldous (2001). (b, c): 20-tipped tree shapes generated with the β-splitting model, with parameter $\beta = -1.99$ (which generates a *caterpillar* tree, b), and with parameter $\beta = 20$ (which generates an almost perfectly balanced tree, c). (d): Maximum likelihood estimate of the β parameter for a data set of bird family level phylogenies (see Chapter 1 of this thesis). Values below 0 (dotted line) indicates that the phylogenies are more imbalanced than what is expected under a birth-death model. We can see that, even though the variance is very high for small phylogenies, the values cluster a common value of -1 for trees that are large enough.

Numbers of other statistics have also been used to measure tree imbalance (reviewed in Shao and Sokal, 1990; Agapow and Purvis, 2002), such as the Sackin index (Sackin, 1972) – defined as $\sum_{i=1}^{n} N_i$, where n is the number of tips in the tree and N_i is the number of internal nodes between the root of the phylogeny and tip i – and the Coless index (Colless, 1982; Heard, 1992) – defined as $\frac{2}{(n-1)(n-2)} \sum_{i=1}^{n-1} |r_i - s_i|$, where r_i and s_i are the number of tips on the right and left sides of bifurcation i.

Compared to other imbalance indices, computing β requires the tree to be fully resolved, without polytomies. Yet, it has the advantage not to depend on the number of species included in the analysis, allowing direct comparison between phylogenies of different sizes. Additionally, the generating model encompasses well-known distributions, enabling to compare the tree imbalance to classical models' expectations. For the constant rate birth-death model (see Section 1.3) – and, as a matter of fact, for all cladogenetic models considering homogeneous diversification rates across lineages –, the expected value of β is 0 (Aldous, 1996, 2001; Lambert and Stadler, 2013). $\beta = 0$ also corresponds to the case in which all unlabelled topologies are equaly likelly. The case in which all labelled topologies are equally likely gives more weight to unbalanced unlabelled topologies, because many labels permutation lead to the same labelled topology if the tree is well balanced, while this is

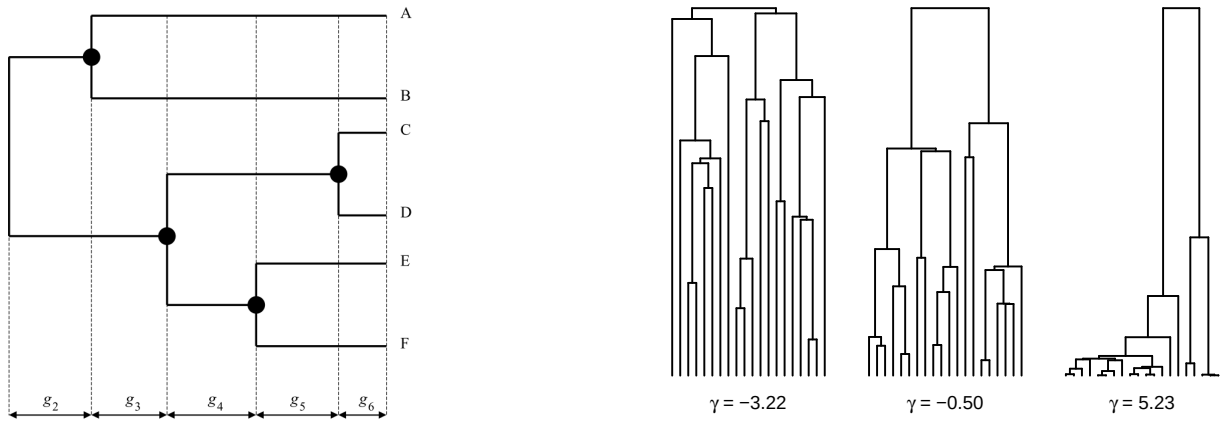
not the case for an unbalanced tree. This case, generally referred to as the proportional to different arrangements model, corresponds to an expected β value of -1.5. Yet in empirical data, β values fall between the two, generally clustering around -1 (Aldous, 2001; Blum and François, 2006, Fig 2d). This finding that empirical trees are less balanced than expected under a birth-death model is also found using other metrics (*e.g.* Heard, 1992).

Branch lengths. The imbalance of a tree do not take its branch lengths into account. The γ statistic was first introduced by Pybus and Harvey (2000) to measure how close the nodes of a phylogeny are to the root. This statistic aims at determine wether nodes tend to accumulate early or late in a clade's history. It is defined by the formula

$$\gamma = \frac{\frac{1}{n-2}\sum_{i=2}^{n-1}\left(\sum_{k=2}^{i} k g_k\right) - \frac{T}{2}}{T\sqrt{\frac{1}{12(n-2)}}}, T = \sum_{j=2}^{n} j g_j, \tag{2}$$

where n is the number of tips in the phylogeny, and the g_i are the internodes distances (see Fig. 3a). This complicated formula was chosen because the quantity is distributed as a standard normal law for phylogenies generated with the pure birth model, and can thus be used to assess whether the node depths distribution deviates from the one expected from this model. $\gamma < 0$ means that the nodes are closer to the root than under the pure birth model, which is the case when diversification rates decreased through time during the history of the clade (Fig. 3b). On the contrary, $\gamma > 0$ means that nodes are closest to the tips of the tree than under the pure-birth model, which happens as soon as extinction is non-zero (Fig. 3d). Trees with this characteristic are often referred to as *tippy* trees. Negative γ values are also found when not all species are sampled in a pure birth tree. Pybus and Harvey (2000) suggested to use a Monte Carlo procedure in order to account for the fact that species undersampling can lead to inappropriately low γ values.

This statistic has been used in many studies on empirical data (Phillimore and Price, 2008; McPeek, 2008; Rabosky and Lovette, 2008a; Moen and Morlon, 2014), leading to the conclusion



(a) (b) (c) (d)

Figure 3: **The γ statistic.** (a): Definition of the g_i as internode distances. The subscript refers to the number of lineages in the reconstructed phylogeny for each internode interval. From Fig. 1 in Pybus and Harvey (2000). (b - d): 20-tipped trees with increasing γ statistic values. The phylogeny in c was simulated with a pure birth model.

that empirical γ is almost always below 0, a pattern that has often been interpreted as a proof that diversification rates tend to decrease through time. Yet other factors might lead to negative γ values, are more sophisticated, likelihood based approaches (see the time-dependent models in Section 2.1) are a more reliable tool to test for time variability in diversification rates.

Age-richness correlation. The relation between clade age and richness in a group has often been evaluated through the computation of age-richness correlation (*e.g.* Magallón and Sanderson, 2001; Rabosky, 2009b). Under a constant-rate birth-death model, this is expected to be a positive value (at least when crown age are used to evaluate clade age Rabosky et al. 2012 ; if stem age is used the positive relationship is lost Stadler et al. 2014; Sánchez-Reyes et al. 2016). In empirical data, a positive relationship is seldom found (Rabosky 2009b; Rabosky et al. 2012 ; but see McPeek and Brown 2007). A weak age-richness relationship may be caused by a bias in the definition of higher taxa, by diversification rates varying among clades, or by diversity dependent speciation (Rabosky, 2009b; Stadler et al., 2014).

While the age-richness correlation is useful to evaluate whether a group deviate from the constant rate birth-death model expectation, it is computed for multiple phylogenetic trees, that are either independently estimated or pruned from a bigger tree (Magallón and Sanderson, 2001; Ricklefs, 2007b; Rabosky et al., 2012). This can be problematic, as the computed values are sensitive to the definition of higher taxa (Stadler et al., 2014; Sánchez-Reyes et al., 2016). In chapter 1 of this thesis, I present an extension of Aldous β-splitting model (Aldous, 1996, 2001) with an additional parameter controlling for nodes order within the phylogeny, which I began to develop during my master and finished during my PhD. The estimation of this parameter gives a tree-wise measure of the age-richness correlation.

1.3 Estimating diversification rates from phylogenies.

The birth-death model. Most of the methods used to analyze different diversification scenarios from empirical phylogenies (for a review, see Morlon, 2014), are based on the birth-death model of cladogenesis (Nee et al., 1992). It is a lineage-based model (meaning that the model takes lineage as the basal evolutionary units) in which the only possible events are speciation and extinction, which happen independently for each lineage with common per-lineage rates λ and μ that are constant through time. When there is no extinction ($\mu = 0$), this model is also known as the pure-birth model or the Yule model. It has since been extended to include many different hypotheses, some of which are of relevance to this PhD thesis and are described in Part 2 of this introduction.

The method of moments. Under such a simple diversification model as the one described above, the expected number of species at present time in a clade is entirely determined by the age of the clade, t, and the net diversification rate, $\lambda - \mu$, and equals $e^{(\lambda-\mu)t}$. The logarithm of the extant species number divided by the time for speciation thus gives an estimate of the diversification rate. Estimates of diversification rates with this method are biased towards high values because extinct clades are never sampled. Magallón and Sanderson (2001) proposed a correction for this using the *method of moments* approach, which necessitates to assume the relative extinction rate.

This method has been widely used in past literature (Alfaro et al., 2007; Adams et al., 2009). Yet it does not allow the estimation of speciation and extinction separately. It has been argued that it may be less prone to error than more recent, complex methods (Meyer and Wiens, 2018), but this later study is subject to caution and its results are misleading because different methods were

fed with different amount of information (Rabosky, 2017). Even though this method can give an estimate of the net diversification rate under a birth-death model, it uses a very small fraction of the information contained in a phylogeny – discarding all the information about the topology and the branch lengths. It is thus not well suited to study more complex diversification patterns and does not allow extensions to account for heterogeneity in the diversification process.

Lineage Through Time plots. Under a constant-rate pure-birth process, the logarithm of the number of lineages is expected to increase linearly with time, and the speciation rate can be obtained directly from the slope of the Lineages Through Time (LTT) plot – the plot of the logarithm of the number of extant lineages through time. If the extinction rate is non-zero, there is an apparent increase in lineage accumulation towards the present, known as the 'pull of the present' effect (Harvey et al., 1994). This is due to the fact that lineages that originated close to the present have had less time to go extinct than more ancient lineages (Nee et al., 1994a). The slope of the curve near the present is the speciation rate λ, while the slope at the root age of the phylogeny reflects the diversification rate $\lambda - \mu$. This has led to argue that extinction rates can be estimated from molecular phylogenies, even without fossil information (Nee et al., 1994a; Kubo and Iwasa, 1995). Approaches using the LTT plot have been used in several empirical studies (*e.g.* Baldwin and Sanderson, 1998; Ribera et al., 2001; Kozak et al., 2006).

Likelihood approaches. The likelihood of a parameter θ – possibly a vector of parameters – knowing the data x is defined as the probability to obtain the data in a given model knowing the parameter value θ:

$$L_x(\theta) = p(x|\theta) \tag{3}$$

In the traditional statistical approach, also known as the *frequentist* approach, probabilities are interpreted as the frequency the data would be obtained, should the experiment be repeated a large number of times. The likelihood of θ is a measure of how much of the data can be explained by the model parameter value. The Maximum Likelihood Estimate (MLE), $\theta_{\text{MLE}} = argmax_\theta(L_x(\theta))$, is the parameter value that best explains the data.

This approach has been widely used to estimate diversification rates, and has the advantage to allow for complexification of the model in order to accommodate more realistic scenarios (see Section 2.1 of this introduction).

Bayesian approaches. The classical likelihood maximization approach considers that there is a true fixed value for each model parameter, and that we can approach it if we are given a sufficient amount of data. The goal is thus not to try and give a probability to a parameter value, but to approach the real one. Yet, probabilities can also be interpreted as a measure of how reasonable we think an hypothesis may be, considering what we know about the data. This is the point of view of *Bayesian statistics*, that heavily rely on the Bayes theorem:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} \tag{4}$$

In that formula, $p(\theta|x)$ is called the posterior distribution of the model parameters. It is the information we have about the model parameters θ after accounting for the data x. $p(x|\theta)$ is the likelihood of the model, and $p(\theta)$ is the knowledge we had on the model before accounting for the data, known as the prior distribution. $p(x) = \int_\Theta p(x|\theta)p(\theta)$ is a normalisation constant.

One could, as in the frequentist approach, try to maximize the posterior distribution so as to then obtain the more likely parameter set. But this is not very informative about the actual shape of the distribution. An extreme case would be one in which the data are uninformative about our model parameter, so that any value would maximize the model posterior for those data. It is thus more appropriate to try and get better insight about the distribution, which is usually done by sampling from it. An efficient and common way to do so is to use Markov Chain Monte Carlo (MCMC). The chain begins at an arbitrary position of the parameter space. At each iteration, a new parameter θ_{prop} is drawn in the neighborhood of the current position $\theta_{current}$. If $p(\theta_{prop}|x) > p(\theta_{current}|x)$, θ_{prop} becomes the new current position. Else, there is a probability $p(\theta_{prop}|x)/p(\theta_{current}|x)$ that θ_{prop} becomes the new current position. This procedure is iterated until a stopping criteria is reached. The equilibrium distribution of this Markov chain is equal to the posterior parameter density.

Bayesian approaches allow to get good estimates for the uncertainty about the model's parameters values. They have been used to fit diversification models to data, as for example in the widely used Bayesian Analysis of Macroevolutionary Mixtures (BAMM) method (Rabosky, 2014) and SSE methods (for State Speciation and Extinction, Maddison et al., 2007), see Section 2.2 of this introduction.

Model selection. In order to gain insight in the processes that best explain our data, one may be interested in comparing how well a set of model fits the data, rather than finding the parameter set that best fits the data for a given model. This is done differently in the frequentist and in the bayesain frameworks.

Frequentist approach.— The goodness of fit of a model is given by the maximum of the likelihood function $max_\theta L_x(\theta) = L_x(\theta_{\text{MLE}})$. An intuitive way to compare models is to compare their maximal likelihood. Yet, this quantity increases with the dimensionality of the model parameters, leading to the preferential selection of the more complex model, an issue known as *over-fitting*. If the two models are nested (which means that the simplest model is a special case of the more complex one), it is possible to correct for this effect by using a Likelihood Ratio Test. If they are not, it is common practice to compare them using their Akaike Information Criterion (AIC), defined as $\text{AIC} = -2log(L_x(\theta_{\text{MLE}})) + 2log(n)$, where n is the number of parameters. The model with the lowest AIC is selected.

Bayesian approach.— In a Bayesian framework, the value of the parameter θ is not fixed, and the goodness of fit of a model is defined as $\int_{\theta\in\Theta} p(x|\theta)p(\theta)d\theta$. The support for a model over another can be quantified using *Bayes factor*, which is the ratio of the goodness of fit for the two model. An interpretation of the Bayes factor value can be found in Kass and Raftery (1995). Model selection can also be performed at the same time than the inference of the parameter of the model using rejection jump MCMC (rjMCMC, Green, 1995), allowing to automatically assess the probability of a model being favored over the set of models included in the analysis. This is for example the approach followed in the BAMM method to infer the number of diversification rate shifts on a phylogeny (Rabosky, 2014, see also Section 2.2).

1.4 Models of cladogenesis below the species level. Unlike lineage-based model, which consider lineages to be the basal evolutionary units, some models focus at lower level of organization, from populations to gene level. A direct advantage of those models is that it is often straightforward to include complex rules and individual variation relevant to the question being addressed. Individual variations may for example be included as spatial variability, local interactions, phenotypic or genetic variability, plasticity, and resulting fitnesses (a measure of the individuals' reproductive success). These kind of models also allow to study the effect of the speciation mode on the obtained macroevolutionary patterns (Davies et al., 2011; Missa et al., 2016). A downside is that finding an analytical solution for the behavior of the models is very rarely possible (even though exceptions do exist; see Hubbell, 2001; Manceau et al., 2015), and likelihood approaches are thus generally not applicable for those models. One generally needs to study them through simulations, which, for complex models, can be costly in terms of computation time. The realism of the obtained phylogenies can be assessed by comparing the summary statistic's distribution to those observed for empirical trees. Examples of models of cladogenesis below the species level include the famous Neutral Theory of Biodiversity (NTB Hubbell, 2001) and its variations – for example with different speciation modes (Davies et al., 2011; Missa et al., 2016) –, the ecological differentiation model of McPeek (2008) – in which the carrying capacity of a species depends on how far its optimal environmental conditions are from the local environment –, and the geographical speciation model of Pigot et al. (2010) – that model species range size dynamics, with geographical barriers that appear at random and result in the creation of new species.

Individual-based models simulate populations or species as a collection of individuals, each having their own set of traits. Population levels properties – such as population size, mean trait value, immigration rates – emerge from individuals properties and behavior. The definition of species from the obtained genealogies can be a tricky part. The often used point-mutation – each mutation gives instantly rise to a new species (Hubbell, 2001) – or random-fission – at speciation, the individuals are dispatched randomly between the two resulting new species (Davies et al., 2011) – speciation modes create non-monophyletic species, which is problematic when we want to look at the resulting phylogenies (Manceau and Lambert, 2017). A way to obtain monophyletic species, and thus be able to unambiguously construct the phylogeny from the individuals genealogy, has been proposed in Manceau et al. (2015). In this paper, species are defined from individual genealogies with mutations as the smallest monophyletic group of extant individuals such that any two individuals of same genetic type always belong to the same group (see also Section 2.1).

In Chapter 3, I present an individual-based model of of two interacting guilds that we developped during my PhD to study the evolution of bipartite interaction networks and the effect of different interaction types on species diversification, in which we use the mode of speciation proposed by Manceau et al. (2015).

2 Incorporating heterogeneity in diversification models.

The discovery that phylogenetic tree shapes generally deviate from those obtained with a birth-death model – as for example in terms of γ and β statistics – fostered the development of an ever-increasing body of more complex diversification models. Those models contribute to get an insight in what is impacting the tempo of diversification, and thus in the processes that shaped today's patterns. This section is an attempt to review representative examples.

2.1 Heterogeneity through time.

A constant rate birth-death model should theoretically generate LTT plots which display an increase in lineage accumulation speed toward present time, the so called *pull of the present* (Harvey et al., 1994; Nee et al., 1994a). Empirical LTT plots, however, more often show the reverse pattern (Kubo and Iwasa, 1995; Zink and Slowinski, 1995; Lovette and Bermingham, 1999). This observation, coupled to the related observation that empirical phylogenies generally display negative γ statistic values (Phillimore and Price, 2008; McPeek, 2008; Rabosky and Lovette, 2008a; Moen and Morlon, 2014), led to the hypothesis that speciation decreases with time within clades. Additionally, diversification rates may be impacted by changes in the biotic or abiotic environment of a clade.

Time dependency. A common way to incorporate this time signature in cladogenesis diversification models is to assume a continuous functional dependency of diversification rates to time (Rabosky and Lovette, 2008b; Morlon et al., 2011; Hallinan, 2012). A likelihood expression is available for those models, allowing to fit them to empirical phylogenies. Although the likelihood can theoretically accommodate any functional time dependency shape for the speciation and the extinction rate, they are generally assumed to vary as linear or exponential functions to limit the number of parameters needed.

Alternatively, models with discrete forms of time variation have also been developed (Stadler, 2011; Hallinan, 2012). Methods based on those models offer the advantage of not requiring to specify a predefined form for how diversification rates changed through time, and rely on likelihood ratio tests to define the number of rate shifts that occurred along a clade's evolutionary history.

Those two approaches allow to assess whether the diversification tempo changed through time – *e.g.* if there was an increase or a decrease in rates, or a waxing and waning pattern (diversity going up and down) – without making assumptions on the factors that may have driven those changes. Those may be extrinsic factors, linked to change in the abiotic or biotic environment of the clade, or intrinsic to the diversification process, as could be the case if speciation is a protracted process.

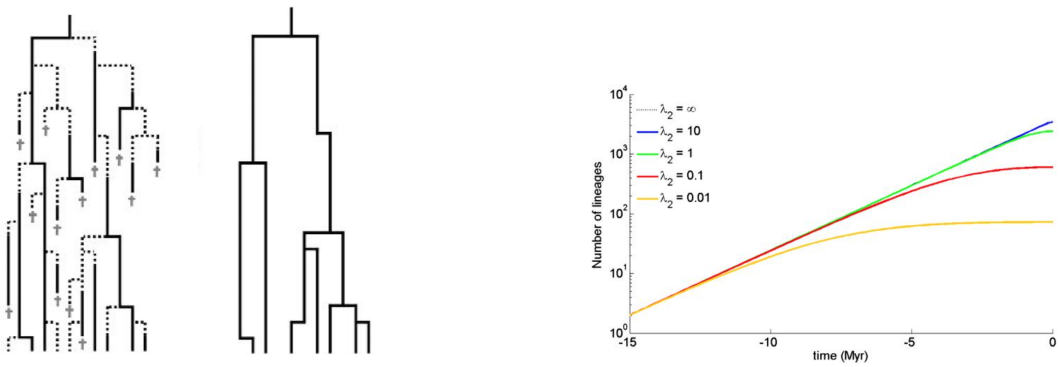
Environmental dependency. Diversification rate changes in the time dependent models may show what the tendencies are, but they are not giving any clues about what have driven them. Further test are needed to get an insight in the processes that are acting. Some of them may be changes in the abiotic (*e.g.* temperature, sea level...) or biotic (*e.g.* predation pressure) environment. If information is known about how those varied through time, it can be incorporated within the birth-death model framework by assuming a functional dependency of diversification rates to the environmental variable (Condamine et al., 2013). As for the time dependent model, a linear or exponential functional dependency is generally assumed. Simulation studies showed that the model is well-behaved and allows recovering the right environmental driver of diversification as long as the dependency is strong enough (Lewitus and Morlon, 2017). The approach has been applied on empirical data, highlighting for example a negative dependence of speciation and extinction rates to temperature in birds (Claramunt and Cracraft, 2015), a positive correlation between sea level and

extinction rates in birdwings butterfly (Condamine et al., 2015), and a positive association between speciation rate and temperature in Cetaceans (Lewitus and Morlon, 2017).

Mass extinction. Past mass extinctions leave a signature in today's biodiversity, showing as a plateau in the LTT plots (Crisp and Cook, 2009). Mass extinction events have been incorporated in the discrete time dependent model (Stadler, 2011), as well as in the continuous one (Höhna, 2015), by setting probability ρ_i to survive at time t_i. The discrete time version was subsequently integrated within a bayesian framework, using reversible jump MCMC to detect the number and positions of rate shifts and mass extinction events (May et al., 2016).

Diversity dependence. In all the models described in the previous paragraphs, species are assumed to be independent from each other. Yet competitive interaction among species is thought to have major impact of species diversity. The fact that diversification rates are often found to decrease with time, and the fact that, for many groups, no relationship is found between clade age and species richness, is often interpreted as an evidence for diversity-dependent speciation (Ricklefs, 2007a; Phillimore and Price, 2008; McPeek, 2008; Rabosky, 2009a), with diversification becoming less and less likely as ecological niches are being filled. This process was first implicitly incorporated in diversification models through decreasing speciation rates with time (Rabosky and Lovette, 2008b), but models with explicit diversity dependent diversification are now available (Rabosky and Lovette, 2008a; Etienne et al., 2011). Etienne et al. (2011) developed a model in which the speciation rate is a function of the number of species in the clade, as well as an inference procedure to fit it on empirical phylogenies. The dependency of the speciation rate to species number is included as a logistic function, with $\lambda(n) = \max\left(0, \lambda_0 - (\lambda_0 - \mu)\frac{n}{K}\right)$, where K is the equilibrium species richness, which is reminiscent of the 'carrying capacity' used in ecological models.

The negative diversity dependent diversification process is thought to be due to the filling of niche space by competitors that impedes diversification. Occasionally, a lineage may escape competition, because of dispersal to a new area or appearance of a key innovation that allows it to enter a



(a) (b) (c)

Figure 4: **The protracted speciation model.** (a, b): A phylogeny simulated with the protracted speciation model. From Fig. 1 in Etienne and Rosindell (2012). (a): The complete phylogeny, showing extinct and incipient species (doted lineages). (b): The corresponding reconstructed phylogeny. (c): LTT plots obtained for different values of the speciation completion rate λ_2. Low λ_2 allows to generate LTT curves that are close to those obtained in empirical data. Other parameters are $\lambda_1 = 0.5$, $\mu_1 = \mu_2 = 0$. From Fig.2 in Etienne and Rosindell (2012).

novel adaptive zone. This can be added to the diversity-dependent diversification model framework as a decoupling of the diversity-dependent dynamics of a subclade from the main clade's diversity-dependent dynamics (Etienne and Haegeman, 2012), allowing to model the whole adaptive radiation diversification, from the first high speciation phase to the slowing in diversification rate phase.

Protracted speciation. Mechanisms other than a dependency to time – or to factors that correlate with it – can lead to a signature of decreasing speciation through time. This is the case if speciation is protracted. In the classical lineage-based models of diversification, speciation is modeled as an instantaneous process. Yet, the mechanisms leading to the formation of distinct species is a complex one that necessitates the completion of several steps, and it may take a non-negligible time to go from initial lineage splitting to the actual completion of speciation (Avise et al., 1998; Gavrilets et al., 2000; Benton and Pearson, 2001; Norris and Hull, 2012). This idea that speciation takes time to complete, referred to as *protracted* speciation, has been incorporated within the birth-death framework by Etienne and Rosindell (2012). In their model, species give birth to new, incipient species at rate λ_1 (which is the rate of speciation initiation), which can subsequently become good species at rate λ_2 (which is the rate of speciation completion), or give birth to another incipient species at rate λ_3 (Fig. 4a). Each stage has its own extinction rate μ_i. The authors show that protracted speciation provides an explanation for the shape of empirical LTT plots (Fig. 4c). Additionally, taking $\lambda_1 < \lambda_3$ generated reconstructed phylogenies with β values below 0, as is generally seen in empirical data.

In Etienne and Rosindell (2012), the likelihood of the model was only available for the pure

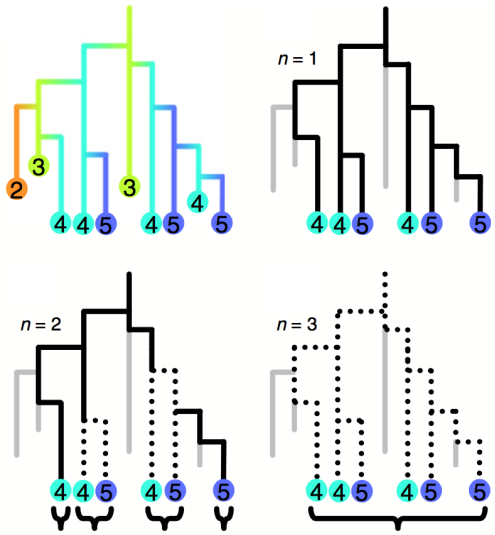


Figure 5: **Species definition in the United Theory of Ecology and Macroevolution model.** From Fig. 1 in Rosindell et al. (2015). Top-left: population genealogy. A color change indicates the presence of a mutation. Top-right: Species phylogeny for $n = 1$ (where n is the minimal number of mutations needed between two populations for them to be part of different species, see text). In that case the speciation mode is equivalent to a point mutation mode of speciation. Down row: Species phylogeny for $n > 1$, leading to a protracted mode of speciation. For $n = 3$, all lineages merge into one good species.

birth case (with $\mu_1 = \mu_2 = 0$). An approximate likelihood expression was latter derived for the model with non-zero extinction (Lambert et al., 2015), allowing to estimate the mean time needed to complete speciation in the model from empirical data (Etienne et al., 2014).

Etienne and Rosindell (2012) model defines protracted speciation at the lineage level, but it can also arise from the way species are defined from individuals genealogies. In Rosindell et al. (2010), the authors studied the effect of adding a protracted mode of speciation to the neutral theory of biodiversity. In their model, species are formed through a point mutation mode of speciation, but they are considered incipient species until τ generations occurred, assuming a fixed time for the completion of speciation. They showed that it added realism to the prediction of the neutral theory, especially when it comes to species lifetime, speciation rate and the number of rare species. Other models allow the mutation process to build up until individuals are considered as being part of different species, using lower level (individual or population level) genealogy with mutation to define species (Rosindell et al., 2015; Manceau et al., 2015). In the United Theory of Ecology and Macroevolution (UTEM Rosindell et al., 2015), two populations are part of the same species if there are less than n mutations along the genealogical path between them, or if other extant individuals bridge the gap between them (Fig. 5). As soon as $n > 1$, speciation takes time to happen and happens as a protracted process. Note that species defined this way are very rarely monophyletic, even for $n = 1$ (Fig. 5). In the Speciation by Genetic Differenciation model (SDG Manceau et al., 2015), individuals follow a birth-death process, with occasional mutation events arising at a constant rate. Species are defined from individual genealogies with mutations as the smallest monophyletic group of extant individuals such that any two individuals of same genetic type always belong to the same group (Fig. 6). This approach was used to define species in Chapter 3 of this thesis. Although neither the UTEM nor the SDG model incorporate evolutionary rate slowdown in their specification, they both generate phylogenies with more realistic species accumulation curves than the constant rate birth-death model.

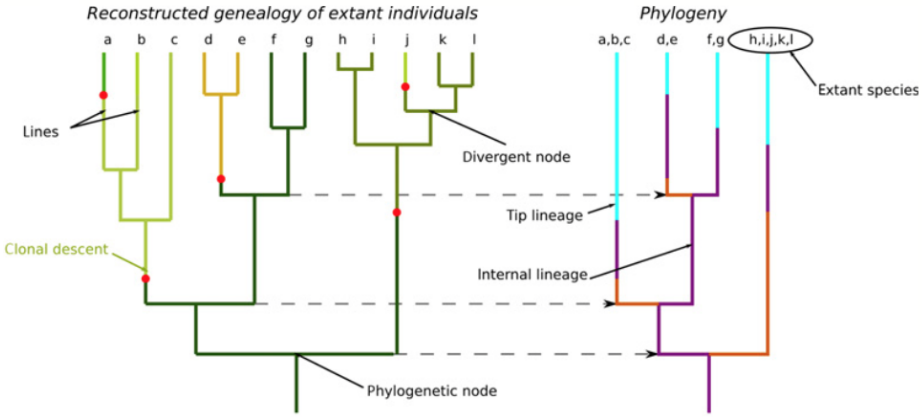


Figure 6: **Species definition in the Speciation by Genetic Differentiation model.** From Fig 1. in Manceau et al. (2015). Species are defined from the genealogy of extant individuals with mutations (left tree, with red dots denoting mutation events) as the smaller monophyletic group of extant individuals such that any two individuals of same genetic type (indicated with different color on the left tree) always belong to the same group. The right tree shows the obtained phylogeny. Even though only one mutation is needed to define a species, the monophyly condition makes the speciation mode to be protracted.

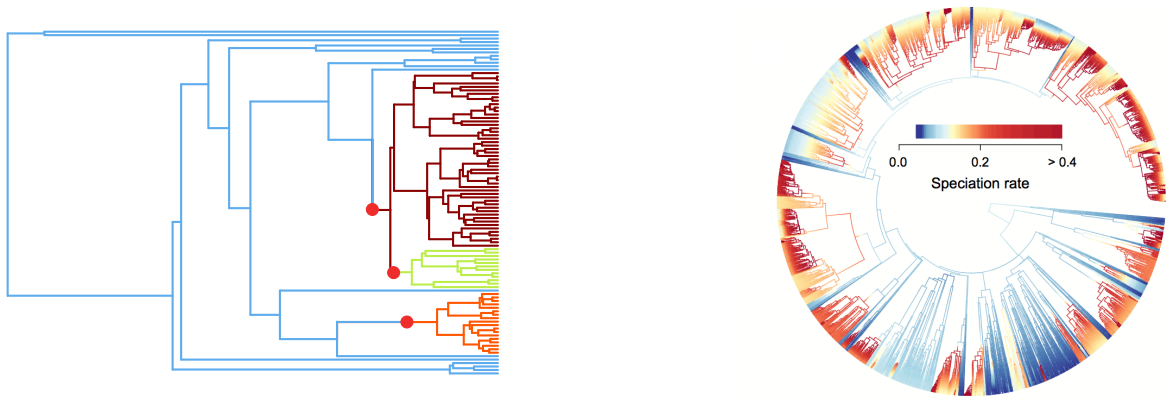
2.2 Heterogeneity across lineages.

In the models described in Section 2.1, diversification rates are hypothesized to be homogeneous within the clade. While this seems a reasonable assumption when considering only a limited set of taxa, large scale phylogenies encompass very different species, with different characteristics and evolutionary histories. There are many reasons for species not to diversify at the same speed. As an example, key innovations may happen at the basis from given clades, giving them an advantage by allowing them to temporally escape predation or competition pressure, or by allowing them to colonize a new environment, thus enhancing their diversification rates. Also, groups of species living in different biogeographical areas, or varying in essential traits such as those affecting reproductive isolation – *e.g.* reproduction mode (Goldberg et al., 2010), or pollination and dispersal syndromes (Onstein et al., 2017) – are likely to diversify at different paces.

All diversification models for which diversification rates are constant within the clade generate trees with topologies identical to those generated with the Yule model, with an expected β statistic equal to 0 (Lambert and Stadler, 2013). This differs much from topology observed in empirical data (Aldous, 2001; Blum and François, 2006), comforting us in the idea that the diversification process is far from being homogeneous in nature. Being able to position those changes on phylogenies, quantify them and link them to changes in species characteristic is of particular interest to evolutionary biology. Another important, but more technical aspect is the fact that not accounting for the possibility that rates may have varied across a clade's lineages leads to bias in diversification rates estimates. It especially leads to extinction rates estimates close to 0, opening a debate about whether extinction rates should be estimated from molecular phylogenies only (Rabosky, 2010; Beaulieu and O'Meara, 2015; Rabosky, 2016a). Yet including even a few rate shifts within the phylogeny can bring back extinction rates estimations close to those obtained from the fossil record (Morlon et al., 2011).

Rate shift detection. Two systematic methods have been proposed to position diversification rate shifts on phylogenies, both building on a lineage-based model in which diversification happens as an homogeneous process, with occasional shift events happening on a lineage. Once a shift happens, the clade descending from this lineage diversifies with its own diversification rate (Fig. 7a). The first of those methods, MEDUSA (Alfaro et al., 2009), uses a stepwise AIC (Akaike Information Criterion) procedure to detect how many shifts occurred during the history of a clade as well as their position and magnitude. A simple constant rate birth-death model is first fitted to the tree. Then a model in which there is one shift at a node is fitted for every possible position of the shift. The AIC of these models are computed, and the model with one shift is selected if the difference in AIC between the constant rate model and the best model with one shift is more than AIC_{crit} (in the first implementation, AIC_{crit} was set to 4, but the more recent version has a AIC_{crit} that depends on the number of tips in the tree). If the one shift model is selected, support for a two shifts model is tested using the same procedure, and so on until no more shifts are selected. A backward elimination procedure is then performed.

BAMM (for Bayesian Analysis of Macroevolutionary Mixtures, Rabosky, 2014, Fig. 7) is another, more recent method that has been developed for detecting diversification rate shifts on a reconstructed phylogeny. This approach uses reversible jump Monte Carlo Markov Chain (rjMCMC). Compared to MEDUSA, it has the advantage of allowing speciation and extinction rates to vary through time between rate shifts, an important feature since a decrease in diversification is commonly observed on empirical phylogenies (McPeek, 2008; Phillimore and Price, 2008; Moen and Morlon, 2014). In addition, the rjMCMC framework allows to simultaneously select the most



(a) (b)

Figure 7: **The rate shift model for the inference of heterogeneous diversification rates.** (a): A tree generated with the rate shift model used in MEDUSA and BAMM. Rate shift positions are indicated with a red dot, and different branch coloring indicates different diversification regimes. (b): Figure 1A from Rabosky (2016b). An example of an application of BAMM to an empirical phylogeny of birds from Jetz et al. (2012), with branch colors corresponding to the inferred instantaneous speciation rate of speciation at each point in the tree. Speciation rates vary by approximately 3000% over the clade history.

probable model (the number of shifts) and the model parameters (the position and the amplitude of the shifts), thus allowing to explore (unlike MEDUSA) uncertainty around the number and location of shifts.

These two methods are very similar in the way they envision diversification rate shifts. They both rely on the hypothesis that rate changes are large and uncommon (Fig. 7a), both implement a model selection procedure, and are both based on the same likelihood expression (with additional trend parameters in the case of BAMM). The major differences are on the way this selection procedure is performed (stepwise AIC for MEDUSA, rjMCMC for BAMM), and the possibility for BAMM to account for time variable rates. The ability of both method to accurately recover the number of shifts and their position on the phylogenies has recently been put into question. In MEDUSA, it come from the fact that the use of the AIC to determine the number of shifts is somewhat arbitrary. In BAMM, a problem is that the number of inferred shifts is higly dependent on the prior used (Moore et al. 2016; but see (Rabosky, 2017)). Additionally, the likelihood used for the inference of both methods use the assumption that no diversification shift occurred within unobserved lineages, and this recently fueled a controversy about their statistical performances (May and Moore, 2016; Moore et al., 2016; Rabosky et al., 2017; Rabosky, 2017; Mitchell and Rabosky, 2017; Meyer and Wiens, 2018).

In Chapter 2 of this PhD thesis, I propose an alternative method, built from a model in which diversification rates vary in a more gradual way across the phylogeny.

Character dependent diversification. The above described methods give an insight on how and where diversification rates varied in the history of a clade, but without making any assumption on why they would actually vary. One question of particular interest to biologists is whether they depend on species characters, and on which of them (Jablonski, 1987; Slowinski and Guyer, 1993; Barraclough et al., 1998). This has traditionally been answered through sister clade analyses, which

consist in comparing the species diversity in to clades descending from a single common ancestor, in which species have different character states (Mitter et al., 1988; Barraclough et al., 1998; Barraclough, 1998). Yet this approach has several caveats, in that it cannot distinguish between increased speciation or decreased extinction – being based on the principle of the method of moments, cf subsection 1.3 of the present introduction –, nor differentiate between increased diversification for one of the character states or asymmetrical transition rates (*i.e.* when the transition from one character state to the other is more likely in one direction than in the other one Maddison, 2006), and prevent using clades with mixed character states for the analysis.

To counter those caveats, a likelihood based approach, the binary-state speciation and extinction model (BiSSE, Fig. 8), has been proposed in Maddison et al. (2007). In their model, lineages are either in state 0 or 1 of the parameter of interest. The possible events for a lineage in state i are speciation (with rate λ_i), extinction (with rate μ_i), or transition to state $i \neq j$ with rate q_{ij}. They derive the likelihood of a phylogeny with observed character state at tips under this model and use it to test whether the character had an impact on diversification (by model selection against a birth-death process) and estimate the six model parameters.

This model has then been subject to many extensions, all known as the SSE methods (for state speciation and extinction). They now allow to include the possibility to account for unsampled lineages or unknown present character state (FitzJohn et al., 2009), multiple states characters or interaction between several characters (FitzJohn, 2012, MuSSE), quantitative traits (FitzJohn, 2010, QuaSSE), geographic characters (Goldberg et al., 2011, GeoSSE), cladogenetic character evolution (the changes in character state happen in conjunction with a speciation event; Goldberg and Igić, 2012, ClaSSE).

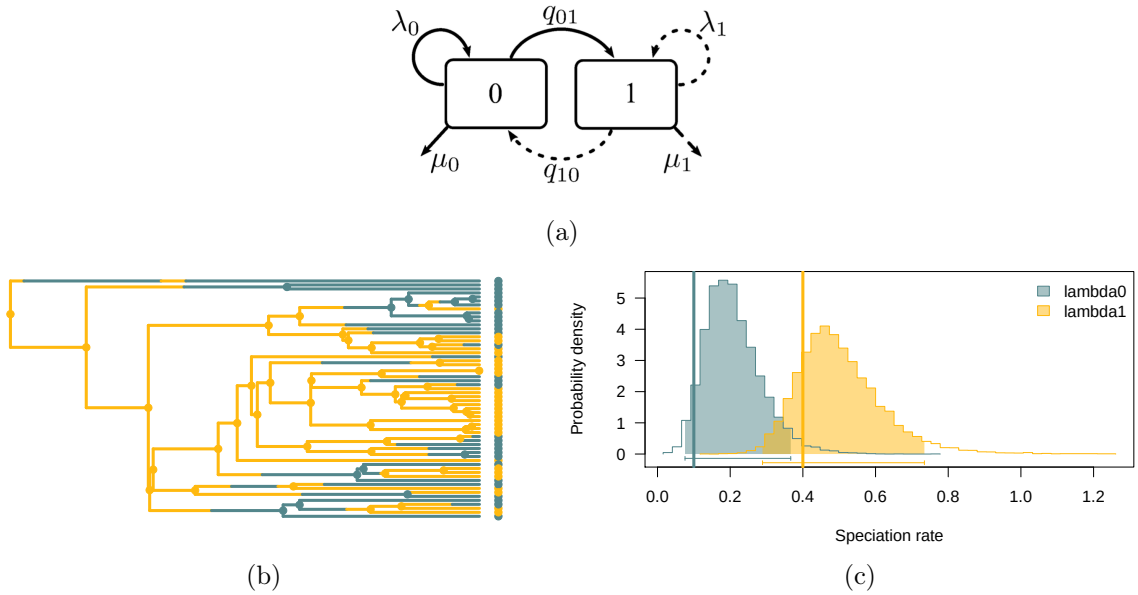


Figure 8: **The binary-state speciation and extinction model (BiSSE).** (a): Figure 1(a) from Goldberg et al. (2011). States and transitions in the BiSSE model. (b): A tree generated with the BiSSE model, with the R-package diversitree (FitzJohn, 2012). Grey is state 0, yellow is state 1. The parameter used are $\lambda_0 = 0.1$, $\lambda_1 = 0.4$, $\mu_0 = \mu_1 = 0$, $q_{01} = 0.05$, $q_{10} = 0.2$. (c): Marginal distributions of λ_0 (grey) and λ_1 (yellow) estimated by MCMC for the tree in (b). The vertical lines show the parameter values used to simulate the tree.

It has recently been shown that these approaches suffer from a high Type I error rate – traits that have evolved neutrally along the branches of a phylogeny are very often shown to have had an impact on diversification (Rabosky and Goldberg, 2015). This likely comes from the fact that the null model against which the state dependent model is compared, a birth-death model, is too simplistic. Empirical phylogenies generally display evidence for diversification rate variation – as shown by the β statistic – and that leads to rejection of the birth-death model by the method, rather than to acceptation of the character effect. New approaches have been proposed to correct for this bias. The HiSSE model (Hidden State Speciation and Extinction; Beaulieu and O'Meara, 2016) is an extension of BiSSE that adds the possibility that a hidden character – whose state at present is unknown – also had an effect on diversification. Another method uses lineage specific diversification rates estimate (obtained from BAMM, Rabosky, 2014) and a permutation procedure on the trait present states to assess whether independence between diversification rates and trait values can be rejected (Rabosky and Huang, 2015). Both those methods allow to get reasonable Type I error rates. Yet BAMM provides subclade specific rather than lineage specific rate estimates. The approach we propose in Chapter 2 of this thesis offers a way to get to lineage specific estimates, which could be useful to use in this context.

Age dependent diversification rates. Another reason for why diversification rates could differ between lineages at a given point in time is the possibility to have age-dependent speciation or extinction rates. In those models, speciation (or extinction) rates are assumed to change with the age a species (Venditti et al., 2010; Hagen et al., 2015; Alexander et al., 2015). This may arise for example from speciation being the result of many small events (Venditti et al., 2010), or as the result of varying ecological pressure or population size throughout a species age (Hagen et al., 2015; Alexander et al., 2015), resulting in non-exponentially distributed branch lengths. Those processes can either be symmetrical – both daughter species have their age reset to 0 at a speciation event – or asymmetrical – one species inherits the age of the ancestor, the other daughter species age is reset to 0. Age-dependent speciation can lead to tree shapes that are more imbalanced than for trees generated with a Yule model, but this seems not to be the case for age-dependent extinction (Hagen et al., 2015).

3 Effects of species interspecific interactions on diversification.

3.1 Biotic drivers of species diversity.

The red queen hypothesis. The *court jester hypothesis* postulates that extinction, speciation and evolution happen in response to random and unpredictable environment changes (Barnosky, 2001). The competing interpretation, the *red queen hypothesis*, postulates that biotic interactions are sufficent for evolutionary changes to happen, even in the absence of abiotic environmental changes, because of the ever changing biotic environment that requires constant adaptation (Van Valen, 1973). According to this hypothesis, an evolutionary change that gives an advantage to one species gives a disadvantage to others, which result in a continuous evolutionary race between species in the community. Although it can reasonably be assumed that both biotic and abiotic factors have a part to play, the relative weight of biotic and abiotic factors on species evolution and diversification is still a much debated issue (Barnosky, 2001; Benton, 2009; Voje et al., 2015). One common prevalent view is that, even though biotic forces act at small spatial and temporal scales, macroevolutionary processes are dominated by abiotic transitions (Barnosky, 2001; Benton, 2009).

The *zero-sum assumption* of the red queen hypothesis – that the gain in fitness of one species should be balanced by losses of fitness of others – has been heavily criticized (Maynard Smith, 1976; Stenseth and Maynard Smith, 1984). But other evolutionary models, not relying on this assumption, allow for continuous evolution and diversification even in the absence of abiotic changes (reviewed in Voje et al., 2015). Models of food web evolution show that antagonistic interactions may lead to unceasing trait evolution and species turnover, with occasional collapse in species number (Loeuille and Loreau, 2005; Guill and Drossel, 2008; Takahashi et al., 2013; Allhoff et al., 2015).

Coevolutionary diversification. Since Ehrlich and Raven (1964) study on the diversification of butterflies and plants, the idea that species coevolution can lead to codiversification has received great interest. In their verbal model, they hypothesized that antagonistic interactions between the plants and the butterfly larvae that feed on them could promote species diversity through an *escape and radiate* mechanism, with new defenses strategies appearing in the resource guild, allowing them to escape their enemies and diversify quickly as long as those did not develop another key innovation that enable them to colonize the new resource clade and radiate in turn.

Subsequent work on coevolutionary diversification has shown stronger evidence for diversification driven by antagonistic interactions than mutualistic ones, in both theoretical and empirical studies (for a review of the effect of different interaction types on diversification, see Hembry et al., 2014). Several theoretical models showed that antagonistic interaction can lead to increased trait diversity among populations (Gandon, 2002; Yoder and Nuismer, 2010). Antagonism driven speciation has been demonstrated in several empirical systems. In North American milkweeds for instance, investment in defense traits resulted in higher diversification rates (Agrawal et al., 2009). In butterflies, major host shifts resulted in bursts of diversification, in concordance with the *escape and radiate* hypothesis (Ehrlich and Raven, 1964; Fordyce, 2010).

In mutualistic communities, modeling works usually predict a reduction of trait diversity (Kopp and Gavrilets, 2006; Yoder and Nuismer, 2010). Empirical evidences for increased species diversity through mutualistic interactions are limited and mainly restricted to two kinds of interactions (Hembry et al., 2014). One of them is resource symbiose, that can facilitate the invasion of new adaptive zones. In gall inducing midges for example, those that have evolved symbiotic associations with fungi are able to use a broader range of host-plant taxa, promoting their diversification by making host-shifts more probable (Joy, 2013). Another case of mutualism inducing diversification is that of pollinator-plant interaction, that acts on reproductive isolation (Ramsey et al., 2003; Kay, 2006; Cruaud et al., 2012). But in Yucca, no association between specialized pollination and elevated diversification rate was found (Smith et al., 2008).

Testing the effect of interactions on the diversification process. While the environment-dependent diversification models (see Section 2.1) offer a tool to test for the effect of abiotic factors on diversification, only few methods are available to test for the effect of biotic interactions on an empirical phylogeny, probably because of the difficulty to account for non-independence between lineages in the birth-death framework. To my knowledge, the only method allowing to account for lineage non-independence within a clade in this framework is the diversity-dependent model (Rabosky and Lovette, 2008a; Etienne et al., 2011, see also Section 2.1). In Etienne et al. (2011), the approach was applied to five empirical phylogenies, that all favored the diversity-dependent model over a constant rate birth-death model. This suggests that negative diversity dependence is common in empirical data. Results obtained with this method have yet to be taken with care, as parameter estimates tend to be strongly biased, and model selection against a diversity independent

birth-death model is subject to a high Type I error (Etienne et al., 2016).

To test for inter-clades interactions effects on diversification, one possibility is to use the environmental dependent model and use the diversity-through-time of one group as an explanatory variable for the diversification of the second group (Lewitus and Morlon, 2017). The method revealed a positive effect of ostracod diversity on the diversification of Cetaceans, possibly due to their role as a food source. Yet this approach does not allow to test for both clade impacting the diversification of the other, as would be expected for example from the escape and radiate scenario.

If we are interested in diversification in trait space, a few other tools are available. The general comparative phylogenetic methods framework can be used to incorporate diversity-dependent trait evolution (Mahler et al., 2010; Weir and Mursleen, 2013), or to test for the reciprocal effect of species similarity in trait value, either within a clade or between the species of two different interacting clades (Drury et al., 2016; Manceau et al., 2016). Applied to the Greater Antillean *Anolis* lizards radiation enabled Drury et al. (2016) to highlight the impact of species competition on trait evolution. Applying this type of method to two interacting clades require to have an idea about what species interacted with whom during the clades history, and thus to have knowledge about how ecological communities assemble and evolve. In the following, I focus on the description of ecological communities structure through the use of ecological networks.

3.2 Bipartite interactions.

Bipartite networks. A common and useful way to describe an interacting community is through the use of interaction networks, that can be represented using the formalism of graph theory. A graph is a pair (V, E) composed of a set of *nodes* (also known as *vertices*) V and a set of *edges* (often referred to as *links* in the ecological literature) $E \subset V^2$. The graph can either be *undirected* – meaning that links have no orientation, the edge (x, y) and (y, x) are identical – or *directed* – in which case edges have orientations, (x, y) is directed from x to y and is different from (y, x) that is directed from y to x. A weighted network is a network in which a weight is assigned to each edge (Fig. 10a).

In the case of ecological networks, V is the set of species in the community. A species pair (x, y) belongs to E if x and y interact together. The links can be ordered – *e.g.* for a food-web, $(x, y) \in E$ means that x feeds on y – or unordered – *e.g.* in the case of symmetrical competition –, and may or may not be weighted by interaction strength.

If there exist two species subsets A and B such as $A \cap B = \emptyset$, $A \cup B = V$ and $\forall (x, y) \in E, (x, y) \in A \times B$ or $(x, y) \in B \times A$ (*i.e.* all interaction links are between a species from A and one from B), the network is said to be *bipartite* (Fig. 10a). Those networks are useful to describe ecological communities in which there are two distinct sets of organisms, such as plant and pollinators, host and parasites, or plants and herbivores.

Bipartite networks may also be equivalently described with their *adjacency matrix*, M, which has $|A|$ rows and $|B|$ columns (where $|X|$ refers to the number of element in the set X). For unweighted networks, $M_{i,j} = 1$ if the i^{th} species from A interacts with the j^{th} species from B and $M_{i,j} = 0$ otherwise. For weighted networks, $M_{i,j}$ is set to the interaction strength between species i and species j (Fig. 9a,b, 10c,d).

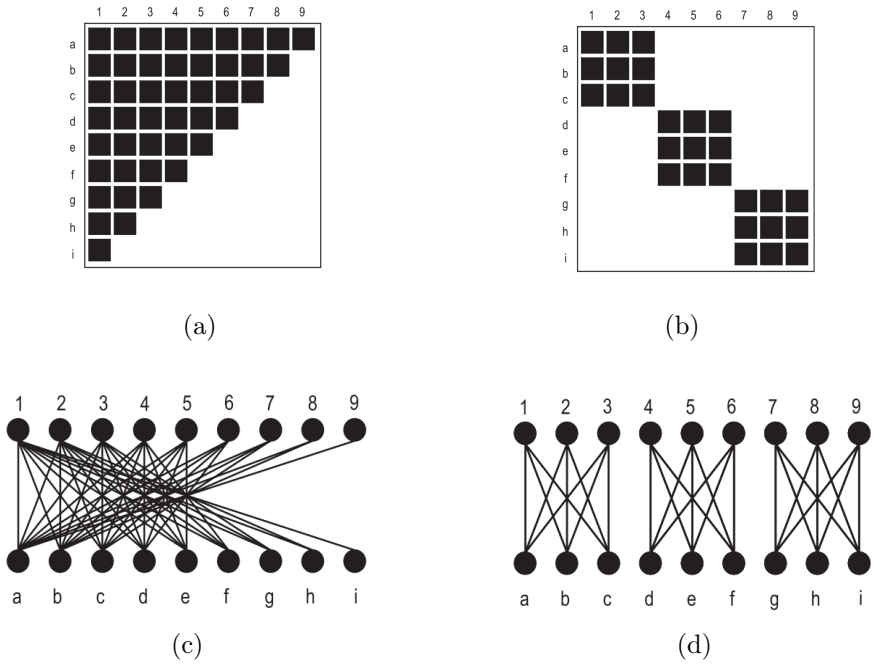


Figure 9: **Two well studied binary network structures,** shown with the matrix representation (a, b) and with the network representation (c, d). From Fig. 1 in Fontaine et al. (2011). (a, c): A perfectly nested bipartite interaction network. Species with few interactions interact with a subset of the interacting partners of the more generalist species. (b, d): A modular bipartite interaction network, displaying 3 modules.

Describing bipartite network structures. Summary statistics are once again a useful tool to be able to compare different ecological networks and find what common properties they share with each other. I will describe here those that are the more often looked at in the literature, how they are distributed in empirical networks, and how modeling studies have allowed to better understand why they do so (Fig. 10).

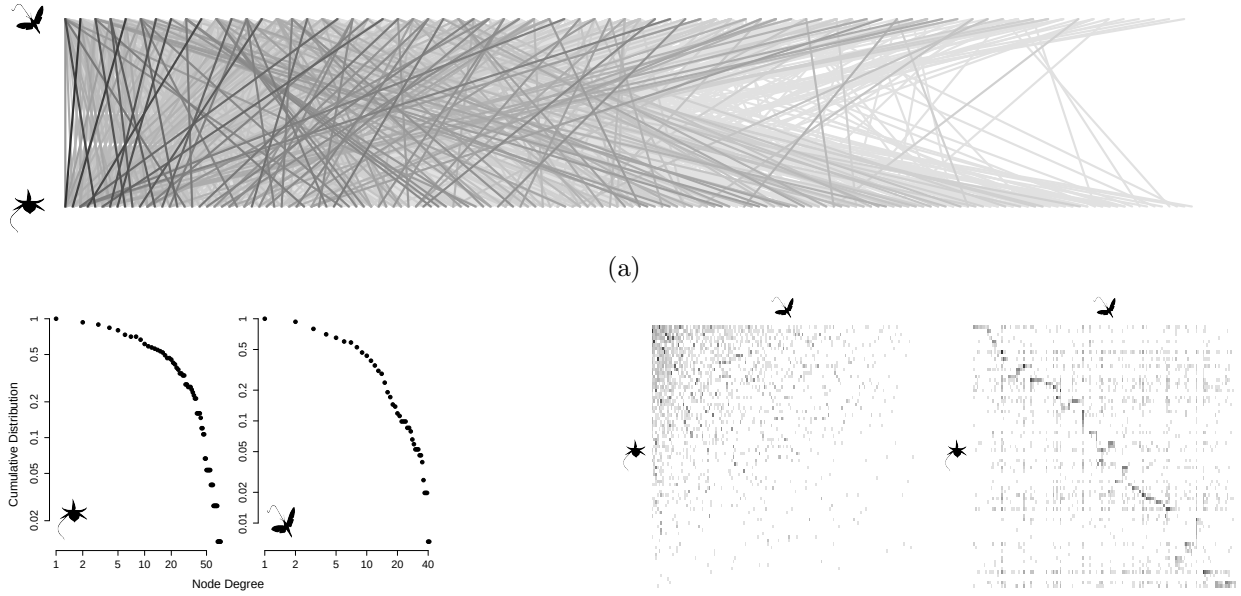
Connectance.— The connectance C of a network is defined as the realized proportion of all possible links. It is one of the simplest summary statistic that can be extracted for a network. For a bipartite interaction network, $C = \frac{|E|}{|A||B|}$. It is closely linked to the link density, which is defined for a given guild as the mean number of link per species (*i.e.* $\frac{|E|}{|A|}$ for guild A).

In empirical networks, the connectance is generally low and decreases as a power of network size (Briand, 1983; Schoenly et al., 1991; Santamaría and Rodríguez-Gironés, 2007; Fortuna et al., 2010; Canard et al., 2012). For a given network size, it is generally higher for mutualistic than for antagonistic networks (Fig. 11a ; Thébault and Fontaine, 2010).

Node degree distribution.— The number of links connected to a node is called the *degree* of this node. In complex networks of various nature, such as scientific collaboration networks, social networks, and chemical-reaction networks, the distribution of node degree (Fig. 10b) is often *scale-free* (meaning that it decays as a power law, *i.e* $P(k) \sim k^{-\gamma}$, where $P(k)$ is the probability that a node has at least degree k, and γ is a positive constant; Barabási and Albert, 1999). Amaral et al. (2000) showed that it could also fall into one of the two other following categories: *broad-scale* (*i.e.* decaying as a truncated power law) or *single-scale* (*i.e.* with a fast decaying tail, such as the exponential or gaussian distributions).

The pervasiveness of the scale-free distribution across network types suggests that there is something in common in the way large networks assemble. It can be obtained from the *preferential attachment* random graph model. In this model, networks are constructed by starting with a small number of nodes m_0 and sequentially adding new nodes, that are connected to $m \leq m_0$ existing ones. The nodes to which the new one is attached are selected with a probability that is proportional to the current nodes degrees (Barabási and Albert, 1999).

Most bipartite ecological networks have broad-scale – or truncated power law – node degree distributions (Jordano et al., 2003; Montoya et al., 2006). Incompatible biological attributes of species result in *forbidden links*, and those constraints on the attribution of new links may be the reason for the departure from the scale-free topology, preventing the 'rich gets richer' dynamics in this context (Montoya et al., 2006). In their data, Jordano et al. (2003), found that 80% of



(a) (b) (c) (d)

Figure 10: **Different representations of a quantitative bipartite interaction network, obtained by simulation from the interaction model in chapter 3.** (a): Graph representation of the network. A link is drawn between two species if they interact together. The link's shade of gray indicates the interaction strength, with light gray for weak interactions and dark gray for strong interaction. (b): Log-log plot of the cumulative node degree distribution for each of the two guilds. (c): Matrix representation of the network in (a), with species ordered to highlight the network's nested structure. (d): Matrix representation of the same network, with species ordered to highlight the network's modular structure.

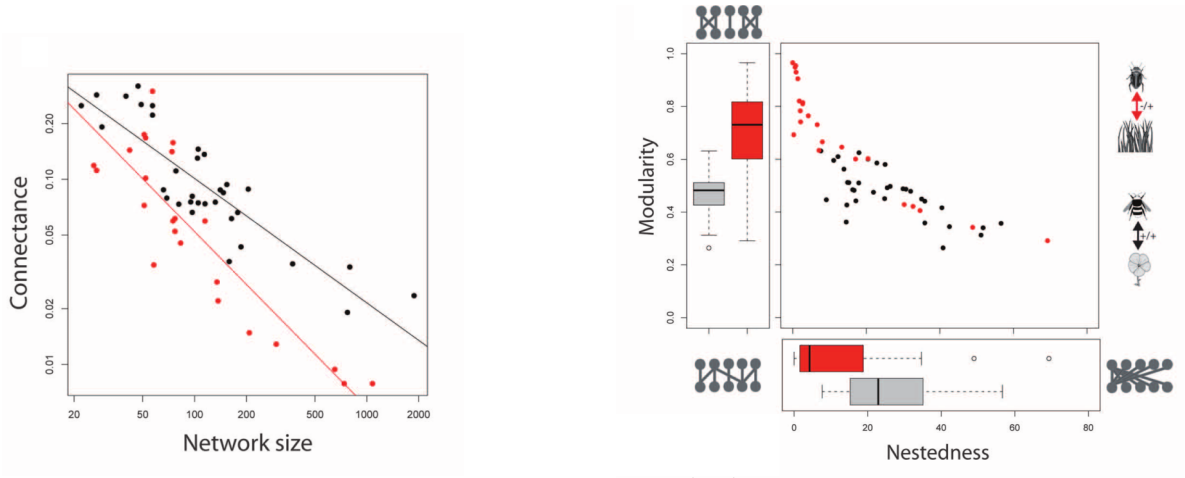
the unobserved interaction could indeed be explained by uncoupling in either species phenology, size restriction or structural constraints. Alternative explanations for the presence of a cutoff in ecological networks node degree distribution involve asymmetrical species interaction frequencies (Vázquez, 2005) or finite size artifacts (Jordano et al., 2003; Keitt and Stanley, 1998).

Nestedness.— A network is said to be *nested* when species with few partners are interacting with a subset of the partners of the more generalist species. It also means that specialist species from one guild tend to interact more with generalist species from the other guild (Fig. 9a,c, 10c).

Many measure of nestedness have been proposed in the literature (for a review of existing metrics, see Ulrich et al., 2009). Among the most commonly used are the temperature metric (Atmar and Patterson, 1993), that measures the number of unexpected presence or absence of links around an isocline of perfect nestedness, and the NODF metrics (for Nestedness metric based on Overlap and Decreasing Fill, Almeida-Neto et al., 2008), based on the interaction overlap between pairs of species with decreasing interaction number. This last measure was extended to be applied to quantitative networks (the wNODF metric, Almeida-Neto and Ulrich, 2011).

Whatever metric is being used, the absolute value of the nestedness is dependent on the number of species in the network as well as on its connectance. To be able to compare nestedness for networks with different size and fill, it is thus recommended to standardize the rough value by using a null model of network organization. This is often summarized with the Z-value, defined as $Z = \frac{N_{obs} - \mathbb{E}(N_{null})}{\sqrt{\text{var}(N_{null})}}$ which would be distributed as a standard normal distribution were the interaction matrix generated with the null model, and thus gives a measure of how much (and in which direction) the observed value deviates from the null model. The choice of the null model depends on what is expected to constrain the system (Gotelli and Ulrich, 2012). A common choice is to generate matrices of the same size as the empirical one, with an interaction probability that is equal to the matrix connectance C for all species pairs (conditioned to the fact that each species have at least one interaction Bascompte et al., 2003). Another choice is to take the probability of two species interacting to be proportional to the product of the interaction degree of the two species, with or without also controlling for the network connectance (Vázquez, 2005; Thébault and Fontaine, 2010). Finally, the observation that empirical network connectance may be due to the high asymmetry in species abundance distribution led to propose a null model in which the probability of two species interacting together is proportional to the product of their abundances (Moore and Swihart, 2007; Blüthgen et al., 2008; Nuismer et al., 2013).

Empirical bipartite mutualistic networks are generally nested (Fig. 11b ; Bascompte et al., 2003; Thébault and Fontaine, 2010). They usually display higher nestedness values than antagonistic networks, even though these also occasionally display significant nestedness (Bascompte et al., 2003; Vázquez et al., 2005). Multiple explanations have been proposed for this pattern, which fall in three categories. For the first family of explanations, the high nestedness values in empirical network reflects the fact that communities that are assembled at random generally beget unstable ecological dynamics (May, 1972). In randomly assembled mutualistic networks, preferential extinction leads to increased nestedness values (Thébault and Fontaine, 2010). Similarly, nested mutualistic networks are stable – in the sense that they allow the coexistence of all species – for a larger range of ecological parameters – interaction and competition strength, growth and death rate – than their non nested counterparts (Rohr et al., 2014; Saavedra et al., 2016). The second family of explanations considers that this high level network structure stems from species level interaction rules, that are mediated by individual traits. Because not all species are able to interact together, the presence



(a) (b)

Figure 11: **Network metrics patterns in empirical bipartite networks.** Adapted from Fig. 3 in Thébault and Fontaine (2010). Data for antagonistic networks are shown in black, while those for mutualistic networks are shown in red. (a): Network connectance as a function of network size, displayed on a log-log scale. For both interaction types, connectance decreases as a power of network size. For a given network size, the connectance is lower in mutualistic communities (red dots and line) than in antagonistic communities (black dots and line). (b): Modularity and nestedness values in empirical communities. Antagonistic communities (in red) are more modular and less nested than mutualistic communities (in black).

of forbidden links determines the structure of the network (Jordano et al., 2003). Species level interaction rules can lead to a nested network structure whether interactions are controlled by trait matching (*e.g.* phenology, color preferences) or trait differences (*e.g.* insect proboscis / flower corolla lengths) mechanisms, especially if there is a phylogenetic signal in those traits (Santamaría and Rodríguez-Gironés, 2007; Rezende et al., 2007). The last family of hypotheses supports the idea that nested network structures come from the highly asymmetric species abundance distributions observed in natural communities. If individuals interact at random with one another, following a neutral scenario, it is unlikely to observe interactions between two rare species (suggesting the concept of neutral forbidden links Canard et al., 2012). Individuals from rare specialist species are more likely to meet and interact with individuals from abundant generalist species (Vázquez, 2005; Santamaría and Rodríguez-Gironés, 2007; Krishna et al., 2008; Vázquez et al., 2009; Coelho and Rangel, 2018). This idea is supported by the fact that, although binary interaction networks are nested, the signal is not as strong when interaction strengths are corrected by abundances to get information about species preferences (Staniczenko et al., 2013). Yet the fact that generalist species are also the most abundant does not necessarily imply that generalism is explained by abundance (although there is some empirical evidence for the relationship going this way, see Fort et al., 2016), and the reverse may be true (Santamaría and Rodríguez-Gironés, 2007; Fontaine, 2013).

Modularity.— Complementary to nestedness, the modularity structure of ecological networks has been well studied in empirical data. A *module* is a group of species that are more strongly connected with each other than with other species in the network (Fig. 9b). The propensity of the network to display modules is called its *modularity* (Fig. 10d). A common metric used to measure it in ecological communities was introduced by Barber (2007) for binary networks, and extended to weighted networks by Newman (2004). It is defined as

$$Q = \frac{1}{2m} \sum_{i,j} (A_{ij} - K_{ij}) \delta(c_i, c_j) \tag{5}$$

where A is the interaction matrix, $m = \sum_{i,j} A_i j$, $\delta(c_i, c_j) = 1$ if species i and j belong to the same networks and $\delta(c_i, c_j) = 0$ if they do not, and K is the matrix of expected interaction weights under a given null model – which equals the interaction probabilities for binary networks. The modularity of the network is the maximum value of Q over all possible modules decompositions. Several algorithms exist to perform the maximization (Barber, 2007; Dormann and Strauss, 2014).

Modularity has been found in a great variety of ecological interaction networks, like food webs (Krause et al., 2003) or pollination networks (Olesen et al., 2007). When looking at bipartite interactions, antagonistic networks generally show a clearer modularity signature than mutualistic ones (Fig. 11b ; Thébault and Fontaine, 2010; Fontaine et al., 2011).

Combining phylogenetical and network information : phylogenetic signal in interaction partners. For interactions that are mediated by traits, closely related species are expected to interact with similar interaction partners if those traits display phylogenetic signal – the fact that closely related species tend to display similar trait values. A phylogenetic signal in interaction partners identity has been proposed as an explanatory factor for the nested structure of bipartite networks (Rezende et al., 2007), and can be measured through different techniques. Ives and Godfray (2006) modeled the interaction probability between a consumer and a prey species as being proportional to the product of a consumer trait X and a prey trait Y, and measured the phylogenetic signal in X and Y. Yet because increasing the trait value for one species increases its interaction probability with all species from the other guild indifferently, their metric is more a measure of phylogenetic signal in number of partners than in partner identity. For modular networks, another way is to first assign species to modules and assess whether the phylogeny explains module affiliation (Gómez et al., 2010), for instance using the metric from Maddison and Slatkin (1991). Finally, phylogenetic signal in interaction partners can also be assessed by using a Mantel test (Mantel, 1967). This test, that has been used in many empirical studies (see for instance Bersier and Kehrli, 2008; Elias et al., 2013; Fontaine and Thébault, 2015), relies on the comparison of two matrices of distances between species, one of them being the phylogenetic distance matrix and the second one a measure of distance of interaction partner sets. The correlation between those two distances is compared to those obtained from permutations of one of those matrix rows and columns, which allows to assess its significance.

In empirical data, ecological interactions have been shown to be generally conserved across the tree of life (Gilbert and Webb, 2007; Bersier and Kehrli, 2008; Gómez et al., 2010). For bipartite communities, the signal depends on the interaction type and is stronger in antagonistic communities than in mutualistic ones (Fontaine and Thébault, 2015). Within antagonistic networks, conservatism is stronger for resource species than for consumer species (Cagnolo et al., 2011; Elias et al., 2013; Fontaine and Thébault, 2015). This is also the case for food webs (Bersier and Kehrli, 2008), and

could be explained by a strong selection for related consumers to shift resource because of competition. Within plant pollinator networks, animal species tend to display a stronger phylogenetic signal in interaction partners than plant, although the difference is much lower than between consumer and resource species in antagonistic networks (Fontaine and Thébault, 2015). Those results generally show that ecological interactions are constrained by evolutionary processes and call for the integration of both ecological and evolutionary processes when studying interacting communities.

Very few modeling studies are looking at the simultaneous emergence of phylogenies and network structures. An exception is Poisot and Stouffer (2016). They proposed a lineage-based model in which new species inherits the interactions of its parent species, but may lose some of them. Yet, even though they assume an effect of phylogenetic relationships on the ecological interactions' structure, speciation rates are taken to be homogeneous in each group and independent of the number or identity of the species interactors. In Chapter 3, we propose an individual-based model with two interacting guilds, that generates both ecological (*e.g.* network structures, species abundance distributions) and evolutionary (*e.g.* phylogenies, trait distributions) patterns, with the possibility of reciprocal interactions of ecological and evolutionary processes on each other.

4 Thesis outline.

In this PhD thesis, I focus on three different approaches to the study of diversification patterns, each based on a stochastic model. In Chapter 1, I present a new metric for the description of phylogenetic tree shape, and more precisely the age-richness relationship within a single tree (see subsection 1.2). While this approach is not explicitly referring to the diversification process, the expected value of our new metric under an homogeneous speciation model is well known, and inferring its value on an empirical dataset of family level bird phylogenies highlights how they differ in their shape from diversification models expectations.

Chapter 2 handles more directly the issue of within clade diversification rate variations, by proposing a new method to infer heterogeneous diversification rates on a phylogeny. The underling model is a lineage-centered model based on the birth-death model, but in which newborn lineages are attributed their own speciation rates, sampled from a distribution that depends on the rates of their parent lineage. We develop a procedure to infer the model's parameter, enabling to quantify rate heritability on a phylogeny and measure branch specific speciation rates. Applying it to bird family-level phylogenies brings out the fact that within-clade rate variations are the same order of magnitude that between-clades variations, justifying the use of this type of model for this dataset.

Finally, Chapter 3 examines one of the possible causes for diversification rates variations through the incorporation of ecological interactions in the evolutionary process. We consider an individual-based model of two interacting guilds, and study how the variations in the interaction type and strength impact the final diversity of the system, as well as the final community structure. We show that simple evolutionary mechanisms suffice to obtain differences in network structure between antagonistic and mutualistic communities that qualitatively match those observed in empirical data.

Chapter 1 : A new index for the clade age-richness relationship

Characterizing the shape of empirical phylogenies allows to highlight the ways in which they differ from models expectations, without needing to have previous expectation on the processes that may have led to this divergence. This is for example the case with imbalance, that is higher in data than expected under homogeneous speciation models. Another expectation of homogeneous rates diversification models is that clade size is positively related to clade age. This relationship have been looked for in empirical studies, and a positive correlation was rarely found. However, classical approaches imply computing the correlation between age and size for a set of clades, which may be obtained by pruning them for a larger scale tree. This necessitates to define which subclades are to be included in the analysis, and the results are sensible to this choice as well as to whether clade age is computed as the stem or the root age. In this chapter, we present a new metric for the clade age-richness relationship that can be computed at the level of one phylogeny by looking at its relative node order. This metric is based on an extension of Aldous's one parameter β-splitting model. We added a novel parameter, α, that controls for the order in which the splits happen. The pair (β, α) characterize the ranked tree shape – tree shape with ordered node depth –, with $(\beta, \alpha) = (0, 1)$ generating the same distribution on ranked tree shape than the birth-death model. We also consider an additional third parameter, η, quantifying the relation between relative abundance – or any other tip data summing to 1 for the whole tree – and richness of subclades, here taken as a proxi for extinction risks. We study the effect of our model's parameter on the loss of Phylogenetic Diversity (PD, the sum of the branch lengths of a clade) during an extinction crisis. This enables to highlight a parameter zone for which PD loss is more rapid than species loss, that we term the 'danger zone'. We develop an inference procedure for our model parameter and apply it to a data set of 120 family level bird phylogenies, with range sizes as tip data. Consistently with previous studies, we find that β values cluster around -1, highlighting the higher than expected imbalance in tree shapes. The inferred α values show that small subclades tend to be deeper in the tree than in a birth-death model in bird family phylogenies, which ca reflect past variations in diversification rates during the evolutionary history of birds. The inferred η values show that species with small range size tend to cluster in small subclades, possibly due to resource limitation affecting both the density of individuals and the species number. The combination of parameters found in the birds dataset show that bird phylogenies fall very close to the danger zone, making they prone to important losses of PD.

This chapter is a work that was initiated during my master thesis with Amaury Lambert and Fanny Gascuel, and was subsequently developed and finished during this PhD. It is now published in *Systematic Biology* (Maliet et al., 2018).

Syst. Biol. 0(0):1–20, 2018

DOI:10.1093/sysbio/syy030

Ranked Tree Shapes, Nonrandom Extinctions, and the Loss of Phylogenetic Diversity

ODILE MALIET[1,2,*], FANNY GASCUEL[1,2,3] AND AMAURY LAMBERT[3,4]

[1]*Institut de Biologie de l'École Normale Supérieure (IBENS), École Normale Supérieure, CNRS, INSERM, PSL Research University, Paris, France;*
[2]*ED 227, Sorbonne Universités, Paris, France;*
[3]*Center for Interdisciplinary Research in Biology (CIRB), Collège de France, CNRS, INSERM, PSL Research University, Paris, France; and*
[4]*Laboratoire Probabilités, Statistique et Modélisation (LPSM), Sorbonne Université, CNRS, Paris, France*
**Correspondence to be sent to: Institut de Biologie de l'École Normale Supérieure, équipe 'Modélisation de la biodiversité', École Normale Supérieure, 46 rue d'Ulm, 75005 Paris, France;*
E-mail: odile.maliet@orange.fr.

Received 3 June 2016; reviews returned 30 March 2018; accepted 8 April 2018
Associate Editor: Erika Edwards

Abstract.—Phylogenetic diversity (PD) is a measure of the evolutionary legacy of a group of species, which can be used to define conservation priorities. It has been shown that an important loss of species diversity can sometimes lead to a much less important loss of PD, depending on the topology of the species tree and on the distribution of its branch lengths. However, the rate of decrease of PD strongly depends on the relative depths of the nodes in the tree and on the order in which species become extinct. We introduce a new, sampling-consistent, three-parameter model generating random trees with covarying topology, clades relative depths, and clades relative extinction risks. This model can be seen as an extension to Aldous' one parameter splitting model (β, which controls for tree balance) with two additional parameters: a new parameter α quantifying the relation between age and richness of subclades, and a parameter η quantifying the relation between relative abundance and richness of subclades, taken herein as a proxy for overall extinction risk. We show on simulated phylogenies that loss of PD depends on the combined effect of all three parameters, β, α, and η. In particular, PD may decrease as fast as species diversity when high extinction risks are clustered within small, old clades, corresponding to a parameter range that we term the "danger zone" ($\beta < -1$ or $\alpha < 0$; $\eta > 1$). Besides, when high extinction risks are clustered within large clades, the loss of PD can be higher in trees that are more balanced ($\beta > 0$), in contrast to the predictions of earlier studies based on simpler models. We propose a Monte-Carlo algorithm, tested on simulated data, to infer all three parameters. Applying it to a real data set comprising 120 bird clades (class Aves) with known range sizes, we show that parameter estimates precisely fall close to the danger zone: the combination of their ranking tree shape and nonrandom extinctions risks makes them prone to a sudden collapse of PD. [Beta-splitting model; biodiversity; broken stick; field of bullets model; macroevolution; phylogenetic tree; rarefaction; sampling distribution; self-similar fragmentation]

As it becomes increasingly clear that human activities are causing a major extinction crisis (Leakey and Lewin, 1995; Glavin, 2007; Wake and Vredenburg, 2008; Barnosky et al., 2011), several theoretical studies have aimed at characterizing how the evolutionary legacy of parts of the tree-of-life, and hence also the genetic diversity able to drive future evolution, will decrease in the face of forthcoming extinctions. This evolutionary component of biodiversity can be measured by the phylogenetic diversity (PD), defined as the sum of the branch lengths of the phylogeny spanned by a given set of taxa (Faith, 1992). This metric is increasingly being used to measure biodiversity and to identify conservation strategies (Veron et al., 2015).

Nee and May (1997) were the first to formally investigate the expected loss of PD in the face of species extinctions, by simulating species trees using the Kingman coalescent. They found that 80% of the PD can be conserved even when 95% of species are lost. Further studies showed that the loss of PD is in fact much higher when trees are generated through other models of species diversification, such as the Yule or the birth–death models (Morlon et al., 2011b; Mooers et al., 2012; Lambert and Steel, 2013). These models indeed generate longer pendant edges (i.e., branches that lead to the tips), hence lower phylogenetic redundancy, than in the standard Kingman coalescent (used by Nee and May, 1997). However, Nee and May (1997) also showed that PD is very sensitive to the shape of the species tree (also called its "topology"), with extremely unbalanced trees ("caterpilar trees") losing much more PD than balanced trees ("bush trees"), due to a lack of phylogenetic redundancy (i.e., the presence of recently diverged sister species). Overall, these results highlighted the sensitivity of the loss of PD in response to species extinctions to both edge lengths and tree shape.

In this line, we also expect the correlation between the species richness of clades and their relative ages to have a significant impact on the loss of PD ("clade" standing here for any subtree within the full phylogeny). Here the age of a clade, also called "stem age," denotes the depth (measured from the present) of its root node (i.e., the node where this clade is tied to the rest of the tree). Under random extinction, since smaller clades are more likely to become extinct first, the consequence of their total extinction on PD will depend on the lengths of pendant edges in these clades compared to those in larger clades. Under models with diversification rates that are constant through time and homogeneous across lineages, the time for speciation hypothesis states that the size of clades is correlated to their age (Magallon and Sanderson, 2001), yet several empirical studies found no correlation between the two (Ricklefs 2007; Rabosky et al. 2012; Sánchez-Reyes et al. 2016; but see McPeek and Brown 2007). The effect of such correlation on the loss of PD has not yet been explored, but should be particularly

important in unbalanced phylogenetic trees (exhibiting large variation in the species richness of clades), which dominate empirical data (e.g., Guyer and Slowinski, 1991; Heard, 1992; Guyer and Slowinski, 1993; Slowinski and Guyer, 1993; Mooers, 1995; Purvis, 1996; Mooers and Heard, 1997; Blum and François, 2006).

Besides, the loss of PD was shown to be influenced by the distribution of extinction risks within species trees. Several studies showed that accounting for realistic scenarios of species extinctions (considering that species with higher extinction risk as per the International Union for Conservation of Nature (IUCN) Red List status are more likely to go extinct first) predicts proportionately higher losses in PD than scenarios with random extinction risks (e.g., and review, Purvis et al., 2000a; von Euler, 2001; Purvis, 2008; Veron et al., 2015). Extinctions may for example be clustered within certain clades (Bennett and Owens, 1997; McKinney, 1997; Russell et al., 1998; Purvis et al., 2000a; Baillie et al., 2004; Bielby et al., 2006; Fritz and Purvis, 2010), correlated to the age of clades (von Euler, 2001; Johnson et al., 2002; Redding and Mooers, 2006), to the species richness of clades (Russell et al., 1998; Hughes, 1999; Purvis et al., 2000a; Schwartz and Simberloff, 2001; von Euler, 2001; Johnson et al., 2002; Lozano and Schwartz, 2005; Vamosi and Wilson, 2008, assuming in some studies a correlation between rarity and extinction risks), or to speciation rates (Heard and Mooers, 2000). In contrast, most of the theoretical analyses of predictions based on model trees (Nee and May, 1997; Mooers et al., 2012; Lambert and Steel, 2013) have all been based so far on the field of bullets model, which considers equal extinction probabilities across species (Raup et al., 1973; Van Valen, 1976; Nee and May, 1997; Vazquez and Gittleman, 1998; but see Heard and Mooers 2000). One can assume extinction events are independent but not identically distributed across species, as considered in the generalized field of bullets model (Faller et al., 2008). In an exchangeable phylogenetic model in which extinction probabilities are themselves random and independent with the same distribution, this would not affect the overall loss of PD (as both models are stochastically equivalent, Lambert and Steel, 2013). However, as stated by Faller et al. (2008), it is essential to explore models that weaken the strong assumption in the (generalized) field of bullets models that extinction events are randomly and independently distributed among the tips of phylogenetic trees.

Here, we hence investigate how the loss of PD is influenced by the two abovementioned factors: (i) the ranked shape of the species tree, characterized by the relation between the richness of clades and their age or depth in the tree and (ii) nonrandom extinctions, characterized by the relation between the richness of clades and the extinction risks within them. Here, "ranked shape" refers to the shape of the tree combined with the additional knowledge of relative depths—the order in which nodes appear in the tree, but to the exclusion of the actual divergence times (e.g., Lambert et al., 2017).

We introduce a three-parameter model generating random ranked tree shapes endowed with random numbers summing to one at the tips, interpreted as relative abundances (or geographic ranges) of contemporary species. This model can be seen as an extension to Aldous' β-splitting model (Aldous, 1996, 2001) with two additional parameters: a parameter α quantifying the relation between the richness of a clade and its relative age (i.e., the rank of appearance of its root node) termed "age-richness index" hereafter, and another parameter η quantifying the relation between the richness of a clade and its relative abundance or frequency (i.e., the sum of abundances of the species it encompasses divided by the sum of abundances of all extant species in the phylogeny), termed "abundance-richness index" hereafter. When $\beta=0$ and $\alpha=1$, the ranked shape of the tree is the same as the ranked shape of a standard coalescent tree or of a Yule tree stopped at a fixed time (see Proposition 1 in Supplementary Appendix S1 available on Dryad). We further assume that extinctions of contemporary species occur sequentially in the order of their abundances, starting with the least abundant species, which roughly reduces to the field of bullets model when $\eta=1$ (see Proposition 2 in Supplementary Appendix S1 available on Dryad).

The parameters of the model are not supposed to map onto biological processes. Our aim is to produce and describe a broad range of ranked tree shapes and extinction risk distributions. The model nevertheless reflects relevant biological patterns. The imbalance and node order of phylogenies may be affected by diversification rates varying with time (Rabosky and Lovette, 2008b; Moen and Morlon, 2014), with species age (Doran et al., 2006; Hagen et al., 2015; Alexander et al., 2015), or among lineages (Cardillo et al., 2005; Maddison et al., 2007; Alfaro et al., 2009; Morlon et al., 2011a). Extinction risk may cluster within the phylogeny if it is correlated with some species characteristic, such as body size (Gaston and Blackburn, 1995; Johnson et al., 2002) or habitat use (Johnson et al., 2002). Finally, tree imbalance, node order and extinction risk clustering are likely to interact if differences in diversification rates and extinction risks are driven by the same trait, or are the result of a common process (e.g., geographic speciation, Pigot et al. 2010).

We explore the rate of decrease of PD as species sequentially become extinct, based on simulated data under variation in all three parameters over a significant range of their possible values. Interestingly, the joint variation of the parameter η with the ranked shape of species trees (set by parameters β and α) affects the clustering of extinction risks and the relationship between extinction risks and clade age (determined by the similarity or dissimilarity of the direction of deviations of α and η from 1). Therefore, considering simultaneous variation in β, α, and η allows us to explore the effects on the loss of PD of the different patterns of nonrandom extinctions observed in empirical

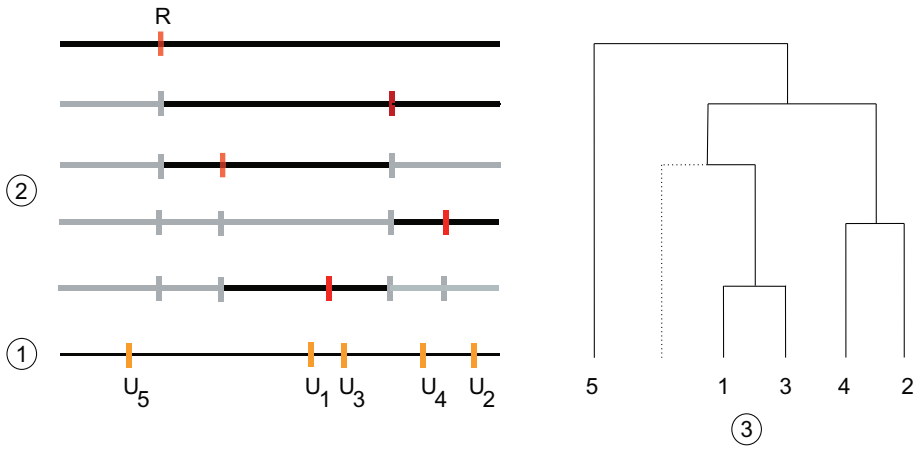


FIGURE 1. Illustration of the model generating ranked tree shapes. Construction of the ranked shape of a tree containing $N=5$ species. (1) Five random marks $(U_i)_{i\in\{1,\dots,5\}}$ are drawn uniformly in the interval $[0,1]$ (the marks on the bottom line). (2) At each time step (time flowing downwards), we randomly select one interval X, with each interval X_j having a weight $|X_j|^\alpha$ (in black). Then, we draw a random variable R in a beta distribution with parameters $(\beta+1, \beta+1)$, and split the selected interval X into two subintervals, X_{left} of size $R|X|$ and X_{right} of size $(1-R)|X|$ (dark red mark). (3) Repeating this process over time until all intervals X_j contain only one mark leads a tree with a ranked shape. Dotted branches correspond to unsampled subtrees (i.e., there is no mark in the corresponding interval).

data. We therefore provide general predictions on the sensitivity of the evolutionary legacy of clades to extinction, as a function of three simple statistics summarizing tree balance, ranked tree shape and the distribution of extinction risks across clades.

Following this exploration, we then propose a Monte-Carlo inference algorithm enabling maximum likelihood estimation of the parameters β, α, and η from real data sets. When tested against simulated data, this algorithm performs reasonably well over a wide range of parameter values for phylogenies with 50 tips or more. The estimates of parameters (beta, alpha, eta) on a real data set of bird family phylogenies and their range size distributions finally reveal empirical patterns clustered within a given parameter zone which make these clades particularly prone to strong loss of PD.

MATERIALS AND METHODS

Modeling Ranked Tree Shapes

The first version of the model we present allows one to generate random ranked tree shapes, that is tree shapes endowed with the additional knowledge of node ranks. Usually, one can generate random ranked tree shapes by time-continuous branching processes stopped at some fixed or random time, where particles are endowed with a heritable trait influencing birth and death rates. In these models, it is generally not possible to characterize the distribution of the tree shape (for an exception, see Sainudiin and Véber, 2016) or to relate it to known distributions whenever it does not have the shape of the Yule tree (i.e., the tree generated by a pure-birth process). Also, since the same trait is usually responsible for both the tree shape and the order of nodes, it is impossible to disentangle the roles of either of these characteristics on the behavior of the tree in the face of current extinctions. Last, these models do not fulfill a very important property called sampling consistency (usually considered in combination with exchangeability, i.e., ecological equivalence between species). This property ensures that one can equivalently draw a random tree with n tips from the distribution or draw a tree with $n+1$ tips and then remove one tip at random.

The model we propose here has two parameters: $\beta\in(-2,+\infty)$ (tree balance index) determines the balance of the tree, similarly as in Aldous' β-splitting model (Aldous, 1996, 2001), and $\alpha\in(-\infty,+\infty)$ (age-richness index) sets the relation between the species richness of a clade and its relative age (Fig. 2).

The construction of a tree according to this model is done by following the steps indicated hereunder (illustrated on Fig. 1). We start with n uniform, independent random variables $(U_i)_{i\in\{1,\dots,n\}}$ in the interval $[0,1]$. For each i, the mark U_i is associated to the tip species labeled i in the phylogeny. The procedure consists in sequentially partitioning $[0,1]$ into a finite subdivision thanks to random variables independent of the marks $(U_i)_{i\in\{1,\dots,n\}}$, until all marks are in distinct components of the partition. At each step, the new point added to the subdivision corresponds to a split event in the tree. In the beginning, there is only one component in the partition (the interval $[0,1]$ itself).

1. Each interval X of the partition containing at least two marks among the $(U_i)_{i\in\{1,\dots,n\}}$ is given a weight equal to $|X|^\alpha$, where $|X|$ denotes the width of the interval X. One of these intervals is selected with a probability proportional to its weight.

2. Draw a random variable R in a Beta distribution with parameters $(\beta+1,\beta+1)$. The selected interval X of width $|X|$ is then split into two disjoint subintervals, X_{left} and X_{right}, with widths $|X_{\text{left}}|=R|X|$ and $|X_{\text{right}}|=(1-R)|X|$. Each subinterval contains a distinct subset of the marks. The

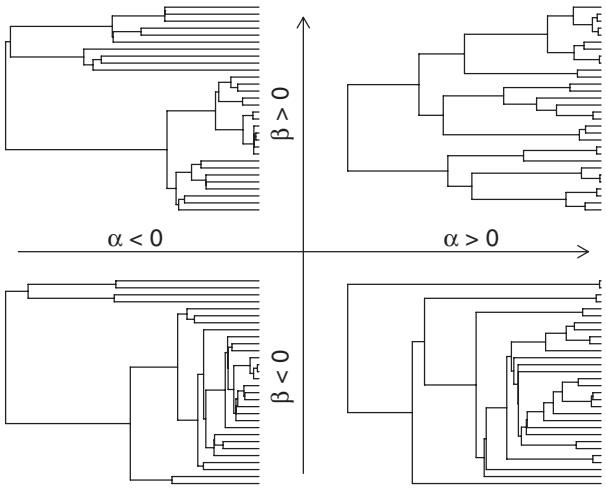


FIGURE 2. Phylogenetic trees simulated for different values of β (tree balance) and α (age-richness index). Node depths are set as in a Yule pure-birth process. Parameter values: $\beta=-1.5$ (bottom) or 10 (top), $\alpha=-10$ (left) or 10 (right), number of species $N=30$, approximation parameter (see Supplementary Appendix S1 available on Dryad) $\varepsilon=0.001$.

marks in the subinterval X_{left} determine the tips in the left subtree of the phylogeny, and the marks in the subinterval X_{right} determine the tips in the right subtree. This step is performed even if one subinterval contains no mark among the $(U_i)_{i\in\{1,\dots,n\}}$, which corresponds to a subtree with no sampled species (i.e., species which are seen in the phylogeny). The order in which the splitting subintervals are selected sets the order of branching events (i.e., nodes) in the tree.

3. If no interval contains more than one mark, the process is stopped. Otherwise, go to Step 1.

We can relate the tree shape in this model to well-known distributions. Because α has no impact on the way we refine the subdivision, the tree shape generated with our model coincides exactly with the tree shape obtained in Aldous' β-splitting model (Aldous, 1996, 2001) with the same parameter β. For small values of β, the intervals are often split close to one of their extremities, and the resulting tree is unbalanced, converging to the perfectly unbalanced "caterpilar" tree as $\beta\to-2$. On the contrary, for large values of β, the intervals are often split close to their middle, and the resulting tree is balanced. We stress that unlike most models, α can be tuned independently of β, allowing node ranks to vary while keeping the same tree shape. For small values of α (in particular $\alpha<0$), the smallest subintervals have a higher probability of being selected, so smaller clades tend to be older. On the contrary, for large values of α, the largest subintervals have a higher probability of being selected, so smaller clades tend to be younger. We notice that as β gets close to -2 the effect of α vanishes, since most of the time there is only one subinterval containing more than one mark, so only one subinterval to split. In maximally unbalanced tree shape ($\beta=-2$), there is only one ranked tree shape and the order of nodes is fixed, so α plays no role.

As is well-known, the tree obtained with $\beta=0$ has the same shape has the tree generated with the Yule process (Yule, 1925) or the Kingman coalescent (Kingman, 1982) after ignoring node ranks (Nee, 2006; Lambert and Stadler, 2013). When $\alpha=1$ in addition to $\beta=0$, we show in Supplementary Appendix S1 (Proposition 1) available on Dryad at http://dx.doi.org/10.5061/dryad.mv980, that our model generates the same tree shape with node ranks as Yule trees, which is actually known to be the same as the ranked tree shape of the Kingman coalescent tree.

The version of the model we present here only allows simulation of trees with $\beta>-1$, as the beta distribution is only defined for positive parameter values. Actually, our model coincides with the ranked tree in a self-similar, binary fragmentation with self-similarity index α (which is the very age-richness index α) and with fragmentation measure $\int_0^1 \delta_{(x,1-x,0,0,\dots)} x^{\beta+1}(1-x)^{\beta+1}dx$ (as defined in Bertoin, 2002, 2006), which makes sense as soon as $\beta>-2$. In Supplementary Appendix S1 (Proposition 3) available on Dryad, we present an algorithm based on fragmentation processes equivalent to that presented above (using one additional approximation parameter ε, the maximal frequency of unsampled clades with insignificant richness, consistently set to 0.001). Albeit less intuitive, this method allows us to simulate trees for all $\beta>-2$.

Last, it is important to notice that our model is both exchangeable and sampling consistent. It is exchangeable because labels can be swapped without changing the distribution of the tree, since marks all have the same distribution. It is sampling consistent because removing tip labeled $n+1$ (or any other tip, by exchangeability) amounts to removing mark U_{n+1}, which does not modify the ranked tree shape obtained from marks $(U_i)_{i\in\{1,\dots,n\}}$.

Incorporating Nonrandom Extinctions

In order to parametrize the relation between the richness of a clade and its relative abundance, we now add to the model a new parameter $\eta\geq 0$ called "abundance-richness index." Each time an interval X is split into two subintervals, X_{left} and X_{right} with widths $|X_{\text{left}}|=R|X|$ and $|X_{\text{right}}|=(1-R)|X|$, each of the two subtrees is granted a part of the relative abundance A_X of the parental clade equal to

$$A_{X_{\text{left}}}=\frac{|X_{\text{left}}|^\eta}{|X_{\text{left}}|^\eta+|X_{\text{right}}|^\eta}A_X=\frac{R^\eta}{R^\eta+(1-R)^\eta}A_X$$

$$A_{X_{\text{right}}}=\frac{|X_{\text{right}}|^\eta}{|X_{\text{left}}|^\eta+|X_{\text{right}}|^\eta}A_X=\frac{(1-R)^\eta}{R^\eta+(1-R)^\eta}A_X.$$

This way of allocating frequencies to taxa is reminiscent of the "broken stick model" (MacArthur, 1957; MacArthur and Wilson, 1967; Colwell and Lees, 2000),

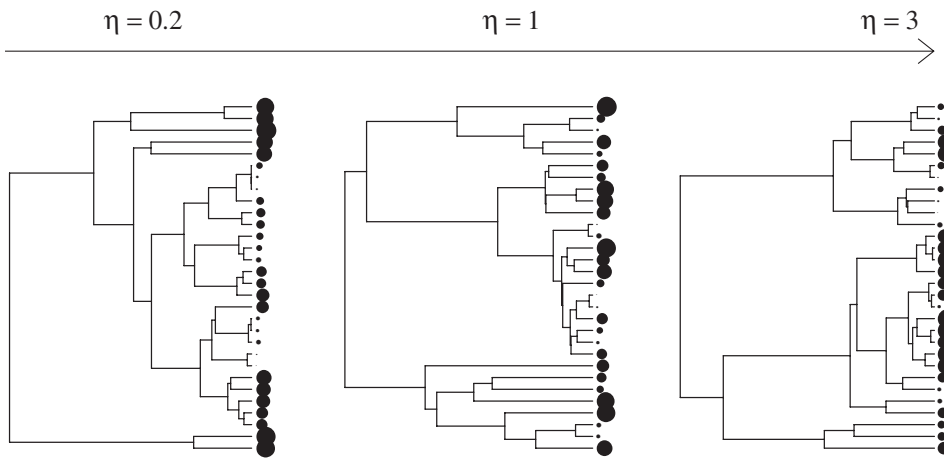


FIGURE 3. Distribution of species frequencies across the tips of phylogenetic trees for different values of abundance-richness index η. Dot sizes sort species according to their frequency (larger dots for more abundant species). Parameter values: $\eta = 0.2$, 1, or 3 (from left to right), $\beta = 0$, $\alpha = 0$, number of species $N = 30$, approximation parameter $\varepsilon = 0.001$. Results with $\beta = -1.9$ are shown in the Supplementary Appendix S1 available on Dryad.

where the unit interval is broken into subintervals each representing the frequency or resource share of each species or clade in the community, even though in our case the allocation of relative abundances has no mechanistic interpretation. This is usually done by throwing uniform points independently in the interval or by throwing the points sequentially, always to the right of the last one, leading to the Poisson–Dirichlet distribution appearing in mathematical population genetics (Feng, 2010; Ewens, 2012) as well as in the neutral theory of biodiversity (Hubbell, 2001).

The model remains sampling-consistent insofar as each A_X is interpreted as the relative abundance of an entire clade, that is the sum of relative abundances of all species, sampled or not (seen or not in the phylogeny), belonging to this clade. Sampling consistency now means that generating a ranked tree shape with relative abundances on n tips is equivalent to the following process: generate a ranked tree shape with relative abundances on $n+1$ tips, remove one tip at random and sum the abundance of the removed tip to that of its sister clade (i.e., the clade descending from the interior node connected to the removed tip by a pendant edge).

In the extinction numerical experiment, we determine the order of species extinctions deterministically based on their abundance: rarest species become extinct first, whereas most abundant species become extinct last. The distribution of relative abundances across species is captured in the abundance-richness index η (Fig. 3 and in the Supplementary Appendix S1 available on Dryad): when $\eta = 1$, all tip species have the same relative abundance in expectation ($\frac{1}{n}$), such that the correlation between the relative abundance of a clade and its richness would also be close to 1.0. When we simulate extinction with $\eta = 1$, all species have an approximately equal chance of being removed, and we approach a field of bullets model (see Proposition 2 in Supplementary Appendix S1 available on Dryad; in the case $\beta = -1$, the equivalence is exact). For $\eta > 1$, species in larger clades have larger abundances on average (and thus lower extinction risks), and for $\eta < 1$, species in larger clades have smaller abundances (and thus higher extinction risks) on average. This modeling approach allows us to tune the sign and strength of the relation between the richness of a clade and the extinction risk of its tip species.

Testing the Effect of β, α, and η on PD Loss

The effect of all three model parameters on the relationship between species loss and PD loss is studied in a systematic way by simulation. We considered values of β in $(-2, 10]$, values of α in $[-3, 3]$ and η in $[0.1, 3]$. Because our model specifies how interior nodes are ranked in time but not their actual timing, we use a pure-birth process to generate node depths, adding the latter on top of ranked tree shapes. The use of another model for generating node depths leads to qualitatively similar results, albeit quantitatively different (as an illustration, we show results with edge lengths set as in the Kingman coalescent in Supplementary Appendices S4 and S6 available on Dryad).

For each set of parameter values, we generated 100 trees with 100 tips ($N = 100$). We sequentially removed extinct species from these trees (in the order of increasing species abundances, as explained earlier), and computed the remaining PD (sum of all branch lengths; Faith, 1992) for increasing fractions of extinct species.

Parameter Inference

We inferred the parameters β, α, and/or η from simulated or empirical data sets by maximum

likelihood. As is already well-known (Aldous, 1996; Blum and François, 2006; Lambert et al., 2017), the likelihood of a labeled tree shape under Aldous' β-splitting model is explicit. Since the likelihood of the tree shape under our model is the same as in Aldous' model (and in particular independent of α and η), we can use it to estimate β. In contrast, computing the likelihood of the ranked tree shape requires to follow through time the lengths of all intervals of the partition containing marks, which may decrease without separating marks (unsampled species). Given that the likelihood of the ranked tree (with or without tip abundances) with the additional knowledge of interval lengths is explicit, we use a Monte-Carlo data augmentation procedure, in which the augmentation data are the numbers and sizes of unsampled splits on each branch (which allow us to reconstruct the interval lengths through time). The likelihood of the ranked tree with tip abundances is then computed by averaging over augmentations and is optimized over possible values of (α, η).

We first tested our ability to infer the model parameters on simulated trees. To do so we simulated trees with 20, 50 and 100 tips for all possible combinations of α in $\{-1, 0, 1, 2\}$, β in $\{-1, 0, 1\}$ and η in $\{0.2, 0.5, 1, 1.5, 2\}$. For each tree size and parameter combination, we simulated 20 trees with tip abundances, for a total number of 3600 trees.

We then inferred the model parameters on these trees and compared them to the values used in the simulations. The inference of the parameter β was straightforward, being computed as the maximum likelihood estimate on the interval $[-2, 10]$ with the function `maxlik.betasplit` from the R-package `apTreeshape` (Bortolussi et al., 2006). The parameters α and η were estimated with the method introduced hereabove, with values respectively constrained in the intervals $[-4, 4]$ and $[0.1, 10]$. The value of ε (approximation parameter, see Supplementary Appendix S1 available on Dryad) was here again fixed to 0.001.

After validating the estimation procedure, we applied it to real bird family trees. We used the Maximum Clade Credibility (MCC) tree from Jetz et al. (2012), and pruned it to keep family level phylogenies. We kept only the phylogenies that included at least 50 species, and used range sizes from Map of Life (https://mol.org/) as proxies for relative abundances. The value of ε and the constraints on parameter ranges were here the same as in the test on simulated phylogenies.

The model was coded—and the analyses of phylogenetic trees were performed—using `R` (R Development Core Team, 2012) and the R packages `cubature` (Johnson and Narasimhan, 2013), `ape` (Paradis et al., 2004), `sads` (Prado et al., 2015), `apTreeshape` (Bortolussi et al., 2006), and `picante` (Kembel et al., 2014). The code is available in the R-package `apTreeshape` (Bortolussi et al., 2006). The list of available functions is given in Supplementary Appendix S10 available on Dryad.

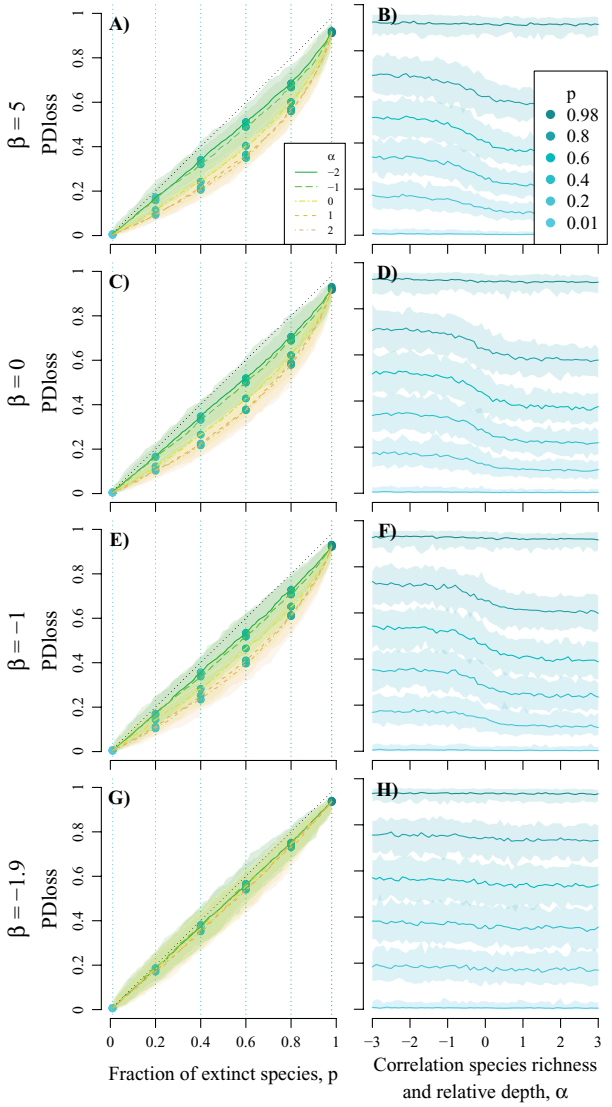


FIGURE 4. Influence of the ranked tree shape (tree balance β and age-richness index α) on PD loss, for increasing fractions of species extinctions. Tree balance β changes from 10 (top row, "bush trees") to -1.9 (bottom row, "caterpilar trees"). Results are shown either as a function of the extinction fraction p (left column; for different α values) or as a function of α (right column; for different extinction fractions p). Extinction fraction p increases from 0.01 to 0.98 (from left to right in a, c, e, g; from light to dark color in b, d, f, h). The dotted lines in a, c, e, g show the bisector. Results are based on 100 simulation replicates: plain lines give median values and light areas give 95% confidence intervals. Other parameter values: number of species $N = 100$, approximation parameter $\varepsilon = 0.001$.

RESULTS

Influence of Ranked Tree Shape on PD Loss

Here we only address the influence of α on PD loss, assuming a field of bullets model for species extinctions ($\eta = 1$). The expected PD loss is then a convex function of the fraction p of extinct species (as proved mathematically for any binary tree under the field of bullets model, see Eq (34) in Lambert and Steel, 2013), always lying below p (Fig. 4A,C,E,G).

Consistently with previous studies (Nee and May, 1997; von Euler, 2001) we find that when the relation between age and richness of clades is similar as that in Yule trees ($\alpha = 1$), very unbalanced trees (caterpilar-like trees) lose more PD in the face of species extinctions than Yule or more balanced trees (Fig. 4G,H vs. A–D, with $\alpha = 1$). The effect is nonlinear in β: the tree shape has little influence on the loss of PD when $\beta \geq -1$, but increases sharply as β decreases from -1 to -1.9 (results as a function of β in Supplementary Appendix S2 available on Dryad). Unbalanced tree shapes are associated with the presence of long edges leading to evolutionary distinct species (Fig. 2). These edges constitute an important fraction of the PD in unbalanced species trees, so that their extinction generates a significant drop in PD. As β gets closer to -2 (case of the "caterpilar tree"), the expected PD loss approaches the fraction of extinct species (Fig. 4G).

Considering ranked tree shapes shows, however, that the order of nodes has a significant influence on the loss of PD, and on the effect of β on this loss. If the age and richness of clades are positively correlated ($\alpha > 0$), the loss of PD is reduced, especially at intermediate extinction fractions (Fig. 4A–F). This is because the smallest subtrees, more prone to early extinction, are younger and hence contain a lower fraction of the PD (Fig. 2). If the age and richness of clades are negatively correlated ($\alpha < 0$), the loss of PD rises, especially at intermediate extinction fractions. The smallest subtrees, prone to extinction, are older and hence contain more evolutionary distinct species (Fig. 2). This generates losses of PD similar to those observed when the tree shapes are very unbalanced (PD loss equal to the fraction of extinct species).

As expected, the effect of α is evened out in very unbalanced trees (β close to -2; Fig. 4G,H), for which the loss of PD remains close to its highest value whatever the value of α. In the case of the maximally unbalanced tree shape, there is only one ranked tree shape and the order of nodes is fixed.

All these effects of ranked tree shapes on the loss of PD are qualitatively conserved if node depths are distributed as in the Kingman coalescent (instead of the Yule process). In the case of Yule trees, PD loss slightly increases with the initial size of the tree, an effect which is due to more efficient sampling of large values in the common (exponential) distribution of node depths. Yet the results presented above are qualitatively conserved if the size of phylogenetic trees changes (analyses performed with number of species $N = 50$ and $N = 200$; see the Supplementary Appendices 3 and 4 available on Dryad).

Influence of Nonrandom Extinction Risks on PD Loss

The strength of the relation parameterized by η, between the richness of a clade and its relative abundance (here directly influencing the extinction risk of its species) may have a paramount influence on the loss of PD in the face of extinctions (Fig. 6). In trees with ranked tree shapes similar to Yule trees ($\beta = 0$, $\alpha = 1$), the concentration of high extinction risks in small clades ($\eta > 1$) increases the loss of PD, by promoting the extinction of entire clades (Fig. 5). In contrast, when extinction risks are higher in species-richer clades ($\eta < 1$), phylogenetic redundancy (and hence the likelihood of conserving at least one species per subtree) limits the loss of PD until high extinction levels.

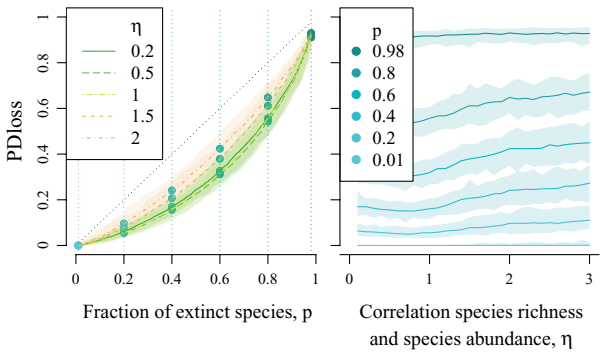


FIGURE 5. Effect of abundance-richness index η on PD loss in Yule trees, for increasing fractions of species extinctions p. Results are shown either a) as a function of the extinction fraction p (for different η values, with dotted lines showing the bisector) or b) as a function of η (for extinction fractions p increasing from 0.01 to 0.98 from light to dark color). Results are based on 100 simulation replicates: plain lines give median values and light areas give 95% confidence intervals. Parameter values: $\beta = 0$, $\alpha = 0$, number of species $N = 100$, approximation parameter $\varepsilon = 0.001$.

The effect of η is modified by the ranked shape of species trees. The strength of the age-richness relation (set by α) modulates the additional loss of PD induced by $\eta > 1$ (i.e., lower abundances in smaller clades; Fig. 6A–F). When $\alpha < 0$, smaller clades are not only more prone to extinction but also have deeper nodes, hence more evolutionary distinct species, which increases even further the loss of PD. Unlike in the field of bullets model, the expected PD loss as a function of the fraction p of extinct species can even change from convex to concave, and so take values larger than p (Fig. 6C,E). When $\alpha > 0$, smaller clades are more prone to extinction but have shallower nodes, which counteracts the increase of PD loss due to $\eta > 1$. To summarize, PD loss is increased when $\eta > 1$ compared to $\eta = 1$, with a maximal effect for negative values of α, progressively flattening as α grows.

We call the "danger zone" the region of parameters corresponding to the theoretical phylogenies that suffer a maximal rate of PD loss straight from the first few extinction events, that is, close to 1% of PD lost for the first 1% of species lost. In the plane (α, η), the "danger zone" corresponds to $\{\alpha < 0, \eta > 1\}$. As testified by Fig. 6, phylogenies in this zone can even suffer a rate of PD loss which is larger than 1:1 from the first extinction and sustains itself above 1:1 throughout the extinction crisis.

In contrast, α has little effect on the decrease in PD loss induced by $\eta < 1$ (i.e., higher abundances in small clades). Indeed, when $\eta < 1$, the deepest nodes are always

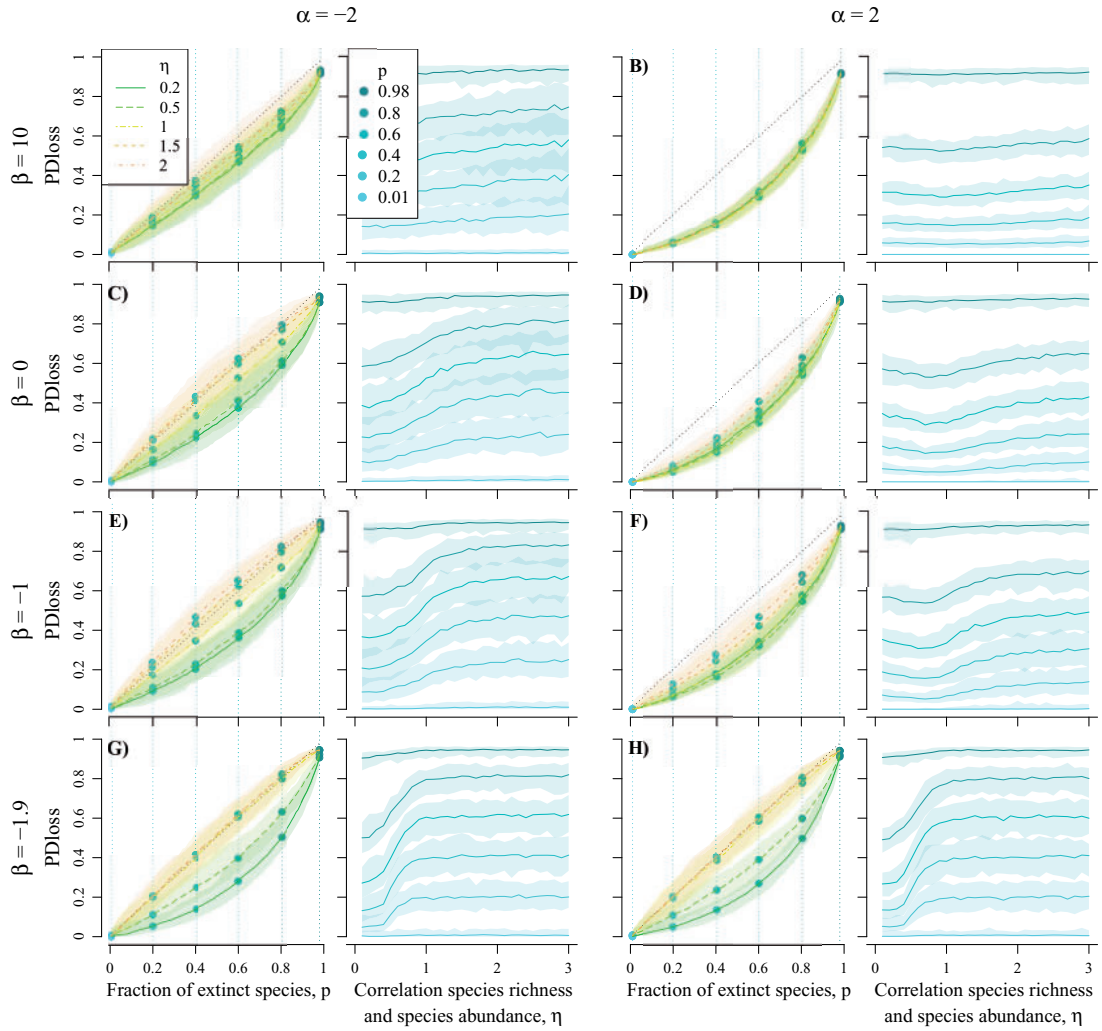


FIGURE 6. Effect of abundance-richness index η on PD loss, for different ranked tree shapes and increasing fractions of species extinctions. Tree balance β ranges from 10 (top row, "bush trees") to −1.9 (bottom row, "caterpilar trees"), and age-richness index α ranges from −2 (A, C, E, G) to 2 (B, D, F, H). Results are shown either as a function of the extinction fraction p (left side; for different η values, and with dotted lines showing the bisector) or as a function of η (right side; for extinction fractions p increasing from 0.01 to 0.98 from light to dark color). Results are based on 100 simulation replicates: plain lines give median values and light areas give 95% confidence intervals. Other parameter values: number of species $N = 100$, approximation parameter $\varepsilon = 0.001$.

protected regardless of the value of α: when $\alpha < 0$ the deepest nodes are in small clades which are protected from extinctions by the high relative abundances of its species (due to $\eta < 1$); when $\alpha > 0$, the deepest nodes are in large clades which are protected by phylogenetic redundancy.

The influence of η on PD loss is amplified by unbalanced tree shapes ($\beta < 0$; Fig. 6E–H) and buffered by balanced tree shapes ($\beta > 0$; Fig. 6A,B), because lower values of β enhance richness inequalities between clades and raise in turn the influence of η on PD loss. This interaction between parameters η and β overwhelms the influence of α (Fig. 6). In the plane (β, η), the "danger zone" is $\{\beta < -1, \eta > 1\}$ and the previous remark thus implies that in the 3D parameter space, the danger zone is $\{\alpha < 0$ or $\beta < -1; \eta > 1\}$.

Interestingly, the effect of β is highly dependent on how extinction risks are distributed within the phylogeny (Fig. 7, and results with other α values in Supplementary Appendix S7 available on Dryad). For $\eta = 1$, we recover the well-known pattern of decreased PD loss as the tree gets more balanced. However, for $\eta < 1$ we see the reverse pattern, that is PD loss increases with the balance of the tree. Recall that $\eta < 1$ buffers PD loss, because extinction risks are clustered in the species-richer clades which also display higher phylogenetic redundancy (smaller pendant edges). When the tree is maximally unbalanced, $\eta < 1$ causes the longest pendant edge to subtend the tip with the largest abundance (and hence to be the last to become extinct). Therefore, the order of extinctions coincides exactly with the increasing order of pendant edge lengths, which results in minimal

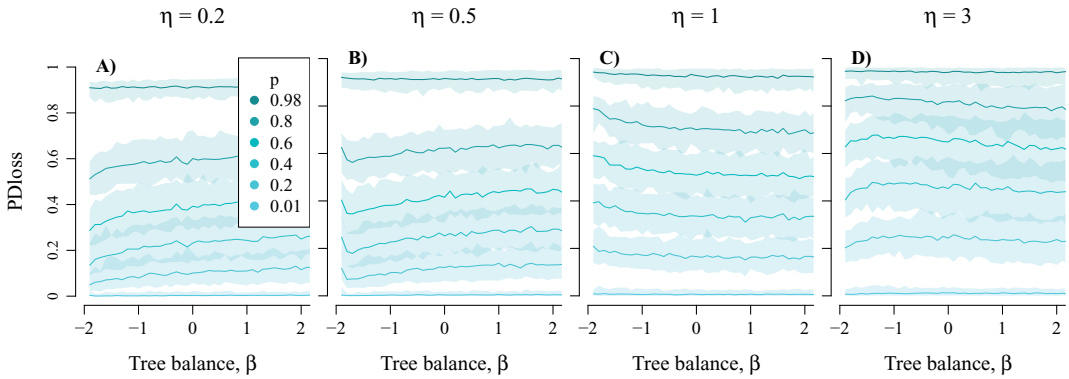


FIGURE 7. Effect of tree balance β on PD loss, for different values of clades abundance-richness indexes η, and increasing fractions of species extinctions p. The clades abundance-richness index η ranges from 0.2 (left) to 3 (right), and the extinction fraction p increases from 0.01 to 0.98 (from light to dark color). Results are based on 100 simulation replicates: plain lines give median values and light areas give 95% confidence intervals. Other parameter values: clades age-richness index $\alpha = -2$, number of species $N = 100$, approximation parameter $\varepsilon = 0.001$.

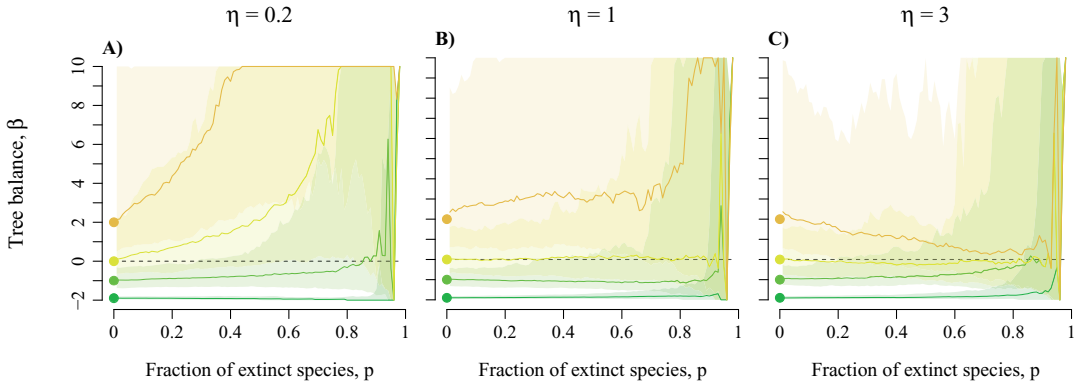


FIGURE 8. Effect of abundance-richness index η on the balance of phylogenetic trees after extinctions (Maximum Likelihood Estimate $\hat{\beta}$ of β). Initial tree balance β ranges from 10 (brown dots and lines, "bush trees") to −1.9 (green dots and lines, "caterpilar trees"). Extinction fraction p increases from 0.01 to 0.98 (from left to right). Results are based on 100 simulation replicates: plain lines give median values and light areas give 95% confidence intervals. Other parameter values: number of species $N = 100$, approximation parameter $\varepsilon = 0.001$, $\alpha = 0$.

PD loss for any given level of extinction. In a more balanced phylogeny, the distribution of clade sizes is more even and the buffering effect of the clustered extinction on PD loss is reduced.

For $\eta > 1$, we again recover the well-known pattern of decreased PD loss with increasing β. However, when we also have $\alpha < 0$, the relationship between PD loss and β is not monotonic, that is for any particular level of extinction, the maximal PD loss is reached for trees with intermediate balance. Recall that $\alpha < 0$ causes small clades to be relatively older and so to contribute more to PD. The maximal loss of PD thus occurs when extinction risks cluster in small clades. And indeed, when $\eta > 1$, at each splitting event the species-richer subtree gets a bigger abundance than the species-poorer subtree. However, within a given clade, the abundance of a species should decrease with the number of nodes (splitting events) on its lineage. This latter effect is stronger in unbalanced trees; in balanced trees, extinction risks cannot cluster in small clades, due to the absence of small clades. Trees with intermediate balance do display small clades, and these small clades are large enough to share their low abundance ($\eta > 1$) into a few species with very low abundance. These species go extinct first, resulting in maximal PD loss.

Effect of Species Extinctions on Tree Shape

We study the effect of species extinctions on tree shape, seeking in particular to check if the influence of η on the patterns of PD loss can be explained by changes in tree shape as species become extinct. Figure 8 shows the balance (defined here as the maximum likelihood estimate $\hat{\beta}$ of the parameter β) of the species tree estimated after a fraction p of its species have become extinct. When $\eta = 1$, tree balance is very little altered by extinctions except in very balanced trees, as predicted by the sampling consistency of the model ($\eta = 1$ amounts to removing species at random except when $\beta \gg 1$, see

Supplementary Appendix S1 available on Dryad). When $\eta < 1$, trees tend to become more and more balanced as p increases ($\hat{\beta}$ increases with p), whereas when $\eta > 1$ trees tend to become more and more similar to Yule trees ($\hat{\beta} \to 0$ as $p \to 1$). The effect of η on PD loss cannot be reduced to its effect on changes in tree shape due to extinctions. On the one hand, η mostly affects the shape of trees with $\beta > -1$ (Fig. 8), whereas tree shape has most effect on PD loss when β varies between -2 and -1 (Fig. 4A,C,E with $\alpha = 0$). In addition, if the effect of η on tree shape had a significant influence on PD loss, $\eta > 1$ should increase this loss when $\beta > 0$ (by decreasing the balance of trees; Fig. 8D) and decrease it when $\beta < 0$ (by increasing the balance of trees). Yet, the changes we observe in the effect of $\eta > 1$ on PD loss for different β values are the reverse of this prediction. Therefore, the indirect effects of η (through changes in tree shape) are negligible compared to its direct effects (through nonrandom distribution of extinction risks).

As reported above results on the effects of nonrandom extinctions on the loss of PD are conserved when node depths are distributed as in the Kingman coalescent, or when the size of phylogenetic trees changes (analyses performed with $N = 50$ and $N = 200$; see Supplementary Appendices S5 and S6 available on Dryad).

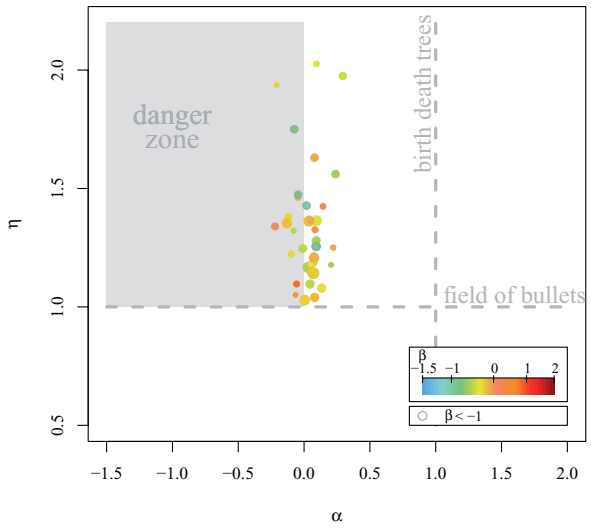


FIGURE 9. Inferred model parameters on bird family trees of 50 tips or more. Maximum posterior estimate of age-richness index α (x-axis), maximum posterior estimate of abundance-richness index η (y-axis) and maximum likelihood estimate of tree balance β (point color). Point sizes are proportional to the number of tips in trees, N. The dashed vertical line shows the value of α for trees generated by a birth–death model, and the dashed horizontal line shows the value of η for which extinction probabilities are distributed within the tree as in a field of bullets model. For all inferences, approximation parameter ε was set to 0.001.

Parameter Inference

When tested against simulated data, the Monte-Carlo inference algorithm by data augmentation performs reasonably well on phylogenies with more than 50 tips for a wide range of parameters (see Supplementary Appendix S8 available on Dryad). As expected, the estimation of β on trees with at least 50 tips is accurate, since the likelihood formula of the unranked tree is explicit, and this accuracy increases as β decreases. The inference algorithm also returns overall good estimates of η and α whenever $\eta > 0.3$.

The inference of α is unbiased except in the cases where $\beta < 0$ and $\eta < 0.3$. This corresponds to cases where the unsampled clades are numerous because β is small, and they have a strong impact on the reconstruction of intervals because η is small. The inferred η is overestimated for trees with only 50 tips. For $\beta < 0$ and $\alpha \geq 0$, η is slightly overestimated whatever the tip number. For $\beta > 0$ and $\alpha \leq 0$, inferences are good for trees with at least 100 tips.

Empirical Values

Estimates of parameter values on real data shows consistent patterns across all bird family trees. Unsurprisingly, we find negative β values, mostly comprised between 0 and -1, corresponding to unbalanced trees (see Supplementary Appendix S9 available on Dryad). Since the estimation of β is quite accurate for low true values of β and is biased towards larger estimates than the true value otherwise, these estimates can be taken with confidence. The estimates of η vary between 1 and 1.5. This indicates that, within bird families, species in small clades tend to have smaller range sizes than species in larger clades. The above study showed that low η values can be difficult to detect in unbalanced trees. Yet when this is the case, η is found to be close to the maximal value allowed in the inference (here 10), which is not the case here. We can therefore be confident that these values do not reflect a bias in the inference, but reflect a true pattern in the distribution of range sizes within the phylogenies. Finally, the estimates of α are clustered around 0, indicating that there is no correlation between clade sizes and clade depths within each bird phylogeny. This in contrast with what is expected in most explicit models of diversification, where larger clades take more time to diversify, resulting in a strong positive correlation between the depth and the size of clades.

When jointly inferring α and η the choice to use range size to infer η is likely to have an impact on the inferred α (because the values of the intervals are reconstructed using tip values, inappropriate tip values would lead to incorrect α). Therefore, we also ran the inference of α: we find fairly similar results between values obtained with the inference of α only compared to the full inference (the median of the inferred α for trees with at least 50 tips is 0.19 when α is inferred alone and 0.05 when both α and η are inferred), indicating that tree shape is indeed driving the result (Supplementary Appendix S9 available on Dryad equivalent as Fig. 9 with the α inferred without knowledge of the tip range sizes).

Discussion

A New Integrative Measure of the Age-Richness Relation, α

We introduced here a new model for random ranked tree shapes with a fixed, arbitrary number of tips. This model features two parameters, β and α tuning respectively the balance of the tree and the relative depths or ages of its nodes. Trees with $\beta \leq 0$ are unbalanced and trees with $\beta > 0$ are balanced. Whatever the value of α, the shape of the tree is the same as in Aldous' β-splitting model (Aldous, 1996, 2001). Large clades coalesce deep in the tree when $\alpha > 0$ and are shallower than smaller clades when $\alpha < 0$. When $\beta = 0$ and $\alpha = 1$, the tree has the same ranked shape as the Kingman coalescent and the Yule tree. Our model does not explicitly incorporate biological processes but enables to generate a broad range of tree shapes by decoupling tree balance and the relation between age and size of clades. In addition, this model is the first model (except the two aforementioned models and the trivial case of the "caterpilar tree") for ranked tree shapes satisfying sampling-consistency, in the sense that a tree with n tips has the same distribution as a tree with $n+1$ tips with one tip removed at random. This property is essential to ensure the robustness of the model with respect to incomplete taxon sampling (Heath et al., 2008; Cusimano et al., 2012; Stadler, 2013).

Predictions from this model highlight the importance of accounting for node ranks to understand forthcoming changes in macroevolutionary patterns of PD. They show in particular that the relationship between the species richness of a clade and its relative depth in the tree, set by parameter α in the model, can have profound impacts on the rate of PD loss (Fig. 4). This parameter α constitutes a new index quantifying the age-richness relation of subclades within a phylogeny. A large number of studies have already considered the age-size correlation, assessing its existence (significance, sign, and pattern) across multiple phylogenetic trees, based on one value of species richness and crown or stem age per phylogeny (e.g., Magallon and Sanderson, 2001; Bokma, 2003; Ricklefs, 2006; McPeek and Brown, 2007; Rabosky et al., 2012; Sánchez-Reyes et al., 2016). These studies notably aimed at testing the hypothesis of time-limited diversity patterns, versus hypotheses of diversity set by diversification rates or by limits to diversity (McPeek and Brown, 2007; Ricklefs, 2007; Rabosky, 2009; Barraclough, 2010; Sánchez-Reyes et al., 2016). Our new index α is different in that it can be measured by maximizing the likelihood on a single phylogeny, implicitly integrating over all subclades of this phylogeny. An interesting consequence is that one does not have to choose which clades to include in the analysis. For example, α is not sensitive to the definition of higher taxa (Stadler et al., 2014; Sánchez-Reyes et al., 2016). Moreover, similarly to what can be done with the index β (compared to other measures of tree balance; Kirkpatrick and Slatkin, 1993; Aldous, 1996, 2001), a measure of the age-richness relation in a given phylogeny is provided by the maximum likelihood estimate of the model-based parameter α. Last, we stress that our model does not require the precise knowledge of node datings in the phylogeny but only the relative positions of nodes in time, which preserves α estimates from the inaccuracies of time calibrations (Kumar, 2005; Welch and Bromham, 2005; Pulquério and Nichols, 2007; Forest, 2009; Schwartz and Mueller, 2010).

Ranked Tree Shapes and the Loss of PD

Our results confirm that in the field of bullets model very unbalanced trees undergo stronger loss of PD than balanced trees, under equal fraction of species extinctions. This property was already well known (Nee and May, 1997) but is important to recall given the predominance of unbalanced phylogenetic trees in nature (β values being often close to −1; e.g., Guyer and Slowinski, 1991; Heard, 1992; Guyer and Slowinski, 1993; Slowinski and Guyer, 1993; Mooers, 1995; Purvis, 1996; Mooers and Heard, 1997; Blum and François, 2006). However, our results also show that the temporal order of nodes among subtrees (set by the parameter α) may have even stronger effects than tree balance (set by parameter β; compare the effect of the latter in Supplementary Appendix S8 available on Dryad to that of α on Fig. 4). Besides, α values below 0 cause drops of PD almost as abrupt as those observed with "caterpilar" shapes (β close to −2, with $\alpha = 1$; Fig. 4D,H). It is therefore essential to consider the ranked shapes of species trees to understand the expected patterns of loss of PD.

Values of α deviating from 1 may arise from differences in stages of diversification among subtrees, resulting from heterogeneity in biotic or abiotic factors acting on diversification processes in different parts of the species tree. This could be due to bursts of diversification in certain subtrees (e.g., following from key innovations or from migration to empty spatial or ecological space), either recently (resulting in $\alpha < 0$) or early in the history of clades (resulting in $\alpha > 0$). Alternatively, α values deviating from 1 could be linked to changes in extinction rates in distinct parts of the tree (e.g., due to changes in the biotic or abiotic environment of phylogenetically related species sharing similar ecological niches). Age-dependent speciation (Venditti et al., 2010; Hagen et al., 2015) and extinction (Pearson, 1992; Alexander et al., 2015) are also likely to make node ranking deviate from what is expected in a homogeneous birth–death model. Heterogeneity in diversification rates across the species tree associated with asymmetric competition among species (e.g., evolutionary advantage to previously established species) could limit diversification in younger subtrees, hence leading to $\alpha > 0$. Last, α can be found negative due to the presence of relictual lineages, that is, old clades harboring few species surviving to the present.

Modeling Nonrandom Extinctions: η and the Loss of PD

The incorporation of the abundance-richness index η within the framework provided by Aldous' β-splitting model allowed us to go beyond the field of bullets assumption. The model allows one to simulate a trait covarying with the shape of the phylogeny. We have so far interpreted this trait as species relative abundance or relative range size, but it could be any species trait with values summing to 1 at the level of the phylogeny. In particular, the trait value of a species can be interpreted as a probability of sampling this species, which is consistent with the initial interpretation of the trait as a relative abundance. In passing, we devised a model of abundance distributions (equivalently interpreted as range size distributions) covarying with the phylogeny, in the broken-stick tradition (MacArthur, 1957; MacArthur and Wilson, 1967). When $\eta > 1$, the most abundant species are in species-rich clades whereas when $\eta < 1$ the most abundant species are in species-poor clades. When $\eta = 1$ all species have the same abundance on average. Here, extinctions are assumed to occur sequentially in the order of increasing abundances. By doing this, we only consider the part of extinction risks that is due to the rarity of a species. In nature, relative extinction risk indeed depends on species abundance, but also on many other features (e.g., dynamics of population growth or decline, fragmentation into subpopulations, biotic or abiotic changes; IUCN, 2012), and may have a significant stochastic component. The simple framework we use to determine extinctions allows us to focus on the direct impact of the distribution of ranked abundances within trees on the loss of PD. This framework can easily be modified to include extrinsic causes of extinctions by taking extinction risk to be a function of abundance and additional known factors.

Previous studies concluded that PD loss is increased if extinction risks are clustered in the phylogeny (Davies and Yessoufou, 2013), but that this effect is not substantial (Parhar and Mooers, 2011). Our model shows that the effect on PD loss depends on the way these extinction risks are distributed among clades: PD loss is increased by $\eta > 1$ (i.e., higher extinction risks in small clades; Fig. 6). Such a distribution of extinction risks may arise from subtrees having low species richness because of higher extinction rates, either due to intrinsic factors (species features that would make them more susceptible to extinction; e.g., long generation time, or low variance or phenotypic plasticity of key ecological traits providing resistance to perturbations or evolutionary advantages in relation to biotic interactions; Purvis et al., 2000c; Johnson et al., 2002) or to extrinsic factors (threats affecting the spatial or ecological space shared by species of the subtree; e.g., Russell et al., 1998; Hughes, 1999; Purvis et al., 2000c; von Euler, 2001; Johnson et al., 2002). Higher extinction risks in small subtrees could also be due to resource limitation affecting simultaneously the density of individuals and the diversity of species, and hence demographic stochasticity (the island effect, Mooers et al., 2009); or to stabilizing selection (e.g., due to competition or to the absence of available spatial or ecological space in the surrounding environment), limiting adaptation and increasing species vulnerability in the face of perturbations (Purvis et al., 2000c; Purvis, 2008).

In contrast, $\eta < 1$ buffers the loss of PD. Higher extinction risks in larger subtrees could result from a trade-off between species richness and average species abundance, provided constrained metacommunity size (with variation along this trade-off following for instance from landscape structure and dynamics, such as geographical isolation affecting the occurrence of allopatric speciation events), from recent speciation events associated with a decrease in average species abundance, geographical range or niche width, or from recent extinction events that removed the most extinction-prone species from certain clades (leaving the latter smaller and with less extinction-prone species; Schwartz and Simberloff, 2001; Lozano and Schwartz, 2005).

Hence, η is expected to vary across clades according to the metacommunity structure and the underpinning diversification dynamics. Given its striking effects on PD loss, this factor should also be accounted for to understand potential future losses of PD.

Combined Effects of β, α, and η: Reversing Some Expected Patterns of PD Loss

The influence of η on the loss of PD is enhanced by $\alpha < 0$ (small clades containing evolutionary distinct species) and $\beta < 0$ (more variability in clades richness) (Fig. 6). However, a stronger clustering of extinction risks does not necessarily lead to higher loss of PD (e.g., if extinctions occur first in species-richer subtrees—which contain more phylogenetic redundancy—as in the case when $\alpha > 0$ and $\eta < 1$).

These interactions between the effects of β, α, and η may reverse two well-known patterns of variation in the loss of PD (Nee and May, 1997). First, the increase in PD loss with tree imbalance can be hampered by η values deviating from one (Fig. 7 and Supplementary Appendix S7 available on Dryad). In particular when $\eta < 1$, this pattern results from the preferential extinction of phylogenetically redundant species in more unbalanced trees when extinction risks are clustered in large clades. Second, when $\eta > 1$ and $\alpha < 0$ the loss of PD proceeds faster than that of species diversity (turning their relationship from convex to concave, except in very balanced or very unbalanced trees; Fig. 6C,E). This pattern (also highlighted in Heard and Mooers, 2000) is caused by the preferential extinction in small subtrees containing evolutionary distinct species. The only other cases where such high loss of PD is reached is when $\beta < -1$ and $\eta > 1$. This parameter zone is a macroevolutionary danger zone, where particular phylogenetic shapes combine with clustered patterns of extinction probabilities to produce large and rapid losses

of evolutionary history, much like ice and wind bringing down thick branches of trees in winter. In this region of the parameter space ($\beta < -1$ or $\alpha < 0$; $\eta > 1$), phylogenies are prone to a sudden collapse of PD.

Loss of PD in Bird Family Phylogenies

Our inference study shows that the phylogeny of bird families tend to exhibit β values comprised between -1 and 0. A similar result was found in many macroevolutionary studies, commonly observing values of β clustering around -1 in real phylogenies (e.g., Guyer and Slowinski, 1991; Heard, 1992; Guyer and Slowinski, 1993; Slowinski and Guyer, 1993; Mooers, 1995; Purvis, 1996; Mooers and Heard, 1997; Blum and François, 2006). With these topologies, we expect both α and η to play a major role in determining the potential losses of PD (Fig. 6E,F).

We observed α values clustering around zero, consistently with several empirical studies that found no positive relation between clade age and clade size (Ricklefs, 2007, 2009; Rabosky et al., 2012). These values contrast with the value of 1 expected in Yule trees, and make phylogenies very sensitive to PD loss.

Our estimates of η values, based on the distribution of range sizes in bird family phylogenies, all fall between 1 and 1.5. This indicates that species in small clades tend to have smaller ranges than species in bigger clades. Range size has been shown to be one of the most important correlates of extinction risks and is one of the IUCN red list criteria (Purvis et al., 2000b; Cardillo et al., 2006; Lee and Jetz, 2011; IUCN, 2012; Arbetman et al., 2017). Expectedly, range sizes in our data set are significantly different among IUCN status classes (see Supplementary Appendix S9 available on Dryad). This pattern of extinction risks clustering in species-poor clades, highlighted in our data by η values above 1, has also been shown in plants (Vamosi and Wilson, 2008), and in birds and mammals for past extinctions (Russell et al., 1998; Mooers et al., 2009).

Considering the three parameters together, we find that bird family trees are situated close to the region of the parameter space termed "danger zone," for which we find the loss of PD to be at least as fast as the loss of species diversity. In particular, the combination of negative α values with $\eta > 1$ leads to higher extinction risks for evolutionary distinct species. We can expect such a pattern as a result from evolutionary mechanisms acting simultaneously on different features of trees. For example, subtree-specific susceptibility to extinction, or stabilizing selection generating relictual lineages, are both expected to beget small subtrees with high divergence times also endowed with high species extinction risks. This pattern has been already found for past extinctions in birds using a species level measure of evolutionary distinctiveness; authors observed in that case a similar loss of species and PD (von Euler, 2001; Szabo et al., 2012). Evolutionary distinct bird lineages were also shown to be more threatened by agricultural expansion and intensification than more recent lineages in Costa Rica (Frishkoff et al., 2014). This was also found in other taxa, such as marsupial mammals (Johnson et al., 2002) and *Sebastes* (Magnuson-Ford et al., 2009) (but see Davies et al. 2011, who show that in South African plants, extinction risks cluster in younger, fast-diversifying genera).

A striking result of our inference study relates to the narrow range of α values obtained as soon as the trees are large enough for the inference to be accurate (see Supplementary Appendix S9 available on Dryad for inferred parameter values as a function of the tip number in the phylogenies). This value, which differs from what is found in birth–death models, adds a new conundrum concerning the shape of empirical trees.

Branch Lengths in Empirical Phylogenies

The parameter α of the model shapes the order in which speciations take place, but does not instantiate the actual times between two consecutive speciation events, that is, edge lengths. In the numerical investigations of PD loss, we considered two models for edge lengths: the pure-birth process (Yule, 1925) and the Kingman coalescent (Kingman, 1982). Using either of these models did not affect our results qualitatively, but affected them quantitatively (compare Figs. 4 and 6 to figures provided in Supplementary Appendices S4 and S6 available on Dryad). Our modeling framework allows easy exploration of predictions under different models of edge lengths. This is interesting as many empirical phylogenies are not time-calibrated, or imprecisely. Besides, empirical phylogenetic trees were shown to often exhibit a decrease in branching tempo, that is, in the rate of lineage accumulation through time (characterized in particular by estimates of the statistic $\gamma < 0$; e.g., Nee et al., 1992; Zink and Slowinski, 1995; Lovette and Bermingham, 1999; Pybus and Harvey, 2000; Rüber and Zardoya, 2005; Kozak et al., 2006; Seehausen, 2006; Weir, 2006; McPeek, 2008; Phillimore and Price, 2008; Rabosky and Lovette, 2008a; Jønsson et al., 2012; Moen and Morlon, 2014). Hence, quantitative predictions on the loss of phylogenetic diversity in the face of species extinctions could be further increased by accounting for real branch lengths. Moreover, several theoretical studies suggested that the branching tempo of species trees may change with clade age, decreasing in particular in younger clades (the "out of equilibrium" hypothesis, proposed to explain the negative values of γ often observed in real phylogenies; Liow et al., 2010; Gascuel et al., 2015; Manceau et al., 2015; Missa et al., 2016; Bonnet-Lebrun et al., 2017). Taking into account such correlations between the age of clades and their branching tempo would also affect the expected loss of PD.

The EDGE program ("Evolutionary Distinct and Globally Endangered"; Isaac et al., 2007) encourages conservation priorities aiming at preserving most evolutionary history within the tree of life, by proposing a ranking of species based on combined criteria of evolutionary distinctiveness and extinction

risk. Although our approach is not species-based but clade-based, it also investigates the preservation of evolutionary history based on principles linked to species evolutionary distinctiveness (related to the depths of subtrees, which depend on α) and to the distribution of extinction risks in the tree (which depends on η). Accordingly, to conserve most evolutionary history and evolutionary potential for further diversification and/or survival, priority could be given to clades that would undergo higher loss of PD in the face of species extinctions, that is, clades in the danger zone ($\eta > 1$ and either $\beta < -1$ or $\alpha < 0$), and although not shown but only discussed herein, with $\gamma < 0$ (decreasing branching tempo; Pybus and Harvey, 2000).

Beyond Losses of PD

As we have seen earlier, the trait value of a species initially interpreted as its relative abundance can also be seen as a probability of sampling it. So an additional tool provided by the parameter η is a sampling distribution where sampling is not independent of the phylogeny. In particular, our results could be interpreted in the light of rarefaction experiments (Nipperess and Matsen, 2013), which study the way phylogenetic patterns in a metacommunity change as sampling decreases. Previous studies already pointed out strong impacts of nonrandom taxon sampling on the macroevolutionary patterns that we observe (e.g., Cusimano and Renner, 2010). Our results provide insights on the effects of nonrandom sampling on PD and phylogenetic tree topology. They reveal how, when the rarest species are not sampled, the discrepancy between observed and real PD depends on tree balance and abundance-richness relation (this discrepancy being larger in particular in the danger zone; Fig. 6E); and how the discrepancy between observed and real tree shape depends on η (real trees being more unbalanced if $\eta < 1$, and diverging from Yule trees towards more balance or more unbalance if $\eta > 1$; Fig. 8). These effects of incomplete sampling on macroevolutionary patterns should be particularly important to understand biodiversity patterns in bacterial and archeal phyla, which remain poorly known in particular because they likely harbor rare species having high chances to remain unnoticed.

Conclusion

This new stochastic model of phylogenetic trees spans a large range of binary trees endowed with node rankings and species abundances/range sizes/extinction risks, based on three parameters only and interpolating other well-known models. We showed that ranked tree shapes, nonrandom extinctions and the interactions thereof, may have a strong impact on the loss of PD in the face of species extinctions, potentially reversing some expected patterns of variation in PD. The simplicity of the model allows one to infer the parameters on empirical phylogenies. Applying our inference procedure on bird family phylogenies we found that, in this data set, the parameters fall within a narrow range of the parameter space; and that the inferred values make the PD of these trees very sensitive to species extinctions.

Supplementary Material

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.mv980.

Acknowledgments

The authors are very grateful to Mike Steel, Ana S.L. Rodrigues, and Hélène Morlon for their feedbacks on an earlier version of this manuscript. They are indebted to Arne Mooers and to Walter Jetz for providing the data on bird range sizes. They wish to thank the Associate Editor and the referees for their careful reviews. They thank the *Center for Interdisciplinary Research in Biology* (Collège de France, CNRS) for funding.

References

Aldous D. 1996. Probability distributions on cladograms. In: Aldous, D., Pemantle, R., editors. Random discrete structures. New York: Springer. p. 1–18.
Aldous D. 2001. Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. Stat. Sci. 16:23–34.
Alexander H.K., Lambert A., Stadler T. 2015. Quantifying age-dependent extinction from species phylogenies. Syst. Biol. 65:35–50.
Alfaro M., Santini F., Brock C., Alamillo H., Dornburg A., Rabosky D., Carnevale G., Harmon L. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proc. Natl. Acad. Sci. USA 106:13410–13414.
Arbetman, M.P., Gleiser G., Morales C.L., Williams P., Aizen M.A. 2017. Global decline of bumblebees is phylogenetically structured and inversely related to species range size and pathogen incidence. Proc. Roy. Soc. Lond. B 284:20170204.
Baillie, J.E.M., Hilton-Taylor C., Stuart S.N. 2004. A global species assessment. Gland, Switzerland: IUCN.
Barnosky, A.D., Matzke N., Tomiya S., Wogan G.O.U., Swartz B., Quental T.B., Marshall C., McGuire J.L., Lindsey E.L., Maguire K.C., Mersey B., Ferrer E.A. 2011. Has the Earth's sixth mass extinction already arrived? Nature 471:51–57.
Barraclough, T.G. 2010. Evolving entities: towards a unified framework for understanding diversity at the species and higher levels. Philos. Trans. R. Soc. Lond. B 365:1801–1813.
Bennett, P.M., Owens I.P.F. 1997. Variation in extinction risk among birds: chance or evolutionary predisposition? Proc. R. Soc. Lond. B. 264:401–408.
Bertoin, J. 2002. Self-similar fragmentations. Ann. I.H. Poincaré 38:319–340.
Bertoin, J. 2006. Random fragmentation and coagulation processes. Cambridge: Cambridge University Press.
Bielby, J., Cunningham A.A., Purvis A. 2006. Taxonomic selectivity in amphibians: ignorance, geography or biology? Anim. Conserv. 9:135–143.
Blum, M., François O. 2006. Which random processes describe the tree of life? A large-scale study of phylogenetic tree imbalance. Syst. Biol. 55:685–691.
Bokma, F. 2003. Testing for equal rates of cladogenesis in diverse taxa. Evolution 57:2469–2474.
Bonnet-Lebrun, A.-S., Manica A., Eriksson A., Rodrigues A.S. 2017. Empirical phylogenies and species abundance distributions are consistent with preequilibrium dynamics of neutral community models with gene flow. Evolution 71:1149–1163.

Bortolussi N., Durand E., Blum M., François O. 2006. apTreeshape: statistical analysis of phylogenetic tree shape. Bioinformatics 22:363–364.
Cardillo, M., Mace G.M., Gittleman J.L., Purvis A. 2006. Latent extinction risk and the future battlegrounds of mammal conservation. Proc. Roy. Soc. Lond. B 103:4157–4161.
Cardillo M., Orme C.D.L., Owens I.P.F. 2005. Testing for latitudinal bias in diversification rates : an example using New World birds. Ecology 86:2278–2287.
Colwell R.K., Lees D.C. 2000. The mid-domain effect: geometric species richness. Trends Ecol. Evol. 15:70–76.
Cusimano N., Renner S.S. 2010. Slowdowns in diversification rates from real phylogenies may not be real. Syst. Biol. 59:458–464.
Cusimano N., Stadler T., Renner S.S. 2012. A new method for handling missing species in diversification analysis applicable to randomly or nonrandomly sampled phylogenies. Syst. Biol. 61:785–792.
Davies T.J., Smith G.F., Bellstedt D.U., Boatwright J.S., Bytebier B., Cowling R.M., Forest F., Harmon L.J., Muasya A.M., Schrire B.D., Steenkamp Y., van der Bank M., Savolainen V. 2011. Extinction risk and diversification are linked in a plant biodiversity hotspot. PLoS Biol. 9:1–9.
Davies T.J., Yessoufou K. 2013. Revisiting the impacts of non-random extinction on the tree-of-life. Biol. Lett. 9:20130343.
Doran N.A., Arnold A.J., Parker W.C., Huffer F.W. 2006. Is extinction age dependent? Palaios 21:571–579.
Ewens W.J. 2012. Mathematical population genetics 1: theoretical introduction. New York: Springer Science & Business Media.
Faith D.P. 1992. Conservation evaluation and phylogenetic diversity. Biol. Conserv. 61:1–10.
Faller B., Pardi F., Steel M. 2008. Distribution of phylogenetic diversity under random extinction. J. Theor. Biol. 251:286–296.
Feng S. 2010. The Poisson–Dirichlet distribution and related topics: models and asymptotic behaviors. Heidelberg: Springer Science & Business Media.
Forest F. 2009. Calibrating the tree of life: fossils, molecules and evolutionary timescales. Ann. Bot. 104:789–794.
Frishkoff L.O., Karp D.S., M'Gonigle L.K., Mendenhall C.D., Zook J., Kremen C., Hadly E.A., Daily G.C. 2014. Loss of avian phylogenetic diversity in neotropical agricultural systems. Science 345:1343–1346.
Fritz S.A., Purvis A. 2010. Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. Conserv. Biol. 24:1042–1051.
Gascuel F., Ferrière R., Aguilée R., Lambert A. 2015. How ecology and landscape dynamics shape phylogenetic trees. Syst. Biol. 64:590–607.
Gaston K.J., Blackburn T.M. 1995. Birds, body size and the threat of extinction. Philos. Trans. R. Soc. Lond. B 347:205–212.
Glavin T. 2007. The sixth extinction: journeys among the lost and left behind. New York: Thomas Dunne Books.
Guyer C., Slowinski J.B. 1991. Comparisons of observed phylogenetic topologies with null expectations among three monophyletic lineages. Evolution 45:340–350.
Guyer C., Slowinski J.B. 1993. Adaptive radiation and the topology of large phylogenies. Evolution 47:253–263.
Hagen O., Hartmann K., Steel M., Stadler T. 2015. Age-dependent speciation can explain the shape of empirical phylogenies. Syst. Biol. 64:432–440.
Heard S.B. 1992. Patterns in tree balance among cladistic, phenetic, and randomly generated phylogenetic trees. Evolution 46:1818–1826.
Heard S.B., Mooers A.O. 2000. Phylogenetically patterned speciation rates and extinction risks change the loss of evolutionary history during extinctions. Proc. R. Soc. Lond. B. 267:613–620.
Heath T.A., Hedtke S.M., Hillis D.M. 2008. Taxon sampling and the accuracy of phylogenetic analyses. J. Syst. Evol. 46:239–257.
Hubbell S. 2001. The unified neutral theory of biodiversity and biogeography. Princeton, NJ: Princeton University Press.
Hughes A.L. 1999. Differential human impact on the survival of genetically distinct avian lineages. Bird Conserv. Int. 9:147–154.
Isaac N.J.B., Turvey S.T., Collen B., Waterman C., Baillie J.E.M 2007. Mammals on the EDGE: conservation priorities based on threat and phylogeny. PLoS One 2:e296.
IUCN. 2012. IUCN red list categories and criteria: version 3.1. 2nd ed. Gland, Switzerland: IUCN.
Jetz W., Thomas G., Joy J., Hartmann K., Mooers A. 2012. The global diversity of birds in space and time. Nature 491:444–448.
Johnson C.N., Delean S., Balmford A. 2002. Phylogeny and the selectivity of extinction in Australian marsupials. Anim. Conserv. 5:135–142.
Johnson S.G., Narasimhan B. 2013. R package "cubature": adaptive multivariate integration over hypercubes. Available from: https://cran.r-project.org/web/packages/cubature/index.html.
Jønsson, K.A., Fabre P.-h., Fritz S.A., Etienne R.S., Ricklefs R.E., Jørgensen T.B. 2012. Ecological and evolutionary determinants for the adaptive radiation of the Madagascan vangas. Proc. Natl. Acad. Sci. USA 109:6620–6625.
Kembel S.W., Ackerly D.D., Blomberg S.P., Cornwell W.K., Cowan P.D., Helmus M.R., Morlon H., Webb C.O. 2014. R package "picante": R tools for integrating phylogenies and ecology. Available from: https://cran.r-project.org/web/packages/picante/index.html.
Kingman J.F.C. 1982. The coalescent. Stoch. Proc. Appl. 13:235–248.
Kirkpatrick M., Slatkin M. 1993. Searching for evolutionary patterns in the shape of a phylogenetic tree. Evolution 47:1171–1181.
Kozak K.H., Weisrock D.W., Larson A. 2006. Rapid lineage accumulation in a non-adaptive radiation: phylogenetic analysis of diversification rates in eastern North American woodland salamanders (Plethodontidae: Plethodon). Proc. R. Soc. Lond. B. 273:539–546.
Kumar S. 2005. Molecular clocks: four decades of evolution. Nat. Rev. Genet. 6:654–662.
Lambert A., Stadler T. 2013. Birth-death models and coalescent point processes: the shape and probability of reconstructed phylogenies. Theor. Popul. Biol. 90:113–128.
Lambert A., Steel M. 2013. Predicting the loss of phylogenetic diversity under non-stationary diversification models. J. Theor. Biol. 337:111–124.
Lambert A. et al. 2017. Probabilistic models for the (sub) tree (s) of life. Braz. J. Probab. Stat. 31:415–475.
Leakey R.E., Lewin R. 1995. The sixth extinction: patterns of life and the future of humankind. New York: Doubleday.
Lee T.M., Jetz W. 2011. Unravelling the structure of species extinction risk for predictive conservation science. Proc. Roy. Soc. Lond. B. 278(1710):1329–1338
Liow L.H., Quental T.B., Marshall C.R. 2010. When can decreasing diversification rates be detected with molecular phylogenies and the fossil record? Syst. Biol. 59:646–659.
Lovette I.J., Bermingham E. 1999. Explosive speciation in the New World Dendroica warblers. Proc. R. Soc. Lond. B. 266: 1629–1636.
Lozano F.D., Schwartz M.W. 2005. Patterns of rarity and taxonomic group size in plants. Biol. Conserv. 126:146–154.
MacArthur R., Wilson E. 1967. The theory of island biogeography. Princeton, NJ: Princeton University Press.
MacArthur R.H. 1957. On the relative abundance of bird species. Proc. Natl. Acad. Sci. USA 43:293–295.
Maddison W.P., Midford P.E., Otto S.P. 2007. Estimating a binary character's effect on speciation and extinction. Syst. Biol. 56:701–710.
Magallon S., Sanderson M.J. 2001. Absolute diversification rates in Angiosperm clades. Evolution 55:1762–1780.
Magnuson-Ford K., Ingram T., Redding D.W., Mooers A.Ø. 2009. Rockfish (sebastes) that are evolutionarily isolated are also large, morphologically distinctive and vulnerable to overfishing. Biol. Conserv. 142:1787–1796.
Manceau M., Lambert A., Morlon H. 2015. Phylogenies support out-of-equilibrium models of biodiversity. Ecol. Lett. 18:347–356.
McKinney M.L. 1997. Extinction vulnerability and selectivity: combining ecological and paleontological views. Annu. Rev. Ecol. Syst. 28:495–516.
McPeek M.A. 2008. The ecological dynamics of clade diversification and community assembly. Am. Nat. 172:E270–E284.
McPeek M.A., Brown J.M. 2007. Clade age and not diversification rate explains species richness among animal taxa. Am. Nat. 169:E97–106.
Missa O., Dytham C., Morlon H. 2016. Understanding how biodiversity unfolds through time under neutral theory. Philos. Trans. R. Soc. Lond. B 371:1–12.

Moen D., Morlon H. 2014. Why does diversification slow down? Trends Ecol. Evol. 29:190–197.
Mooers A., Gascuel O., Stadler T., Li H., Steel M. 2012. Branch lengths on birth-death trees and the expected loss of phylogenetic diversity. Syst. Biol. 61:195–203.
Mooers A., Heard S. 1997. Inferring evolutionary process from phylogenetic tree shape. Q. Rev. Biol. 72:31–54.
Mooers A.O. 1995. Tree balance and tree completeness. Evolution 49:379–384.
Mooers A. Ø., Goring S.J., Turvey S.T., Kuhn T.S. 2009. Holocene extinctions and the loss of feature diversity. Turvey S. editor. Holocene extinctions. Oxford: Oxford University Press. p. 279–338
Morlon H., Parsons T.L., Plotkin J.B. 2011a. Reconciling molecular phylogenies with the fossil record. Proc. Natl. Acad. Sci. USA 108:16327–16332.
Morlon H., Schwilk D.W., Bryant J.A., Marquet P.A., Rebelo A.G., Tauss C., Bohannan B.J., Green J.L. 2011b. Spatial patterns of phylogenetic diversity. Ecol. Lett. 14:141–149.
Nee S. 2006. Birth-death models in macroevolution. Annu. Rev. Ecol. Evol. Syst. 37:1–17.
Nee S., May R.M. 1997. Extinction and the loss of evolutionary history. Science 278:692–694.
Nee S., Mooers A.O., Harvey P.H. 1992. Tempo and mode of evolution revealed from molecular phylogenies. Proc. Natl. Acad. Sci. USA 89:8322–8326.
Nipperess D.A., Matsen F.A. 2013. The mean and variance of phylogenetic diversity under rarefaction. Methods Ecol. Evol. 4:566–572.
Paradis E., Claude J., Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290.
Parhar R.K., Mooers A.Ø. 2011. Phylogenetically clustered extinction risks do not substantially prune the tree of life. PLoS One 6:e23528.
Pearson P.N. 1992. Survivorship analysis of fossil taxa when real-time extinction rates vary: the paleogene planktonic foraminifera. Paleobiology 18:115–131.
Phillimore A.B., Price T.D. 2008. Density-dependent cladogenesis in birds. PLoS Biol. 6:e71.
Pigot A.L., Phillimore A.B., Owens I.P.F., Orme C.D.L. 2010. The shape and temporal dynamics of phylogenetic trees arising from geographic speciation. Syst. Biol. 59:660–673.
Prado P.I., Miranda M.D., Chalom A. 2015. R package "sads": maximum likelihood models for species abundance distributions. Available form: http://search.r-project.org/library/sads/html/fitsad.html.
Pulquério, M.J.F., Nichols R.A. 2007. Dates from the molecular clock: how wrong can we be? Trends Ecol. Evol. 22:180–184.
Purvis A. 1996. Using interspecies phylogenies to test macroevolutionary hypotheses. In: Harvey, P., Leigh Brown, A., Maynard Smith, J., Nee, S., editors. New uses for new phylogenies. Oxford: Oxford University Press. p. 153–168.
Purvis A. 2008. Phylogenetic approaches to the study of extinction. Annu. Rev. Ecol. Evol. Syst. 39:301–319.
Purvis A., Agapow P.-M., Gittleman J.L., Mace G.M. 2000a. Nonrandom extinction and the loss of evolutionary history. Science 288:328–330.
Purvis A., Gittleman J.L., Cowlishaw G., Mace G.M. 2000b. Predicting extinction risk in declining species. Proc. Roy. Soc. Lond. B 267:1947–1952.
Purvis A., Jones K.E., Mace G.M. 2000c. Extinction. BioEssays 22:1123–1133.
Pybus O.G., Harvey P.H. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. Proc. R. Soc. Lond. B. 267:2267–2272.
R Development Core Team. 2012. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
Rabosky D.L. 2009. Ecological limits and diversification rate: alternative paradigms to explain the variation in species richness among clades and regions. Ecol. Lett. 12:735–743.
Rabosky D.L., Lovette I.J. 2008a. Density-dependent diversification in North American wood warblers. Proc. R. Soc. Lond. B. 275:2363–2371.
Rabosky D.L., Lovette I.J. 2008b. Explosive evolutionary radiations: decreasing speciation or increasing extinction through time? Evolution 62:1866–1875.
Rabosky D.L., Slater G.J., Alfaro M.E. 2012. Clade age and species richness are decoupled across the eukaryotic tree of life. PLoS Biol. 10:e1001381.
Raup D.M., Gould S.J., Schopf T.J.M., Simberloff D.S. 1973. Stochastic models of phylogeny and the evolution of diversity. J. Geol. 81:525–542.
Redding D.W., Mooers A.Ø. 2006. Incorporating evolutionary measures into conservation prioritization. Conserv. Biol. 20:1670–1678.
Ricklefs R. 2009. Speciation, extinction and diversity. In: Butlin, R., Bridle, J., Schluter, D., editors. Speciation and patterns of diversity. Cambridge: Cambridge University Press. p. 257–277.
Ricklefs R.E. 2006. Global variation in the diversification rate of passerine birds. Ecology 87:2468–78.
Ricklefs R.E. 2007. History and diversity: explorations at the intersection of ecology and evolution. Am. Nat. 170:S56–S70.
Rüber L., Zardoya R. 2005. Rapid cladogenesis in marine fishes revisited. Evolution 59:1119–1127.
Russell G.J., Brooks T.M., McKinney M.M., Anderson C.G. 1998. Present and future taxonomic selectivity in bird and mammal extinctions. Conserv. Biol. 12:1365–1376.
Sainudiin R., Véber A. 2016. A beta-splitting model for evolutionary trees. R. Soc. Open Sci. 3:160016.
Sánchez-Reyes L.L., Morlon H., Magallón S. 2016. Uncovering higher-taxon diversification dynamics from clade age and species-richness data. Syst. Biol. 66:367–378.
Schwartz M.W., Simberloff D. 2001. Taxon size predicts rates of rarity in vascular plants. Ecol. Lett. 4(5):464–469.
Schwartz R.S., Mueller R.L. 2010. Branch length estimation and divergence dating: estimates of error in Bayesian and maximum likelihood frameworks. BMC Evol. Biol. 10:1–21.
Seehausen O. 2006. African cichlid fish: a model system in adaptive radiation research. Proc. R. Soc. Lond. B. 273:1987–1998.
Slowinski J., Guyer C. 1993. Testing whether certain traits have caused amplified diversification—an improved method based on a model of random speciation and extinction. Am. Nat. 142:1019–1024.
Stadler T. 2013. Recovering speciation and extinction dynamics based on phylogenies. J. Evol. Biol. 26:1203–1219.
Stadler T., Rabosky D.L., Ricklefs R.E., Bokma F. 2014. On age and species richness of higher taxa. Am. Nat. 184:447–455.
Szabo J.K., Khwaja N., Garnett S.T., Butchart S.H. 2012. Global patterns and drivers of avian extinctions at the species and subspecies level. PLoS One 7:e47080.
Vamosi J.C., Wilson J.R. 2008. Nonrandom extinction leads to elevated loss of angiosperm evolutionary history. Ecol. Lett. 11:1047–1053.
Van Valen L. 1976. Ecological species, multispecies, and oaks. Taxon 25:233–239.
Vazquez D.P., Gittleman J.L. 1998. Biodiversity conservation: does phylogeny matter? Curr. Biol. 8:379–381.
Venditti C., Meade A., Pagel M. 2010. Phylogenies reveal new interpretation of speciation and the Red Queen. Nature 463:349–352.
Veron S., Davies T.J., Cadotte M.W., Clergeau P., Pavoine S. 2015. Predicting loss of evolutionary history: where are we? Biol. Rev. 0:0–0.
von Euler F. 2001. Selective extinction and rapid loss of evolutionary history in the bird fauna. Proc. R. Soc. Lond. B. 268:127–130.
Wake D.B., Vredenburg V.T. 2008. Are we in the midst of the sixth mass extinction? A view from the world of amphibians. Proc. Natl. Acad. Sci. USA 105:11466–11473.
Weir J.T. 2006. Divergent timing and patterns of species accumulation in lowland and highland neotropical birds. Evolution 60:842–855.
Welch J.J., Bromham L. 2005. Molecular dating when rates vary. Trends Ecol. Evol. 20:320–327.
Yule G. 1925. A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis, F.R.S. Philos. Trans. R. Soc. Lond. B 213:402–410.
Zink R., Slowinski J. 1995. Evidence from molecular systematics for decreased avian diversification in the Pleistocene Epoch. Proc. Natl. Acad. Sci. USA 92:5832–5835.

Supplementary material for Chapter 1

Appendix 1. Mathematical details about the model.

Trees with $\beta = 0$ and $\alpha = 1$ are uniform labelled ranked tree shapes.

Here we want to show the following proposition.

Proposition 1.1. *The law of the ranked tree shape under our model with $\beta = 0$ and $\alpha = 1$ is the uniform distribution on binary trees with ranked nodes and n labelled leaves.*

The distribution mentioned in the proposition is called URT in (Lambert and Stadler, 2013), for 'uniform ranked tree'. To be more concise, we will call *uniform tree*, a random ultrametric tree which follows URT.

Note that the distribution mentioned in the proposition is the reference model used in phylogenetics and population genetics. Indeed, it is known (*e.g.*, Lambert et al., 2017) that the standard Kingman coalescent tree (backward time) started from n labelled tips follows URT, and that the Yule tree (forward time) stopped when it has n leaves which are then uniformly labelled, also follows URT. In the proof, we will use a third characterization of this distribution using *oriented trees*.

Proof. Recall that in our model, we start with n uniform, independent random variables $U_{i\in\{1,\dots,n\}}$ in the unit interval $[0,1]$ and sequentially partition the unit interval thanks to random variables independent of $U_{i\in\{1,\dots,n\}}$ by iterating the following procedure until all variables $U_{i\in\{1,\dots,n\}}$ are in distinct components of the partition: 1) select one interval of the current partition proportionately to its length to the power α, and 2) throw in this interval, say (a,b) a variable $V = a+(b-a)R$, where R is an independent random variable following the Beta distribution with parameters $(\beta+1,\beta+1)$, so that V splits the interval into two new components of the partition. Now assume that $\alpha = 1$ and $\beta = 0$. In this case R is uniform so it is clear that the two steps of the procedure boil down to throwing one point V which is uniformly distributed in the unit interval, to split the partition. Let us call sequentially these independent, uniform variables $V_1, V_2, \dots$ and let $(U'_i)_{i\in\{1,\dots,n\}}$ denote the order statistics of the $(U_i)_{i\in\{1,\dots,n\}}$. Now set $J_0 := 0$ and define inductively on $k \geq 1$

$$J_k := \min\{j > J_{k-1} : V_j \in (U'_i, U'_{i+1}) \text{ for some } i \in \{1,\dots,n-1\} \setminus \{I_1,\dots,I_{k-1}\}\},$$

and I_k the unique integer $i \in \{1,\dots,n-1\} \setminus \{I_1,\dots,I_{k-1}\}$ such that $V_{J_k} \in (U'_i, U'_{i+1})$. Then the tree can be built by following sequentially the splits induced by $V_{I_1},\dots,V_{I_{n-1}}$ in this order.

An *oriented tree* is a ranked tree embedded in the plane where tips are arranged on a line from left to right as in Figure 2a in (Lambert and Stadler, 2013) and edges do not cross. In particular it has $n-1$ pairwise distinct node depths also arranged from left to right, that can be put in this order into a vector. A *uniform oriented tree* is an oriented tree where the order of node depths is given by a uniform permutation. Now we will use the following fact (Lambert and Stadler, 2013). Fact: The tree obtained from a uniform oriented tree after labeling uniformly the tips and forgetting the orientation follows URT. Then to show that the ranked labelled tree constructed in the previous paragraph follows URT, it is sufficient to show that the vector $(I_1,\dots,I_{n-1})$ is a uniform permutation of $\{1,\dots,n-1\}$. Indeed if this is the case then the oriented tree whose tip i is marked by U'_i is a uniform oriented tree. And since the vector (U_i) is a uniform permutation of the vector

(U'_i), our initial tree is obtained from the uniform oriented tree by relabeling uniformly its tips (and forgetting the orientation as implicitly done in the definition).

So it only remains to prove that $(I_1, \dots, I_{n-1})$ is a uniform permutation of $\{1, \dots, n-1\}$. To do that, it is sufficient to show that for any $k \in \{1, \dots, n-1\}$ conditional on $(I_1, \dots, I_{k-1})$, I_k is uniformly distributed in $\{1, \dots, n-1\} \setminus \{I_1, \dots, I_{k-1}\}$. Now conditional on $J_k = j$, the vector $S := (W_i)_{i \in \{1,\dots,n+j-1\}}$ defined as the uniform relabeling of $\{U_1, \dots, U_n\} \cup \{V_1, \dots, V_{j-1}\}$ has the same law as $n+j-1$ independent uniforms, so that each interval between two consecutive points of S has the same probability to be the one welcoming a new independent uniform. Now it can be proved that conditional on J_k and conditional on $(I_1, \dots, I_{k-1})$, the labels i such that $W_i \in \{U'_i, i \neq I_1, \dots, I_{k-1}\}$ are independent of the order statistics of S, which shows that I_k is uniformly distributed conditional on J_k, and so that it is uniformly distributed. □

The field of bullets model of extinction and the $\eta = 1$ case.

In our model involving η, recall that each tip is endowed with a non-negative real number interpreted as the frequency of the corresponding species in the phylogeny (relative abundance or range size). Since these frequencies sum up to 1, they can also be interpreted as a sampling distribution over the tips of the tree, that is each species can be drawn with a probability equal to its frequency. For example, a species is drawn to survive the contemporary mass extinction with a probability equal to its frequency. This is not exactly how we proceed in the main text to model the extinction experiment, where we rather keep the species with the largest frequencies.

In the field of bullets model of extinction, each species is drawn uniformly, independently of the phylogeny. In our model, as soon as all species have the same expected frequency (equal to $1/n$) conditional on the phylogeny, this is also what happens, regardless of whether we draw the species with the largest frequencies or whether we draw them proportionately to their frequency. Thanks to the recursive construction of the tree, this is also equivalent to saying that at the basal split, if X denotes the size of the left subclade and K denotes the number of tip species in it, then

$$E\left(\frac{X^\eta}{X^\eta + (1-X)^\eta}\middle| K = k\right) = \frac{k}{n}. \tag{6}$$

We do the calculation in the case when $\eta = 1$.

Proposition 1.2. *For any $\beta > -2$,*

$$E(X|K=k) = \frac{k+\beta+1}{n+2(\beta+1)}. \tag{7}$$

Proof. Write $B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}$ for the beta function, so that (when $\beta > -1$)

$$P(X \in dx, K = k) = B(\beta+1, \beta+1)^{-1} x^\beta (1-x)^\beta \binom{n}{k} x^k (1-x)^{n-k} dx.$$

As a consequence (and this is is true even if $\beta \in (-2, -1]$),

$$P(X \in dx | K = k) = \frac{x^{\beta+k}(1-x)^{\beta+n-k}}{\int_0^1 y^{\beta+k}(1-y)^{\beta+n-k}\, dy}\, dx.$$

In particular,

$$\begin{aligned} E(X|K=k) &= \frac{\int_0^1 x^{\beta+k+1}(1-x)^{\beta+n-k}dx}{\int_0^1 y^{\beta+k}(1-y)^{\beta+n-k}\,dy} \\ &= \frac{B(\beta+k+2,\beta+n-k+1)}{B(\beta+k+1,\beta+n-k+1)} \\ &= \frac{\Gamma(\beta+k+2)\,\Gamma(\beta+n-k+1)}{\Gamma(2\beta+n+3)}\frac{\Gamma(2\beta+n+2)}{\Gamma(\beta+k+1)\,\Gamma(\beta+n-k+1)} \\ &= \frac{\Gamma(\beta+k+2)}{\Gamma(\beta+k+1)}\frac{\Gamma(2\beta+n+2)}{\Gamma(2\beta+n+3)}, \end{aligned}$$

where $\Gamma(x)=\int_0^\infty t^{x-1}e^{-x}dt$ denotes the gamma function. The result follows from the well-known fact that $\Gamma(x+1)=x\Gamma(x)$. □

Except when $\beta=-1$, Equations (6) (with $\eta=1$) and (7) are not equivalent but for all practical purposes β only takes values between -2 and 2, so the approximation of (7) $E(X|K=k)\approx k/n$ holds whenever $k\gg 1$. This means that when $\eta=1$ the extinction experiment is roughly equivalent to the field of bullets model. To be more precise, observe that when $\beta=-1$, the r.h.s. exactly equals k/n, which yields the rigorous equivalence in this case with the field of bullets model. In the other cases, it is easy to check that for any k,

$$k<\frac{n}{2}\Longleftrightarrow E(X|K=k)>\frac{k}{n}$$

if $\beta>-1$ and

$$k<\frac{n}{2}\Longleftrightarrow E(X|K=k)<\frac{k}{n}$$

if $\beta<-1$. So if we compare the case $\eta=1$ to the field of bullets model, smaller clades in our model have a slight demographic advantage over larger clades when $\beta>-1$ and conversely when $\beta<-1$.

Simulating trees with $\beta\le -1$.

Let ν be a measure on (0,1) such that $\int_{(0,1)} x(1-x)\,\nu(dx)<\infty$. We are interested in the binary self-similar fragmentation (Bertoin, 2002, 2006) with index α, zero erosion and fragmentation measure ν (or more rigorously the push-forward of ν by the mapping which maps $s\in(0,1)$ to the infinite sequence $(\max(s,1-s),\min(s,1-s),0,0\ldots)$), as mentioned in the main text.

We will assume that ν is symmetric, in the sense that for any non-negative function f, $\int_{(0,1)} f(x)\,\nu(dx)=\int_{(0,1)} f(1-x)\,\nu(dx)$, and further that ν has a density denoted g. In the application given in the paper, we will only be interested by the case $g(x)=x^\beta(1-x)^\beta$ for $\beta>-2$.

We start at time 0 with a fragment of size x containing the n uniformly distributed marks $U_{i\in\{1,\ldots,n\}}$. We apply the fragmentation process to this fragment, until the first time τ when the n marks are *not* in the same fragment. For any $0\le t<\tau$, we can define X_t the size at time t of the fragment containing the n marks. By definition, at time τ the fragment is split into two fragments with sizes say $RX_{\tau-}$ and $(1-R)X_{\tau-}$ containing respectively K and $n-K$ marks. We seek to characterize the joint distribution of $(\tau,X_{\tau-},R,K)$. We write $P_{x,\alpha}$ to emphasize the dependence on the initial size x of the fragment and on the index α of self-similarity.

Proposition 1.3. *The pair (R, K) is independent of $(\tau, X_{\tau-})$ and has law given by*

$$P(R \in dr, K = k) = a_n^{-1} \binom{n}{k} r^k (1-r)^{n-k}\, g(r)\, dr,$$

where we have set

$$a_n := \int_{(0,1)} \left(1 - r^n - (1-r)^n\right)\, g(r)\, dr. \tag{8}$$

Under the law $P_{0,1}$ of the homogeneous fragmentation ($\alpha = 0$) started with a fragment of size 1, the process $(X_s; 0 \le s < \tau)$ has the same law as $\left(e^{-\hat{\xi}_s}; 0 \le s < \sigma\right)$, where

- *$\hat{\xi}$ is the subordinator with Lévy measure $\hat{\Lambda}$, where $\hat{\Lambda}(dx) = 2e^{-(n+1)x} g\left(e^{-x}\right)\, dx$.*
- *σ is an independent exponential time with parameter a_n given by (8).*

Thanks to (9), the law of $(\tau, X_{\tau-})$ under $P_{x,\alpha}$ is the same as $\left(x^{-\alpha} \int_0^\sigma e^{\alpha\hat{\xi}_s} ds, x e^{-\hat{\xi}_\sigma}\right)$.

Proof. The first part of the proposition is due to an elementary application of the compensation formula.

Now we will use the fact that self-similar fragmentations can be obtained by time-changing homogeneous fragmentations (case $\alpha = 0$). More specifically, if , then for any non-negative or bounded, bivariate function f

$$E_{x,\alpha}\left[f(\tau, X_{\tau-})\right] = E_{1,0}\left[f\left(\int_0^\tau (xX_s)^{-\alpha} ds, xX_{\tau-}\right)\right]. \tag{9}$$

So now we seek to compute $E_{1,0}[F(X_s; 0 \le s < \tau)]$, where F is any non-negative or bounded measurable functional of the trajectory $(X_s; 0 \le s < \tau)$, in particular F can of course be of the form $f\left(\int_0^\tau (xX_s)^{-\alpha} ds, xX_{\tau-}\right)$.

We will take advantage of the fact that the law of a fragment tagged by one single mark is already known, see Section 3.2.2 in (Bertoin, 2006). We denote by M the size of the fragment tagged by one mark and set $\xi := -\ln(M)$. Then we can use Theorem 3.2 in (Bertoin, 2006), which states that ξ is a subordinator with Laplace exponent Φ given by

$$\Phi(q) = \int_{(0,1)} \left(1 - s^{q+1} - (1-s)^{q+1}\right)\, \nu(ds) \qquad q \ge 0.$$

Recall this means that $E(\exp(-q\xi_t)) = e^{-t\Phi(q)}$. By symmetry of ν, we have

$$\Phi(q) = \int_{(0,1)} \left(s - s^{q+1} + (1-s) - (1-s)^{q+1}\right)\, \nu(ds) = 2\int_{(0,1)} s\left(1 - s^q\right)\, \nu(ds).$$

Writing $s = e^{-x}$, it can readily be seen that the Lévy measure Λ of ξ can be written as $2e^{-x}\,\mu(dx)$, where μ is the push-forward of ν by the mapping $s \mapsto -\ln(s)$. If $\nu(ds) = g(s)\, ds$, this yields

$$\Phi(q) = 2\int_0^1 s\left(1 - s^q\right)\, g(s)\, ds = 2\int_0^\infty e^{-2x}\left(1 - e^{-qx}\right)\, g\left(e^{-x}\right)\, dx, \tag{10}$$

that is $\Lambda(dx) = 2e^{-2x} g\left(e^{-x}\right)\, dx$.

Now X can be seen as M with $n-1$ additional marks, so τ is the unique jump time t of ξ such that these $n-1$ additional marks are all in M_{t-} but not all in M_t. By the compensation formula, letting $U_1, \ldots, U_{n-1}$ denote independent uniform $(0,1)$ random variables,

$$\begin{aligned}
E_{1,0}(F) &= E \sum_{t:\Delta\xi_t>0} F\left(M_s; 0 \le s < t\right) \mathbb{1}_{\forall i, U_i \le M_{t-}}\, \mathbb{1}_{\exists i, U_i > M_t} \\
&= E \sum_{t:\Delta\xi_t>0} F\left(M_s; 0 \le s < t\right) \left(\mathbb{1}_{\forall i, U_i \le M_{t-}} - \mathbb{1}_{\forall i, U_i \le M_t}\right) \\
&= E \sum_{t:\Delta\xi_t>0} F\left(M_s; 0 \le s < t\right)\, M_{t-}^{n-1} \left(1 - \left(M_t/M_{t-}\right)^{n-1}\right) \\
&= E \int_0^\infty dt\, F\left(M_s; 0 \le s < t\right)\, M_{t-}^{n-1} \int_0^\infty \Lambda(dx) \left(1 - e^{-(n-1)x}\right)
\end{aligned}$$

Note that the last integral equals $\Phi(n-1)$ which was denoted earlier a_n, see (8). Also recall that the Lebesgue measure of jump times of a subordinator is a.s. 0, so that

$$E_{1,0}(F) = a_n\, E \int_0^\infty dt\, F\left(M_s; 0 \le s < t\right)\, M_t^{n-1} = a_n\, E \int_0^\infty dt\, F\left(e^{-\xi_s}; 0 \le s < t\right)\, e^{-(n-1)\xi_t}.$$

In particular, since $E(\exp(-q\xi_t)) = e^{-t\Phi(q)}$ and $\Phi(n-1) = a_n$,

$$\begin{aligned}
E_{1,0}(F) &= a_n\, E \int_0^\infty dt\, F\left(e^{-\xi_s}; 0 \le s < t\right)\, e^{-(n-1)\xi_t} \\
&= \int_0^\infty dt\, a_n e^{-a_n t}\, E\, F\left(e^{-\xi_s}; 0 \le s < t\right)\, e^{-(n-1)\xi_t + t a_n} \\
&= \int_0^\infty dt\, a_n e^{-a_n t}\, E\, F\left(e^{-\hat{\xi}_s}; 0 \le s < t\right),
\end{aligned}$$

where $\hat{\xi}$ is the Markov process obtained by h-transform from ξ via the positive martingale $\left(e^{-(n-1)\xi_t + t a_n}; t \ge 0\right)$. It is then straightforward to prove that $\hat{\xi}$ is the subordinator with Laplace exponent

$$\hat{\Phi}(q) = \Phi(q+n-1) - \Phi(n-1),$$

which can be written either in the form

$$\hat{\Phi}(q) = \int_{(0,1)} \left(1 - s^{q+n} - (1-s)^{q+n}\right)\, \nu(ds) - a_n,$$

or in the form

$$\hat{\Phi}(q) = \int_0^\infty e^{-(n-1)x} \left(1 - e^{-qx}\right)\, \Lambda(dx),$$

that is, $\hat{\xi}$ is the subordinator with (zero drift and) Lévy measure $\hat{\Lambda}$ given by

$$\hat{\Lambda}(dx) = e^{-(n-1)x}\, \Lambda(dx), \tag{11}$$

which can be expressed as in the statement $\hat{\Lambda}(dx) = 2e^{-(n+1)x} g\left(e^{-x}\right)\, dx$. □

Let us try to apply this to random, ranked binary tree shapes. Let g be defined by $g(r) = r^\beta(1-r)^\beta$. Here

$$a_n = \int_{(0,1)} (1 - r^n - (1-r)^n)\; r^\beta(1-r)^\beta\, dr.$$

To simulate the subordinator $\hat{\xi}$, we need to fix a cutoff parameter ε. Let $(N_\varepsilon(s); s \geq 0)$ be a homogeneous Poisson process with parameter

$$\lambda_\varepsilon := \int_\varepsilon^\infty \hat{\Lambda}(dx) = 2\int_\varepsilon^\infty e^{-(\beta+n+1)x}\left(1-e^{-x}\right)^\beta\, dx.$$

Let (Y_i) be independent random variables with density

$$\frac{2}{\lambda_\varepsilon} e^{-(\beta+n+1)x}\left(1-e^{-x}\right)^\beta \tag{12}$$

and set

$$Z_\varepsilon(s) := \sum_{i=1}^{N_\varepsilon(s)} Y_i.$$

Then as $\varepsilon \downarrow 0$, the process Z_ε converges to $\hat{\xi}$. Since one has to stop at the time σ which is exponentially distributed with parameter a_n, one can first simulate σ and then conditional on $\sigma = t$, draw a Poisson number $N_\varepsilon(t)$ with parameter $\lambda_\varepsilon t$ of jump times, which are then uniformly distributed in $[0,t]$.

So when ε is sufficiently small, we should have a good approximation of $(\tau, X_{\tau-})$ by

$$\left(x^{-\alpha}\int_0^t e^{\alpha Z_\varepsilon(s)} ds, xe^{-Z_\varepsilon(t)}\right).$$

The small trick is that the smaller β the smaller ε has to be chosen.

This approach generates identical results as the one presented in the main text (Fig. A1).

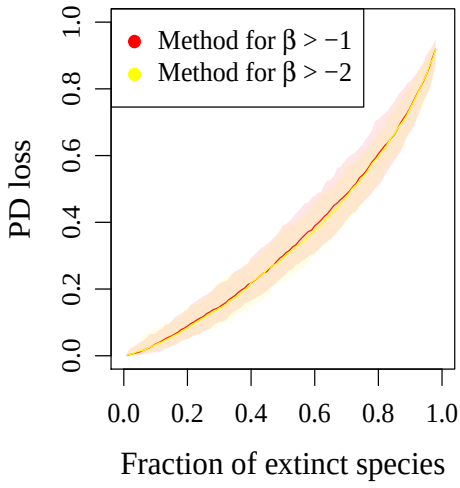


Figure A1: **Comparison of results with both methods used to simulate ranked tree shapes.** The (more intuitive) method for $\beta > -1$ is explained in the main text, whereas the method for $\beta > -2$ is explained hereabove. Results are based on 100 simulation replicates: lines give median values and light areas give 95% confidence intervals. Parameter values: $\beta = 0$, $\alpha = 0$, $\eta = 1$, number of species $N = 100$, $\epsilon = 0.001$.

Species abundance distributions.

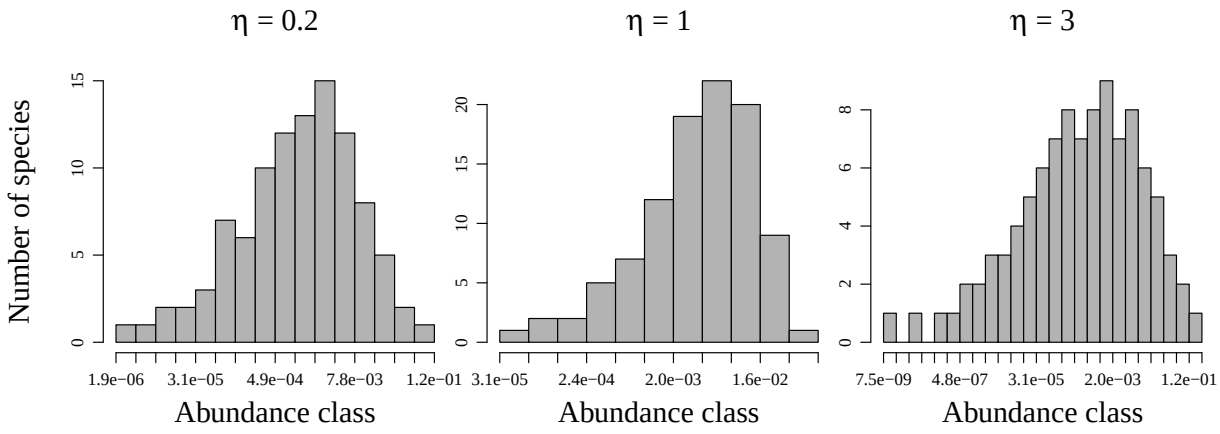


Figure A2: **Preston plots of the species abundance distributions generated by the model, for different clades abundance-richness indices η.** Absolute values of species frequencies are not used in the analyses, but relative values set the order of species extinctions (starting from rarer species). Parameter values: $\eta = 0.2$, 1 or 3 (from left to right), $\beta = 0$, $\alpha = 0$, number of species $N = 100$, $\epsilon = 0.001$.

Relative species abundances in 'comb trees'.

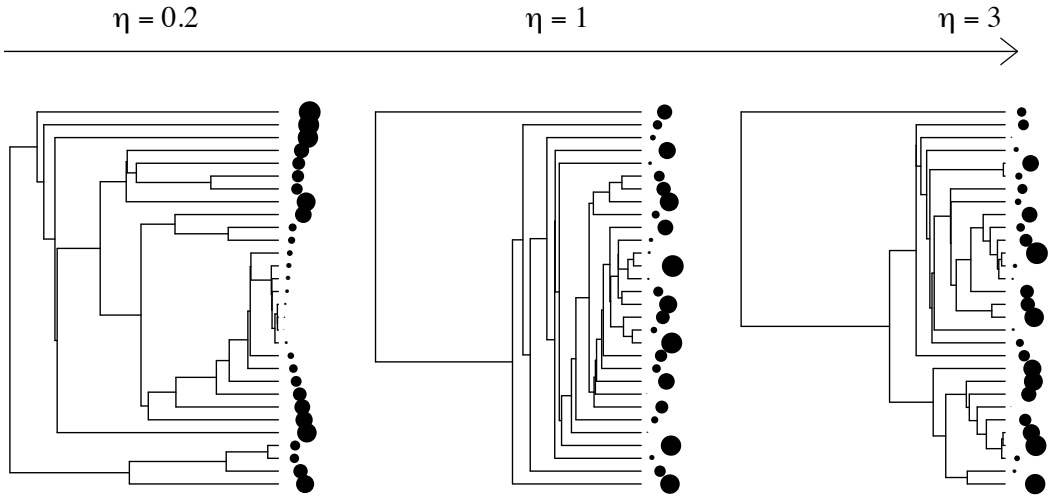


Figure A3: **Distribution of species frequencies across the tips of phylogenetic trees under variation in the clades abundance-richness index η, when topologies are very unbalanced.** Dot sizes rank species according to their abundance (larger dots for more abundant species). Parameter values: $\eta = 0.2$, 1 or 3 (from left to right), $\beta = -1.9$, $\alpha = 0$, number of species $N = 30$, $\epsilon = 0.001$.

Appendix 2. Effect of the ranked tree shape (β, α) on PD loss viewed as a function of β.

Fig. A4 shows the well known effect of decreasing PD loss for increasing β for different α values. This effect is dampened for low α values.

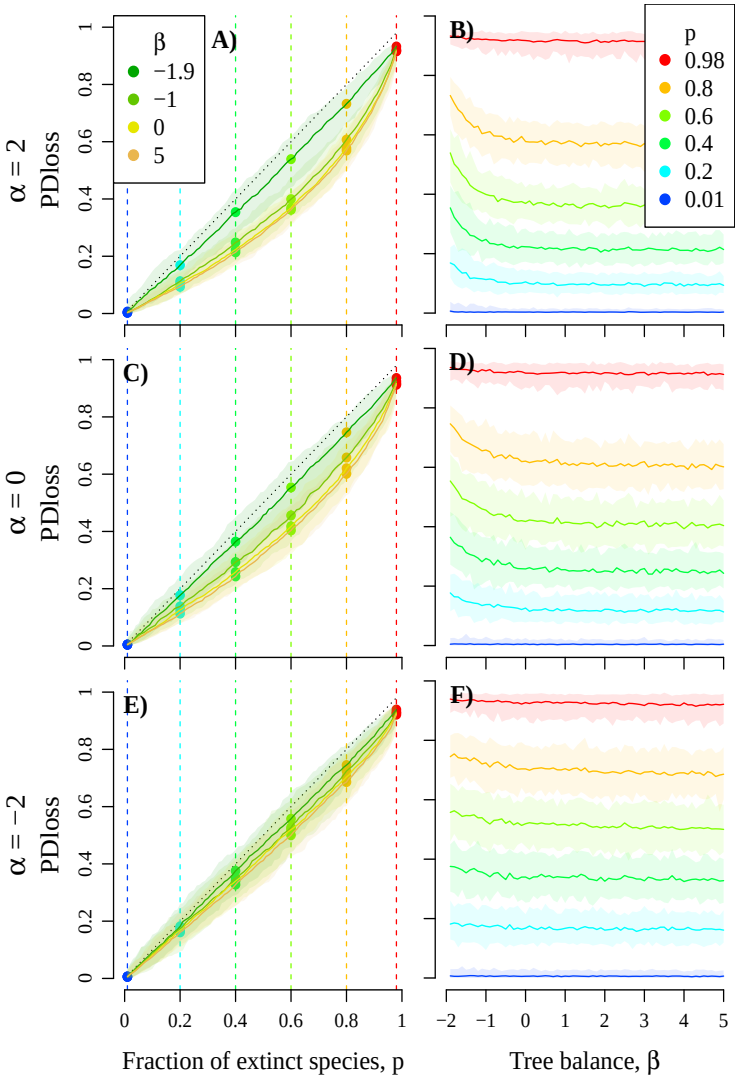


Figure A4: **Influence of the ranked tree shape (tree balance β and clades age-richness index α) on PD loss, for increasing fractions of species extinctions p.** The clades age-richness index α changes from 2 (top row, larger clades being older) to -2 (bottom row, smaller clades being older). Results are shown either as a function of the extinction fraction p (left column; for different β values) or as a function of β (right column; for different extinction fractions p). Extinction fraction p increases from 0.01 to 0.98 (from left to right in A, C, E; from blue to red in B, D, F). The dotted lines in A, C, E show the bisector. Results are based on 100 simulation replicates: plain lines give median values and light areas give 95% confidence intervals. Other parameter values: number of species $N = 100$, $\epsilon = 0.001$.

Appendix 3. Sensitivity of the effect of the ranked tree shape (β, α) on PD loss to tree size N.

Fig. A5 shows that tree size has little effect on the predictions of the model.

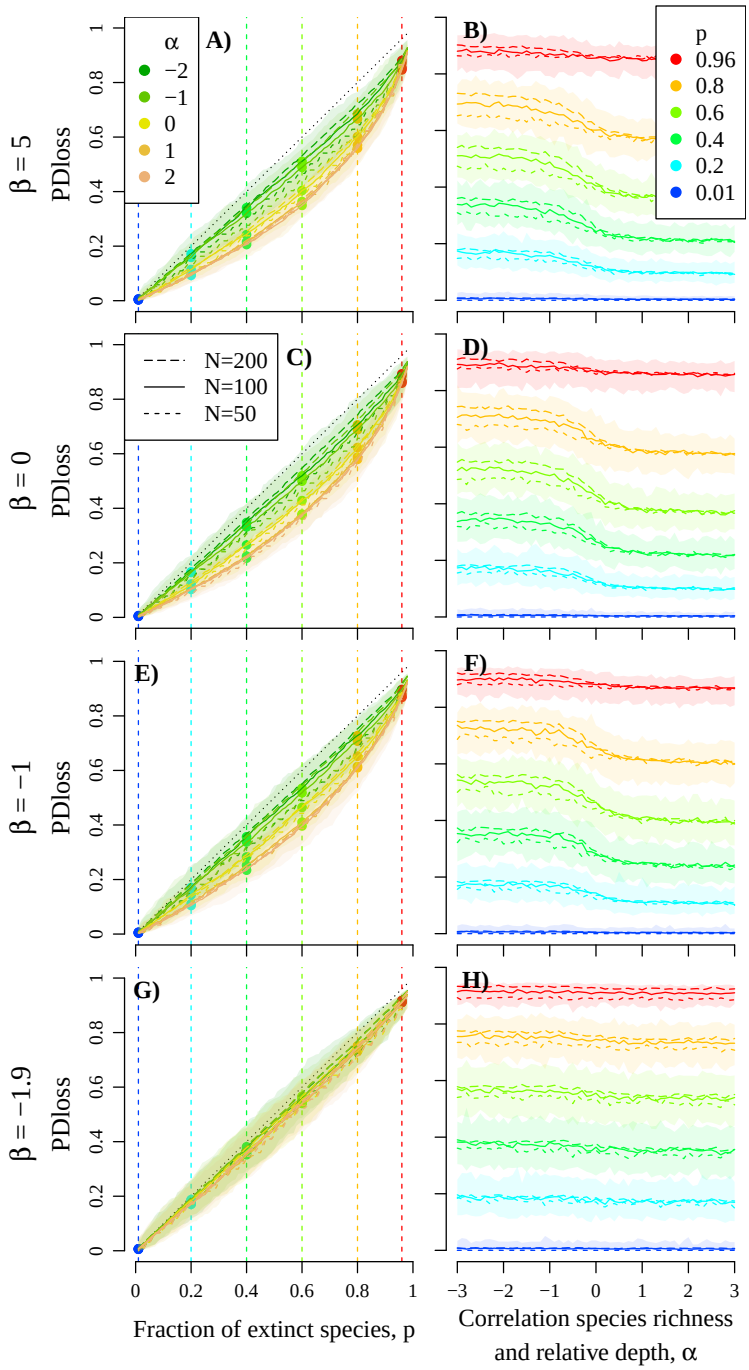


Figure A5: **Influence of tree size on the effect of the ranked tree shape (tree balance β and clades age-richness index α) on PD loss, for increasing fractions of species extinctions p.** Tree balance β changes from 5 (top row, 'bush trees') to -1.9 (bottom row, 'comb trees'). Tree size varies from $N = 50$ (short dashes) to $N = 200$ (long dashes); results with $N = 100$ (plain lines) are those shown on Figure 4. Results are shown either as a function of the extinction fraction p (left column; for different α values) or as a function of α (right column; for different extinction fractions p). Extinction fraction p increases from 0.01 to 0.96 (from left to right in A, C, E, G; from blue to red in B, D, F, H). The dotted lines in A, C, E, G show the bisector. Results are based on 100 simulation replicates: plain lines give median values and light areas give 95% confidence intervals. Other parameter values: $\epsilon = 0.001$.

Appendix 4. Sensitivity of the effect of the ranked tree shape (β, α) on PD loss to the model of node depths.

We tested for the influence of the model of node depths used to obtain ultrametric trees from ranked tree shapes. In our default model, node depths are set as in the Yule (birth) process (Yule, 1925). Here, similarly to Nee and May (1997), we modelled node depths as in the Kingman coalescent (Kingman, 1982). Figure A6 provides example of the resulting ultrametric trees.

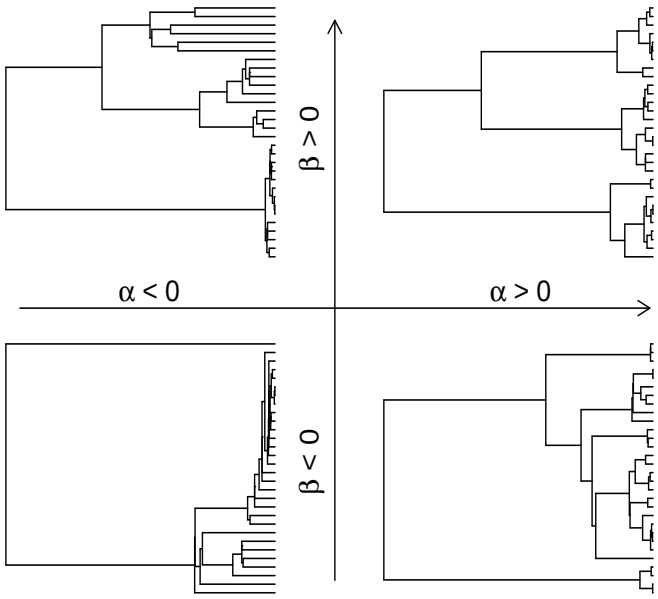


Figure A6: **Phylogenetic trees simulated for different values of β (tree balance) and α (clades age-richness index), with node depths set as in the Kingman coalescent.** Parameter values: $\beta = -1.5$ (bottom) or 10 (top), $\alpha = -10$ (left) or 10 (right), number of species $N = 30$, $\epsilon = 0.001$. Figure to be compared with Figure 2.

Comparing Figure A7 to Figure 4 shows, as already found by previous studies (Mooers et al., 2011; Lambert and Steel, 2013), that PD loss is increased when node depths follow the Kingman coalescent model, compared to the Yule process. However, this comparison also shows that the effects of the ranked tree shape (β, α) on PD loss are qualitatively similar with both models of node depths.

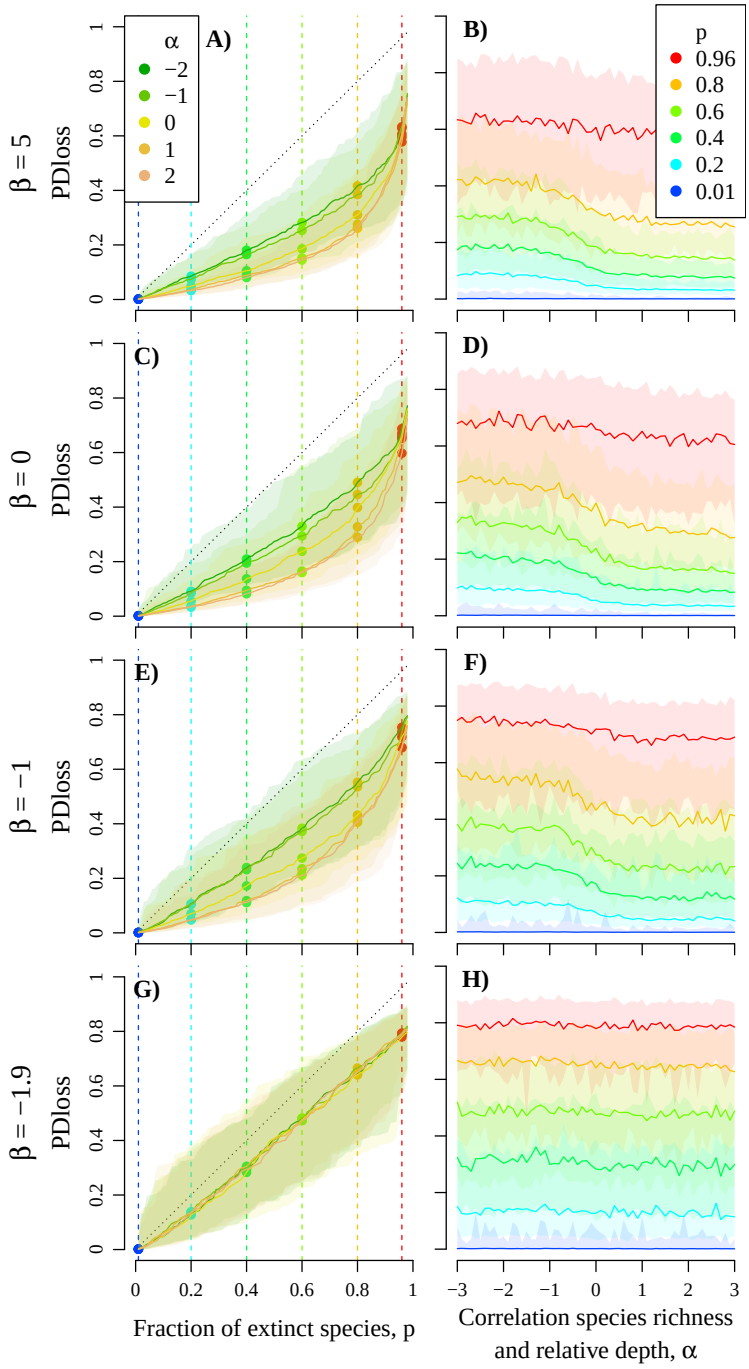


Figure A7: **Influence of the ranked tree shape (tree balance β and clades age-richness index α) on PD loss, for increasing fractions of species extinctions p, with node depths set as in the Kingman coalescent.** Tree balance β changes from 5 (top row, 'bush trees') to -1.9 (bottom row, 'comb trees'). Results are shown either as a function of the extinction fraction p (left column; for different α values) or as a function of α (right column; for different extinction fractions p). Extinction fraction p increases from 0.01 to 0.98 (from left to right in A, C, E, G; from blue to red in B, D, F, H). The dotted lines in A, C, E, G show the bisector. Results are based on 100 simulation replicates: plain lines give median values and light areas give 95% confidence intervals. Other parameter values: number of species $N = 100$, $\epsilon = 0.001$. Figure to be compared with Figure 4.

Appendix 5. Sensitivity of the effect of η on PD loss and on tree balance after species extinctions to tree size N.

We performed sensitivity analyses to test for the effect of tree size, simulating trees with N=50 and N=200 species. Fig. A8 and Fig. A9 show that tree size has little effect on the predictions of the model.

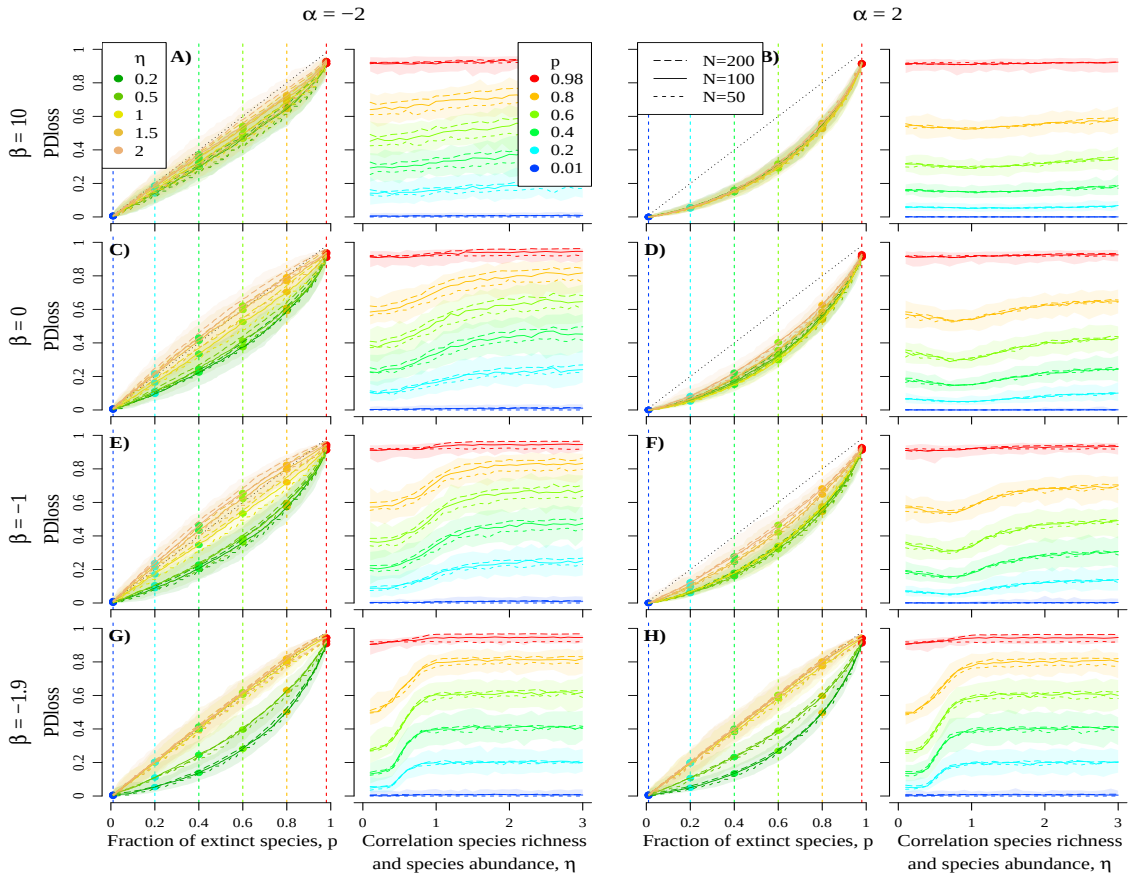


Figure A8: **Influence of tree size on the effect of η (clades abundance-richness index) on PD loss, for different ranked tree shapes and increasing fractions of species extinctions p.** Tree balance β ranges from 10 (top row, 'bush trees') to -1.9 (bottom row, 'comb trees'), and clades age-richness index α ranges from -2 (A, C, E, G) to 2 (B, D, F, H). Tree size varies from $N = 50$ (short dashes) to N=200 (long dashes). Results with $N = 100$ (plain lines) are those shown on Figure 5. Results are shown either as a function of the extinction fraction p (left side; for different η values, and with dotted lines showing the bisector) or as a function of η (right side; for extinction fractions p increasing from 0.01 to 0.98 from blue to red). Results are based on 100 simulation replicates: plain lines give median values and light areas give 95% confidence intervals. Other parameter value: $\epsilon = 0.001$.

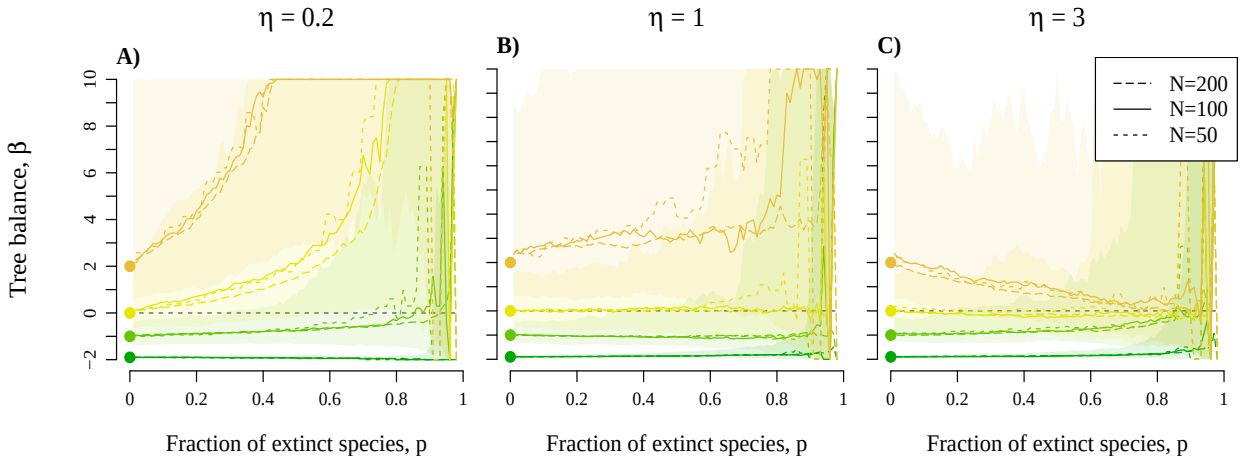


Figure A9: **Influence of tree size on the effect of η (clades abundance-richness index) on the balance of phylogenetic trees after extinctions.** Initial tree balance β ranges from 10 (brown dots and lines, 'bush trees') to -1.9 (green dots and lines, 'comb trees'). Tree size varies from $N = 50$ (short dashes) to N=200 (long dashes). Results with $N = 100$ (plain lines) are those shown on Figure 6. Extinction fraction p increases from 0.01 to 0.98 (from left to right). Results are based on 100 simulation replicates: plain lines give median values and light areas give 95% confidence intervals. Other parameter values: $\epsilon = 0.001$, $\alpha = 0$.

Appendix 6. Sensitivity of the effect of η on PD loss to the model of node depths.

Considering node depths as in the Kingman coalescent (similarly to Appendix 4) has quantitative effects on PD loss (reducing the latter as a result of shorter pendant edges), but no qualitative effects on the influence of η and its interactions with the ranked shape of species trees (Fig. A10).

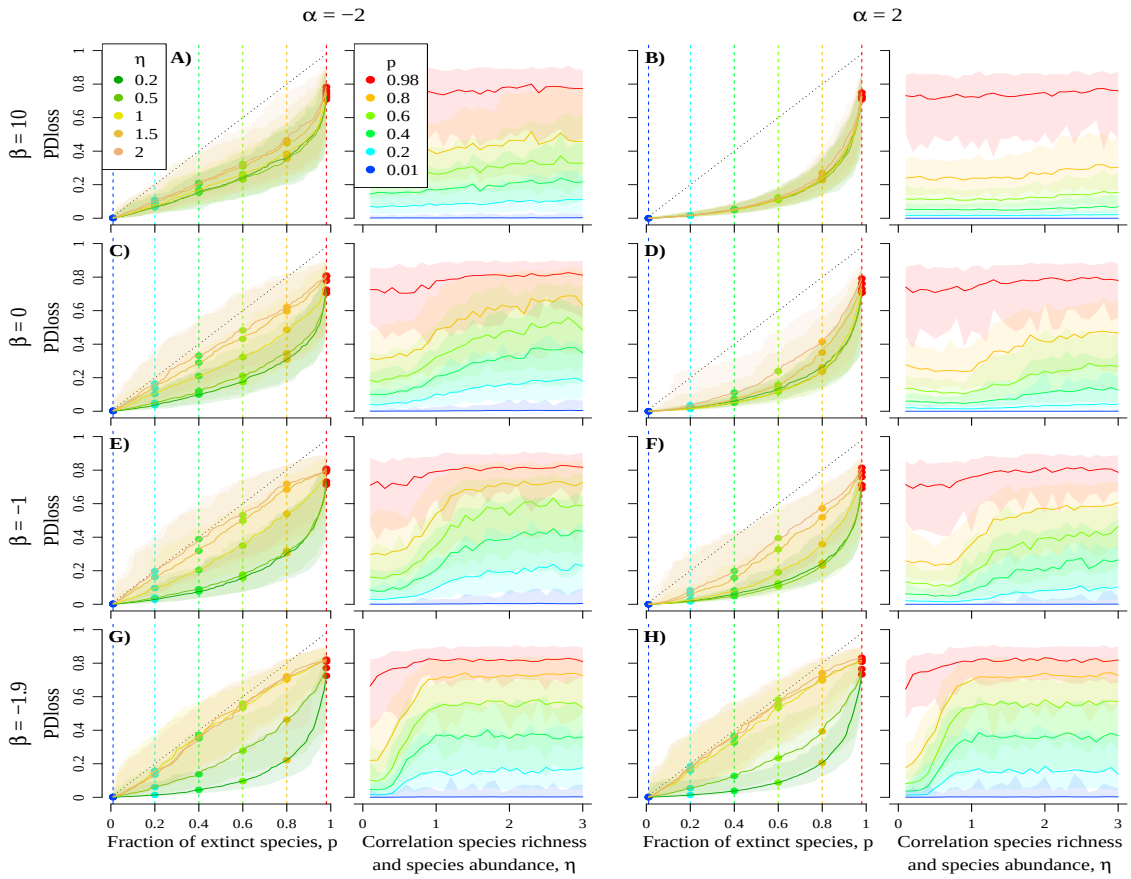


Figure A10: **Effect of η (clades abundance-richness index) on PD loss, for different ranked tree shapes and increasing fractions of species extinctions p, with node depths set as in the Kingman coalescent.** Tree balance β ranges from 10 (top row, 'bush trees') to -1.9 (bottom row, 'comb trees'), and clades age-richness index α ranges from -2 (A, C, E, G) to 2 (B, D, F, H). Results are shown either as a function of the extinction fraction p (left side; for different η values, and with dotted lines showing the bisector) or as a function of η (right side; for extinction fractions p increasing from 0.01 to 0.98 from blue to red). Results are based on 100 simulation replicates: plain lines give median values and light areas give 95% confidence intervals. Other parameter values: number of species $N = 100$, $\epsilon = 0.001$.

Appendix 7. Effect of the ranked tree shape (β, α) on PD loss, for different values of η.

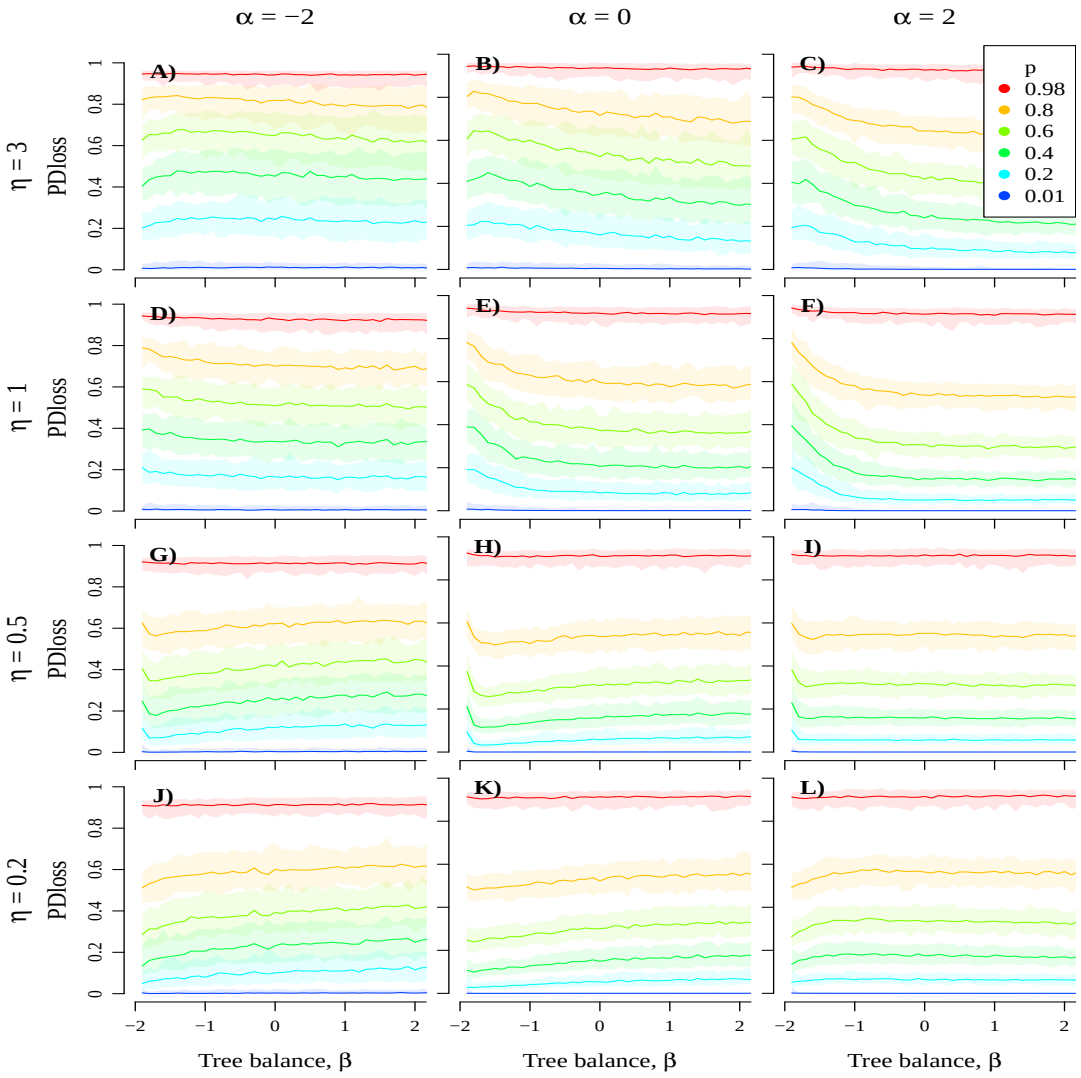


Figure A11: **Effect of tree balance β on PD loss, for different clades age-richness index α and clade abundance-richness index η, and increasing fractions of species extinctions p.** The clades abundance-richness index η ranges from 3 (top row) to 0.2 (bottom row), and the clades age-richness index α ranges from -2 (left column) to 2 (right column). Extinction fraction p increases from 0.01 to 0.98 (from blue to red). Results are based on 100 simulation replicates: plain lines give median values and light areas give 95% confidence intervals. Other parameter values: number of species $N = 100$, $\epsilon = 0.001$.

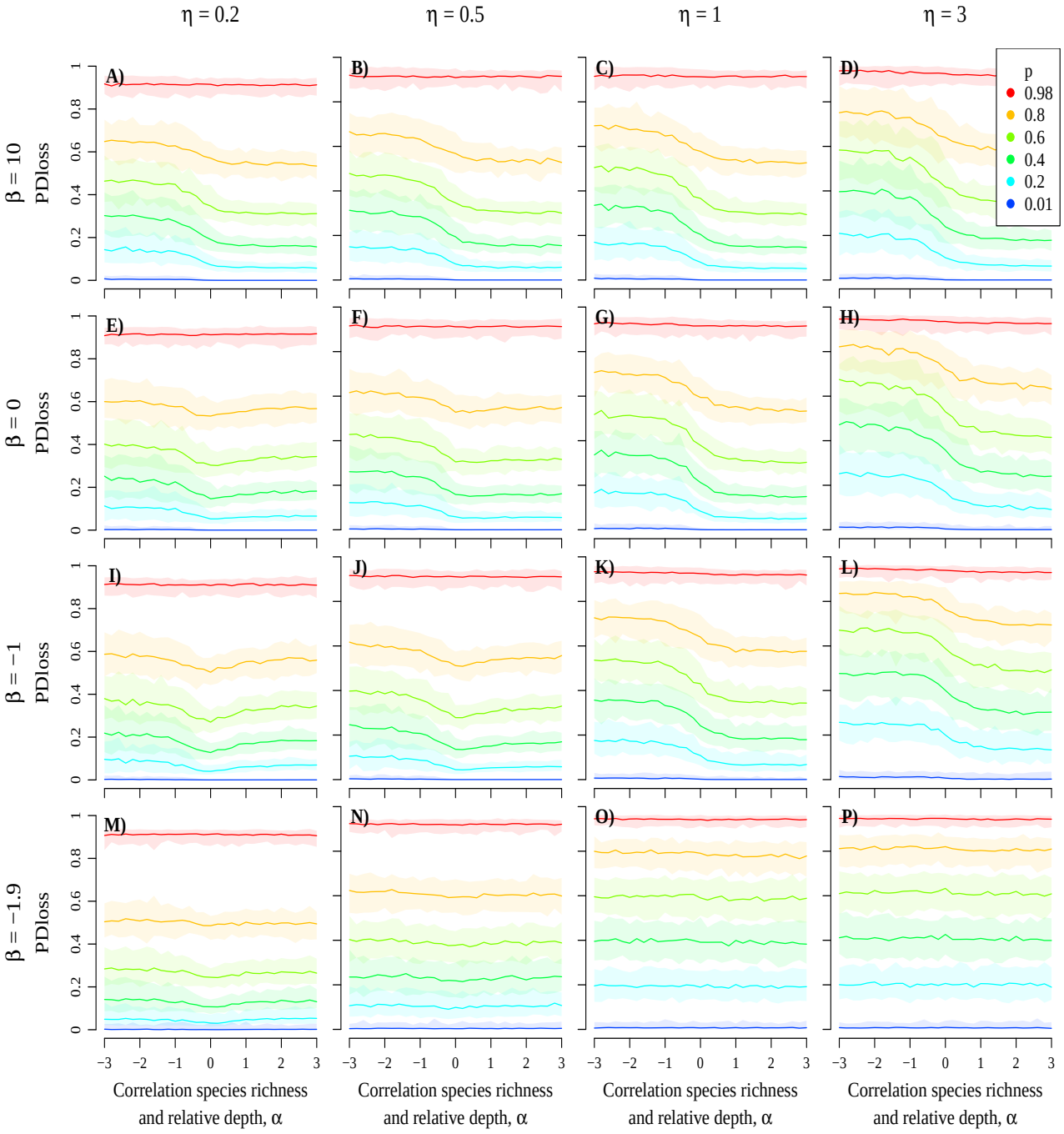


Figure A12: **Effect of the clades age-richness index α on PD loss, for different levels of tree balance β and clades abundance-richness index η, and increasing fractions of species extinctions p.** Tree balance β ranges from 10 (top row) to -1.9 (bottom row), and the clades abundance-richness index η ranges from 0.2 (left column) to 3 (right column). Extinction fraction p increases from 0.01 to 0.98 (from blue to red). Results are based on 100 simulation replicates: plain lines give median values and light areas give 95% confidence intervals. Other parameter values: number of species $N = 100$, $\epsilon = 0.001$.

Appendix 8. Parameter inference on simulated trees.

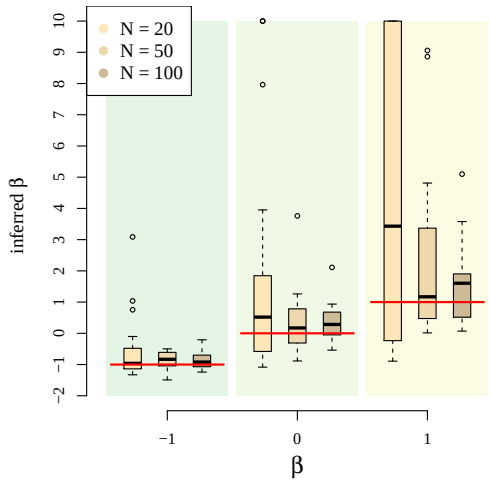


Figure A13: **Inference of the β parameter on simulated trees with tip abundances.** Simulations were run for all possible combinations of tree balance β in $\{-1, 0, 1\}$, cclades age-richness index α in $\{-1, 0, 1, 2\}$, clades abundance-richness index η in $\{0.2, 0.5, 1, 1.5, 2\}$, and tip number N in $\{50, 100, 200\}$. Results are based on 20 simulation replicates per parameter set. The inferred β is the maximum likelihood estimates of Aldous beta-splitting model (Aldous, 1996), which does not depend on α nor η, thus all trees with similar β values and tip number in the simulations are put together. The red bar indicates the true value used in simulations, the bloxes bars are the interquartile of the infered values, and whisquers show the most distant infered values that are less than ($1.5\times$box size) away from the box. In all simulations ϵ was set to 0.001.

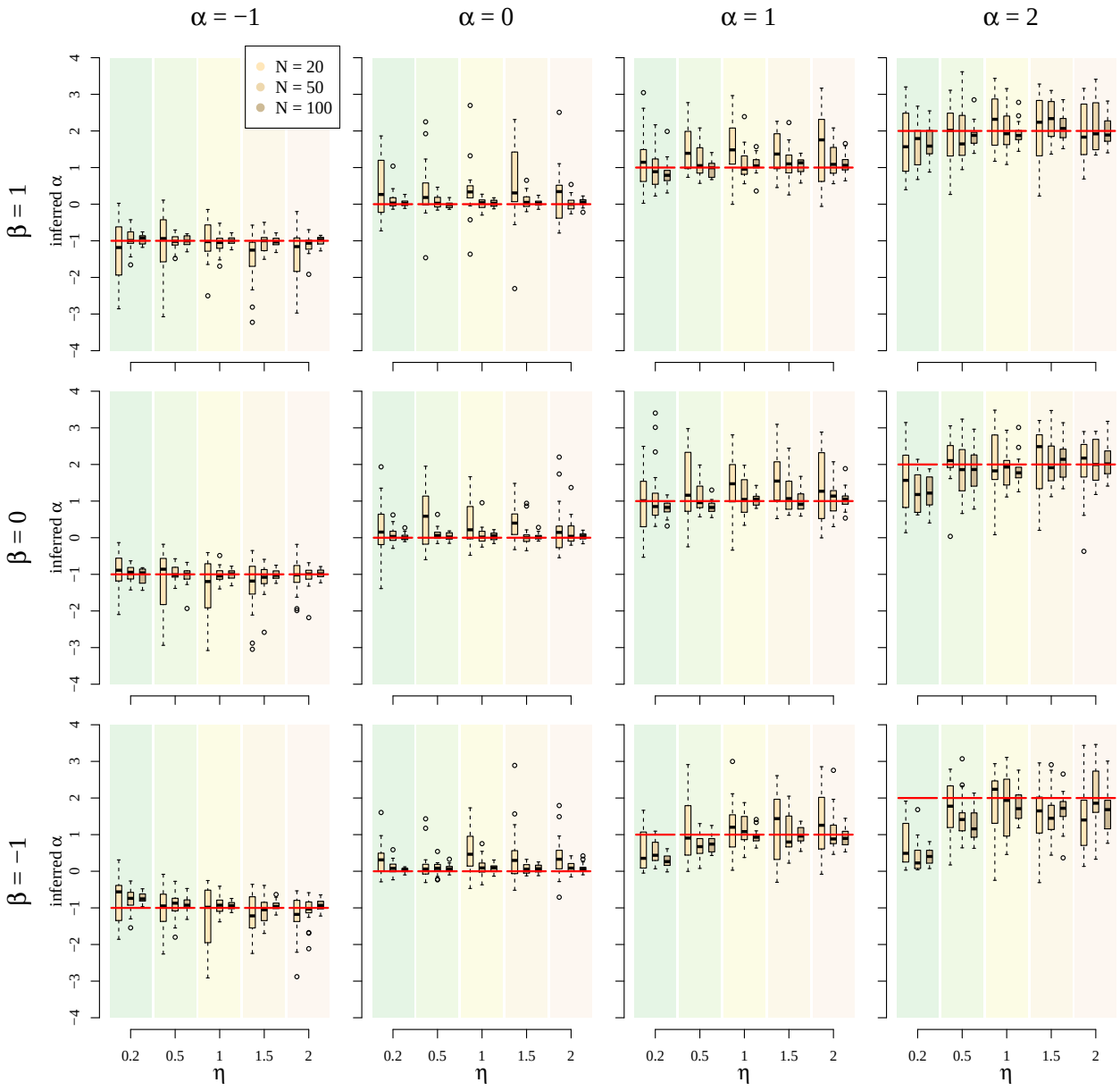


Figure A14: **Inference of the α parameter on simulated trees with tip abundances.** Simulations were run for all possible combinations of tree balance β in $\{-1, 0, 1\}$, clades age-richness index α in $\{-1, 0, 1, 2\}$, clades abundance-richness index η in $\{0.2, 0.5, 1, 1.5, 2\}$ and tip number N in $\{50, 100, 200\}$. Results are based on 20 simulation replicates per parameter set. The red bar indicates the true value used in simulations, the bloxes bars are the interquartile of the infered values, and whisquers show the most distant infered values that are less than ($1.5 \times$ box size) away from the box. In all simulations and inferences, ϵ was set to 0.001.

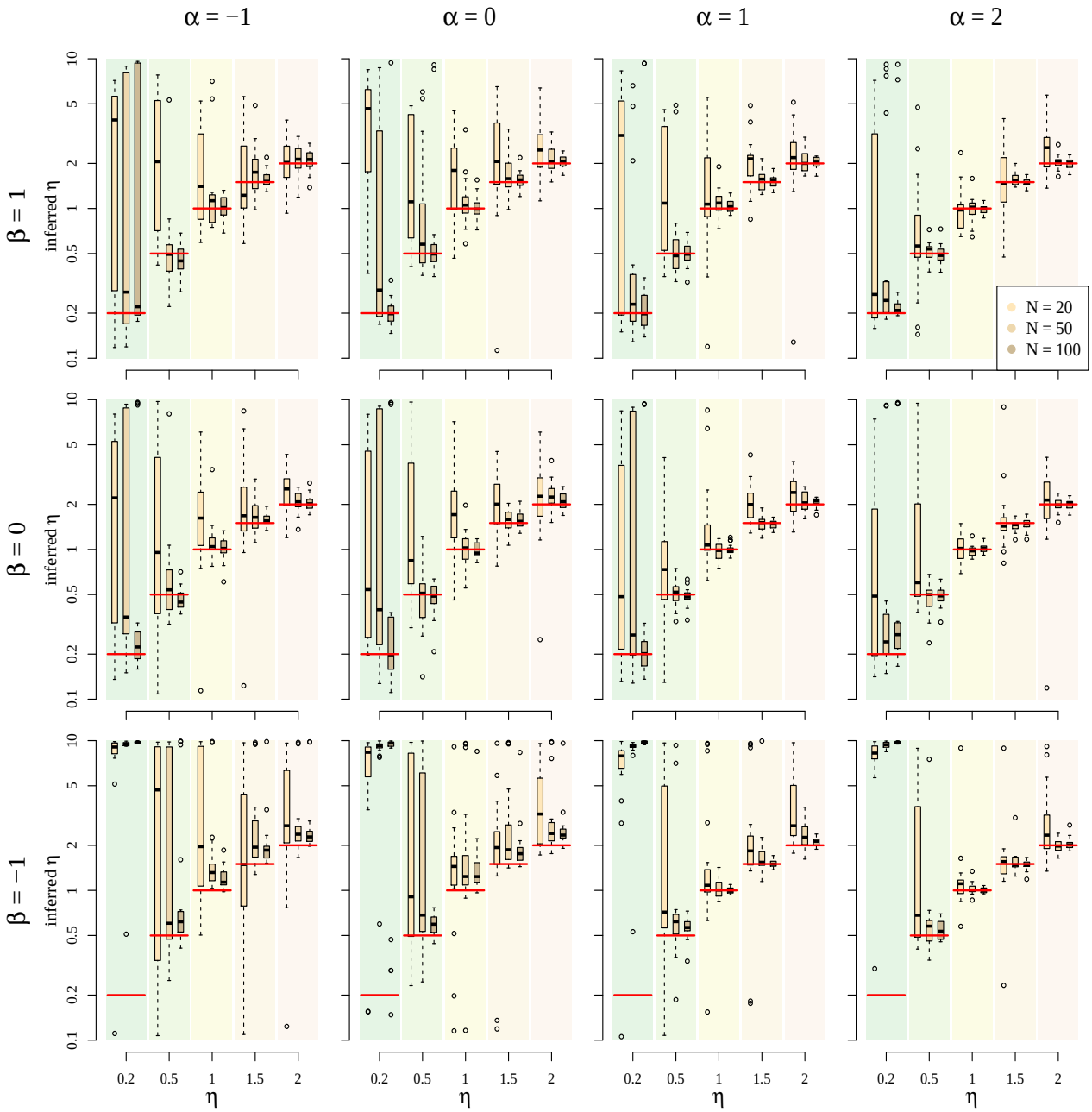


Figure A15: **Inference of the η parameter on simulated trees with tip abundances.** Simulations were run for all possible combinations of tree balance β in $\{-1, 0, 1\}$, clades age-richness index α in $\{-1, 0, 1, 2\}$, clades abundance-richness index η in $\{0.2, 0.5, 1, 1.5, 2\}$ and tip number N in $\{50, 100, 200\}$. Results are based on 20 simulation replicates per parameter set. The red bar indicates the true value used in simulations, the bloxes bars are the interquartile of the infered values, and whisquers show the most distant infered values that are less than (1.5 × box size) away from the box. In all simulations and inferences, ϵ was set to 0.001.

Appendix 9. Parameter inference on bird family trees.

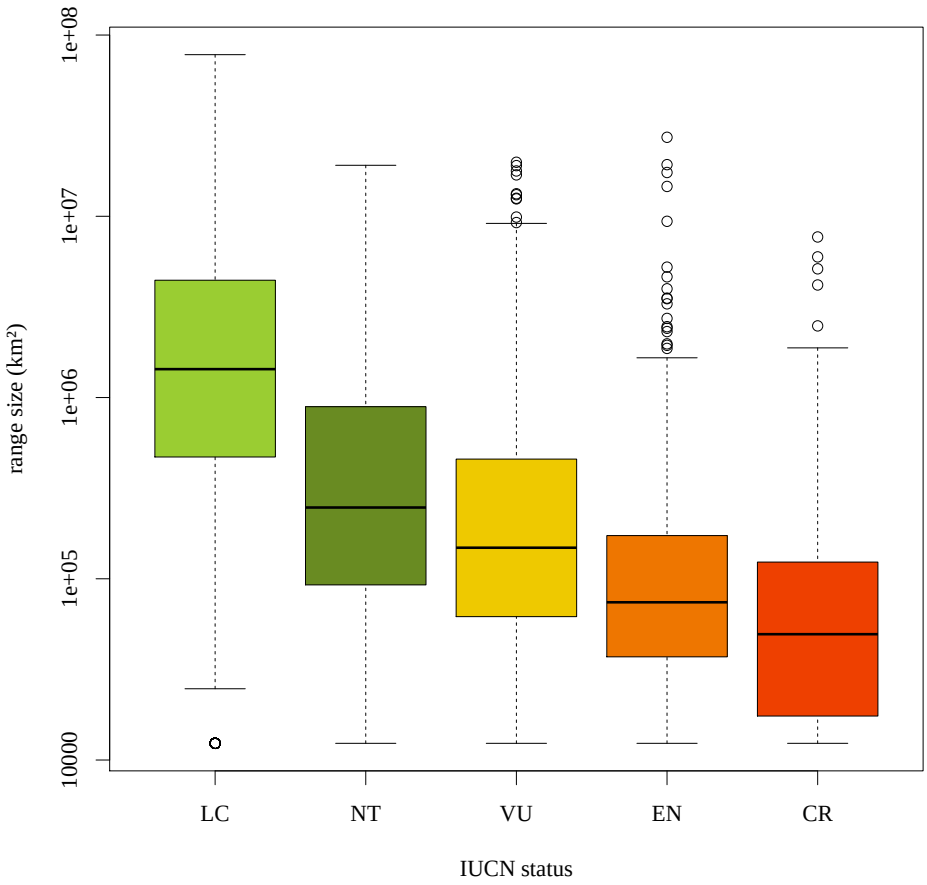


Figure A16: **Range sizes of bird species with respect to IUCN status.** The boxes show the quartiles of the range size distribution, and whisquers show the most distant points that are less than ($1.5 \times$ box size) away from the box. Nearly all means are significatly different from one another (p-value of the ANOVA test 2e−16, p-values of the Tukey SHD test all are bellow 1e−10, except between EN and CR for which it is 0.054). The tests where performed using the functions aov and TukeySHD for the R package `stats` (R Development Core Team, 2012). Species for which we have no information about their IUCN status are not shown on this figure.

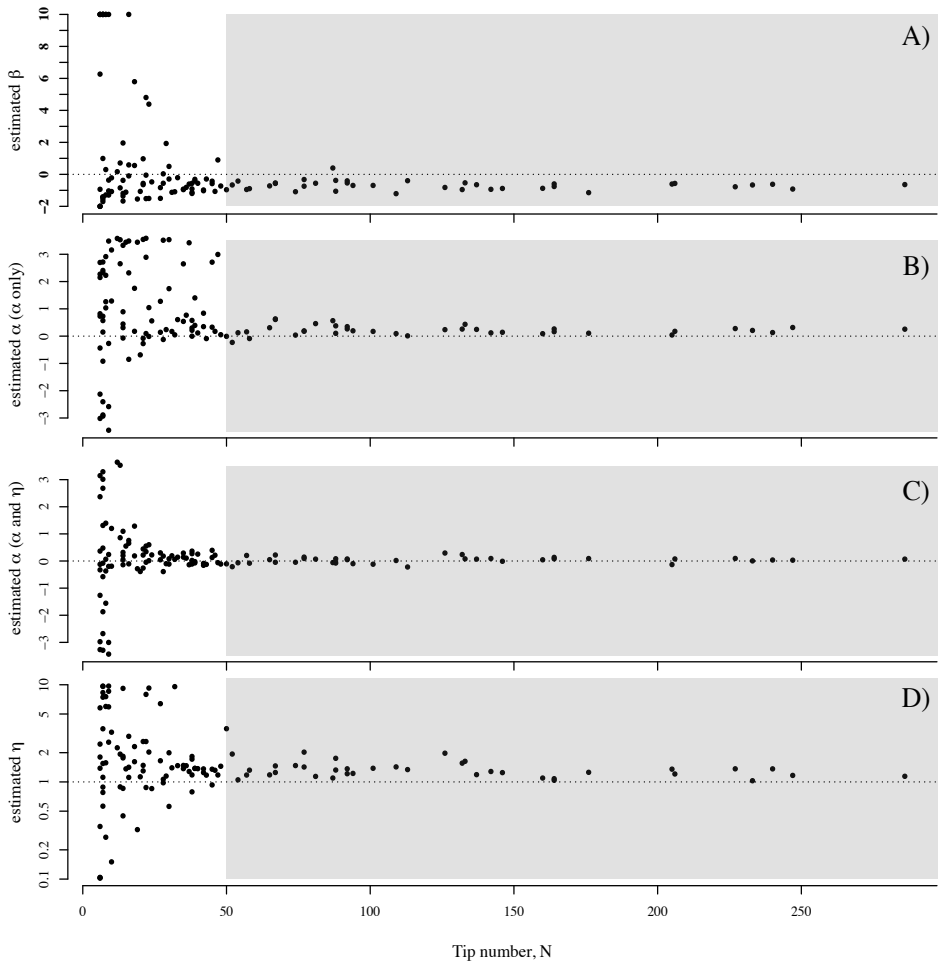


Figure A17: **Inference of the model parameters on bird family trees according to number of tips in trees,** N**.** Panels show the maximum likelihood estimates of β **A**), the posterior maximum of α when α is inferred alone **B**), the posterior maximum of α when α is inferred together with η **C**), and the posterior maximum of η **D**). The trees shown in Fig. 9 in the main text are the one within the grey zone (trees with at least $N = 50$ tips). The red bar indicates the true value used in simulations, and the black bars are the interquartile of the infered values. In all inferences, ϵ was set to 0.001.

Appendix 10. List of R functions.

The following R functions are available in the R package `apTreeshape` (Bortolussi et al., 2006) to simulate our model and infer its parameter on phylogenies. More details are given in the `apTreeshape` documentation.

simulate_tree	Simulates a tree topology for a given β, α and η.
simulate_kingman	Simulates a tree with topology drawn for a given β, α and η and node depths Kingman's coalescent.
simulate_yule	Simulates a tree with topology drawn for a given β, α and η and node depths from a birth-death process.
get_PD_sample	Computes the proportion of conserved phylogenetic diversity as a function of the proportion of conserved species, in trees simulated by the model.
get_tree_beta	Computes the maximum likelihood estimate of the parameter β in trees simulated by the model, as a function of the proportion of conserved species.
mcmc_alpha	Inference function for the α parameter.
mcmc_eta	Inference function for the α and η parameters.

Chapter 2 : Quantifying diversification rate heterogeneity in empirical phylogenies

Empirical phylogenies are typically much more imbalanced that what is expected from homogeneous speciation models. While the hypothesis of common diversification rates along lineages makes sense when the studied clade is small enough, it more difficult to believe for large groups within which species are likely to have very different life history traits and evolutionary histories. Previous phylogenetic approaches for detecting changes in diversification rates across a phylogenetic tree have focused on 'major' rate shifts, with the underlying idea that few rare events, such as key innovations, facilitate the invasion of new adaptive zones, a drastic impact on diversification rates. Another view of evolution is that speciation and extinction rates may vary gradually across lineages as a response to the particular biotic and abiotic environment experienced by each lineage. Such changes in diversification rates likely occur far more frequently than key innovations, resulting in heterogeneous diversification rates at much finer taxonomic scales. In this paper, we develop a new Bayesian approach for estimating lineage specific diversification rates. Our approach is based on a birth-death diversification process where diversification rates are inherited at speciation, but with a shift. We test it on a large simulated dataset of simulated phylogenies to assess its statistical performances, and show that it is able to accurately infer both the way shift happen at speciation events – how constrained they are by their parental value, and whether there is a trend in rates evolution – and the lineage specific diversification rates. The latter is a critical step to our understanding of the processes that lead some species groups to diversify faster than others. We then apply our method to time-calibrated phylogenies for 42 birds clades. This analysis reveals a pervasive pattern of declines in speciation rates over time congruent with previous studies, together with a remarkable heterogeneity in speciation rates. We show that the variability is comparable within and between clades, suggesting that rate variation may be much more gradual than currently thought and implemented in existing models. Our results emphasize the need to consider diversification models that embrace the pervasive heterogeneity of the evolutionary process.

This Chapter corresponds to a work done with during this PhD together with Hélène Morlon and Florian Hartig. The manuscript is in the final preparation stage and will be submited very soon.

Many gradual versus few punctual shifts in diversification during evolution

Odile Maliet, Florian Hartig, Hélène Morlon

Abstract

Understanding how and why diversification rates vary through time, space, and across species groups is key to understanding the emergence of today's biodiversity. Phylogenetic approaches aimed at identifying variations in diversification rates during the evolutionary history of clades have focused on exceptional shifts subtending evolutionary radiations. While such shifts have undoubtedly affected the history of life (Alfaro et al., 2009), they may hinder the importance of smaller but more frequent changes. We develop ClaDS, a new Bayesian approach for estimating branch-specific diversification rates on a phylogeny, that relies on a model with continuous changes in diversification rates at each speciation event. We show using Monte-Carlo simulations that the approach performs well at inferring both gradual and abrupt changes in diversification. Applying our approach to bird phylogenies covering the entire avian radiation, we find that diversification rates are remarkably heterogeneous within evolutionary restricted species groups. Some groups such as *Accipitridae* (hawks and allies) cover almost the full range of speciation rates found across the entire bird radiation. As much as 76% of the variation in branch-specific rates across this radiation is due to intra-clade variation, suggesting that the evolution of life proceeds with many gradual changes rather than few punctuated ones.

Manuscript

Understanding if, how and why speciation and extinction rates vary through the tree of life is key to our understanding of the diverse processes that shaped today's biodiversity. These questions have fostered the development of several phylogenetic approaches for detecting when and on which lineages diversification rates have changed during the evolutionary history of clades (Chan and Moore, 2004; Alfaro et al., 2009; Morlon et al., 2011; Rabosky, 2014). Previous phylogenetic approaches for detecting changes in diversification rates across a phylogenetic tree have focused on 'major' rate shifts, with the underlying idea that few rare events, such as key innovations, facilitate the invasion of new adaptive zones, with a drastic impact on diversification rates (Miller, 1949; Hunter, 1998). In these models, outside of few remarkable events, diversification rates are assumed to be homogeneous. However, while major rate shifts linked to key innovations have undoubtedly affected the history of life (Alfaro et al., 2009), they are not the only – nor necessarily the most important – source of variation in diversification rates.

Another view of evolution is that shifts in diversification rates are widespread. Speciation and extinction rates may vary gradually across lineages as a response to the particular biotic and abiotic environment experienced by each lineage (Benton, 2009); they may also vary as a response to traits that affect reproductive isolation such as reproduction mode Goldberg et al. (2010) or pollination and dispersal syndromes Onstein et al. (2017). Such changes in diversification rates

likely occur far more frequently than key innovations, resulting in heterogeneous diversification rates at much finer taxonomic scales (Jetz et al., 2012). Accounting for such finer scale heterogeneity is crucial if we want to obtain refined estimates of lineage-specific diversification rates and to better understand the processes subtending heterogeneity in the diversification of life. Here, we develop a new Bayesian approach (ClaDS) for estimating lineage-specific diversification rates on a phylogeny that better accounts for the diverse sources of variation in diversification rates that occur during the evolutionary history of clades. Using Monte Carlo simulations, we quantify the ability of ClaDS to faithfully recover both gradual and abrupt changes in diversification rates. Finally, we apply the method to time-calibrated phylogenies for 42 bird clades to evaluate the extent to which differences in the pace of diversification across the entire avian radiation result from few punctual versus many gradual events.

A gradual model of diversification rate variation.

We consider a birth-death diversification process, the cladogenetic diversification rate shift (ClaDS) model, where diversification rates are inherited at speciation, but with a shift (Fig. 1). At the beginning of the process, the clade is composed of one lineage with speciation rate λ_0 and extinction rate μ_0. At each speciation event, the two daughter lineages inherit new diversification rates (λ_{i1}, λ_{i2}) and (μ_{i1}, μ_{i2}) sampled from a joint probability distribution ν parameterized by the parental rates λ_i and μ_i. If the change in speciation and extinction rates are assumed to be independent, the λ_i are sampled from a distribution ν_λ, the μ_i are sampled from a distribution ν_μ, and $\nu = \nu_\lambda \times \nu_\mu$. Moreover, we allow for the possibility that some extant species are missing by assuming that each extant species is observed with probability $f \leq 1$. We derive the probability density of a reconstructed phylogeny under this general model (Materials and Methods & SI Appendix).

We then consider several scenarios in ClaDS where: i) ν_λ is a lognormal distribution with parameters $\log(\alpha * \lambda)$ and σ; the latter ensures that the relative change in rate at speciation λ_i/λ is independent from the parental rate with a mean m given by $\alpha \exp(\sigma^2/2)$; σ controls how constrained daughter rates are (highly constrained for small σ values) and α controls the trend at speciation (i.e. whether daughter rates tend to be higher or lower than parental rates) ii) extinction rates are either negligible ($\mu_i = 0$ for all lineages, ClaDS0), homogeneous across all lineages in the clade ($\mu_i = \mu_0$ for all lineages, ClaDS1) or vary across lineages, but with a constant turnover ε (i.e. $\mu_i/\lambda_i = \varepsilon$ for all lineages, ClaDS2). We use Monte Carlo simulations under ClaDS1 and ClaDS2 (Materials and Methods) to verify that our likelihood expression is

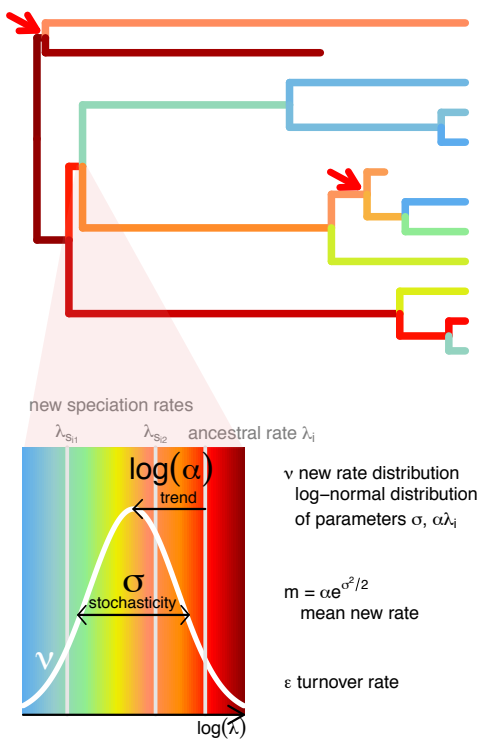


Figure 1: Illustration of the cladogenetic diversification rate shift model (ClaDS). Upper panel: cartoon phylogeny simulated under ClaDS, with branches colored according to their speciation rate (red: high rate, blue: low rate). Speciation rates are inherited at speciation with a shift determined by the probability distribution ν_λ (here taken to be a lognormal distribution, insert). Red arrows indicate speciation events (and associated diversification rate shifts) that are hidden in the reconstructed phylogeny as a result of extinction.

correct (SI Appendix, Fig. S3 & S4). Finally, we implement a Monte Carlo Markov Chain (MCMC) sampler that, given a reconstructed phylogeny, simultaneously estimates both the parameters of ClaDS (λ_0, α, σ, and either μ_0 or ε) and the speciation rates λ_i for each branch i of the phylogeny (Materials and Methods). Branch-specific extinction rates μ_i are given by μ_0 for ClaDS1 and by $\varepsilon * \lambda_i$ for ClaDS2.

Under these scenarios of the ClaDS process, heterogeneity in speciation rates across lineages is determined on the one hand by a stochastic component (controlled by σ), and on the other hand by a trend component (controlled by m). When the expected daughter rate is equal to the parental rate ($m = 1$), the resulting trees are relatively imbalanced and tippy (SI Appendix, Fig. S1 & S2): lineages that by chance have high speciation rates early in clade's history spread, leading to rates that are heterogeneous across lineages and average rates that increase through time. This sorting effect is exacerbated when the expected daughter rate is higher than the parental rate ($m > 1$, Fig. S1 & S2), corresponding to a 'niche-piling' scenario where diversity begets diversity (Emerson and Kolm, 2005). To the contrary, when the expected daughter rate is lower than the parental rate ($m < 1$), corresponding to a 'niche-filling' scenario where diversification gets harder as new species arise (Rabosky and Lovette, 2008b; Phillimore and Price, 2008; Moen and Morlon, 2014), the heterogeneity in speciation rates across lineages is reduced, and with a low enough m, the average rate is constant or even decreasing through time (Fig. S1 & S2). Importantly, ClaDS is able to produce the combination of stemmy and imbalanced tree shapes observed in nature, and under a wider set of parameter values for the scenario with constant turnover (ClaDS2) than the scenario with constant extinction rate (ClaDS1, Fig. S1 & S2).

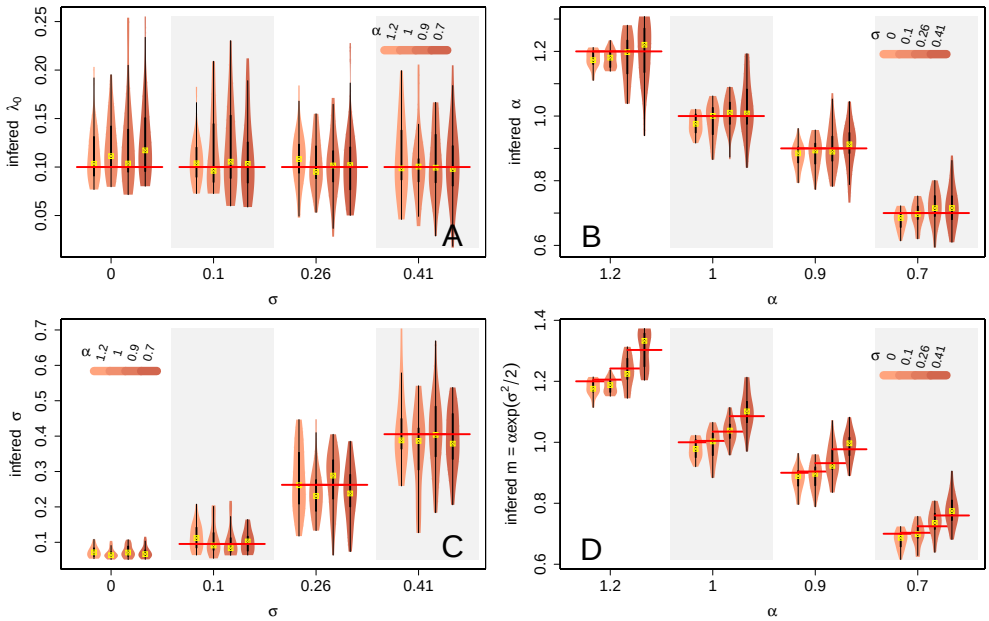


Figure 2: Recovery of ClaDS parameters. Estimated λ_0 (A), α (B), and σ (C) inferred with ClaDS, and (D) resulting estimation of $m = \alpha * \exp(\sigma^2/2)$. Violin plots: distribution of estimated parameters; yellow cross: median; thick black line: quartiles; red lines: values used in the simulations. Different shades of brown correspond to: in A and C, the values of α used in the simulations (1.2 (light), 1, 0.9, 0.7 (dark)) ; in B and D, the values of σ used in the simulations (0 (light), 0.1, 0.26, 0.41 (dark)). Results corresponding to simulated trees of size 200; results for other tree sizes are shown in Fig. S5 to S8.

Performance of ClaDS.

We begin by testing the performance of ClaDS under gradual rate changes and in the absence of extinction (ClaDS0) (Materials and Methods). We find that the approach provides unbiased estimates of all model's parameters for large enough trees (size 200, Fig. 2); the relative change in rate at speciation m is also well estimated (Fig. 2 D). As expected, bias and variability around parameter estimates increase for smaller trees (Fig. S5 to S8).

ClaDS provides reliable estimates of branch specific speciation rates on average: while low rates tend to be slightly overestimated and large rates slightly underestimated, ClaDS can detect regions of the tree with relatively high or low rates (Fig. 3 & Fig. S9-S10).

When considering also extinctions, focusing on the scenario with constant turnover (ClaDS2) as it generally produced tree shapes closer to those observed in nature, we found that estimates remain accurate at low levels of extinction ($\varepsilon = 0.1$) for both model parameters (Fig. S14) and branch-specific speciation rates (Fig. S15). At high levels of extinction ($\varepsilon = 0.9$), σ and, when the mean change in rate at speciation m approaches 1, branch-specific speciation rates, remain well estimated. It is not the case, however, of the turnover rate ε, α, and branch-specific speciation rates when $m < 1$, although accounting for extinction does improve inferences over ignoring it (Fig. S14 & S15). When extinction is not accounted for, estimated branch-specific speciation rates are generally lower than realized ones, but higher than realized net diversification rates (Fig. S15C & D).

Finally, if there are a small number of major rate shifts during the evolution of clades, rather than gradual changes (tested here with a single rate shift, Materials and Methods), ClaDS is still

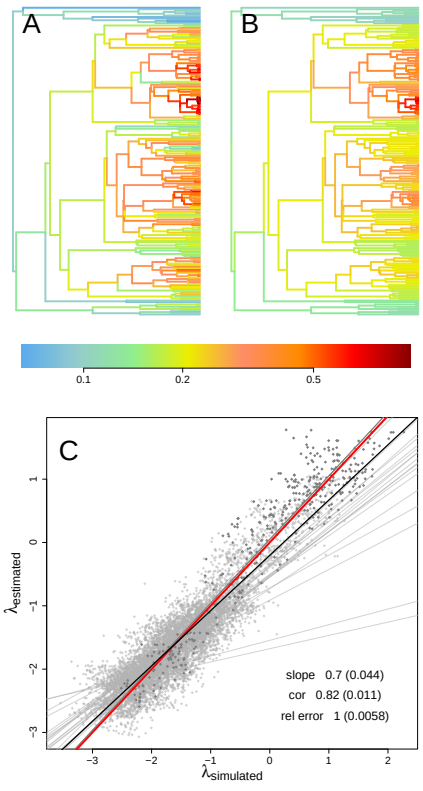


Figure 3: ClaDS performs well in recovering branch-specific speciation rates A) tree simulated under the ClaDS model ($\lambda_0 = 0.1$, $\sigma = 0.18$, $\alpha = 1$, $\varepsilon = 0$, size $N = 200$), with branches colored according to their realized speciation rate B) same tree with branches colored according to inferred speciation rates B) Inferred versus simulated branch-specific speciation rates (on a log scale) for 20 trees simulated with the same parameters and size as the tree from panel A; the darker points highlight rates for the tree shown in panel A. Each regression line (light gray) corresponds to one of the 20 trees, and the black line corresponds to the regression across all trees. The red line displays the 1:1 relationship. Values in the bottom right corner correspond to the mean and standard deviation of the slope and correlation coefficient across the 20 regressions, and those of the relative error in branch-specific speciation rates estimates ($\lambda_{\text{estimated}}/\lambda_{\text{simulated}}$) across all branches from the 20 trees

able to provide reliable estimates of branch-specific rates (Fig. S12). The model is also able to detect when two branches in the tree belong to distinct speciation regimes as soon as the difference in rates between the two regimes is large enough (a two-fold increase or decrease in our simulations) and both regimes are represented by a large enough number of branches in the phylogeny (Fig. S13A). The false detection rate associated to this test is low (Fig. S13B).

Diversification across the avian radiation. When applying ClaDS to major bird clades (Materials and Methods), we found that lineage-specific speciation rates can vary by as much as 2 orders of magnitude within clades (Fig. 4E). In *Accipitridae* (hawks and allies) for example, speciation rates range from 0.013 to 1.2 Mya^{-1}, which almost covers the range found across the entire avian radiation ($0.013 - 5\ \text{Mya}^{-1}$). Comparable within-clade heterogeneities occur in other clades, such as *Muscicapidae* & *Turdidae*, *Tyrannidae* and *Parulidae* (Fig. 4E, in orange). A variance partitioning of speciation rates across the bird radiation (Material and Methods) reveals that intra-clade variance accounts for 76% of the total variance. Yet there are also some clades that

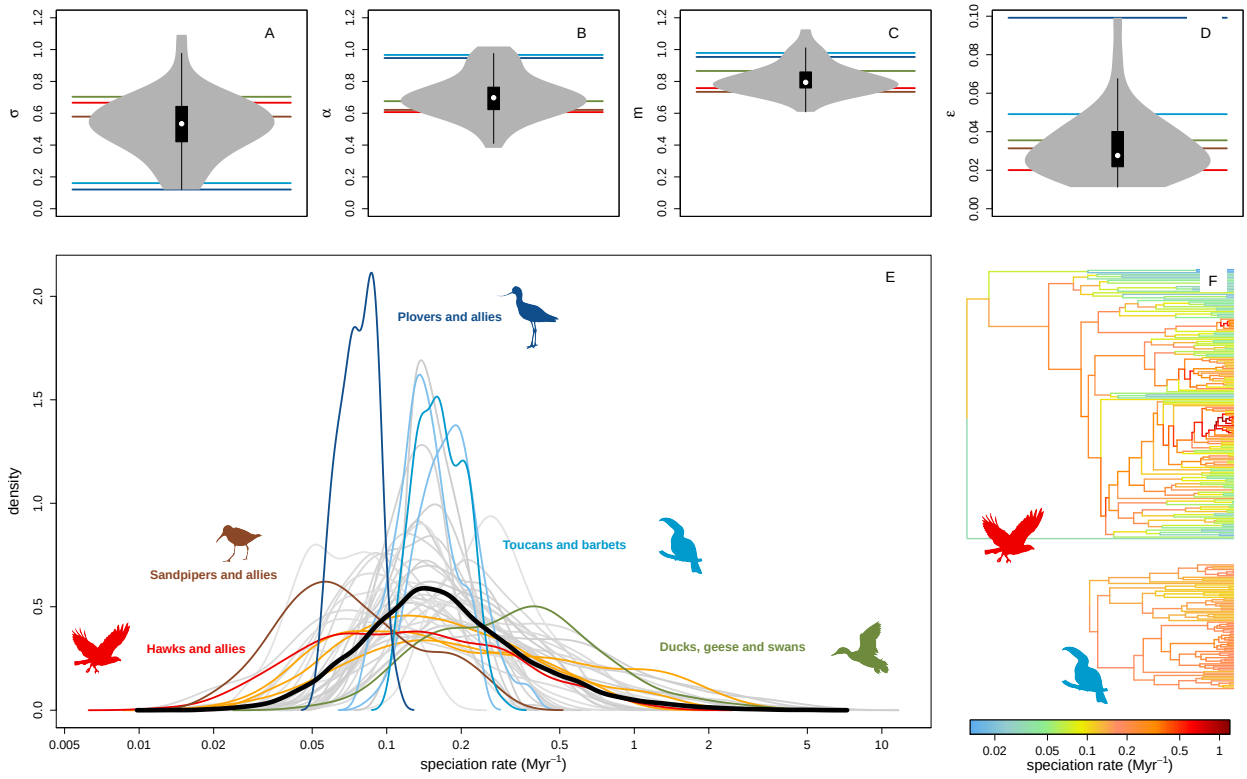


Figure 4: Patterns of diversification across 42 bird clades. Distributions across clades of (A) σ, (B) α, (C) $m = \alpha \exp(\sigma^2/2)$, and (D) ε values estimated with ClaDS. E: Distributions of branch specific speciation rates for each specific clade (grey and colored lines) and all clades pooled together (thick black line). Red: *Accipitridae*; Orange: *Muscicapidae* & *Turdidae*, *Tyrannidae* and *Parulidae*; Dark blue: *Charadrii*; Medium blue: *Ramphastides*; Light blue : *Alcedinidae* and *Phasianidae*); Brown: *Scolopaci*; Green: *Anatinae* F: Exemplar phylogenies colored according to their inferred branch-specific speciation rates, in Myr^{-1}, and plotted on the same time scale. Top panel: the *Accipitridae* phylogeny subtends very variable rates that tend to decrease through time (inferred parameters: $\sigma = 0.67$, $\alpha = 0.61$, $m = 0.76$ and $\varepsilon = 0.02$). Bottom panel: the *Ramphastides* phylogeny subtends rather homogeneous rates ($\sigma = 0.16$, $\alpha = 0.97$, $m = 0.98$ and $\varepsilon = 0.05$).

are quite homogeneous, such as *Ramphastides*, *Alcedinidae*, *Charadrii* and *Phasianidae* (Fig. 4E, in blue). We did not find any significant relationship between the variance in rate values within a clade and the size ($p = 0.49$) or age ($p = 0.93$) of the clade, indicating that rate heterogeneity is not a mere result of time or species richness; rather, rates are pretty constrained in some old and rich clades (e.g. *Phasianidae*) as well as in some younger or less species-rich clades (e.g. *Alcedinidae*), while they can take very different values for distinct species of both old or young clades (e.g. *Parulidae*, *Tyrannidae*). The wide range of σ estimates found across bird clades (Fig. 4A), in comparison with rather tight α and m estimates (Fig. 4B & C), suggests that differences in rate heterogeneity across clades are due to the stochastic component of the model, rather than its trend component. Indeed, α ranges between 0.38 and 1.02 (with a mean of 0.71, Fig. 4B), which indicates a universal tendency for daughter rates to be smaller than ancestral ones, with a decline that is comparable in magnitude across clades. There is only one case when m is clearly above 1 (1.12 in *Campephagidae*); this corresponds to a case when most shifts correspond to rate declines, but the few shifts that correspond to rate increases are much bigger in magnitude.

Discussion. Models of diversification applied to phylogenies of extant taxa are increasingly used to understand the long-term evolution of biodiversity. These approaches have highlighted how much variable diversification rates can be across the tree of life, and the importance of these variations for explaining current patterns of diversity (the so-called 'diversification rate hypothesis', Rosenzweig, 1992). Yet, despite recent advances in phylogenetic approaches for understanding diversification, detecting diversification rate variations and the processes underlying these variations remain a challenge spurring a heated debate (Rabosky and Goldberg, 2015; May and Moore, 2016; Moore et al., 2016; Rabosky et al., 2017; Rabosky, 2017). In this paper, we have developed ClaDS, a new model with gradual variations in diversification rates together with a method to infer branch-specific diversification rates on a phylogeny. We have shown using simulations that ClaDS accurately estimates branch-specific rates. Finally, applying ClaDS to the bird phylogeny, we have shown that gradual changes have been instrumental in shaping global rate variation during the avian radiation.

One of the major advances of our model is to rely on an explicit and exact computation of the likelihood in the presence of extinction. Previous likelihood expressions under diversification models with variable rates were computed with the underlying assumption that shifts do not occur in extinct lineages (Alfaro et al., 2009; Morlon et al., 2011; Rabosky, 2014); this is biologically implausible and can introduce an important bias depending on the intensity of extinction (Moore et al., 2016; Rabosky et al., 2017). In ClaDS we relax this inconvenient assumption by integrating appropriate Ordinary Differential Equations (ODEs, SI Appendix). This allows computing likelihoods accounting for rate shifts on extinct lineages, which has so far only been done through intense and impractical Monte Carlo simulations (Moore et al., 2016). The ODE integration is computationally intensive, but not as much as to prevent running ClaDS on reasonably sized trees, as we illustrated on the bird phylogenies. Despite this significant improvement, our simulations show that estimating extinction remains difficult, in line with the well-known difficulty of estimating extinction from phylogenies of only extant taxa (Rabosky, 2010). This is true even when simulations and inferences are performed under simple models with constant extinction or turnover rate. Hence, even though our ODEs are general, with extinction rates that can shift at each branching event similarly to speciation rates, we considered only the simple models with constant rates in our implementation; this has the advantage of reducing computation time. Despite difficulties in estimation extinction rates, properly accounting for extinctions in the likelihood computation is satisfying on a biological and theoretical standpoint, and, as we have shown, improves the estimation of both model parameters

and branch specific speciation rates.

Another advantage of ClaDS is to avoid using model selection to select the number and location of rate shifts, by assuming that shifts happen at each speciation event. In the frequently-used MEDUSA method (Alfaro et al., 2009), stepwise AIC is used to perform this selection, with associated statistical limitations (May and Moore, 2016). In the approach of Morlon et al. Morlon et al. (2011), likelihood ratio tests are performed to select the number of shifts, but the location of these shifts needs to be fixed *a priori*. Finally, in the popular Bayesian analysis of macroevolutionary mixtures (BAMM, Rabosky, 2014), reversible jump mcmc is used, with a prior on the number and location of shifts that may influence the results (Moore et al., 2016; Mitchell and Rabosky, 2017). ClaDS avoids these limitations, while still performing well in the presence of rare rate shifts with large effects.

Maybe more importantly than these technical aspects, ClaDS represents a view of evolution distinct from that of previous models: existing models focus on a small number of discrete diversification shift events spread across the tree, an idea that fits well with the concept of key innovations driving major diversification shifts (Morlon et al., 2011; Alfaro et al., 2009; Rabosky, 2014); to the contrary, ClaDS allows for more gradual variations linked, for example, to changes in environmental conditions or associations with continuously evolving heritable traits. Accordingly, ClaDS does not aim at identifying specific nodes in a phylogeny subtending major diversification rate shifts. Rather, it assumes that rate shifts happen at each speciation event and focuses on estimating branch-specific diversification rates. In nature, both many shifts with small effects and few shifts with large effects are likely to occur, and so it is reassuring to see that ClaDS can properly estimate branch specific rates under these two evolutionary processes.

Accurately estimating branch specific diversification rates is a critical step for understanding the processes that lead some species groups to diversify faster than others. For example, species' traits can modulate their propensity to diversify, and tests based on assessing the correlation between trait values at a phylogenies' tips and metrics capturing the diversification rate of the corresponding lineages ('tip-rate correlations' tests) have been developed to detect such effects (Freckleton et al., 2008). These types of tests have regained interest lately (see e.g. STRAPP Rabosky and Goldberg (2015), FiSSE (Rabosky and Goldberg, 2017), ES-sim (Harvey and Rabosky, 2017), pNoTO (Bromham et al., 2015; Hua and Bromham, 2016)), as an alternative or complement to state-dependent speciation-extinction (SSE) methods that jointly model diversification dynamics and trait evolution (Maddison et al., 2007; FitzJohn, 2012). However, current metrics of species-level diversification rates have limitations. Some of them are derived from BAMM (Rabosky, 2014) and thus reflect a limited set of diversification rate regimes rather than lineage-specific rates *per se*. Others are summary statistics describing phylogenetic branching patterns, such as the "node density" (Freckleton et al., 2008), the "equal split" (Redding and Mooers, 2006), or the "diversification rate" (Jetz et al., 2012) statistics, and are not rigorously derived from speciation-extinction models. ClaDS provides tip level estimates of diversification rates that should help identifying the specific features of a species that make it more or less prone to diversify. In the future, we could imagine a hybrid between SSE and ClaDS that would account for both trait-dependent diversification and residual rate variation not accounted for by the trait, in the spirit of hidden states models (HiSSE, Beaulieu and O'Meara, 2016). This could for example be done by imputing in ClaDS specific trend parameters α corresponding to trait shifts.

Changes in biotic and abiotic conditions can also modulate the tempo of diversification, leading diversification to be faster during some time periods than others. ClaDS accommodates temporal

trends in rate variation, without the need to specify a specific form for this variation *a priori* as in time-dependent diversification models (Nee et al., 1994b; Rabosky and Lovette, 2008b; Morlon et al., 2011), and with more flexibility than models where a discrete rate shift at a given time point affects the whole clade (Stadler, 2011). In the future, the trend parameter α could depend on measured environmental variables; this would allow directly testing for an effect of these environmental variables on diversification, as in environment-dependent diversification models (Condamine et al., 2013; Lewitus and Morlon, 2017), while accounting for residual rate variation.

Our ClaDS analysis of the avian radiation reveals a series of compelling results. First, and even though these estimates need to be taken with caution, we find significant (non-zero) turnover rates. Second, we find a pervasive pattern of declines in speciation rates over time congruent with previous studies (Rabosky and Lovette, 2008b; Phillimore and Price, 2008; Moen and Morlon, 2014). Third, we find a remarkable heterogeneity in speciation rates, with per-lineage rates that vary by two orders of magnitude ($0.01 - 5$ Mya^{-1}), peaking around 0.15 Mya^{-1}. Fourth, we find that variability in speciation rates can be as high within than between clades, suggesting that rate variation may be much more gradual than currently thought and implemented in existing models. Finally, we highlight a remarkable difference across clades in terms of how constrained their diversification rates are, with plovers and allies on one extreme, and hawks and allies on the other extreme of a continuum between rates that vary less than 2 fold to more than 80 folds (Fig. 4E, F). These differences in how constrained diversification rates are striking and remain to be explained: these could be linked to differences in genetic architecture, developmental constraints, or biogeographies, for example.

Together, our results refute the idea that speciation may be clock-like (Hedges et al., 2015) and emphasize the need to consider diversification models that embrace the pervasive heterogeneity of the evolutionary process. Further, they promise a bright future for approaches, such as ours, that relax the speciation clock similarly to the way the molecular clock has been relaxed (Thorne et al., 1998; Huelsenbeck et al., 2000; Lartillot et al., 2016): quite alike molecular rates, diversification rates vary according to many gradual shifts.

Material and Methods

Likelihood, simulation and Bayesian implementation of ClaDS.

Likelihood.— We derived the probability density of observing a reconstructed phylogeny with branches delimited by the times $(t_i, s_i)_{i\in[\![1,N]\!]}$ and with branch specific speciation and extinction rates λ_i and μ_i under the cladogenetic diversification rate shift model (SI Appendix). We note Θ the parameters of the new rate distribution ν. The probability density can be derived from three main probability functions: $\Phi_{\Theta,\lambda,\mu}(t)$, the probability that a lineage alive at time t has speciation and extinction rates λ and μ and no descendant in the reconstructed phylogeny; $\chi_{\Theta,\lambda,\mu}(t)$, the probability that a lineage alive at time t has speciation and extinction rates λ and μ and exactly one descendant species sampled in the reconstructed phylogeny; and $\xi_{\Theta,\lambda,\mu}(t, s, \lambda_1, \lambda_2, \mu_1, \mu_2)$, the probability that a lineage alive at time t has speciation and extinction rates λ and μ and gives birth at time s to two daughter lineages that respectively have speciation rates λ_1 and λ_2 and extinction rates μ_1 and μ_2. We obtained ordinary differential equations (ODEs) to solve for Φ, χ and ξ by considering the different events that can happen during a short time interval Δ_t and making Δ_t

tend to 0 (SI Appendix). Under a pure birth model and for a completely sampled phylogeny, the ODEs can be solved analytically (SI Appendix). In the presence of extinction and/or if there are missing taxa in the phylogeny, Φ, χ and ξ are computed by integrating the ODEs numerically, which is more computationally intensive (SI Appendix).

Simulation.— We implemented a simulation algorithm of ClaDS in the R-package RPANDA (Morlon et al., 2016, function sim_ClaDS) (SI Appendix). In this implementation, the speciation rates of daughter lineage are drawn independently from a distribution ν_λ. Their extinction rates are either drawn from a distribution ν_μ, given by μ_0 (constant extinction rate scenario, ClaDS1), or given by $\varepsilon * \lambda_{s_{i,1}}$ and $\varepsilon * \lambda_{s_{i,2}}$ (constant turnover scenario, ClaDS2). ν_λ and ν_μ can be normal, log-normal, or uniform distributions. The simulations are continued until a stopping criterion is met, either a fixed time or a fixed number of species. In addition, sim_ ClaDS takes as one of its arguments a parameter p controlling the probability that a shift happens at each speciation event (the default value $p = 1$ corresponds to the model investigated here), and a parameter n, controlling a maximum number of shifts (the default value $n = \text{INF}$ corresponds to the model investigated here; if n takes a finite value, then p switches to 0 as soon as n switches have occurred).

Bayesian implementation.— We implemented a Bayesian inference approach for fitting ClaDS to reconstructed phylogenies in the R-package RPANDA (Morlon et al., 2016, function fit_ClaDS)(SI Appendix). In order to fit ClaDS0 (no extinction), we use a Metropolis within Gibbs MCMC (Monte Carlo Markov Chain) sampler with a Bactrian proposal (Yang and Rodríguez, 2013), and convergence is monitored by running three MCMC chains in parallel and computing Gelman statistics (Gelman et al., 2014). In order to fit ClaDS1 and ClaDS2 (i.e. in the presence of extinction), and/or if there are missing taxa in the phylogeny, we use the faster blocked Differential Evolution (DE) MCMC sampler, with sampling from the past of the chains (**?**). We also ran three chains.Each estimate was computed as the mean over the iterations and the three chains.

Testing the performance of ClaDS.

We performed intensive simulations to test the performance of ClaDS. We tested both the performance of ClaDS under data generated by this model, and its performance for data generated with a discrete speciation rate shift. In order to assess the performance of ClaDS under a large parameter set and for a variety of tree sizes, we considered primarily the pure birth model with completely sampled phylogenies. We also considered the model with extinction and/or missing taxa, but only in a limited, computationally tractable, set of simulations.

Gradual rate variation (ClaDS model).— For each combination of the following parameter values, we simulated 20 pure birth trees, stopping the simulation when a target tip number of 50, 100 and 200 was reached. λ_0 was fixed at 0.1, σ was taken in $\{0, 0.1, 0.18, 0.26, 0.34, 0.41\}$, and α in $\{1.2, 1.1, 1, 0.95, 0.9, 0.7\}$. We recorded the realized speciation rate on each branch in each of these simulations. We then ran ClaDS on each simulated tree using our fit_ClaDS function. Lastly, we compared the retrieved estimates of λ_0, σ and α to their simulated values; we also compared the retrieved estimates of branch-specific speciation rates for each tree to their realized values by performing linear regressions and computing relative errors (ratio of estimated versus realized rates).

In order to explore the model accounting for extinction, we simulated 5 trees of size 100 under 4 scenarios with constant turnover rate (ClaDS2), and for each condition either low ($\varepsilon = 0.1$) or high ($\varepsilon = 0.9$) turnover (8 scenarios in total). We focused on the scenario with constant turnover, because

this scenario produced tree shapes similar to those of empirical trees under a wider set of parameter values than the alternative scenario with constant extinction rate (Fig. S1 versus S2). Maintaining a balance where extinction is neither negligible nor driving clades to extinction is also easier under ClaDS2. The four scenarios were as follows: i) high heterogeneity and decreasing rates : $\lambda_0 = 0.1$, $\sigma = 0.7$, $\alpha = 0.7$ (mean relative change $m = 0.9$), ii) no heterogeneity and constant rates (equivalent to constant rate birth-death trees) : $\lambda_0 = 0.1$, $\sigma = 0$, $\alpha = 1$ ($m = 1$) iii) Low heterogeneity and no average change in rate at speciation : $\lambda_0 = 0.1$, $\sigma = 0.2$, $\alpha = 0.98$ ($m = 1$) iv) Low heterogeneity and decreasing rates : $\lambda_0 = 0.1$, $\sigma = 0.2$, $\alpha = 0.88$ ($m = 0.9$). We recorded the realized speciation rate on each branch in each of these simulations. We then ran ClaDS on each simulated tree using our fit_ClaDS function, both accounting and not accounting for extinction, the latter to evaluate the bias resulting from not accounting for extinction when it occurs. Lastly, we compared the retrieved estimates of σ, α, m and ε for each tree to their simulated values. We did not compare the retrieved estimates of λ_0 to the simulated values, because the estimates correspond to the speciation rate at the crown while the simulated values correspond to the speciation rate at the stem. These two rates can be very different in the presence of extinction. We also compared the retrieved estimates of branch-specific speciation rates and net diversification rates (speciation minus extinction) for each tree to their realized values by performing linear regressions and computing relative errors.

Discrete rate shift.— We also tested the behavior of ClaDS under a 'key innovation' scenario with only a single large rate shift during the history of the clade. In order to simulate this scenario, we used our sim_ClaDS function with λ_0 (the background rate in this case) fixed at 0.1, p (the probability that a rate shift happens at each speciation event) fixed at 0.02, and n (the maximum number of shifts) fixed at 1. The new speciation rate took a series of values from lower (uniformly drawn in $[0.025, 0.03]$, $[0.03, 0.05]$, $[0.05, 0.1]$) to higher (uniformly drawn in $[0.1, 0.15]$, $[0.15, 0.2]$, $[0.2, 0.3]$, $[0.3, 0.4]$, $[0.4, 1]$) than the background rate. For each of these rate values, we simulated phylogenies of size 200 until we had a good coverage of subclade new rate/size combination (from 300 to 500 phylogenies per parameter set). In such simulations, there are only two distinct rates across the tree: the background rate and the new rate. We then ran ClaDS on each simulated tree using our fit_ClaDS function and compared the retrieved estimates of branch-specific speciation rates for each tree to their simulated values by performing linear regressions and computing relative errors. Finally, we tested whether the model is able to detect if two branches in the tree belong to the same or distinct speciation regime(s): two branches were considered to have significantly different rates (distinct regime) if the difference in the estimated speciation rates between the two branches was of constant sign on at least 95% of the MCMC chains. We assessed the significance of speciation rate differences (and the corresponding sign) for all pairs of branches in the simulated trees. Finally, we quantified the 'proper detection' rate as the proportion of pairs for which a significant difference was inferred when the two branches indeed belonged to distinct speciation regimes (i.e. one had the background speciation rate and the other one had the new rate), and the 'false detection' rate as the proportion of pairs for which a significant difference was inferred, while the two branches actually belonged to the same speciation regime (i.e. both had either the background speciation rate or the new rate).

Diversification of the avian radiation. We applied ClaDS, accounting for extinction (ClaDS2, model with constant turnover) and incomplete sampling, to bird phylogenies. We used the MCC trees from Jetz et al. (Jetz et al., 2012) with only the species for which there was molecular data, along with the associated sampling fractions provided by the authors. Most of these are family level phylogenies, with some spawning two or a few more families. We ran the model on the 42 bird phylogenies with more than 50 species. We report the distribution of branch-specific speciation rates across the 42 clades, as well as individual distributions for each clade. We partitioned the total variance of the logarithm of the branch specific speciation rates ($\sum_i \left(\ln(\lambda_i) - \overline{\ln(\lambda)}\right)^2$, where $\overline{\ln(\lambda)}$ is the mean of the log of the speciation rates for all branches in all clades) between the intra-clade ($\sum_i \left(\ln(\lambda_i) - \overline{\ln(\lambda_{c_i})}\right)^2$, where c_i is the clade to which branch i belongs and $\overline{\ln(\lambda_c)}$ is the mean of the log of the speciation rates for all branches in clade c) and inter-clade variance ($\sum_i \left(\overline{\ln(\lambda_{c_i})} - \overline{\ln(\lambda)}\right)^2$). We also tested for a potential correlation between the variance in rates and the size (number of tips) and age (crown age) of clades using PGLS (Grafen, 1989) on the Hacket backbone phylogeny provided in Jetz et al. (2012).

Supplementary material for Chapter 2

Appendix 1: Supplementary Methods.

1.1 Simulation of phylogenies under ClaDS.

We initialize each simulation by defining a root lineage with speciation rate λ_0 and death rate μ_0. At any iteration of the simulation, the time for the next event is drawn from an exponential distribution with mean $1/r$, where the rate r is given by the sum of the speciation and extinction rates of all lineages alive at this stage. Next, the nature of the event (speciation or extinction of a given lineage) is drawn according to their respective probabilities μ_i/r (extinction) and λ_i/r (speciation). When a speciation occurs in lineage i, we draw the speciation rate of each daughter lineage independently from ν_λ. Their extinction rates are either drawn from a distribution ν_μ, given by μ_0 (constant extinction rate scenario), or given by $\varepsilon * \lambda_{i1}$ and $\varepsilon * \lambda_{i2}$ (constant turnover scenario). The process is continued until a stopping criterion is met, either a fixed time or a fixed number of species. Our simulation algorithm, implemented in the R-package RPANDA (Morlon et al., 2016, function sim_ClaDS), can take ν_λ and ν_μ to be a normal, log-normal, or uniform distribution. In addition, sim_ ClaDS takes as one of its arguments a parameter p controlling the probability that a shift happens at each speciation event (the default value $p = 1$ corresponds to the model investigated here), and a parameter n, controlling a maximum number of shifts (the default value $n = \text{INF}$ corresponds to the model investigated here; if n takes a finite value, then p switches to 0 as soon as n switches have occurred).

In order to explore the shape of phylogenetic trees arising from ClaDS, we simulated 100 trees under each of the following parameter combinations: $\lambda_0 = 0.1$, $\sigma = 0.8$, $m \in \{0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3\}$, $\mu_0 \in \{0.01, 0.09\}$ (for constant extinction) or $\varepsilon \in \{0.1, 0.9\}$ (for constant turnover), and $\lambda_0 = 0.1$, $\sigma \in \{0, 0.05, 0.1, 0.2, 0.3, 0.5, 0.8, 1\}$, $m = 0.9$, $\mu_0 \in \{0.01, 0.09\}$ or $\varepsilon \in \{0.1, 0.9\}$. Next, we summarized tree shape by the classical gamma statistic that measures the root to tip distributions of nodes in a phylogeny (Pybus and Harvey, 2000), and the beta statistic that measures the imbalance of a phylogeny (Aldous, 1996, 2001). Empirical phylogenies are often characterized by negative gamma and beta values (Mooers and Heard, 1997; Blum and François, 2006; Phillimore and Price, 2008; McPeek, 2008).

1.2 Derivation of the likelihood and posterior distribution.

In the following text, we denote by λ_0 the initial speciation rate, μ_0 the initial extinction rate, Θ the parameters of the distribution determining how rates are inherited between lineages, f the probability to sample a species alive in the present, $\nu_\Theta(\lambda, \mu, .)$ the distribution of new speciation and extinction rates for a lineage whose parent has speciation rate λ and extinction rate μ, Λ the diversification rates space, λ_i the branch specific speciation rate of branch i at t_i, μ_i the branch specific extinction rate of branch i at t_i, and i_1 and i_2 the two daugther lineages of i. The time goes from the present to the past and equals 0 at present, and we note t_i the birth time of branch i in the reconstructed phylogeny, s_i the end time of branch i in the reconstructed phylogeny (which is the branching time for internal branches, and 0 for terminal branches).

The likelihood of the model is defined as

$$\begin{aligned} L(\Theta, \lambda_0, \mu_0) &= p((t_i), (s_i) | \Theta, \lambda_0, \mu_0) \\ &= \int_{(\lambda_i)_{i>0} \in \Lambda^n, (\mu_i)_{i>0} \in \Lambda^n} p((t_i), (s_i), (\lambda_i)_{i>0}, (\mu_i)_{i>0} | \Theta, \lambda_0, \mu_0) d\lambda_1 ... d\lambda_n d\mu_1 ... d\mu_n \end{aligned}$$

The term inside the integral can be computed (see the following sections), but the high dimensionality of the space over which we have to integrate it makes the evaluation of the likelihood very difficult, making the estimation of Θ and λ_0 in a maximum likelihood framework impractical.

Numerically, this problem is easier to evaluate through MCMC sampling in a Bayesian framework. We thus aim to sample from the posterior distribution

$$p(\Theta, \lambda_0, (\lambda_i)_{i>0}, (\mu_i)_{i>0}|(t_i), (s_i)) \propto p((t_i), (s_i), (\lambda_i)_{i>0}, (\mu_i)_{i>0}|\Theta, \lambda_0, \mu_0)p(\Theta, \lambda_0, \mu_0)$$

where $p(\Theta, \lambda_0, \mu_0)$ is the prior distribution of our model parameters. The computation of $p((t_i), (s_i), (\lambda_i)_{i>0}, (\mu_i)_{i>0}|\Theta, \lambda_0, \mu_0)$ is detailed in Section 1.3 for the general case with extinction and potentially missing species and in Section 1.4 for the case with no extinction and complete sampling.

1.3 Model with extinction and incomplete sampling.

As in the case of the model without extinction and complete sampling, for the model with extinction and incomplete sampling $p((t_i), (s_i), (\lambda_i)_{i>0}, (\mu_i)_{i>0}|\Theta, \lambda_0, \mu_0)$ is given by the product of the probability density of each branch of the tree and rate changes at speciation event. To compute these densities, we first need to be able to compute the probability that a lineage has no descendants in the sample. We can then compute the probability density of internal and external branches.

Probability to have no extant descendants.

We call $\Phi_{\Theta,\lambda,\mu}(t)$ the probability that a species with speciation rate λ and extinction rate μ at time t has no sampled extant descendants. For $t = 0$, $\Phi_{\Theta,\lambda,\mu}(0)$ is the probability that the species was not sampled, which equals $1 - f$.

We have

$$\Phi_{\Theta,\lambda,\mu}(t) = \mathbb{P}\left(\begin{array}{l|l}\text{a lineage has no descendant} & \text{the lineage is alive and} \\ \text{in the sample} & \text{has speciation rate } \lambda \\ & \text{and extinction rate } \mu \text{ at} \\ & \text{time } t\end{array}\right)$$

$$\begin{aligned}\Phi_{\Theta,\lambda,\mu}(t+\Delta t) =& \mathbb{P}\begin{pmatrix}\text{the lineage goes extinct in } (t, t+ \\ \Delta t)\end{pmatrix} \\ &+ \mathbb{P}\begin{pmatrix}\text{the lineage does not go extinct, but speci-} \\ \text{ates and neither of the resulting lineages} \\ \text{(which each have new speciation and ex-} \\ \text{tinction rates) has descendants in the sam-} \\ \text{ple}\end{pmatrix} \\ &+ \mathbb{P}\begin{pmatrix}\text{no extinction nor speciation but the lin-} \\ \text{eage has no descendants in the sample}\end{pmatrix} \\ &+ o(\Delta t) \\ =& \mu\Delta t \\ &+ (1-\mu\Delta t)\lambda\Delta t\left(\int_\Lambda \nu_\Theta(\lambda, \mu, \lambda_1, \mu_1)\Phi_{\Theta,\lambda_1,\mu_1}(t)d\lambda_1 d\mu_1\right)^2 \\ &+ (1-\mu\Delta t)(1-\lambda\Delta t)\Phi_{\Theta,\lambda,\mu}(t) + o(\Delta t)\end{aligned}$$

By subtracting $\Phi_{\Theta,\lambda,\mu}(t)$, dividing by Δt and taking $\Delta t \to 0$ we get

$$\frac{\partial \Phi_{\Theta,\lambda,\mu}}{\partial t}(t) = \mu(1 - \Phi_{\Theta,\lambda,\mu}(t)) + \lambda \left(\left(\int_\Lambda \nu_\Theta(\lambda, \mu, \lambda_1, \mu_1) \Phi_{\Theta,\lambda_1,\mu_1}(t) d\lambda_1 d\mu_1 \right)^2 - \Phi_{\Theta,\lambda,\mu}(t) \right)$$

which is then integrated numerically (see section 4: Implementation in R for details).

Probability density of internal branches.

For $t > s$ we define

$$\xi_{\Theta,\lambda,\mu}(t, s, \lambda_1, \lambda_2, \mu_1, \mu_2) = \mathbb{P}\left(\begin{array}{l|l} \text{a lineage has exactly one descendant} & \text{the lineage is alive} \\ \text{in the reconstructed phylogeny at} & \text{at time } t \text{ and has} \\ \text{time s and speciate at time s to give} & \text{speciation rate } \lambda \text{ and} \\ \text{birth to two lineages with diversifi-} & \text{extinction rate } \mu \text{ at} \\ \text{cation rates } \lambda_1, \mu_1 \text{ and } \lambda_2, \mu_2 & \text{time } t \end{array} \right)$$

For $t = s$ we get

$$\begin{aligned} \xi_{\Theta,\lambda,\mu}(t, s, \lambda_1, \lambda_2, \mu_1, \mu_2) &= \mathbb{P}\left(\begin{array}{l|l} \text{a lineage speciates at time s to give} & \text{the lineage is alive} \\ \text{birth to two lineages with diversifi-} & \text{at time } s \text{ and has} \\ \text{cation rates } \lambda_1, \mu_1 \text{ and } \lambda_2, \mu_2 & \text{speciation rate } \lambda \text{ and} \\ & \text{extinction rate } \mu \text{ at} \\ & \text{time } s \end{array} \right) \\ &= \lambda \nu_\Theta(\lambda, \mu, \lambda_1, \mu_1) \nu_\Theta(\lambda, \mu, \lambda_2, \mu_2) \end{aligned}$$

We have

$$\begin{aligned} \xi_{\Theta,\lambda,\mu}(t + \Delta t, s, \lambda_1, \lambda_2, \mu_1, \mu_2) =& \mathbb{P}\left(\begin{array}{l} \text{no extinction nor speciation in } (t, \Delta t), \text{ and the} \\ \text{lineage has exactly one descendant in the recon-} \\ \text{structed phylogeny at time } s, \text{ which speciate at time} \\ s \text{ to give birth to two lineages with diversification} \\ \text{rates } \lambda_1, \mu_1 \text{ and } \lambda_2, \mu_2 \end{array} \right) \\ &+ \mathbb{P}\left(\begin{array}{l} \text{the lineage does not go extinct, but speciates} \\ \text{and one of the resulting lineage is not present in} \\ \text{the sample, while the other lineage has exactly} \\ \text{one descendant in the reconstructed phylogeny} \\ \text{at time } s \text{ and speciates at time } s \text{ to give birth} \\ \text{to two lineages with diversification rates } \lambda_1, \mu_1 \\ \text{and } \lambda_2, \mu_2 \end{array} \right) \\ &+ o(\Delta t) \\ =& (1 - \mu\Delta t)(1 - \lambda\Delta t) \xi_{\Theta,\lambda,\mu}(t, s, \lambda_1, \lambda_2, \mu_1, \mu_2) \\ &+ 2(1 - \mu\Delta t)\lambda\Delta t \int_\Lambda \nu_\Theta(\lambda, \mu, \lambda_3, \mu_3) \Phi_{\Theta,\lambda_3,\mu_3}(t) d\lambda_3 d\mu_3 \\ &\times \int_\Lambda \nu_\Theta(\lambda, \mu, \lambda_3, \mu_3) \xi_{\Theta,\lambda_3,\mu_3}(t, s, \lambda_1, \lambda_2, \mu_1, \mu_2) d\lambda_3 d\mu_3 \\ &+ o(\Delta t) \end{aligned}$$

By subtracting $\xi_{\Theta,\lambda,\mu}(t,s,\lambda_1,\lambda_2,\mu_1,\mu_2)$, dividing by Δt and taking $\Delta t \to 0$ we get

$$\begin{aligned}\frac{\partial \xi_{\Theta,\lambda,\mu}}{\partial t}(t,s,\lambda_1,\lambda_2,\mu_1,\mu_2) =& -(\lambda+\mu)\xi_{\Theta,\lambda,\mu}(t,s,\lambda_1,\lambda_2,\mu_1,\mu_2)\\ &+ 2\lambda \int_\Lambda \nu_\Theta(\lambda,\mu,\lambda_3,\mu_3)\Phi_{\Theta,\lambda_3,\mu_3}(t)d\lambda_3 d\mu_3\\ &\times \int_\Lambda \nu_\Theta(\lambda,\mu,\lambda_3,\mu_3)\xi_{\Theta,\lambda_3,\mu_3}(t,s,\lambda_1,\lambda_2,\mu_1,\mu_2)d\lambda_3 d\mu_3\end{aligned}$$

Probability density of terminal branches.

We define

$$\chi_{\Theta,\lambda,\mu}(t) = \mathbb{P}\left(\begin{array}{l|l}\text{a lineage has exactly one} & \text{the lineage is alive at time } t\\ \text{descendant in the sam-} & \text{and has speciation rate } \lambda \text{ and}\\ \text{ple} & \text{extinction rate } \mu \text{ at time } t\end{array}\right)$$

For $t=0$, $\chi_{\Theta,\lambda,\mu}(0)$ is the probability that the species was sampled, which equals f.

We have

$$\begin{aligned}\chi_{\Theta,\lambda,\mu}(t+\Delta t) =& \mathbb{P}\begin{pmatrix}\text{no extinction nor speciation in } (t,\Delta t), \text{ and the lin-}\\ \text{eage has exactly one descendant in the sample}\end{pmatrix}\\ &+ \mathbb{P}\begin{pmatrix}\text{the lineage does not go extinct but speciates,}\\ \text{and one of the resulting lineages is not present}\\ \text{in the sample while the other has exactly one}\\ \text{descendant lineage in the sample}\end{pmatrix}\\ &+ o(\Delta t)\\ =&(1-\mu\Delta t)(1-\lambda\Delta t)\chi_{\Theta,\lambda,\mu}(t)\\ &+ 2(1-\mu\Delta t)\lambda\Delta t\int_\Lambda \nu_\Theta(\lambda,\mu,\lambda_1,\mu_1)\Phi_{\Theta,\lambda_1,\mu_1}(t)d\lambda_1 d\mu_1\\ &\times \int_\Lambda \nu_\Theta(\lambda,\mu,\lambda_1,\mu_1)\chi_{\Theta,\lambda_1,\mu_1}(t)d\lambda_1 d\mu_1 + o(\Delta t)\end{aligned}$$

By substracting $\chi_{\Theta,\lambda,\mu}(t)$, dividing by Δt and taking $\Delta t \to 0$ we get

$$\begin{aligned}\frac{\partial \chi_{\Theta,\lambda,\mu}}{\partial t}(t) =& -(\lambda+\mu)\chi_{\Theta,\lambda,\mu}(t) + 2\lambda\int_\Lambda \nu_\Theta(\lambda,\mu,\lambda_1,\mu_1)\Phi_{\Theta,\lambda_1,\mu_1}(t)d\lambda_1 d\mu_1\\ &\times \int_\Lambda \nu_\Theta(\lambda,\mu,\lambda_1,\mu_1)\chi_{\Theta,\lambda_1,\mu_1}(t)d\lambda_1 d\mu_1\end{aligned}$$

Remark that $\xi_{\Theta,\lambda,\mu}(t,s,\lambda_1,\lambda_2,\mu_1,\mu_2)$ and $\chi_{\Theta,\lambda,\mu}(t)$ are solution of the same diferential equation in t, only the inital condition changes.

Likelihood expression.

The joint likelihood is expressed as the product of the probability density of each branch of the tree. We define t_i the time of birth of branch i, s_i its end time , E_{int} the set of internal branches of the tree, E_{term} the set of terminal branches.

To calculate the likelihood for a tree conditional on its root age:

$$p((t_i,\lambda_i,\mu_i)_{i\in[\![1,N]\!]},(s_i)_{i\in[\![0,N]\!]}|\Theta,\lambda_0,\mu_0,t_0) = \frac{\prod_{i\in 0\cup E_{int}}\xi_{\Theta,\lambda_i,\mu_i}(t_i,s_i,\lambda_{i_1},\lambda_{i_2},\mu_{i_1},\mu_{i_2})\prod_{i\in E_{term}}\chi_{\Theta,\lambda_i,\mu_i}(t_i)}{1-\Phi_{\Theta,\lambda_0,\mu_0}(t_0)}$$

To calculate the likelihood for a tree conditional on its its crown age:

$$\begin{aligned}p((t_i,s_i,\lambda_i,\mu_i)_{i\in[\![1,N]\!]}|\Theta,\lambda_0,\mu_0,s_0) = &\\ &\frac{\nu_\Theta(\lambda_0,\mu_0,\lambda_{0_1},\mu_{0_1})\nu_\Theta(\lambda_0,\mu_0,\lambda_{0_2},\mu_{0_2})}{(1-\Phi_{\Theta,\lambda_{0_1},\mu_{0_1}}(s_0))(1-\Phi_{\Theta,\lambda_{0_2},\mu_{0_2}}(s_0))}\\ &\times\prod_{i\in E_{int}}\xi_{\Theta,\lambda_i,\mu_i}(t_i,s_i,\lambda_{i_1},\lambda_{i_2},\mu_{i_1},\mu_{i_2})\prod_{i\in E_{term}}\chi_{\Theta,\lambda_i,\mu_i}(t_i)\end{aligned}$$

1.4 Model without extinction and with complete sampling.

In this section we consider the case with $(\mu_i)_{0\le i\le n}$ fixed to 0 and f fixed to 1. In this case, $\Phi_{\theta,\lambda,0}(t)=0\forall t$ so the differential equations for the 2 other functions reduce to:

$$\begin{aligned}\frac{\partial\xi_{\Theta,\lambda,0}}{\partial t}(t,s,\lambda_1,\lambda_2,0,0) &= -\lambda\xi_{\Theta,\lambda,\mu}(t,s,\lambda_1,\lambda_2,0,0)\\ \frac{\partial\chi_{\Theta,\lambda,0}}{\partial t}(t) &= -\lambda\chi_{\Theta,\lambda,0}(t)\end{aligned}$$

with the initial conditions

$$\begin{aligned}\xi_{\Theta,\lambda,0}(s,s,\lambda_1,\lambda_2,0,0) &= \lambda\nu_\Theta(\lambda,\lambda_1)\nu_\Theta(\lambda,\lambda_2)\forall s\in\mathbb{R}+\\ \chi_{\Theta,\lambda,0}(0) &= 1\end{aligned}$$

This can be solved as

$$\begin{aligned}\xi_{\Theta,\lambda,0}(t,s,\lambda_1,\lambda_2,0,0) &= e^{-\lambda(t-s)}\nu_\Theta(\lambda,\lambda_1)\nu_\Theta(\lambda,\lambda_2)\\ \chi_{\Theta,\lambda,0}(t) &= e^{-\lambda t}\end{aligned}$$

If we note E_{int} the set of internal edges of the tree, the likelihood is then:
For a tree conditioned by its stem age:

$$p((t_i,\lambda_i)_{i\in[\![1,N]\!]},(s_i)_{i\in[\![0,N]\!]}|\Theta,\lambda_0,t_0) = \left(\prod_{i=0}^{n}e^{-\lambda_i(t_i-s_i)}\right)\left(\prod_{i\in 0\cup E_{int}}\lambda_i\nu_\Theta(\lambda_i,\lambda_{i_1})\nu_\Theta(\lambda_i,\lambda_{i_2})\right)$$

For a tree conditioned by its crown age:

$$p((t_i, s_i, \lambda_i)_{i\in[\![1,N]\!]}|\Theta, \lambda_0, s_0) =$$
$$\nu_\Theta(\lambda_0, \lambda_{0_1})\nu_\Theta(\lambda_0, \lambda_{0_2}) \left(\prod_{i=1}^{n} e^{-\lambda_i(t_i-s_i)}\right) \left(\prod_{i\in E_{int}} \lambda_i \nu_\Theta(\lambda_i, \lambda_{i_1})\nu_\Theta(\lambda_i, \lambda_{i_2})\right)$$

1.5 Implementation of the likelihood computation in R.

The implementing of the numerical integration of the differential equations derived in the preceeding part was done in R. We consider the case for which the death rate is uniform for the clade ($\mu_i = \mu_0 \forall i$). In what follows, we consider $\Lambda = [m_\lambda, M_\lambda]$, $\Theta = \sigma, \alpha$ and $\nu_\Theta(\lambda, \lambda_1)$ a lognormal centered on $\alpha \times \lambda$ and restricted to Λ.

We begin by integrating the equation for $\Phi_{\Theta,\lambda}$. To do so we approximate the integral on Λ by a sum:

$$\begin{aligned}
&\int_\Lambda \frac{1}{\lambda_1\sigma\sqrt{2\Pi}} \exp\left(-\frac{(\ln(\lambda_1) - \ln(\lambda) - \ln(\alpha))^2}{\sigma^2}\right) \Phi_{\Theta,\lambda_1}(t) d\lambda_1 \\
&\quad = \int_{\ln(m_\lambda)}^{\ln(M_\lambda)} \frac{1}{\sigma\sqrt{2\Pi}} \exp\left(-\frac{(\lambda_2 - \ln(\lambda) - \ln(\alpha))^2}{\sigma^2}\right) \Phi_{\Theta,\exp(\lambda_2)}(t) d\lambda_2 \\
&\quad \simeq \sum_{i=0}^{\lfloor \frac{\ln(M_\lambda)-\ln(m_\lambda)}{\Delta\lambda} \rfloor} \Delta\lambda \exp\left(-\frac{(\ln(m_\lambda) + i\Delta\lambda - \ln(\lambda) - \ln(\alpha))^2}{\sigma^2}\right) \Phi_{\Theta, m_\lambda e^{i\Delta\lambda}}(t)
\end{aligned}$$

We define the matrix

$$M = \left[\Delta\lambda \exp\left(-\frac{((i-j)\Delta\lambda - \ln(\alpha))^2}{\sigma^2}\right)\right]_{i,j=0..\lfloor \frac{\ln(M_\lambda)-\ln(m_\lambda)}{\Delta\lambda} \rfloor}$$

and the vector $\Phi_\Theta(t) = \left(\Phi_{\Theta, m_\lambda e^{i\Delta\lambda}}\right)_{i=0..\lfloor \frac{\ln(M_\lambda)-\ln(m_\lambda)}{\Delta\lambda} \rfloor}$

We have: $\frac{\partial \Phi_\Theta}{\partial t}(t) \simeq \epsilon\lambda(1 - \Phi_\Theta(t)) + \lambda(M\Phi_\Theta(t) * M\Phi_{\Theta,\lambda}(t) - \Phi_{\Theta,\lambda}(t))$

where $*$ is the scalar product

This equation is then integrated using the ode function in R (package ..., ref). We use the fact that M is a Toeplitz matrix multiplied by a column vector, and use the Fast Fourier Transform method to compute the product of this matrix with a column vector, which allows us to speed the computation.

To evaluate the other two functions, we use the evaluation of $\Phi_{\Theta,\lambda}$. By replacing integrals by sums, as we did before, we get:

$$\begin{aligned}
\frac{\partial \chi_\Theta}{\partial t}(t) &\simeq -(\lambda + \mu)\chi_\Theta(t) + 2\lambda M\Phi_\Theta * M\chi_\Theta(t) \\
&\simeq B_t \chi_\Theta(t)
\end{aligned}$$

With the notations

$$B_t = -\lambda(1+\epsilon)I_{\lfloor \frac{\ln(M_\lambda)-\ln(m_\lambda)}{\Delta\lambda} \rfloor} + 2\lambda\, diag(M\Phi_\Theta(t))M$$

we then use Magnus expansion to compute $\chi_\Theta(t)$.

1.6 Testing the accuracy of our likelihood computation.

We use Monte Carlo simulations to test that our analytical solutions of the key probabilities $\Phi_{\theta,\lambda,\mu}(t)$ and $\xi_{\theta,\lambda,\mu}(t)$ are correct and computed with a low numerical error, for both the scenario with constant extinction rate (ClaDS1) and the scenario with constant turnover rate (ClaDS2). We first randomly draw 500 sets of parameter values uniformly: t is drawn from an exponential distribution with parameter 0.1, λ is uniformly drawn in $[0, 0.5]$, μ in $[0, 1.5 * \lambda]$ for the constant rate scenario and ε in $[0, 1.5]$ for the constant turnover scenario, α in $[0.5, 1.5]$, σ in $[0, 1]$ and f in $[0, 1]$.

Next, for each parameter set, we compute $\Phi_{\theta,\lambda,\mu}(t)$ and $\xi_{\theta,\lambda,\mu}(t)$, both by integrating the ODEs (as described above) and by using Monte Carlo simulations: we simulate N realizations of ClaDS ($N = 10,000$) for a duration t, starting with one lineage characterized by speciation rate lambda. $\Phi_{\theta,\lambda,\mu}(t)$ is then computed as the average over the N realizations of 1 when the process has gone extinct and $(1-f)^n$ when the process has led to n living descendants (i.e., the probability that none of the n descendants has been sampled). $\xi_{\theta,\lambda,\mu}(t)$ is computed as the average over the N realizations of 0 when the process has gone extinct and $nf(1-f)^{n-1}$ when the process has led to n living descendants (i.e., the probability that exactly one of the n descendants has been sampled). Finally, we compare the values found by integrating the ODEs and by Monte Carlo simulations. Our results indicate a good match between simulated and computed values (Fig S3 & S4). Using the same procedure to test for the computation of $\chi_{\theta,\lambda,\mu}(t)$ is tricky, but this function is obtained using the same ODE as $\xi_{\theta,\lambda,\mu}(t)$ – only with different initial conditions–, so we are confident that it is properly computed.

1.7 Implementation of the Bayesian approach.

We use a Metropolis within Gibbs MCMC (Monte Carlo Markov Chain) sampler with a Bactrian proposal (Yang and Rodríguez, 2013) to simultaneously infer the hyperparameters (λ_0, α, σ, and either μ_0 or ε) and the branch specific speciation rates λ_i of the model. When considering also extinctions, resulting in hidden speciation events and thus also hidden rate shifts in the reconstructed phylogeny (Fig. 1), λ_i is chosen by convention to be the rate at the beginning of branch i. We use an inverse gamma prior with shape parameter 1 and rate parameter 0.1 for σ and a flat prior for all other parameters. To improve MCMC convergence, we partly ran the MCMC on transformed parameters. Under a pure birth model and for a completely sampled phylogeny, we ran the MCMC on the logarithm of the relative rates (log(rate/parent rate)) to reduce correlations between the inferred rates. To monitor convergence, we ran three MCMC chains in parallel and computed Gelman statistics (Gelman et al., 2014). We started the MCMC with an adaptation phase of $10,000$, during which the proposal width was adjusted every 1000 iterations with a goal acceptance rate of 0.3 (Yang and Rodríguez, 2013). This initial phase is not recorded. The following first $200,000$ iterations were considered a burn-in phase and also discarded for the analysis. After the burn-in, we recoded another $2,000,000$ MCMC iterations, that we thinned every $200,000$ iterations. We then computed the gelman statistic, and added another $2,000,000$ iterations until the potential

scale reduction factor was smaller than 1.05 for all the parameters. Those iterations were used to calculate posterior densities. The time and number of iterations needed for the convergence of the chains (in the sense of the gelman statistic) was very variable. For a tree of size 200, convergence was typically achieved after a few tens of millions of iterations and within a few hours; however in some (rare) instances convergence could take up to several hundreds of millions of iterations and several days. Parameter estimates, including estimates of the branch specific speciation rates, were computed as the mean over the iterations and the three chains of each parameter.

In the presence of extinction and/or if there are missing taxa in the phylogeny, the computation of the likelihood takes much longer. The initial adaptation phase in particular can be very long. We therefore used a blocked Diferential Evolution (DE) MCMC sampler, with sampling from the past of the chains (Ter Braak, 2006; **?**). This sampler is self-adaptive because proposals are generated from the past of the chains. In this sampler, three chains are run simultaneously for each tree. Block updates were implemented by first drawing the number of parameters to be updated from a truncated geometric distribution with mean 3, then drawing uniformly which parameter to update, and then following the normal DE algorithm. To save computation time, we ran the MCMC directly on the logarithm of the rates instead of the logarithm of the relative rates, which allows us to update only the contribution of the branches for which the rate has changed in the likelihood when we propose a new parameter set. We run the chains for at least 500000 iterations (1500000 likelihood computations). Parameter estimates are computed as the posterior maxima over the second half of the iterations. The computation of the likelihood of the model takes time in the presence of extinction, which is the reason why we were not able to reach MCMC convergence for the full model with as stringent a criterion as the one used for the pure birth model. However, because we tested it on several trees per parameter set and results are similar for all trees within a parameter set, we think it unlikely that chains are stuck on a local optimum. Our inference approach is implemented in the R-package RPANDA (Morlon et al., 2016, functions fit_ClaDS).

Appendix 2: Supplementary Figures and Results.

2.1 Tree shapes.

The general results described in the main text mostly apply whether extinction rate is constant (Fig. S1) or proportional to the speciation rate (Fig. S2). There is only one notable exception: with a high turnover rate, trees become more stemmy with higher σ, indicating that the average per-lineage speciation rate decreases through time (Fig. S2): in this case lineages which happen to have exceptionally high speciation rates are also highly volatile and fail to spread, while those with exceptionally low speciation rates rarely go extinct and are maintained throughout history.

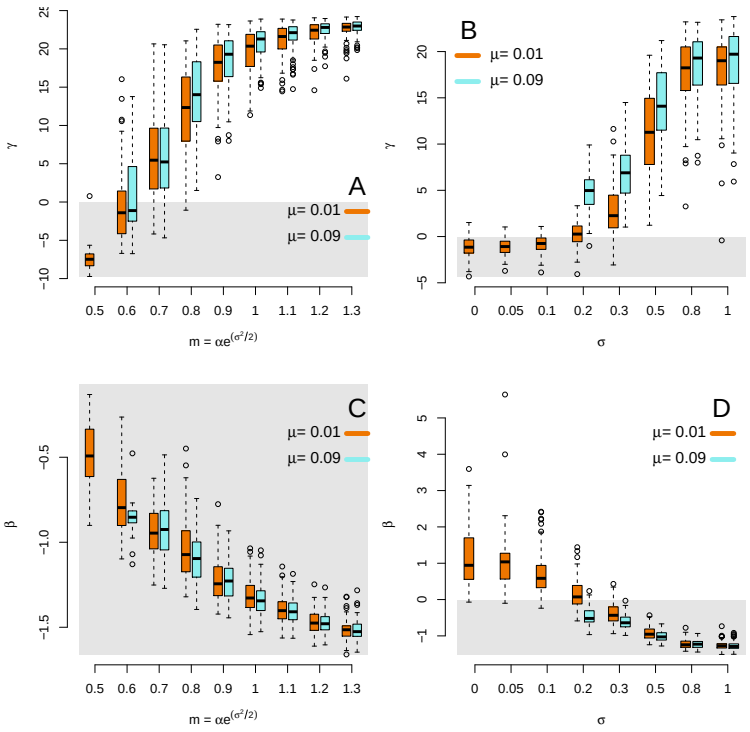


Figure S1: **Shape of trees simulated with ClaDS with constant extinction rate.** Tippiness index γ (A & B) and imbalance index β (C & D) of phylogenies simulated with a common $\sigma = 0.8$ and varying m (A & C) or with a common $m = 0.9$ and varying σ (B & D), for two extinction rate values ($\mu_0 = 0.01$, orange boxes; $\mu_0 = 0.09$, blue boxes). For all simulations, $\lambda_0 = 0.1$. In empirical phylogenies, we generally have $\gamma < 0$ and $\beta < 0$ (gray areas).

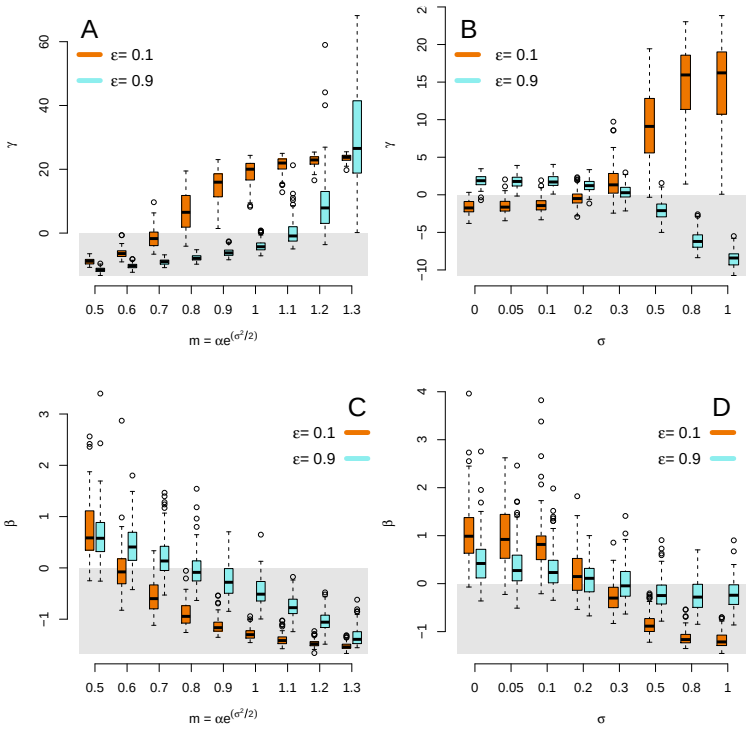


Figure S2: **Shape of trees simulated with ClaDS with constant turnover rates.** Tippiness index γ (A & B) and imbalance index β (C & D) of phylogenies simulated with a common $\sigma = 0.8$ and varying m (A & C) or with a common $m = 0.9$ and varying σ (B & D), for two turnover rate values ($\varepsilon = 0.1$, orange boxes; $\varepsilon = 0.9$, blue boxes). For all simulations, $\lambda_0 = 0.1$. In empirical phylogenies, we generally have $\gamma < 0$ and $\beta < 0$ (gray areas).

2.2 Testing the accuracy of our likelihood computation.

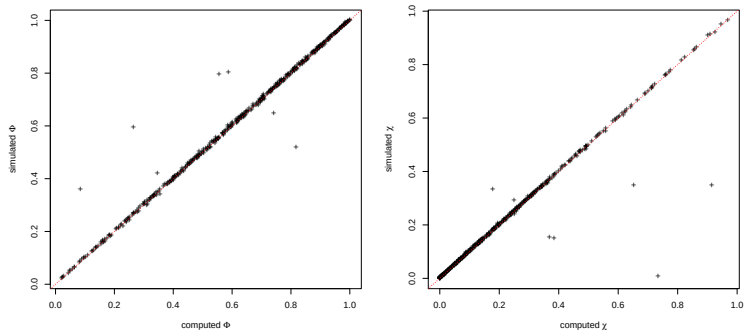


Figure S3: **Comparison between simulated and computed values of the extinction probability of the process, Φ and the probability to get a terminal edge, χ, for the model with constant extinction.** The simulation procedure is explained in section 5. We find good agreement between simulated and computed values for those functions, except for a few parameter sets for which the ode function did not converge.

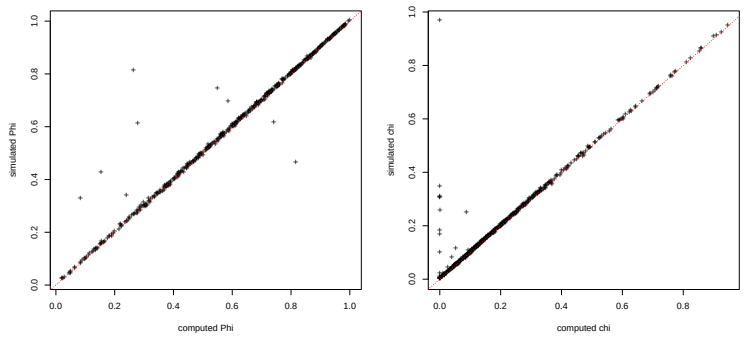


Figure S4: **Comparison between simulated and computed values of the extinction probability of the process, Φ and the probability to get a terminal edge, χ,for the model with constant turnover.** The simulation procedure is explained in section 5. We find good agreement between simulated and computed values for those functions, except for a few parameter sets for which the ode function did not converge.

2.3 Statistical performance of ClaDS _ Pure birth model.

Model's parameters.

The model's parameters are overall well estimated when the tree size is large enough (Fig. S5-S8). We find that a tree of 100 suffices to produce reliable estimates of λ_0 and m (and also α when σ is large enough), while trees of size 200 are required to reliably estimate σ, and also α when σ is small. For small trees (size 50), λ_0 tends to be slightly overestimated (Fig. S5), while α is slightly underestimated (Fig. S6), and small σ values are over-estimated while large values are underestimated (Fig. S7); overall, this results in a slight underestimation of m across parameter space (Fig. S8). Note that these small-sample biases could be due to a small-sample bias in the MLE, but are more likely the result of the increasing prior weight with decreasing data size, which naturally biases estimates towards the prior. The trend parameter α is the only parameter for which bias depend on other parameters values. The bias of the estimation is highest for small σ values, and its variability is highest for large σ values (Fig. S6). When σ is high the across-lineage variation in speciation rates is less predictable, which explains this result.

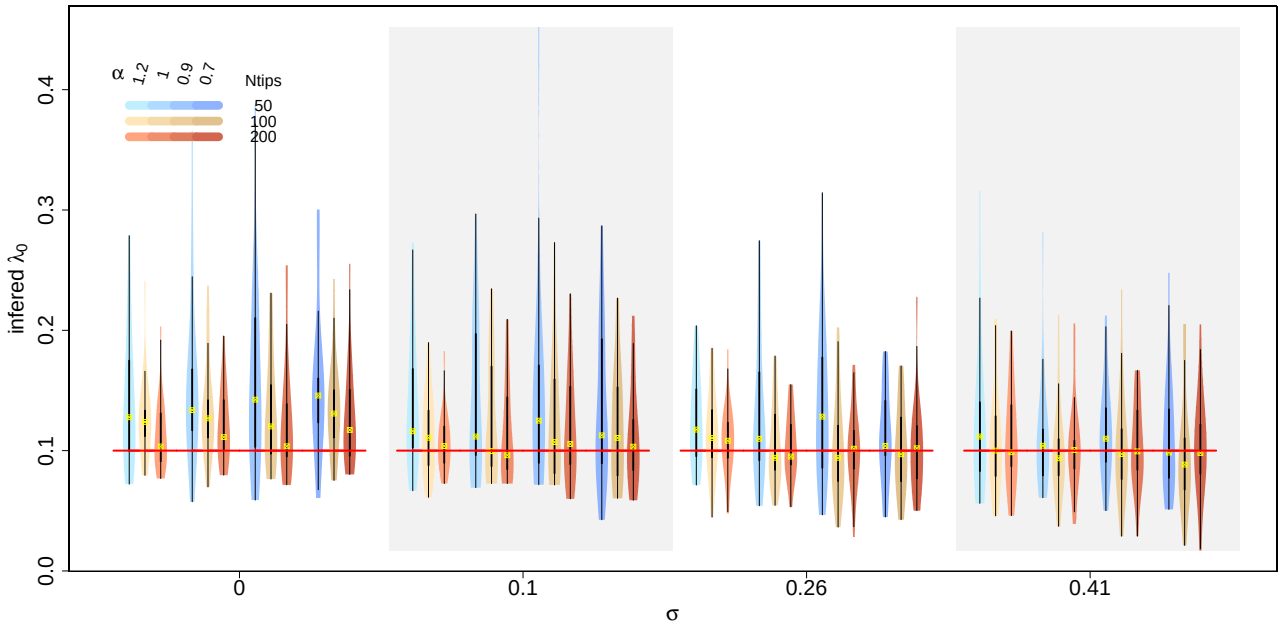


Figure S5: **Inference of λ_0 for different tree sizes.** Violin plots show the distribution of the inferred values for 20 replicates. Yellow crosses show the median values, red bars show the values used for simulating the phylogenies. Different violin colors indicate the tip number (blue for 50 tips, yellow for 100 tips, orange for 200 tips), different color shades indicate different α values used to simulate the phylogenies (from light (0.7) to dark (1.2)).

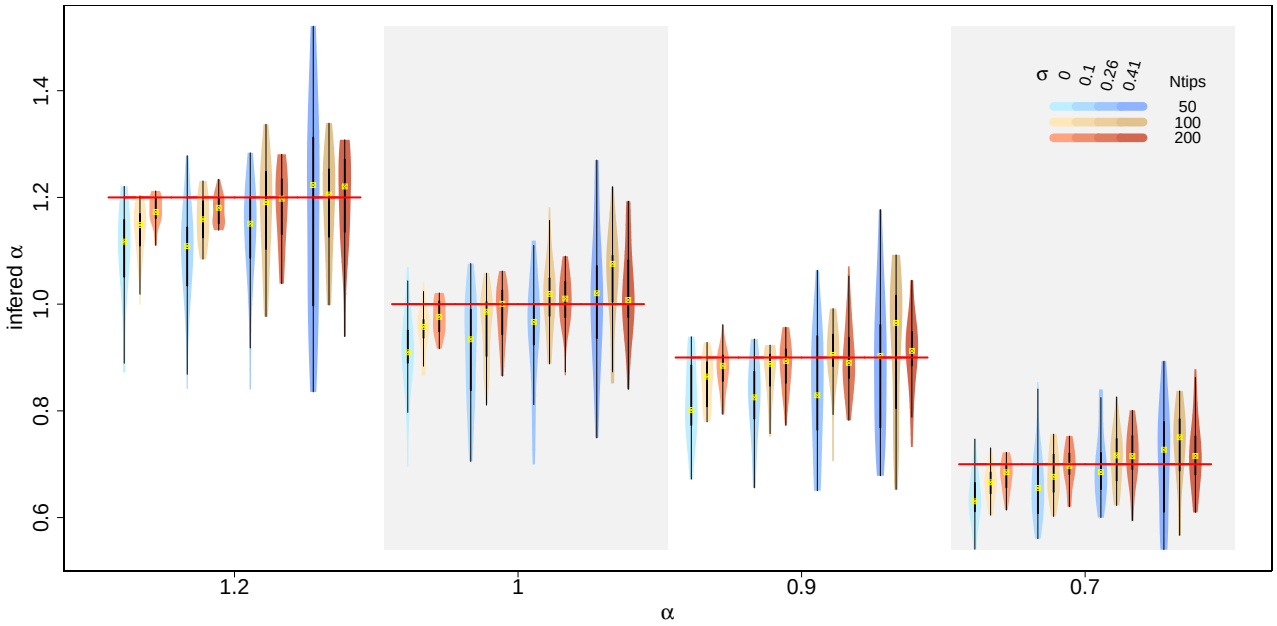


Figure S6: **Inference of α for different tree sizes.** Violin plots show the distribution of the inferred values for 20 replicates. Yellow crosses show the median values, red bars show the values used for simulating the phylogenies. Different violin colors indicate the tip number (blue for 50 tips, yellow for 100 tips, orange for 200 tips), different color shades indicate different σ values used to simulate the phylogenies (from light (0) to dark (0.4)).

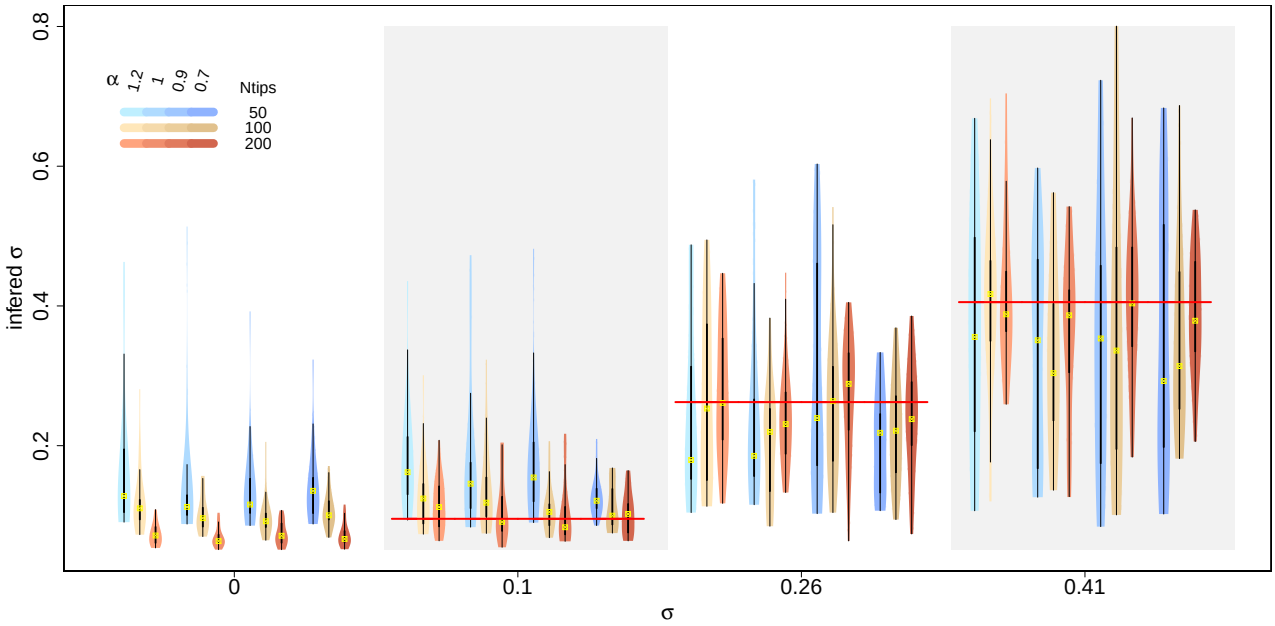


Figure S7: **Inference of σ for different tree sizes.** Violin plots show the distribution of the inferred values for 20 replicates. Yellow crosses show the median values, red bars show the values used for simulating the phylogenies. Different violin colors indicate the tip number (blue for 50 tips, yellow for 100 tips, orange for 200 tips), different color shades indicate different α values used to simulate the phylogenies (from light (0.7) to dark (1.2)).

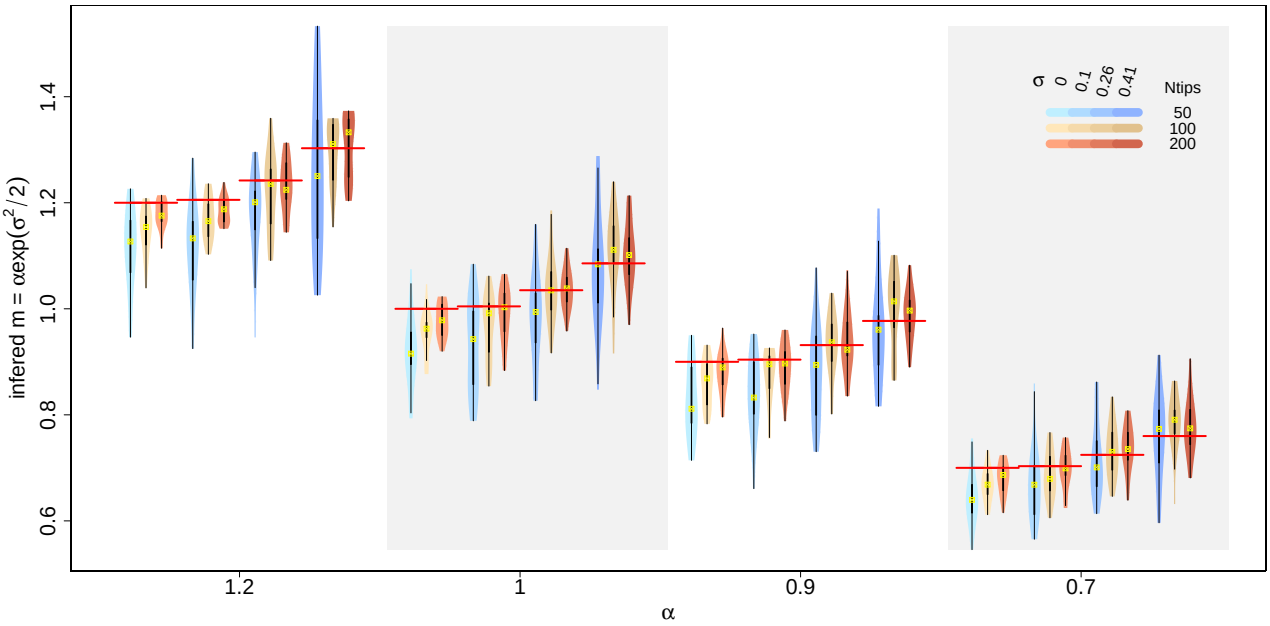


Figure S8: **Mean new rate m infered for different tree sizes.** Violin plots show the distribution of the inferred values for 20 replicates. Yellow crosses show the median values, red bars show the values used for simulating the phylogenies. Different violin colors indicate the tip number (blue for 50 tips, yellow for 100 tips, orange for 200 tips), different color shades indicate diferent σ values used to simulate the phylogenies (from light (0) to dark (0.4)).

Branch specific rates.

ClaDS provides reliable estimates of branch specific speciation rates on average: while low rates tend to be slightly overestimated and large rates slightly underestimated, ClaDS can detect regions of the tree with relatively high or low rates (Fig. S9 & S10). This conservative bias, which is expected from the statistical interpretation of σ as a shrinkage parameter that controls how much freedom is given to change diversification rates at each time step, decreases with increasing tree size (Fig. S9 D-F & Fig. S10 B). The quality of the estimation depends on the amount and source of rate heterogeneity in the tree (Fig. S9). It is best when there is substantial heterogeneity (slight rate variations are difficult to detect) and when this heterogeneity is due to a trend ($\alpha \neq 1$, σ small), and worst when rates are rather homogeneous (α close to 1, σ small). In the presence of a strong trend (α far from 1), the quality of the estimation decreases when the amount of heterogeneity due to stochasticity increases (i.e. with increasing σ). On the contrary, when there is no trend (α close to 1), the quality of the estimation increases with increasing stochastically-driven heterogeneity.

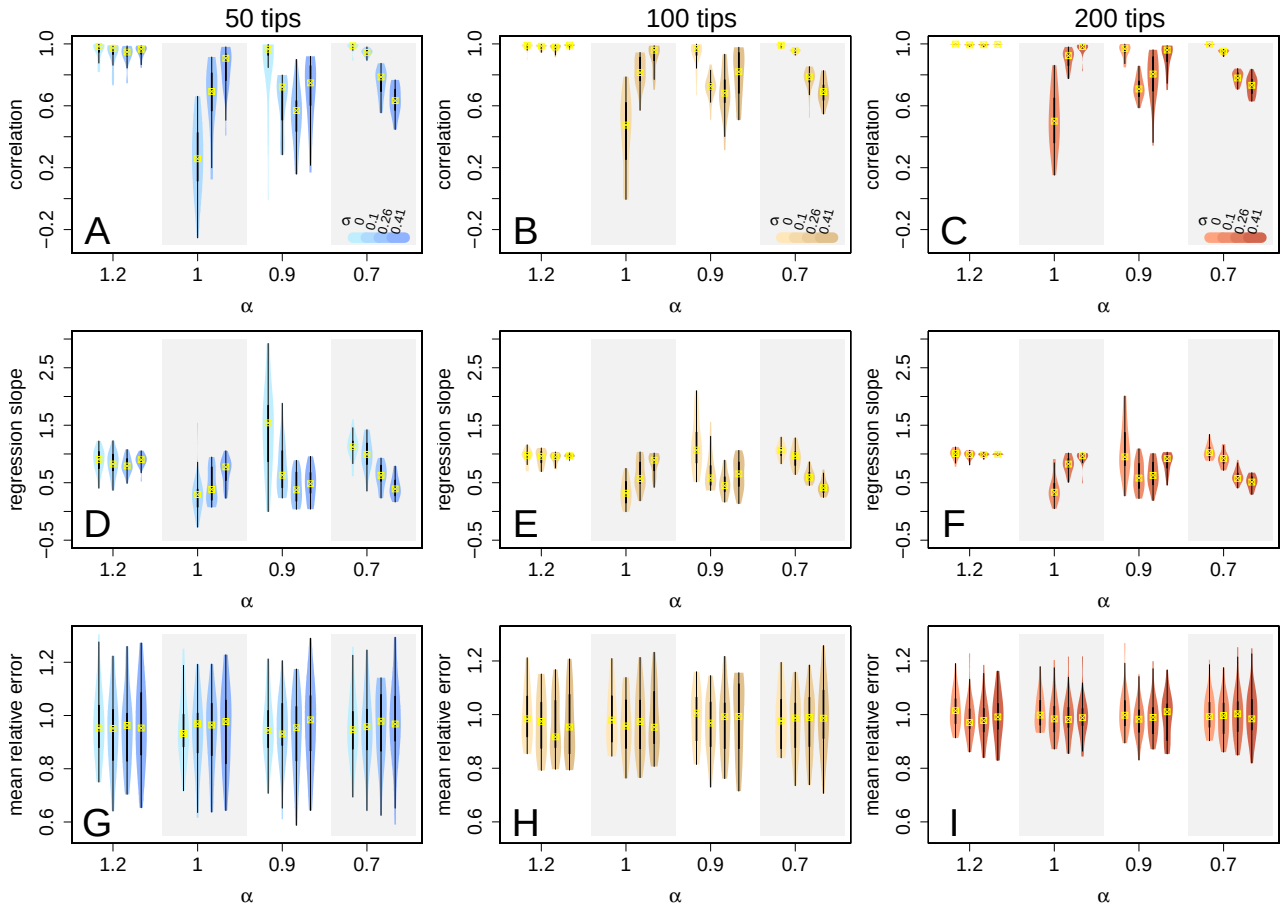


Figure S9: **Different measures of the goodness of fit of branch specific speciation rates for different tree sizes.** Violin plots show the distribution of the metrics for 20 replicates. Yellow crosses show the median values, red bars show the values used for simulating the phylogenies. Different violin colors indicate the tip number (blue for 50 tips (panels A, D & G), yellow for 100 tips (B, E & H), orange for 200 tips (C, F & I)), different color shades indicate different σ values used to simulate the phylogenies (from light (0) to dark (0.4)). We show the correlations between the log of simulated and inferred speciation rates (panels A, B & C), the regression slopes between the log of simulated and inferred speciation rates (panels D, E & F), and the mean relative error on each phylogeny (inferred rate/simulated rate, panels G, H & I))

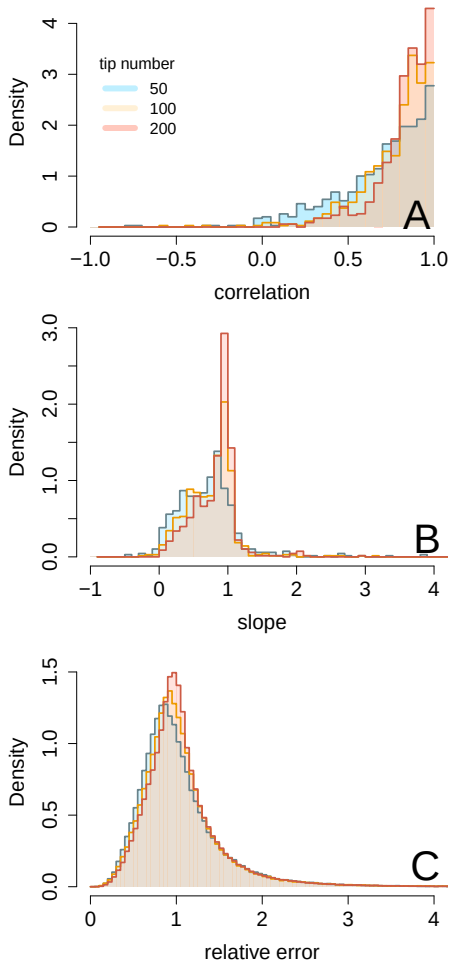


Figure S10: **Different measures of the goodness of fit of branch specific speciation rates for different tree sizes.** Histograms show the distribution of the metrics for all our simulated phylogenies. Different histogram colors indicate the tip number (blue for 50 tips , yellow for 100 tips and red for 200 tips). We show the distributions of the correlations between the log of simulated and inferred speciation rates (panel A), the regression slopes between the log of simulated and inferred speciation rates (B), and the mean relative error (inferred rate/simulated rate, C))

Discrete rate shifts.

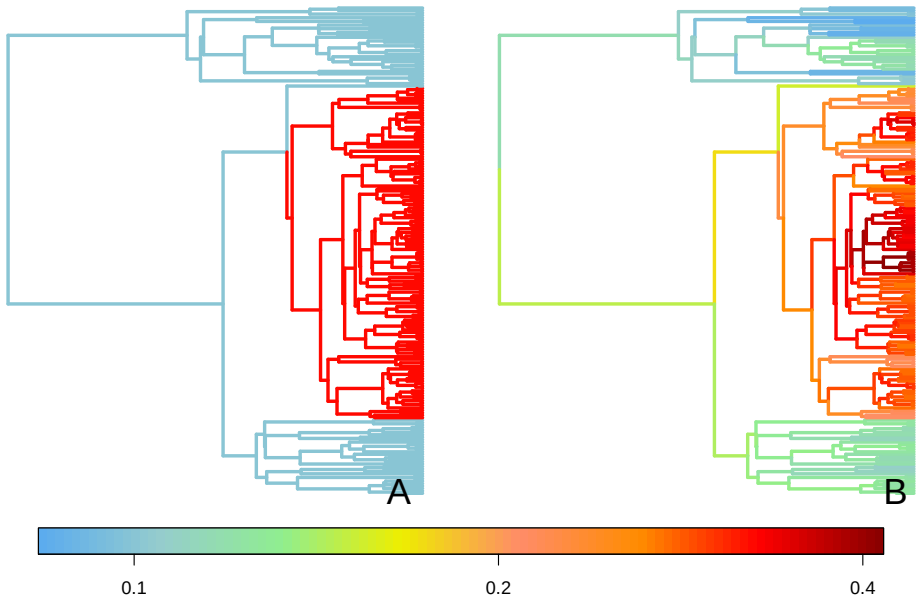


Figure S11: **Result of the inference of our model on an example tree with one discrete rate shift.** The branch colors show the speciation rates as used to simulate the tree (A) and as inferred using ClaDS (B). In the simulation, the background speciation rate is 0.1 (in blue, A), and the new rate is 0.32 (in red, A).

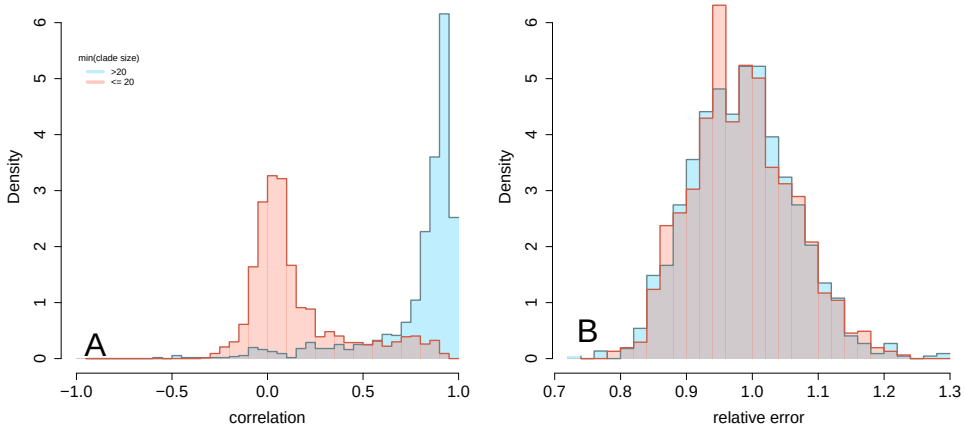


Figure S12: **Result of the test of our inference procedure on trees simulated with one rate shift.** Histograms show the distribution of the goodness of fit metrics for all our simulated phylogenies. Different histogram colors indicate the tip number in the smaller subclade (red if it is lower than 20 tips , blue if it is higher). We show the distributions of the correlations between the log of simulated and inferred speciation rates (panel A) and the mean relative error (inferred rate/simulated rate, B)).

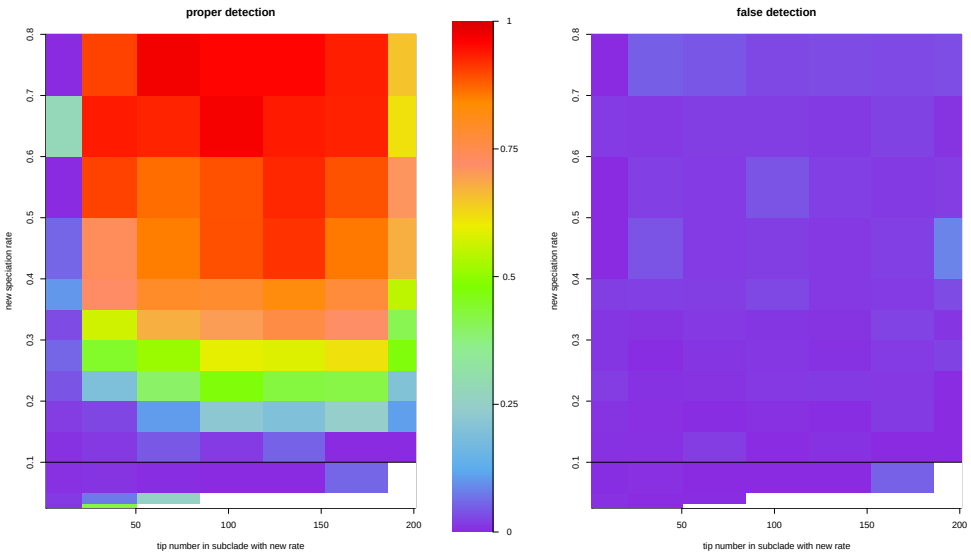


Figure S13: **Result of the test of our inference procedure on trees simulated with one rate shift.** In all the simulations the background rate was 0.1 and trees had 200 tips. The y-axis gives the value of the new speciation rate, and the x-axis indicates the number of tips in the subtree with the new rate. The left panel shows the mean proportion of branch pairs with different speciation rates for which a significant difference in speciation rate was found in the posterior (*proper detection*). The right panel shows the mean proportion of branch pairs with identical speciation rates for which a significant difference in speciation rate was found in the posterior (*false detection*). The white cells in the bottom right corner of each panels correspond to situations that are very unlikely (a subtree with a low speciation rate but a large number of tips) and were not obtained in our simulations.

2.4 Statistical performance of ClaDS _ Birth death model.

Model's parameters.

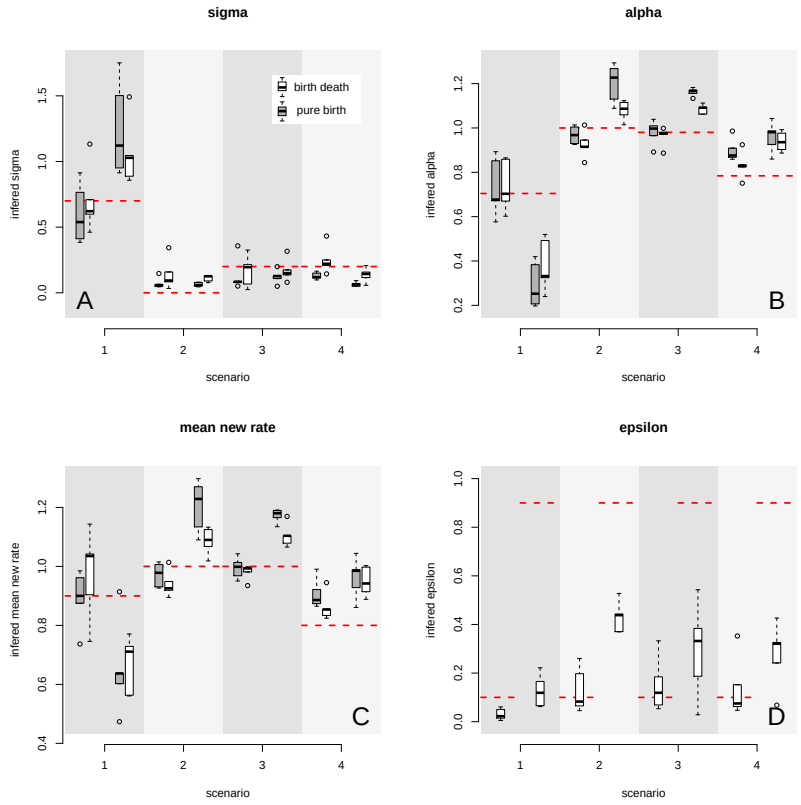


Figure S14: **Inference of the model's parameters for different simulation scenarios.** The boxplots show the distribution of the inferred values of σ (panel A), α (B), m (C) and ε (D) for 5 replicates, the red dotted line show the values used in the simulations. White boxes show the values inferred with the pure birth model, gray boxes those inferred with the birth death model. The scenarios are:1) high heterogeneity and decreasing rates: $\lambda_0 = 0.1$, $\sigma = 0.7$, $\alpha = 0.7$ (mean relative change $m = 0.9$), 2) no heterogeneity and constant rates (equivalent to constant rate birth-death trees): $\lambda_0 = 0.1$, $\sigma = 0$, $\alpha = 1$ (mean relative change $m = 1$) 3) Low heterogeneity and no average change in rate at speciation: $\lambda_0 = 0.1$, $\sigma = 0.2$, $\alpha = 0.98$ (mean relative change $m = 1$) 4) Low heterogeneity and decreasing rates: $\lambda_0 = 0.1$, $\sigma = 0.2$, $\alpha = 0.88$ (mean relative change $m = 0.8$).

Branch specific rates.

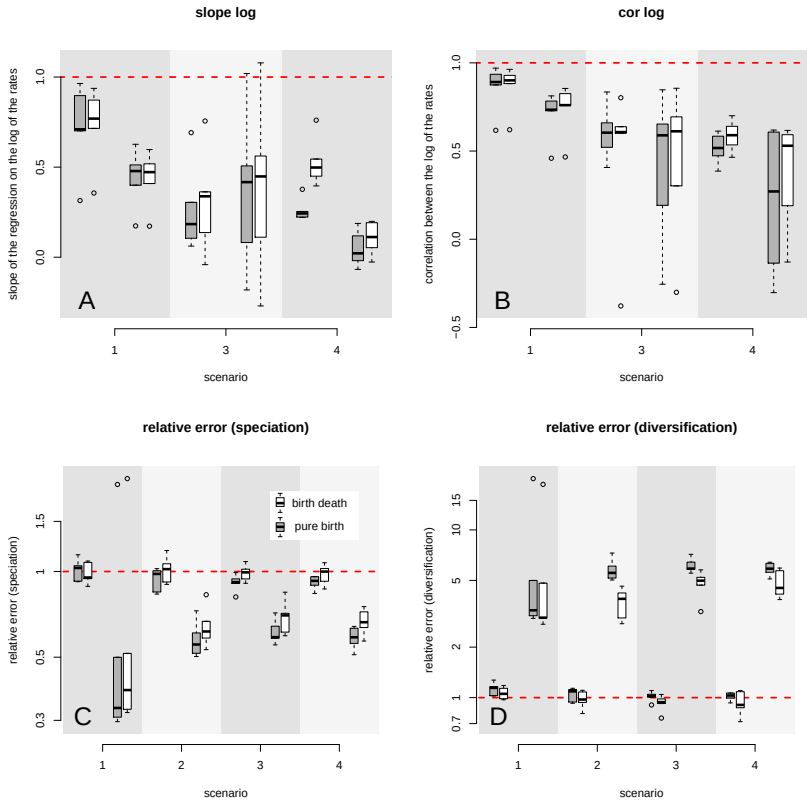


Figure S15: **Different measures of the goodness of fit of branch specific speciation rates for different scenarios.** The boxplots show the distribution of the regression slopes between the log of simulated and inferred speciation rates (panel A), the correlation between the log of simulated and inferred speciation rates (B), the mean relative error (inferred rate/simulated rate) on speciation rates (C) and the mean relative error on diversification rates (speciation - extinction, D) for 5 replicates, the red dotted line show the target values. White boxes show the values inferred with the pure birth model, gray boxes those inferred with the birth death model. The scenarios are:1) high heterogeneity and decreasing rates: $\lambda_0 = 0.1$, $\sigma = 0.7$, $\alpha = 0.7$ (mean relative change $m = 0.9$), 2) no heterogeneity and constant rates (equivalent to constant rate birth-death trees): $\lambda_0 = 0.1$, $\sigma = 0$, $\alpha = 1$ (mean relative change $m = 1$) 3) Low heterogeneity and no average change in rate at speciation: $\lambda_0 = 0.1$, $\sigma = 0.2$, $\alpha = 0.98$ (mean relative change $m = 1$) 4) Low heterogeneity and decreasing rates: $\lambda_0 = 0.1$, $\sigma = 0.2$, $\alpha = 0.88$ (mean relative change $m = 0.8$).

Chapter 3 : On bipartite ecological interactions and their impact on species richness

Ecological interactions are a major source of selection pressure, and are thus likely to have played an important part in past diversification. Yet, very few modeling studies have looked at the simultaneous emergence of communities composition and their interactions structures. In this chapter, I present a mechanistic individual-based model, in which individuals from two interacting guilds die, reproduce and accumulate mutations, eventually leading to the formation of new species. The reproductive success of an individual depends on its interaction with the individuals from the other guild. Our model allows the simultaneous emergence of the community composition – in terms of number of species with their trait distributions and abundances–, and community structure – in terms of network shapes, thus accounting for potential eco-evolutionary feedbacks. It provides the additional advantages to put neutral, antagonistic and mutualistic interactions within a same framework, enabling for direct comparison of the resulting patterns. We studied the behavior of the model through simulation for a large parameter set, in the case of antagonistic, mutualistic and neutral communities, for different niche widths. Consistently with previous modeling works, we find that, compared to neutral interactions, antagonistic interactions promotes diversity, both in terms of trait and species richness, while mutualistic interactions decrease both diversity measures. This result can be explained by the strong stabilizing selection existing for mutualistic communities in our model, while there is an advantage for resource species with rare trait values in antagonistic communities. The obtained network structure patterns are consistent with what is observed in empirical data. All networks tend to display nested structure because of the skewed species abundance distribution, but while trait-matching association rules create a modular structure in antagonistic communities, this does not happen for mutualistic ones, in which trait values are fairly constrained. This overall results in antagonistic communities being more modular and less nested than mutualistic ones. Our results thus globally shows that present communities structures are well explain by simple evolutionary mechanisms that are rarely accounted for when studying ecological networks.

This Chapter corresponds to a work done during this PhD together with Nicolas Loeuille and Hélène Morlon. The manuscript is in preparation and will soon be ready for submission.

An individual based model for the emergence of bipartite ecological networks

Odile Maliet, Hélène Morlon, Nicolas Loeuille

Abstract

Previous theoretical and empirical studies have shown that the structures of ecological interaction networks allowing to maintain a stable and diverse community depend on the nature of the interaction at play. Yet how such structures emerge and are affected by species coevolution at evolutionary time scales remains unclear. Here we build an individual-based eco-evolutionary model for the emergence of mutualistic, antagonistic and neutral bipartite interaction networks. We explore how the type of interaction influences the diversity and structure of the network, as well as the evolutionary conservatism of interaction partners. We find that antagonistic interactions foster species and trait diversity, while mutualistic interactions do not. In antagonistic scenarios, the resulting networks are more modular and less nested than neutral ones, likely due to a higher specialization in both guilds; resource species in these networks often display a phylogenetic signal in interaction partner. In mutualistic scenarios, stabilizing selection leads to networks that are less modular and more nested, with low phylogenetic signal in interaction partners. We discuss how these network properties compare to empirical observations.

Manuscript

In ecological communities, species are generally not isolated but interact with many others, through competition, antagonistic (such as prey predator, host parasite...) and mutualistic interactions (such as plant pollinator, seed dispersal ...). Empirical studies have repeatedly shown that the structure of networks representing bipartite interactions is highly non-random, and depends on the interaction at play. Antagonistic networks often display a modular structure, with subsets of species interacting more strongly among each other than with the rest of the community (May, 1972; Krause et al., 2003). In contrast, mutualistic communities more often display a nested structure, characterized by the fact that specialist species preferentially interact with generalist ones (Thébault and Fontaine, 2010; Lewinsohn et al., 2006; Bascompte et al., 2003; Jordano et al., 2003; Rohr et al., 2014).

Many studies have sought to explain the non-random structure of species interaction networks, a large number of them showing that it increases stability. In a community where species interact at random, diversity is known to have a negative impact on stability (May, 1972; Krause et al., 2003; Jordano et al., 2003; Montoya et al., 2006) but nonrandom network structures are known to counteract this effect. Theoretical studies have shown that the structure allowing the maintenance of a diverse community is dependent on the nature of interactions at play (Fontaine et al., 2011). In a randomly assembled network, community dynamics lead to nonrandom local extinctions that increase modularity in antagonistic communities and nestedness in mutualistic ones (Thébault and Fontaine, 2010). The persistence of species is increased in nested mutualistic networks, especially

when adaptive foraging is allowed (Valdovinos et al., 2016). Similarly, nested mutualistic networks broaden range of ecological parameters (interaction and competition strength, population growth rates) for which species can coexist (Rohr et al., 2014; Saavedra et al., 2016; Grilli et al., 2017).

A second family hypothesis to explain nonrandom, interaction dependent structures, relies on variations in species traits. In simulating studies, traits were either assigned to species at random (Santamaría and Rodríguez-Gironés, 2007) or evolved on empirical phylogenies with a brownian model (Rezende et al., 2007). The authors showed that both phenotypic difference and phenotypic complementarity interaction rules can lead to network that are more nested than random assemblages, especially if several traits are involved.

Finally, the neutrality hypothesis emphasizes the importance of species abundance distributions. If individuals interact at random with one another, asymmetrical species abundances distributions result in nested networks, because individuals from rare specialist species are more likely to meet and interact with individuals from abundant generalist species (Vázquez, 2005; Vázquez et al., 2009; Santamaría and Rodríguez-Gironés, 2007; Krishna et al., 2008; Staniczenko et al., 2013; Coelho and Rangel, 2018). A consequence of this observation is that the nestedness that is commonly observed in mutualistic networks do not need to rely on interaction-specific component, while the often observed modularity of antagonistic networks cannot be explained by neutral processes alone.

Studies seeking to explain the non-random structure of species interaction networks have rarely focused on the emergence of such structures over evolutionary time scales. Rather, they have fixed the ecological context, assuming a given network structure (Thébault and Fontaine, 2010), species number, species abundance distribution (Nuismer et al., 2013) or trait distribution (Santamaría and Rodríguez-Gironés, 2007), disregarding eco-evolutionary feedback loops that may affect such structures through changes in species trait or density. A few recent modeling studies have looked at the emergence of the community and network structure, but they consider neutral dynamics and thus do not allow to study the differences between mutualistic and antagonistic communities (Poisot and Stouffer, 2016; Coelho and Rangel, 2018). How different community structures emerge and are affected by species coevolution remains unclear. Species coevolution affects the feasibility and strength of interactions, which creates opportunities for eco-evolutionary feedback loops. For instance, evolution of specialization degree (Egas et al., 2004; Rueffler et al., 2006) based on adaptive foraging constraints (eg, Kondoh, 2003; Valdovinos et al., 2016) directly affects the distribution degree within the network thereby changing its nestedness. Thus, we may expect different evolutionary signatures for different interaction types, as observed in present day data (Krasnov et al., 2012; Fontaine and Thébault, 2015).

Coevolution is also expected to impact species diversity, which might in turn influence the interaction structure of the community. Different hypotheses imply an effect of species interactions on diversity (Ehrlich and Raven, 1964; Van Valen, 1973; Hembry et al., 2014). A previous modeling study for instance suggests that eco-evolutionary dynamics under antagonistic interactions leads to higher diversity while mutualistic ones are more likely to impede diversity (Yoder and Nuismer, 2010). This study however focuses on trait diversity and the effect of different interaction types on species diversity remains to be explored. An empirical test of the role of interaction type on diversification is very difficult to perform, but many empirical observations suggest that antagonistic communities are indeed very diverse (e.g. the high diversity of plant defense strategies, Futuyma and Agrawal, 2009), while only a few such observations exist for mutualistic communities (Hembry et al., 2014).

Here we present an individual based eco-evolutionary model allowing the emergence of mutu-

alistic, antagonistic and neutral bipartite interaction networks. The model consists of two guilds interacting on a grid, with individual fitness determined by the difference between its trait and that of the individual it interacts with. Starting from a single monomorphic species in each guild, the model allows the network structure to emerge. We characterize how different key components of community structure, such as diversity and network modularity and nestedness, vary depending on the interaction type and strength we assume, and how the signature of coevolution can be seen in the phylogenetic signal in interaction partners in the different scenarios. Our model is complementary to previous works as it (1) allows the simultaneous emergence of network structures and trait distributions through eco-evolutionary dynamics; (2) allows to consider alternatively antagonistic or mutualistic interactions within a single framework, thus allowing a direct comparison; (3) defines explicitly species, so that it is possible to link emergent network properties and phylogenetic signals.

Methods.

An individual-based model for the eco-evolutionary emergence of interaction networks.

We consider a grid of size $n_x \times n_y$, in which there are individuals from two different guilds (guild A and B). Each cell is occupied by one individual of each guild. Each Individual is characterized by a three-dimensional trait value x that determines the effect on its fitness of the interaction with the co-occurring individual from the other guild. The three dimensions were chosen because of a previous study showing that ecological networks are best described by traits with several yet few dimensions (Eklöf et al., 2013), and multidimensional traits are likely to have evolutionary dynamics qualitatively different from one-dimensional ones (Gilman et al., 2012; Ispolatov et al., 2016). Interactions constrained by trait matching are widely observed, as for instance proboscis and floral tube length, color preferences, matching organism sizes or phenology in mutualistic interactions such as plant-pollinator networks and presence of metabolic compounds and the ability to metabolize those in antagonistic interactions such as plant-herbivore networks. We thus use a classical trait matching expression for individual fitness whereby the effect of the interaction is maximal when the traits of the two interacting individuals are identical. For mutualistic interactions, the fitnesses of the individual of both guilds are increased for close trait values (equation (13)) ; for antagonistic interactions, the fitnesses of ndividuals from the consumer clade are increased for close trait values while those of the resource guild are decreased (equation (14)). The interaction change in individual fitness is given by a Gaussian function with standard deviation $1/\alpha$. α thus measures the degree of specialization, with higher α values reflecting more specialized species. The fitness function is further parametrized by a parameter r that measures the ratio between the maximum and minimum fitness. Higher r values thus reflect larger effects of trait values on fitness.

This gives the following expressions:

- In the mutualistic case:

$$\begin{aligned} W_A(x_A, x_B) &= \frac{1}{r_A - 1} + e^{-\|x_A - x_B\|^2 \times (\alpha_A^2/2)} \\ W_B(x_A, x_B) &= \frac{1}{r_B - 1} + e^{-\|x_A - x_B\|^2 \times (\alpha_B^2/2)} \end{aligned} \tag{13}$$

- In the antagonistic case (A being the resource guild and B the consumer guild):

$$W_A(x_A, x_B) = \frac{1}{r_A - 1} + 1 - e^{-\|x_A - x_B\|^2 \times (\alpha_A^2/2)}$$
$$W_B(x_A, x_B) = \frac{1}{r_B - 1} + e^{-\|x_A - x_B\|^2 \times (\alpha_B^2/2)} \tag{14}$$

The neutral case is obtained by tacking $\alpha = 0$ in the mutualistic fitness (tacking $\alpha = 0$ in the antagonistic version would yield to identical simulations since only the relative fitness (individual fitness/ maximal individual fitness) has an effect in our model).

At each time step we update the grid as follows:

- We select one individual from guild A at random and kill this individual. We record the trait value x_B of the individual of guild B present on this cell.
- We compute the fitness of all individuals of guild A present in the grid, should it interact with an individual of trait x_B .
- We select an individual from guild A to replace the killed one with probabilities proportional to the individual fitnesses. To make the model spatially explicit (which is not the case in this study), it would here be possible to select the individual with a probability proportional to the individual fitness time a dispersion kernel.
- The new individual from guild A has a probability μ_A to mutate, in which case its new trait is drawn in a (multivariate) normal distribution centered on the parent trait with standard deviation 1 (All trait values are thus scaled to the mutation size, which is the same for both guilds. Replacing it by a parameter sigma would be equivalent to dividing α_A and α_B by sigma). If no mutation occurs, the new individual inherits the trait value from its parent.
- We repeat the previous steps for guild B.
- We record the genealogy obtained in each of the two guilds, as well as the mutations that occurred on this genealogy.

Based on the resulting genealogies with mutations, we define species following the model of Speciation by Genetic Differentiation (Manceau et al., 2015), except that we allow the number of mutations s needed to belong to the same species to vary (in a way that is reminiscent of (Rosindell et al., 2015). Species are thus the smallest monophyletic group of individuals from the genealogy such that two individuals separated by less than s mutations belong to the same species. With this species definition, we obtain the resulting species-level phylogenies as well as an explicit partitioning of intraspecific and interspecific trait variations.

The interaction network is defined at the scale of the entire grid and is based on individuals' co-occurrence: we consider that two individuals interact if they co-occur in the same cell. Next, at the species level, we consider both a quantitative and a binary network. For the quantitative network, we take the strength of the interaction between two species to be the number of pairs of individuals of these species interacting together. For the binary network, we consider that two species interact if at least one pair of individuals of these species interact (that is if their interaction strength in the quantitative network is nonzero). An interaction network in our model thus results from the cumulative effects of smaller scale interaction events (as in for example in Pillai et al., 2011).

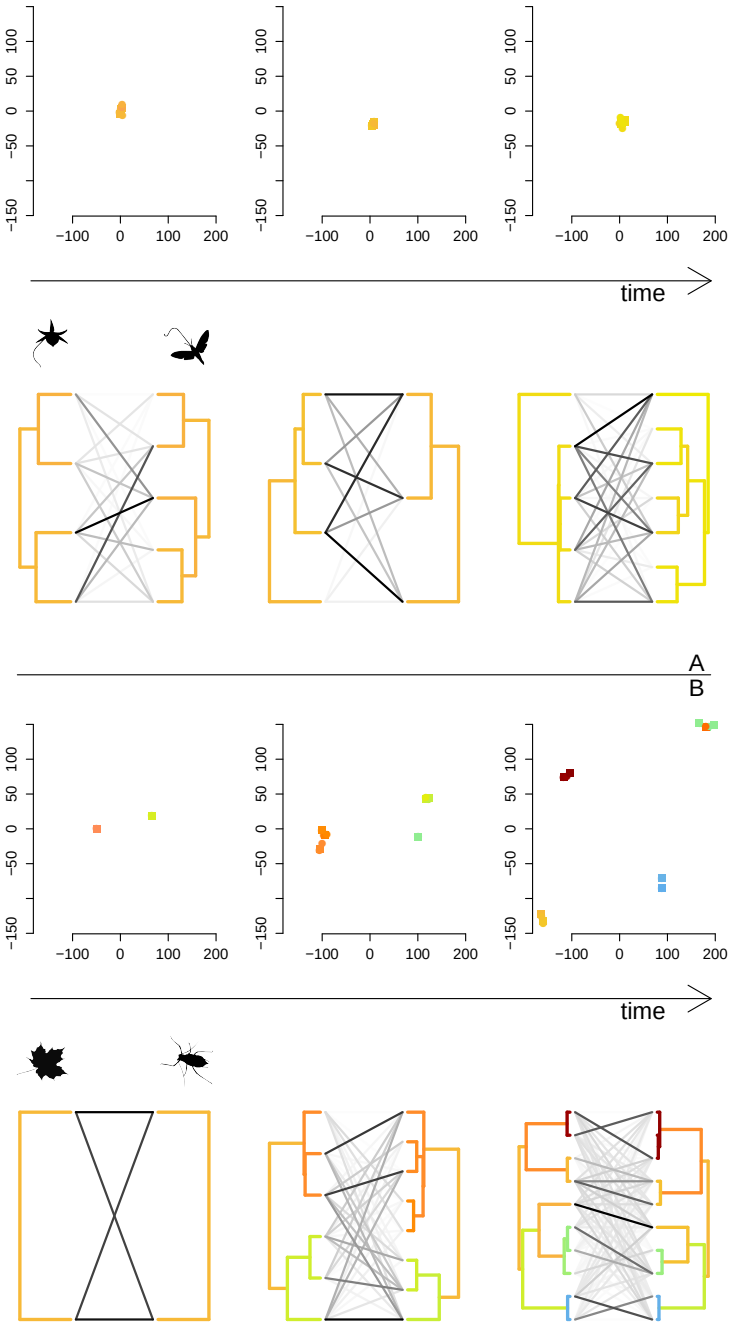


Figure 1: **Example of a model run for one mutualistic (panel A) and antagonistic case (panel B).** In each panel the upper row shows the evolution of trait distribution (in a three-dimensional trait space, point color representing the third trait dimension). The lower row shows the corresponding interaction networks, with darker links corresponding to interactions of higher strength, and associated phylogenies, with branches colored according to the value of the trait's third dimension.

Analyzing emerging patterns.

In order to explore how the type of interaction influences the diversity and structure of the network, as well as the evolutionary conservatism of interaction partners, we run the model for $4e6$ death events for the following parameters:

- 4000 individuals
- $\mu_A = \mu_B = 0.05$
- $r_A = 10$
- $r_B = +\infty$
- α_A and α_B in $0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1$

At the end of each simulation, we build the resulting phylogenies and interaction networks using 6 different species definition thresholds s ($s = 1$, 2, 5, 10, 20 and 50). Next, we compute diversity both in terms of species richness and trait diversity, measured as the variance in trait values on each trait dimension. We compute network nestedness using the weighted (for quantitative networks, method = "weighted NODF" in the nested function of the R-package bipartite (Almeida-Neto and Ulrich, 2011; Dormann et al., 2008)) and unweighted (for binary networks, method = "NODF2" in the nested function of the R-package bipartite) NODF metric (Almeida-Neto et al., 2008). We compute modularity on quantitative and binary networks with the function computeModules from the bipartite package (quanBiMo algorithme, (Dormann and Strauss, 2014)). We compare the resulting binary nestedness and modularity values to the distribution of values obtained for a null model in which the network connectance is kept constant (method "shuffle.web" in the function nullModel from the bipartite R package, that only works for binary networks, called NM1 in the following text). Both the binary and quantitative nestedness and modularity were compared to those obtained under a null model in which row and column sums of the interaction matrix are kept constant, using the method "r2d" in the function nullModel from the bipartite R package (NM2 in the following text). Finally, we compute the phylogenetic signal of interaction partners using a Mantel test that assesses the significance of the correlation between the phylogenetic distance of two species (the length of the path between these species in the phylogeny) and the dissimilarity of their interaction partners, measured as the (weighted) fraction of unshared phylogenetic branch length between the two sets of interaction partners (computed using quantitative uniFrac (Lozupone et al., 2007), "d_1" in the function GUniFrac from the R package GUniFrac (Chen, 2012)). We used a phylogenetic metric for the dissimilarity of interaction partners to limit the effect of the threshold used for species definition on our results. Results with a non-phylogenetic metric were qualitatively similar, although significance was obtained for fewer networks (Supplementary Figures).

Results.

Figure 1 shows two typical simulations (one mutualistic, the other antagonistic), illustrating simultaneously the variations in time of trait values and of network structures and outlining the main results.

Trait diversity. In mutualistic scenarios, trait values stay fairly constant throughout the simulation (Fig. 1A) and trait diversity is always lower than trait diversity in the neutral case in our simulations (Fig. 2A). This is due to the stabilizing selection that maintains species trait from one guild close to the traits of species from the other guild in those scenarios. Trait variance within a guild only depends on that guild niche width (Fig. 2A & Fig. S1A). The correlation between the traits of interacting individuals is slightly positive but stays very low (Fig. 2B). On this figure, we show the results obtained for the first trait dimension. The behavior of the model is identical in each dimension, so results were similar for the other two trait dimensions (results not shown).

For antagonistic scenarios, clusters of traits progressively emerge from the co-evolutionary dynamics (Fig. 1B). A broad parameter range leads to a trait diversity higher than in the neutral case in both guilds (Fig. 2A), and trait diversity is consistently higher than the one observed in mutualistic scenarios. The diversity is very similar between the two guilds, increases with both niche widths but collapses when consumers have a larger niche than resources (Fig. 2A & Fig. S1A). The traits of interacting individuals are positively correlated (Fig. 2B).

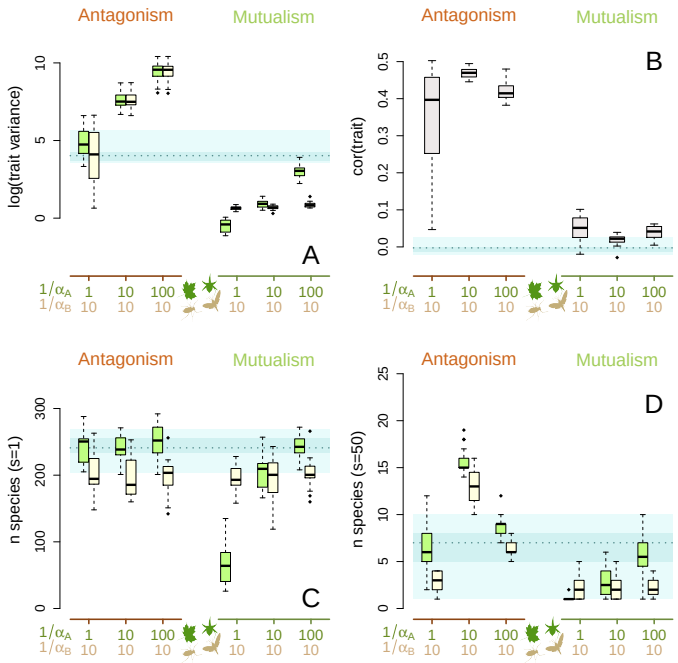


Figure 2: **Measures of diversity.** A : Logarithm of the variance of the first dimension of the species trait in 6 example scenarios. The blue area shows the values taken in the neutral case (range of values, quartiles and median). Yellow boxplots show the result for the guild A, green boxplots show the result for guild B. Antagonistic scenarios commonly lead to more trait variability than in the neutral case, while mutualistic scenarios always display a lower variability than in the neutral case. B : Correlation of the first dimension of the traits of the interacting individuals. Traits are consistently positively correlated in antagonistic scenarios, and correlation values stay pretty low in mutualistic ones. C : Species number in each guild for a species definition threshold $s = 1$. Diversity is higher in antagonistic scenarios than in mutualistic ones but always stays bellow that obtained for neutral scenarios. D : Species number in each guild for a species definition threshold $s = 50$. Diversity is higher in antagonistic scenarios than in mutualistic ones, and a few antagonistic scenarios display a diversity that is higher than in neutral simulations.

Species diversity. Globally, patterns of species diversity are largely consistent with observed variations in trait distributions described above. The emergence of network structure shows a lower species diversification in mutualistic networks (Fig. 1A) compared to antagonistic networks (Fig. 1B). For species richness, global comparison of mutualistic vs antagonistic scenarios shows that diversity is systematically larger in antagonistic scenarios, regardless of the species definition (Fig. 2C, 2D).

In mutualistic scenarios, the species number is always lower than that in the neutral case, whatever the niches width $1/\alpha$, and the results are qualitatively identical for all or species definition threshold s (Fig. 2C, D). As for trait diversity, the number of species within a guild mostly depends on its own niche width (Fig. 2C,D & Fig. S1C,D). When the niche of the clade under consideration becomes large, individuals have approximately the same fitness regardless of who they are interacting with. As trait matching becomes less influential, the behavior gets close to the one of the neutral case. But when their niche is narrow the strong stabilizing selection inhibits speciation.

In antagonistic scenarios, results depend on the species definition threshold s. For a low s, species number is always lower than that in the neutral case (Fig. 2C). Only the niche width of the consumer affects the level of diversity, with species number increasing with the consumer niche width, as in the mutualistic scenarios. Resource species number is identical in all scenarios, and similar to what we get in the neutral case. The niche width of resource species has no impact of the species number of both clades (Fig. S1C).

For a higher s (around 50, Fig. 2D), there are parameter values for which species diversity is higher than in the neutral scenario. This happens in cases where the width of the resource fitness equals or fall below that of the resource fitness, and resource fitness width has an intermediate value (Fig. 2D, S1D). For such a high species definition threshold, both guild display very similar species number, with a few more species for the resource guild.

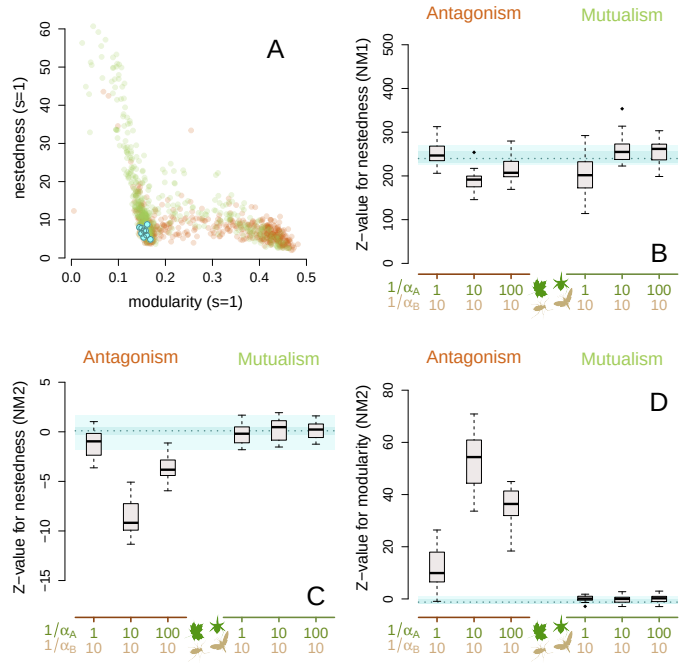


Figure 3: **Networks metrics for a species definition threshold** $s = 1$**.** A : Weighted modularity and nestedness values for all simulations. A point shows the values for one simulation, the point color indicate the interaction type (blue : neutral ; red : antagonistic ; green : mutualistic). The networks that have high nestedness values are mostly mutualistic ones, while those with high modularity values are mostly antagonistic ones. B : Z-values for the binary NODF corrected by null model NM1 (see methods) for 6 example scenarios. The blue area shows the values taken in the neutral case (range of values, quartiles and median). All networks are significantly nested. C : Z-values for the weighted NODF corrected by null model NM2. In this case almost no networks are significantly nested, and a large proportion of antagonistic networks are significantly anti-nested. D : Z-values for the quantitative modularity corrected by null model NM2. Most of the antagonistic networks are significantly modular, while mutualistic networks have modularity values close to those of the null model.

Network structure. Typical dynamics (as shown on figure 1) lead to a modular network for antagonistic scenarios (Fig. 1B) while eco-evolutionary dynamics under mutualistic interactions typically lead to highly nested structures (Fig. 1A). When we look at network metrics without any comparison to a null model, most networks have higher nestedness values than those in the neutral case were generated in mutualistic scenarios, while most of those that have higher modularity values than those in the neutral case were generated in antagonistic scenarios (Fig. 3A).

For all scenarios (including the neutral one), networks are significantly nested when compared to the null model that only conserve connectance (NM1). Z-values are very high (Fig. 3B) and mostly depend on the diversity of the community, with more diverse communities getting higher Z-values.

In the neutral case, we never get significantly modular networks (compared to the values obtained with the neutral model that corrects for abundance, NM2). For antagonistic scenarios, most networks are significantly modular. For mutualistic ones, networks are significantly modular only when the obligate mutualist fitness width is less than a tenth of the facultative mutualist fitness width, all the other scenarios leading to non-significantly modular networks (Fig. 3D, Fig. S2D).

When nestedness is compared to that of the neutral model that corrects for species abundances (NM2), the pattern is opposite to what we see for modularity (Fig. 3C vs Fig. 3D). Networks that are significantly modular are significantly anti-nested (*ie* their NODF values are significantly lower than those obtained with the null model), and network that have non-significant modularity values also have non-significant nestedness values. Thus, when compared to the neutral case, evolved antagonistic networks most often exhibit a modular, anti-nested structure, while the reverse is most often observed for mutualistic networks (Fig. 3C, Fig. S2C).

Phylogenetic signal. In the two example dynamics shown on figure 1, note that the emergent network structure is highly linked to the emergent phylogenies. Compartment in the antagonistic network structures are strongly determined by phylogenies (Fig. 1B). Conversely, in the mutualistic case, interactions are less systematically linked to phylogenies (Fig. 1A). Such observations hold more generally (Fig4A). In mutualistic communities, the phylogenetic signal in interaction partner is weak for all scenarios and species definition threshold for both guilds (Fig. 4B). In antagonistic communities, a signal in interaction partner is more commonly observed (Fig. 4A,B). It is stronger for resource species than for consumer species (Fig. 4B). For both guilds, phylogenetic signal is

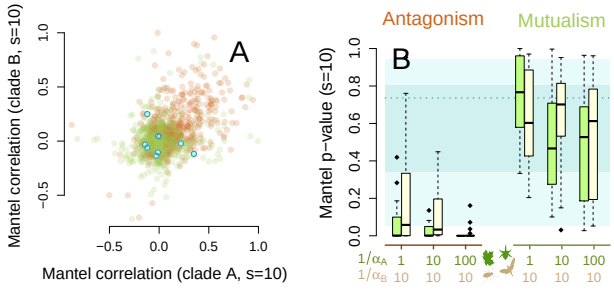


Figure 4: **Phylogenetic signal in interaction partners.** A : Mantel correlation for all simulations. A point shows the values for one simulation, the point color indicate the interaction type (blue : neutral ; red : antagonistic ; green : mutualistic). B : Mantel p-value for 6 example scenarios. The blue area shows the values taken in the neutral case (range of values, quartiles and median). Yellow boxplots show the result for the guild A, green boxplots show the result for guild B.

higher for an intermediate consumer guild fitness width, and increases with the resource guild fitness width (Fig. 4B, Fig. S3B).

Discussion.

Among the metrics we look at, many show a clear difference between values obtained for the different interaction types. Modulating the niche widths of both guilds leads to quantitative differences but the qualitative patterns hold for a large range of the parameter space and general conclusions could be drawn.

In antagonistic simulations, trait diversity was larger than in neutral simulations, as well as species richness in many cases, consistent with previous empirical and theoretical studies (Ehrlich and Raven, 1964; Yoder and Nuismer, 2010; Janz, 2011). In North American milkweeds for example, investment in defense traits resulted in higher diversification rates (Agrawal et al., 2009). On the contrary, both diversity measures (trait diversity and species richness) were decreased by the presence of mutualistic interactions when compared to the neutral scenario. Even though an empirical test that mutualism decreases diversity is hard to come up with, in particular because all natural systems include many different interactions, those results are consistent with theoretical studies (Yoder and Nuismer, 2010) and several empirical examples (for a review, see Hembry et al., 2014). In Yucca-moth interaction for instance, it has been shown that specialized pollination does not increase Yucca diversification (Smith et al., 2008). In another paper, Armbruster and Muchhala (2009) showed that, in several groups of angiosperms, it is diversity that promotes floral specialization – through character displacement – rather than the reverse. Yet other empirical systems support a positive impact of mutualistic interaction on diversification, such as is the case for fig-wasp interactions (Cruaud et al., 2012). Our model makes the strong hypothesis of fixed population sizes. Mutualistic interaction may heighten the density a community can sustain through more efficient feeding and reproduction, or open new adaptive zones, leading to a diversity begets diversity kind of mechanism (Emerson and Kolm, 2005; Joy, 2013). In the case of pollination, mutualistic interaction can also facilitate reproductive isolation (van der Niet and Johnson, 2012). Finally, we considered a non-spatial model, but communities are integrated in a geographical context and it may also be that geographical isolation is necessary for mutualism to promote speciation (Thompson and Cunningham, 2002; Kay and Sargent, 2009). Our results suggest that it is unlikely for diversity to be increased by mutualistic interactions without the help of this kind of mechanisms.

The choice of the null model has a dramatic impact on the conclusions that can be drawn for network structure patterns. When compared to the null model in which only the total connectance was constrained (NM1), all our binary bipartite networks showed a strong nestedness signature. Yet when they are compared to the null model in which species abundances are kept constant (NM2), the nestedness values are either unsignificant or below those obtained for the null model. It is important to remark that NM2 is equivalent to randomly reassigning a position on the grid for all individuals without looking at their trait values, which is also how networks are built in our neutral simulations. Using this null model is a way to correct for species abundances. All the nestedness signal we see in our simulations thus comes from the uneven species abundances distribution, in agreement to the neutral hypothesis (Vázquez, 2005; Vázquez et al., 2009; Santamaría and Rodríguez-Gironés, 2007; Krishna et al., 2008; Staniczenko et al., 2013; Coelho and Rangel, 2018). Most empirical or theoretical studies that found significant netstedness in bipartite networks did not use a null model correcting for species abundances (Thébault and Fontaine, 2010; Lewinsohn et al., 2006; Bascompte et al., 2003). But studies that do correct for it do not find a clear nested pattern (Vázquez, 2005; Staniczenko et al., 2013; Canard et al., 2014).

The differences in network structure we get between antagonistic and mutualistic communities are consistent with empirical observations. In our antagonistic simulations, reciprocal specialization is common between interaction partners, leading to a positive correlation between the interaction partner traits. The identity of the partner thus depends on the traits values as well as on chance, decreasing the nestedness values (compared to those in NM2) and leading to positive modularity Z-values (compared to NM2).

In mutualistic simulations, because trait diversity stays low compared to the fitness width, the interaction partner identity is mostly due to chance (which we can see in the low correlation values between interaction partners' traits in mutualistic simulations), and networks structure stay close to that of neutral interaction networks (and to that of networks from NM2), both for nestedness and modularity. It is remarkable that most of the neutral bipartite interaction model were built to explain the structure of mutualistic rather than antagonistic networks (Vázquez 2005, Santamaría and Rodríguez-Gironés 2007, Krishna et al. 2008, Coelho and Rangel 2018, but see Canard et al. 2012, in which they find nestedness values in their neutral model that are in the upper range of their empirical values). Here we show that evolutionary dynamics leads to nested interaction structure in mutualistic communities that are close to those obtained in neutral communities, even though the evolutionary outcomes (species number, trait diversity...) are much different between mutualistic and neutral communities. That being said, we do find raw nestedness values that are higher in mutualistic scenarios than in neutral (and antagonistic) ones, in a similar way that what can be seen for empirical bipartite networks (Thébault and Fontaine, 2010; Fontaine et al., 2011).

We get different phylogenetic signals in interaction partners for different interaction types. Because traits values are only of marginal importance for the choice of the interaction partner in mutualistic scenarios, given the restricted evolved trait range, the phylogenetic signal in interaction partner stays low compared to what we see in antagonistic scenarios. In those scenarios, resource species generally show a strong signal in interaction partners, while the signal is positive but have lower significance for consumer species. This is once again in accordance to what has been found in empirical communities (Krasnov et al., 2012; Fontaine and Thébault, 2015). In plant herbivore systems, the stronger conservatism seen in plants has sometimes be interpreted as a consequence of the fact that chemical defenses are difficult to evolve and thus much constrained (Fontaine and Thébault, 2015), but in our model we get to a similar pattern without making the assumption of different evolutionary rates among guilds. One of the possible explanation is that for resource species it is always beneficial to change trait to avoid consumption. Only the few directions in trait space in which there are consumers must be avoided, but these represent a very small proportion of the possibilities (this would not be the case in a one-dimensional trait space). But for consumer species, there is a selection pressure to go toward existing resource species, so it is not uncommon to see convergence in traits, even between rather phylogenetically distant species (host shifts), which weakens the phylogenetic signal.

Supplementary figures for Chapter 3

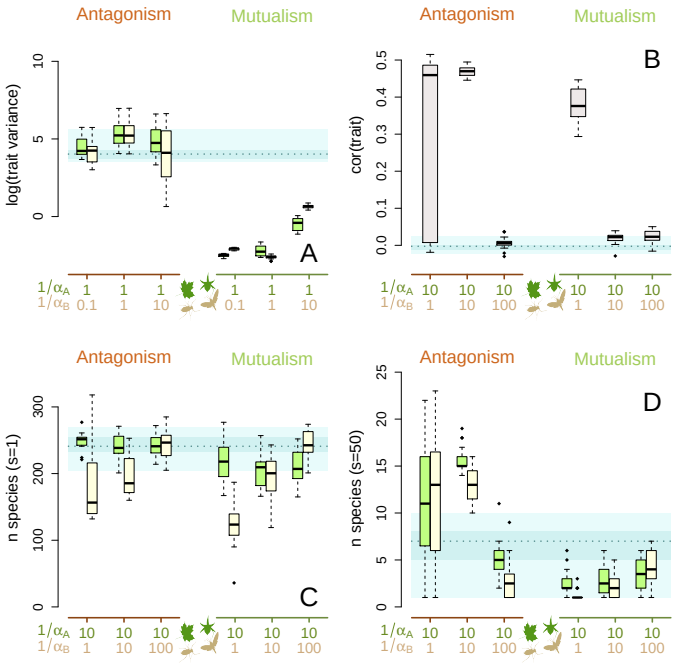


Figure S1: **Measures of diversity.** A : Logarithm of the variance of the first dimension of the species trait in 6 example scenarios. The blue area shows the values taken in the neutral case (range of values, quartiles and median). Yellow boxplots show the result for the guild A, green boxplots show the result for guild B. Antagonistic scenarios commonly lead to more trait variability than in the neutral case, while mutualistic scenarios always display a lower variability than in the neutral case. B : Correlation of the first dimension of the traits of the interacting individuals. Traits are consistently positively correlated in antagonistic scenarios, and correlation values stay pretty low in mutualistic ones. C : Species number in each guild for a species definition threshold $s = 1$. Diversity is higher in antagonistic scenarios than in mutualistic ones but always stays bellow that obtained for neutral scenarios. D : Species number in each guild for a species definition threshold $s = 50$. Diversity is higher in antagonistic scenarios than in mutualistic ones, and a few antagonistic scenarios display a diversity that is higher than in neutral simulations.

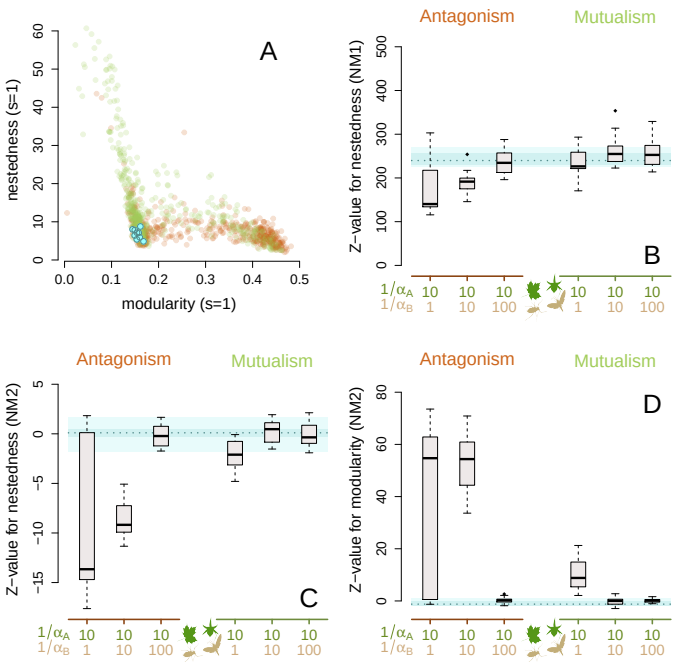


Figure S2: **Networks metrics for a species definition threshold** $s = 1$**.** A : Weighted modularity and nestedness values for all simulations. A point shows the values for one simulation, the point color indicate the interaction type (blue : neutral ; red : antagonistic ; green : mutualistic). The networks that have high nestedness values are mostly mutualistic ones, while those with high modularity values are mostly antagonistic ones. B : Z-values for the binary NODF corrected by null model NM1 (see methods) for 6 example scenarios. The blue area shows the values taken in the neutral case (range of values, quartiles and median). All networks are significantly nested. C : Z-values for the weighted NODF corrected by null model NM2. In this case almost no networks are significantly nested, and a large proportion of antagonistic networks are significantly anti-nested. D : Z-values for the quantitative modularity corrected by null model NM2. Most of the antagonistic networks are significantly modular, while mutualistic networks have modularity values close to those of the null model.

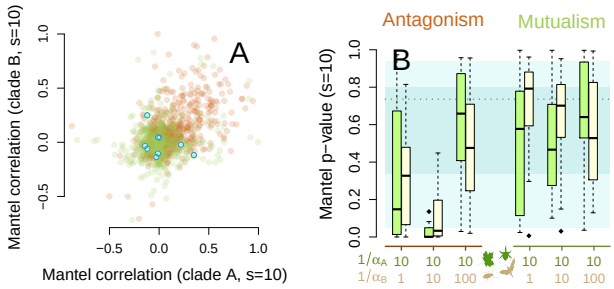


Figure S3: **Phylogenetic signal in interaction partners.** A : Mantel correlation for all simulations. A point shows the values for one simulation, the point color indicate the interaction type (blue : neutral ; red : antagonistic ; green : mutualistic). B : Mantel p-value for 6 example scenarios. The blue area shows the values taken in the neutral case (range of values, quartiles and median). Yellow boxplots show the result for the guild A, green boxplots show the result for guild B.

General discussion

1 Various modeling approaches to study biodiversity evolution.

In the previous chapters, I have presented my work on diversification. Each of them propose a different approach, all model-based, but the model types are very different in nature.

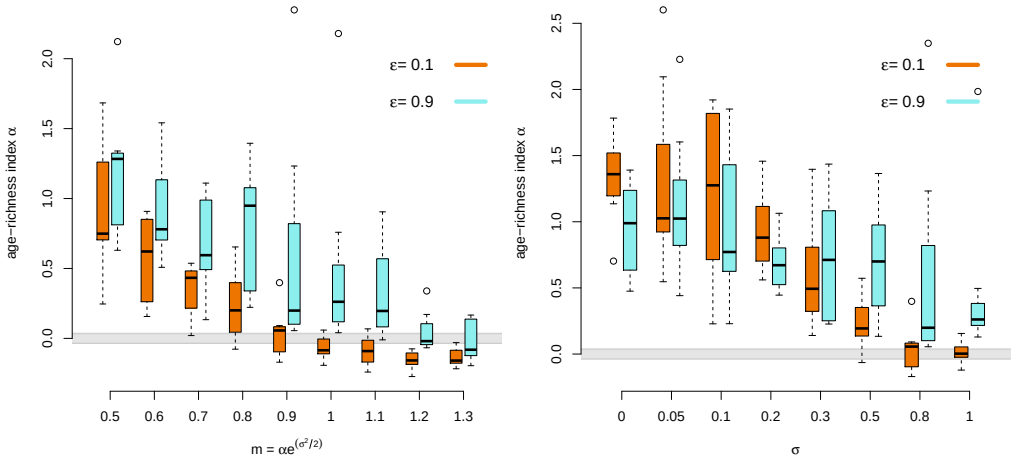
In Chapter 1, we propose a model that generates ranked tree shapes endorsed with relative abundances at their tips. The model has three parameters – β, controlling for tree balance, α, controlling for node order within the tree, and η, controlling for the relation between clades age and richness – but does not propose a biological process for generating those shapes. We study the effect of the different parameters on the relation between species extinction and phylogenetic diversity loss – the phylogenetic diversity (PD) of a group of species is the sum of the branch lengths of the phylogeny that encompass those species. We propose a method to infer our model parameters on empirical phylogenies, test it on a simulated data set and eventually apply it to a data set of family level bird phylogenies.

In Chapter 2, we propose a lineage-based diversification model with non-homogeneous diversification rates. Lineages are characterized by their own speciation and extinction rates, and at each speciation events, the two daughter lineages get new diversification rates sampled in a law that depends on the rates of their parent lineage. We here again propose an inference procedure to determine the more likely model parameters that could have generated a given phylogeny, among which the mean change in speciation rate and the variability of the new rates distribution. This also enables to 'paint' the branches of phylogenies with their most likely speciation rates, highlighting regions in the tree that diversified at different tempos than the rest of the clade. Applying the method to bird phylogenies, we show that the within-clade variability in speciation rates is comparable to the between-clades variability in speciation rates in this data set.

Finally, Chapter 3 presents an individual-based model with two interacting guilds, aiming at studying the effect of distinct bipartite interactions on the diversification process. In this model, individuals are characterized by trait values, and their fitness depends on the difference between their own trait and that of the individual of the other guild they are interacting with. The interaction type is determined by whether having a trait close to that of the interactor is advantageous or disadvantageous. Individuals die at random, and reproduce to replace dead individuals with probability depending on their fitness value. Mutation in the trait value may occur at a reproduction event, and accumulating mutations eventually results in speciation. We studied the effect of our model parameters, and especially interaction type and the width of the interaction niche, on different summary statistics of the system at stable state. Those include species and trait diversity, interaction network nestedness and connectance, and phylogenetic signal in interaction partners. Antagonistic communities are more diverse – both in terms of species and trait diversity –, more modular and less nested in our simulations than mutualistic ones, in agreement with what can be observed from empirical data.

The three models present a gradation in how finely processes are being described and implemented, and on the level of organization that it focuses on. The first two models aim at describing patterns at different levels. The ranked tree shape model (in Chapter 1) is looking at the general shape of the phylogeny, and offers a way to describe it. Although those kinds of non-mechanistic models do not attempt to discriminate between diverse biological processes, they allow to quantify patterns in the data – here the clades age-richness relationship – and to highlight mismatches between the data and what would be expected under given hypothesis – as is the case here in our bird

data set, in which α values cluster around 0, that should be compared to the value of 1 expected from an homogeneous speciation process. One of the possible explanations for this deviation from the Yule model expectation would be the presence of rate heterogeneity between the lineages in the phylogenies (this can be illustrated by the fact that in our model from Chapter 2, adding rate stochasticity make α estimates go from 1 to 0, Fig. 1b).



(a) (b)

Figure 1: **Age-richness index α from Chapter 1 of trees simulated with the ClaDS model of Chapter 2.** Simulations are performed for with a common $\sigma = 0.8$ and varying m (a) or with a common $m = 0.9$ and varying σ (b), and two turnover rate values ($\varepsilon = 0.1$, orange boxes; $\varepsilon = 0.9$, blue boxes). For all simulations, $\lambda_0 = 0.1$, and phylogenies are conditioned by their final tip number $N = 100$. In empirical phylogenies, we found α values clustering around 0 (gray lines). A similar value can be obtained with the ClaDS model when rates are increasing at speciation events (high m a) or when the speciation rates are highly variable (high σ values, 1b).

In the heterogeneous speciation model (Chapter 2), the focus level is that of lineages. Contrary to the first model, it refers to biologically meaningfull events – speciation, extinction, rate changes, and can thus been viewed as a mechanistic model. However no hypothesis is formulated about the finer scale processes that can result in events happening to lineages, and no biological explanations are proposed. As is the case in the first model, the aim is rather to describe and quantify the patterns observed in empirical phylogenies, even though the ultimate motivation would be to understand the underling biological reason for these changes in rates. The method also provides a characterization of tree-level properties, through the estimation of the rate stochasticity σ and trend in rate changes α. The model is able to generate phylogenies with very different shapes. Importantly, it is able to generate tree shapes close to those observed in nature, in terms of imbalance, tippiness (quantified with β and γ, see Fig. S1 and S2 in the Supplementary Material for Chapter 2), and clades' age-richness correlation (Fig. 1).

While the first two model are introducing ways to describe and characterized heterogeneity in diversification rates, the last model (Chapter 3) aims at proposing a mechanism to explain the patterns observed in empirical data. The focus level of organisation is that of individuals, which makes it straightforward to include more complex, ecological mechanisms that what is allowed from the modeling choices made in Chapters 1 and 2. The model enables to simulate data, which can then be compared to our empirical knowledge of the system. The comparison to empirical patterns

is done here in a qualitative way – we observe general tendencies for the studied metrics that are in accordance with the differences observed between mutualistic and antagonistic interaction networks in nature. Because we do not have a likelihood function in this case, a more quantitative analysis would have to be performed through the use of approximate methods like Approximate Bayesian Computation (ABC), but the time needed to simulate the model would prevent to use this approach in our case. Furthermore, being able to find the best parameter set that fits a dataset would doubtfully be of little use here: we certainly do not believe our model to reproduce mechanisms in a realistic way, but we want to evaluate whether simple evolutionary mechanisms suffice to explain general tendencies observed in ecological communities.

2 Limitations and perspectives.

2.1 Goodness of fit.

Fitting a single model to data does not guarantee that it is indeed a good model for those data. This is also true when selecting model among a set of candidates, which gives the best model of the set, with no guarantee that it is a good one. If the model is well suited to our data, we would expect it to be able to generate data that are similar to ours. A standard approach is to compare the divergence from model predictions by generating simulated data with the model and quantifying how they differ from the actual data regarding key summary statistics. The significance of the deviation between the data and model prediction can be assessed using p-values – computed as the proportion of the simulations in which the summary statistic was higher than the observed value –,

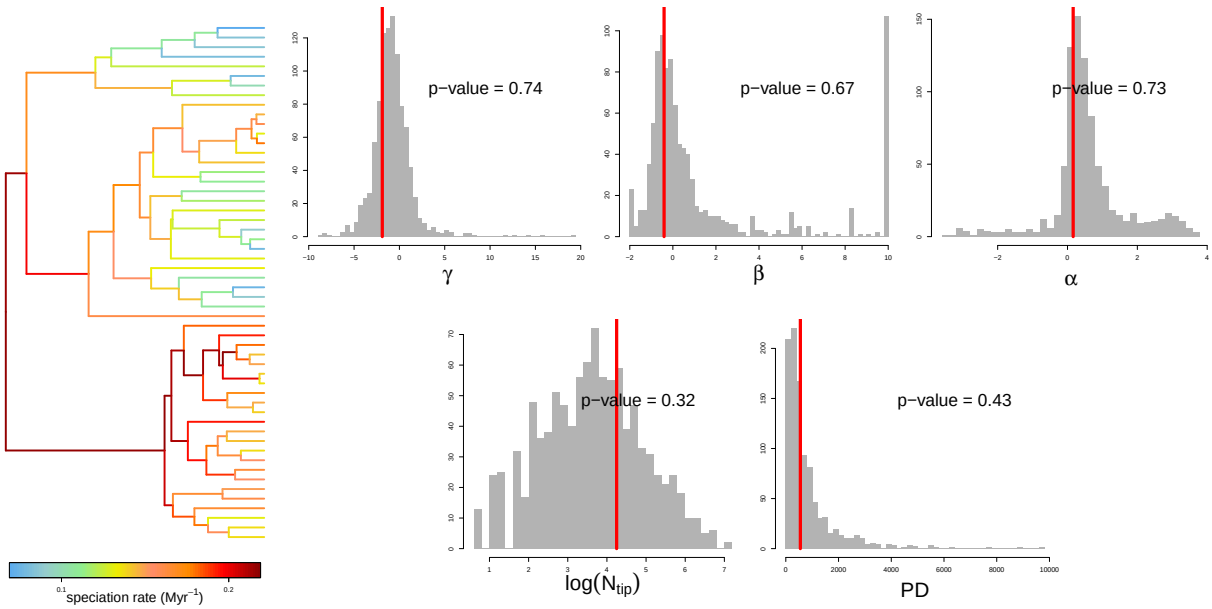


Figure 2: **Posterior predictive distributions of 6 summary statistics for the model in Chapter 2 applied on the phylogeny of *Alcedinidae* (kingfishers).** The histograms show values obtained from 1000 simulations from the posterior predictive distributions. Simulated trees are conditionned on their crown age. The statistics shown are the γ statistic (tree tippiness), the β statistic (tree imbalance), the α statistic from our model in Chapter 1 (age-richness index), the logarithm of the number of tips in the tree, and the phylogenetic diversity (PD, sum of all branch lengths in the tree). The red vertical line shows the value for the empirical phylogeny.

but it is often more informative to plot the observed value against the reference distribution. In the frequentist framework, data are simulated with the model using the Maximum Likelihood Estimates for the parameter values. In the Bayesian framework, this approach is known as *posterior predictive checks* (Gelman et al., 1996), and data are simulated with the generating model with parameters sampled from the posterior distribution.

In figure 2, we provide an illustration of this approach for the method in Chapter 2 and one of the bird trees that were used to illustrate the method in this chapter, the phylogeny of *Alcedinidae* (kingfishers). The models parameters (rate stochasticity σ, trend in rate changes α, species turnover rate ϵ, initial speciation rate λ_0) are sampled from the MCMC chains, a phylogeny is simulated with our model using this parameter set whose tips are sampled with probability $f = 0.56$ (the empirical sampling fraction used for the inference), and summary statistics are computed for the simulated and empirical data. None of the computed summary statistics for the empirical tree deviates substantially from the posterior predictive distribution, showing a good match between the model and the data in this example.

2.2 Incomplete sampling.

In diversification analysis.

Non-uniform species sampling.— A nice property of our ranked tree shape model (Chapter 1) is that it is sampling consistent – meaning that, for a given parameter set, we get the same probability distribution on ranked tree shapes either by generating a n-tipped tree with the model or by sampling n tips from a N-tipped tree generated with the model ($n < N$). In a similar way, the heterogeneous speciation model allows to account for incomplete species sampling, as long as the the probability of being sampled is the same for all living species. Those properties are important when inferring the parameters of both models on empirical data, because empirical phylogenies are very rarely complete. Not taking this into account may lead to bias in parameter inferences and to an apparent decrease in speciation rate (Pybus and Harvey, 2000).

However both these properties rely on the assumption that the sampling is performed uniformly among species. This is often not the case, as phylogenies are generally built so as to be as characteristic of the group as possible, by including early-diverging species or species representative of previous intrataxon classification. Because closely-related species are less likely to be included, this typically results in a more thorough sampling of deep nodes, which can be mistaken for a slowdown in diversification rates through time (Cusimano and Renner, 2010; Brock et al., 2011). Non uniform taxa sampling can also affect the reconstructed phylogeny topology (see for example Fig. 8 from Chapter 1 ; here the deviation from random sampling is measured with the η metrics from our model). Methods have been proposed to correct for the bias in the context of the detection of a slowdown in diversification (Brock et al., 2011; Cusimano et al., 2012), and in that of diversification rates inference (Höhna et al., 2011). Also, even when sampling is done uniformly within subgroups, the sampling fraction may vary among them. For the method we proposed in Chapter 2, this could bias rates estimates towards lower values in subclades in which the sampling fraction is lower than the global fraction. While our method does not yet include such an option, it would be theoretically possible to add it by putting different sampling fractions for different parts of the tree in the likelihood. The sampling fraction for branches that are ancestral to parts of the tree with different sampling intensity would then be set as the average of the sampling fractions, weighted by the number of descending species in each subclade.

Selection for clade size.— A sampling bias can also come at play at the level of the clade. As already mentioned in the introduction, previous clade age-richness relationship measures necessitated to choose how to cut a big tree to select which clade to include in the analysis. This led to different results obtained for different such choices (Stadler et al., 2014; Sánchez-Reyes et al., 2016). Our ranked tree shape model in Chapter 1 offers a way to deal with this problem by looking at the relative order of all the nodes within a given phylogeny. In diversification analysis, a first source of bias is the fact that extinct clades are never included in studies, because they cannot be observed in the present (Magallón and Sanderson, 2001). This is generally corrected for when fitting diversification models to data by conditioning the likelihood on the non-extinction of the process, as we did in our method in Chapter 2. Failing to do so leads to biases in estimated diversification rates. A similar source of bias comes from the fact that species-poor clades are not included in diversification analysis either (especially in adaptive radiation studies, in which speciose groups are more likely to be selected Pennell et al., 2012). Even when adaptive radiation is not the focus of the study, fitting a model to a small tree will not lead to easily interpretable biological results because estimators are much variable for small datasets, and those will thus not appear in studies. When trees are simulated with a birth-death process conditioned on the age of the crown, the largest phylogenies are those that by chance diversified a lot in the beginning of the simulation. Large phylogenies are thus very likely to show a signature of early burst – with the γ statistic or by comparing constant rates and time-dependent diversification models–, even though the rates were constant in time in the simulations (Phillimore and Price, 2008; Pennell et al., 2012). In our empirical study in Chapter 2, we only kept the bird phylogenies with at least 50 tips, because applying the method to smaller phylogenies does not lead to reliable parameter estimates; it is thus possible that selection for clade size biased our results, especially regarding the estimates of the trend parameter α. However, we found no significant correlation between this parameter values and the size of analyzed trees (p-value=0.48). This bias is thus likely not to be very strong, and we do not think it can explain the pervasive pattern of decrease in speciation rates found in this dataset.

In network studies. Ecological networks are of course also subject to the problem of incomplete sampling. The number of possible interactions increases as a power of two of the number of species, making it very difficult and costly to record all the interactions occurring in a community. Incomplete interaction sampling has been shown in a simulation study to have an effect on different measured network metrics, especially if there is a skew in the number of observations per species, and patterns such as the decrease of connectance with network size might be partly explained by those sampling effects (Blüthgen et al., 2008). The way interactions are sampled in empirical communities may also bias measures of network structure. Whether sampling effort is allocated equally among plant species or proportionally to their abundance has a significant effect on the obtained network structure for a given sampling effort, in particular in terms of connectance (Gibson et al., 2011). Other network metrics, however, are more robust to incomplete sampling and show deviation to the true network values for very low sampling effort only (Nielsen and Bascompte, 2007; Rivera-Hutinel et al., 2012). This is the case of the metrics we considered in our study – nestedness, measured as NODF (Almeida-Neto et al., 2008), and modularity.

2.3 Molecular phylogenies as empirical data.

Phylogenetic uncertainty. In both Chapter 1 and Chapter 2, we assumed that phylogenies where perfectly known. In practice, they are reconstructed from molecular data, and are subject to errors and uncertainty, both in terms of topology and branch lengths. When the reconstruction is done through bayesian methods, which is commonly the case, it provides a sample from the posterior distribution of phylogenies rather than 'a best' tree. A possibility to deal with this consists in summarizing this information into one representative single tree topology. Several methods allow to do this (see Heled and Bouckaert, 2013, for a review). One family of methods rely on *consensus* rules. A *strict consensus tree* contains only the splits that are present in all the trees from the posterior sample ; a *majority rule tree* contains only the splits that are present in at least half of the trees from the posterior samples. Consensus trees are often unresolved, with polytomies. The *Clade Credibility* of tree is equal to the product of the posterior density of all the clades in the tree (the posterior density of a clade is the frequency at which it appears in the posterior sample). The tree with the highest clade credibility in the posterior sample is called the *Maximum Clade Credibility* (MCC) tree. Node depths can then be assigned as the mean or the median of the clade age in the posterior sample. Contrarily to consensus methods, this produces a fully bifurcating tree. This is a necessary condition to apply the methods developed in Chapters 1 and 2, so we used MCC trees in our empirical applications in these chapters.

Summarizing the information in one single representative topology can however be misleading. The summary tree is an average of the posterior distribution and is not itself guaranteed to have a high posterior density, and the result of the inference on this tree is not necessarily representative of the rest of the posterior distribution. A good practice is to assess the sensitivity of the method to phylogenetic uncertainty by applying it to many trees from the posterior distribution. In a bayesian framework, one possibility is to sample trees from the posterior distribution at the same time as we sample for the model parameters, thus integrating the inference over the phylogenetic space. In our case however, the inferences of our models parameters are already costly in terms of computation time, preventing us to account for phylogenetic uncertainty in our empirical applications in both Chapter 1 and Chapter 2.

Bias in reconstruction methods. Even when accounting for phylogenetic uncertainty, the phylogeny reconstruction may be biased towards certain topologies (Huelsenbeck and Kirkpatrick, 1996; Holton et al., 2014) or branch length distributions. By simulating trees according to the Yule model together with DNA sequences evolving along their branches, and subsequently reconstructing the phylogenies using those sequences, Huelsenbeck and Kirkpatrick (1996) and Holton et al. (2014) showed that the reconstructed trees tend to be biased towards more imbalanced topologies than the true simulated topologies, especially when evolutionary rates are high. For our method in Chapter 2, the would lead to inappropriately high σ estimate (the rate stochasticity parameter). According to the author, this bias towards unbalanced topologies comes from the fact that, if the phylogenetic signal in the sequences is low, the inferred distribution of phylogenies will get closer to a uniform distribution on labeled tree shapes (tree topologies with species identity at tips), which puts a stronger weight on imbalanced tree (expected β value -1.5) than the distribution generated by the Yule model (expected β value 0; see Section 1.2 in the Introdution). However, it has been argued by Aldous (2001) that tree reconstructed 'by hand' also tend to be more unbalanced than Yule trees, while it is reasonable to assume that people would more likely be biased towards constructing too-balanced trees for the sake of classification. Also, if the bias towards unbalanced topology

reflects the lack of signal in the molecular data, providing posterior distributions that look like the prior distribution, one would expect reconstructed trees to become more and more balanced as reconstruction methods are improving and larger sequences are being used. To my knowledge, no such effect has been reported, comforting us in the belief that reconstructed phylogeny imbalance indeed reflects characteristics of the diversification process rather than methodological artifacts.

A bias is also possible in the estimation of branching times. Underparametrization of models of sequence evolution, for example by ignoring the possibility to have asymmetric nucleotide substitution rates, or ignoring that rates may vary in time or across the phylogeny, can lead to inferred phylogenies with nodes closer to the root than they should be in reality (resulting in strong negative bias in the γ statistic ; Revell et al., 2005; Ho et al., 2005). The topology of the tree may also affect the estimate of diversification times. Tree imbalance can lead to underestimations of deep nodes ages (Duchêne et al., 2015). This is linked to the node density effect (Sanderson, 1990; Hugall and Lee, 2007), an expression referring to the fact that the amount of sequence changes may be underestimated along long branches – that are more common in imbalanced trees –, leading to an apparent increase in sequence evolution in the more diverse region of the tree. All those biases could lead to inappropriately low estimates of the trend parameter α in our model in Chapter 2.

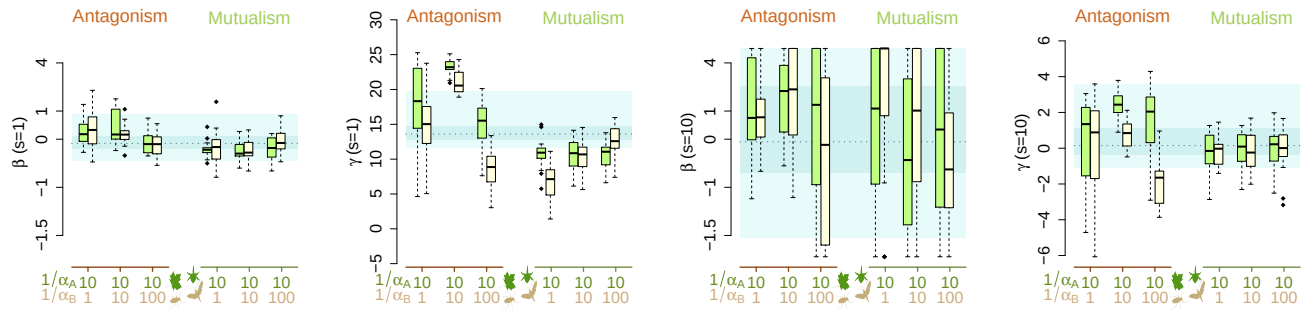
2.4 From patterns to processes.

The first two chapters aims at describing and quantifying patterns present in empirical data. On the contrary, goal of the last chapter is to propose a mechanism for the emergence of empirical patterns, principally in interaction network structures, that are already well known from the literature on empirical networks. Many processes can lead to the same pattern. Nestedness as an example has been obtained from many models, involving community dynamics, neutral processes, or species phenotypes. Yet it is likely that a mechanism proposed to describe a pattern cannot account for all the other. It is thus important to look at many patterns if we are to tell apart the relative importance of different mechanisms (McGill, 2003).

In our model, we focused on the differences between mutualistic and antagonistic communities, and thus at metrics that have been reported to vary between the two of them. We have thus looked at descriptors of the community diversity, both in trait and species diversity, at well described networks measures, nestedness and modularity, as well as at phylogenetic signal in interaction partners. Our results are in good accordance with the tendencies observed in empirical communities, showing that they are well explained by the evolutionary process we considered. However, a good number of different patterns can be obtained from our model, both for ecological and phylogenetic metrics. Studying them would give further insights into the model's behavior and on what aspect it might depart from empirical data. I will present a few patterns which may be of interest to look at.

Tree shape patterns. We get the phylogenies of species from both guilds as a model output. Phylogenetic shape contains information about the diversification process, and, as mentioned in the introduction, is well described in the literature for empirical phylogenies. However, in Chapter 3, we focused on differences between antagonistic and mutualistic communities. Not such systematic differences have been reported for phylogenetic shape, and we do not expect to find any, because in empirical communities all kinds of interactions are present. No phylogeny can be said to be the product of only antagonistic or only mutualistic interactions, contrary to what is the case in our model.

Figure 3 shows the β and γ statistics obtained at the end of our simulations. For low species definition ratchets ($s = 1$, Fig. 3a,b), both β and γ values are too high compared to what is seen in



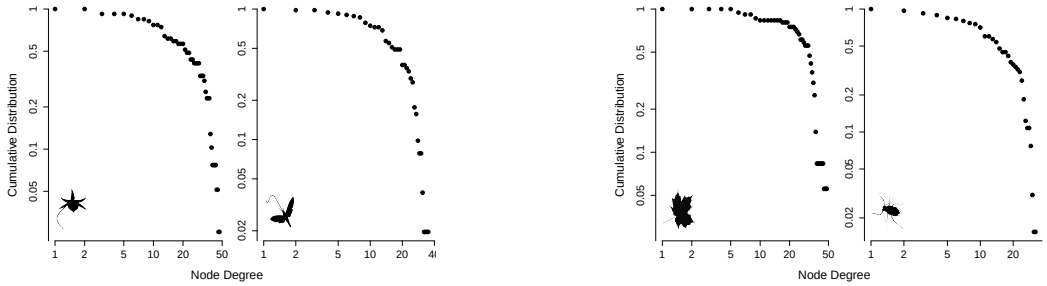
(a) (b) (c) (d)

Figure 3: **Tree shape metrics obtained for simulations from the model in Chapter 3.** The metrics are shown for 6 examples scenarios, for two different species definition ratchets, $s = 1$ (a, b) and $s = 10$ (c, d). In each panel, the blue area shows the values taken in the neutral case (range of values, quartiles and median). (a) : Tree imbalance, measured with the β statistic for $s = 1$. (b) : Tree tippiness, measured with the γ statistic for $s = 1$. (c) : β statistic for $s = 10$. (d) : γ statistic for $s = 10$.

empirical phylogenies for the 6 scenarios shown, indicating that the phylogenies generated by our model are too balanced and tippy. γ, however, is much dependent on the species definition ratchet s, as this parameter controls for the number of mutations needed to belong to different species and thus the time to speciation, making speciation more and more protracted as s gets higher (see the paragraph 'Protracted speciation' in the section 2.1 from the Introduction). For $s = 10$, γ values are closer to empirical values (Fig. 3d). β is less sensitive to this parameter, although its variance increases as the number of species in the guild gets lower. What value of s would be a 'good' value is a tricky issue. The γ values suggest that a fairly high on would be needed, but species number then becomes rather low for the community size we used in our simulations, thus the results on other metrics display high variations among simulations with the same parameter set, making it difficult to reach conclusions.

Differences can be seen on Figure 3 between antagonistic and mutualistic communities, even though the patterns are far from being as clear as those obtained for network or diversity metrics (see Chapter 3). Clades in antagonistic communities show a tendency to be more balanced and tippy than those in mutualistic ones, but further work would be required to understand whether those differences are significant and if they hold for other parameter values. It would also be interesting to see whether a similar trend can be observed in empirical communities, between groups in which the diversification is thought to have primarily driven by mutualistic interactions (*e.g.* fig-wasp systems, Cruaud et al., 2012), ad groups in which it is thought to have been driven by antagonistic interactions (*e.g.* plant-herbivore systems, Ehrlich and Raven, 1964; Futuyma and Agrawal, 2009).

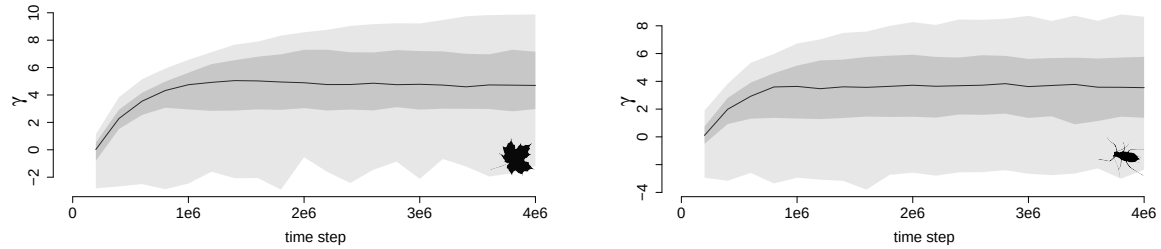
Node degree distribution. Network-wise metrics such as nestedness are much impacted by the node degree distributions. In empirical data, node degree distributions are well fitted by a truncated power law. A first glance at the node degree distributions from two example simulations in our model outputs suggests that we do get this type of distribution (Fig. 4). It would be interesting to see if this holds true for all scenarios, and how the parameters of the fitted distributions vary with our model parameters.



(a) (b)

Figure 4: **Example node degree distributions obtained for simulations from the model in Chapter 3.** The species definition ratchet s is set to 5 in both examples. (a): Log-log plot of the cumulative node degree distribution for each of the two guilds in a mutualistic scenario ($\alpha_A = -0.02$, $\alpha_B = 0.1$). (b): Log-log plot of the cumulative node degree distribution for each of the two guilds in an antagonistic scenario ($\alpha_A = -0.1$, $\alpha_B = 0.5$).

Temporal dynamics. In our study, we looked at the patterns obtained with our model at stationary state. This allows our results to be independent from the starting point and to be able to compare different scenarios at the same state of evolution. But whether ecological communities are at equilibrium or not is a controversial issue. For tree shapes in particular, it has been shown that looking at out-of-equilibrium dynamics generates phylogenies that are closer to empirical ones, in terms of β and γ statistics (Liow et al., 2010; Gascuel et al., 2015; Missa et al., 2016). A preliminary look at our simulations shows that they also present a initial phase with low γ values (Fig. 5), but because we were interested in the stationary phase, we did not record the system state at a fine



(a) (b)

Figure 5: **Temporal dynamics of the γ statistic (tree tippiness) obtained for simulations from the model in Chapter 3.** All antagonistic scenarios are pulled together, for guild A (a) and guild B (b). In the beginning of the simulations, γ values are rather low, but a finer time resolution would be needed to see whether they pass through negative values in the initial phase. The black lines shows the median, the dark gray area the middle quartiles, and the light gray area the $[0.05, 0.95]$ quantiles. The species definition ratchet s is set to 5.

enough time resolution to be able to study this phase. Simulation studies also report that tree shapes take longer to reach equilibrium than species diversity (Missa et al., 2016). It would be interesting to see whether this is also the case for network structure, or if the nestedness or modularity of the community predates the diversity equilibrium.

2.5 Extensions of the bipartite speciation model.

Sensitivity to model's assumptions. In Chapter 3, we studied how the model behavior's changed with the niche width of both guilds. To do so, we had to fix other parameters. Because of the time needed to run the model, we did not carry out a systematic study of their impact on the metrics we looked at, but it would be of interest to assess the sensitivity of our results to their effect.

Other fitness parameters.— The individual fitness in our model depends on two parameters; its width α (which is the one we varied in Chapter 3), and the gain in fitness r (which is the fitness maximum divided by its minimum). Varying the r parameter would allow to modulate how much a guild is dependent on the other, and what it implies for the community structure. For mutualism for example, it would allow to study if we expect to attain different structure if species are facultative or obligate mutualists, and when there is an asymmetry in the dependencies.

Population sizes.— The model was run for a fixed population size of 4000 individuals per guild. This choice was made for computational reasons. While we do expect some of the metrics to depend on population size – principally the diversity measure metrics – the other should be less sensitive to this parameter. Yet the tree metrics mentioned in the previous part suggest that a high species definition ratchet s gives more realistic tree shapes, but the other metrics cannot be studied for a high s because the number of species get very low. Being able to run the model for more individuals would allow to study those metrics for higher species definition ratchet which may be more realistic.

Trait space dimensionality.— The whole study of the model's behavior was performed for a 3-dimensional trait. We chose to use a multi-dimensional trait because the one-dimensional behavior is expected to be very specific. It has been shown indeed that multidimensional traits are likely to have evolutionary dynamics qualitatively different from one-dimensional ones (Gilman et al., 2012; Ispolatov et al., 2016). It is quite intuitive in the case of our model that resource species are likely to be stuck between two consumer species in a one-dimensional trait space, while escape is more likely when the interaction is driven by several traits or a multi-dimensional one. Another reason for this choice is a previous study showing that ecological networks are best described by traits with several yet few dimensions (Eklöf et al., 2013). Yet the dimension choice is a rather arbitrary one and we could try to assess how sensitive our results are to this assumption.

Abiotic selection. In our model the trait space is unbounded, with fitness depending only on the individual's trait and that of its interactor. This assumption makes it straightforward to compare our results to the neutral case. However, it would be reasonable to assume that this is not the case, as most traits can be costly to maintain at extreme values. This could easily be included in the model by multiplying the individual fitness by a gaussian function centered on an abiotic trait optimum, whose standard deviation would be a parameter of the model. We may expect this to impact especially antagonisic scenarios, in which individual spread in trait space – the abiotic niche

would have to be pretty narrow to have an effect for mutualistic communities, in which trait values are already much constrained by stabilizing selection. In antagonistic scenarios, the restriction of niche space should result in rapid niche filling, and subsequently decreased γ statistics. The effect on other tree shape metrics, such as β, is more difficult to forecast, but we can imagine that species stuck at the border of the abiotic niche would be less successful and less likely to speciate than those at its center, which would make phylogenies more imbalanced than what we have for now (Fig. 3). Altogether adding this mechanism in our model could make phylogenetic shapes more realistic. The effect on network metrics are uneasy to forecast and should be evaluated through simulations.

Including space. The model was initially written to include a spatial dimension, which was subsequently abandoned because of the already complex behavior of the non-spatial model. In reality we may expect the spatially explicit version to differ in several ways from the spatially implicit one that we presented. Empirical communities evolve in a complex spatial environment, and spatial dynamics shape the local species assemblages, with a potentially important impact on the resulting diversity and interaction patterns. The Geographic Mosaic Theory of Coevolution (GMTC, Thompson, 1997) highlights the relevance of local species interactions on the coevolutionary process, and suggests that its study necessitate a spatially explicit model. In our model, mutualistic interactions generate stabilizing selection that impedes species and trait diversity. The outcomes may be different if there were opportunities for local coadaptations (similarly to what was obtained in Jabot and Bascompte, 2012). In our model framework, space could be included by multiplying the fitness (which in our case is proportional to the probability to invade a cell) by a dispersion kernel, whose extent will be a parameter of the model. We might expect limited dispersal abilities to have an effect on the interaction network structure, as interaction probabilities between two species would now depend on their local coexistence as much as their trait values. This might contribute to the creation of modularity in our communities. A first step would be to study the effect of the extent of the dispersal kernel on neutral communities, to see what kind of network structures can be generated in this context, and then add non neutral interactions.

2.6 Extensions of the heterogeneous speciation model.

The method presented in Chapter 2 allows to estimate branch specific diversification rates without any prior idea of why they varied. If the ultimate goal is to know whether a given species trait (or environmental condition) is responsible for changes in diversification rates, a possible option is to use a permutation procedure on the trait present states to assess whether independence between diversification rates and trait values can be rejected, in a similar way to what is done in Rabosky and Huang (2015). However, if we have reasons to believe that a particular factor impacted diversification, it might be better to include it in the model. Different possible ways to do this in our model framework are envisioned in the following.

Incorporating environment-dependency. The effect of the abiotic and biotic conditions on diversification tempo of the clade cane be tested using environment-dependent models (Condamine et al., 2013; Lewitus and Morlon, 2017). Yet, these models generate trees that are more balanced than the empirical ones, – as do all homogeneous rate diversification models (Lambert and Stadler, 2013) –, showing that not all changes in diversification tempo are well explained by those model. Including an environment-dependency in our model could allow to add residual variations in diversification rates in environment-dependent models. One of the way to do so would be to take the branch-specific speciation rates $\Lambda_i = \Lambda(T) + \lambda_i$, where $\Lambda(T)$ is a function of the environmental conditions T, and λ_i are the ones obtained from the model presented in Chapter 2, accounting for for

the residual rate variations. The likelihood of this model should be possible to compute numerically using the same approach as in Chapter 2.

Incorporating trait-dependency. The State Speciation and Extinction methods (SSE models, Maddison et al., 2007) allow to check for the effect of a trait on the diversification process, but have been shown to suffer from a high type I error, possibly because they compare the null model they use is to simplistic and do not allow for rate variations due to another factor than the trait considered in the analysis (Rabosky and Goldberg, 2015). One of the proposed solutions to this problem implies the addition of an hidden character, with unknown states at present, that also impacted diversification rates (HiSSE model, for Hidden State Speciation and Extinction; Beaulieu and O'Meara, 2016). We could envision to use our model in a similar way to account both for the effect of a binary trait and of unknown, heritable factors on diversification rates. In this new model, lineages are characterized by their speciation rate and their character state (0 or 1). As in the model in Chapter 2, a shift in speciation rates happens each time there is a speciation event, and is sampled in a lognormal distribution with parameters $\log(\alpha)$ and σ^2. Transitions between character states occur at rate r_{ij} (from state i to state j; $(i, j) \in \{(0, 1), (1, 0)\}$). Each time there is a transition in character state, a shift in speciation rates happen, sampled in a lognormal distribution with parameters $\log(\alpha_{ij})$ and σ^2. As in the previous paragraph, the likelihood of the model should be possible to compute numerically with the approach used in Chapter 2. $\alpha_{ij} > 1$ means that speciation rates tend to increase when there is a transition from state i to state j, and that this transition is advantageous for the lineage in which it occurs. Estimating the α_{ij} parameters of this model would thus by a way to assess the effect of a binary trait on diversification, while accounting for residual variations due to the effect of other, unknown factors.

2.7 A few concluding words.

The work presented in this thesis introduces complementary ways to study heterogeneity in the diversification process. A first step consists in describing the diversity patterns that need to be explained. This may be performed by looking at the general shape of phylogenetic trees. The first chapter falls in this category, together with a rich literature on how to describe phylogenies topologies and branch lengths distributions (Sackin, 1972; Aldous, 1996; Pybus and Harvey, 2000). The method proposed in the second chapter also offer a tool to characterize tree-level properties though the estimation of the model's parameter, that allow to quantify how heritable diversification rates are in a clade and the global temporal trend in rates changes. This method also allows to assess finer scale patterns, by estimating the diversification rates of specific lineages. Being able to do so is a first step to the understanding of the reasons that drove some groups to diversify more than others. Previous approaches allow to automatically assess the presence of shifts in diversification rates on a phylogeny (Alfaro et al., 2009; Rabosky, 2014), yet ours differs in its approach, in that it is based on the idea that rates changes might occur gradually, as a response to changing environment or species traits. Finally, in the third chapter, we more directly focused on one of the possible explanations for changes in diversification rates, which is the presence of ecological interactions between species. In that aim, we proposed an individual-based model for the emergence of bipartite ecological communities, allowing the simultaneous emergence of community composition and interaction structure. In accordance with a previous study on trait diversity (Yoder and Nuismer, 2010), we find a different effect on species diversity for different types of interactions. In particular, antagonistic interactions are promoting trait and species diversity, while mutualistic interactions impede it. We also show that our model is able to generate realistic network structures

from simple evolutionary rules. Other reasons might be invocated to explain diversification rates variations. These include changes in the abiotic environmental conditions experienced by species (Benton, 2009), or variations in key species traits, such as generation time (Baker et al., 2014). Several methods exist to test for the effects of environmental (Condamine et al., 2013; Lewitus and Morlon, 2017) or trait dependency (Maddison et al., 2007; Beaulieu and O'Meara, 2015; Rabosky and Huang, 2015) in diversification rates on an empirical phylogeny. Yet, even though it is possible to fit the environment-dependent model with a biotic factor as an explanatory variable (such as the diversity of competitors, Lewitus and Morlon, 2017), no approach currently exists to test for the reciprocal effect of two interacting clades on each other's diversification. Being able to develop such a tool, using for example a similar framework to that of the diversity dependent model (Rabosky and Lovette, 2008a; Etienne et al., 2011), would allow a better integration of ecological and evolutionary processes, and enable to test our model prediction on the efect of bipartite interaction on species diversity on empirical data.

References

Adams, D. C., C. M. Berns, K. H. Kozak, and J. J. Wiens. 2009. Are rates of species diversification correlated with rates of morphological evolution? Proceedings of the Royal Society of London B: Biological Sciences Pages rspb–2009.

Agapow, P.-M. and A. Purvis. 2002. Power of eight tree shape statistics to detect nonrandom diversification: a comparison by simulation of two models of cladogenesis. Systematic biology 51:866–872.

Agrawal, A. A., M. Fishbein, R. Halitschke, A. P. Hastings, D. L. Rabosky, and S. Rasmann. 2009. Evidence for adaptive radiation from a phylogenetic study of plant defenses. Proceedings of the National Academy of Sciences 106:18067–18072.

Aldous, D. 1996. Probability distributions on cladograms. Pages 1–18 *in* Random discrete structures. Springer.

Aldous, D. J. 2001. Stochastic models and descriptive statistics for phylogenetic trees, from yule to today. Statistical Science Pages 23–34.

Alexander, H. K., A. Lambert, and T. Stadler. 2015. Quantifying age-dependent extinction from species phylogenies. Systematic biology 65:35–50.

Alfaro, M. E., F. Santini, C. Brock, H. Alamillo, A. Dornburg, D. L. Rabosky, G. Carnevale, and L. J. Harmon. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proceedings of the National Academy of Sciences 106:13410–13414.

Alfaro, M. E., F. Santini, and C. D. Brock. 2007. Do reefs drive diversification in marine teleosts? evidence from the pufferfish and their allies (order tetraodontiformes). Evolution 61:2104–2126.

Allhoff, K. T., D. Ritterskamp, B. C. Rall, B. Drossel, and C. Guill. 2015. Evolutionary food web model based on body masses gives realistic networks with permanent species turnover. Scientific reports 5:10955.

Almeida-Neto, M., P. Guimaraes, P. R. Guimarães, R. D. Loyola, and W. Ulrich. 2008. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. Oikos 117:1227–1239.

Almeida-Neto, M. and W. Ulrich. 2011. A straightforward computational approach for measuring nestedness using quantitative matrices. Environmental Modelling & Software 26:173–178.

Amaral, L. A. N., A. Scala, M. Barthelemy, and H. E. Stanley. 2000. Classes of small-world networks. Proceedings of the national academy of sciences 97:11149–11152.

Armbruster, W. S. and N. Muchhala. 2009. Associations between floral specialization and species diversity: cause, effect, or correlation? Evolutionary Ecology 23:159.

Atmar, W. and B. D. Patterson. 1993. The measure of order and disorder in the distribution of species in fragmented habitat. Oecologia 96:373–382.

Avise, J. C. et al. 1998. Pleistocene phylogeographic effects on avian populations and the speciation process. Proceedings of the Royal Society of London B: Biological Sciences 265:457–463.

Baker, T. R., R. T. Pennington, S. Magallon, E. Gloor, W. F. Laurance, M. Alexiades, E. Alvarez, A. Araujo, E. J. Arets, G. Aymard, et al. 2014. Fast demographic traits promote high diversification rates of amazonian trees. Ecology Letters 17:527–536.

Baldwin, B. G. and M. J. Sanderson. 1998. Age and rate of diversification of the hawaiian silversword alliance (compositae). Proceedings of the National Academy of Sciences 95:9402–9406.

Barabási, A.-L. and R. Albert. 1999. Emergence of scaling in random networks. science 286:509–512.

Barber, M. J. 2007. Modularity and community detection in bipartite networks. Physical Review E 76:066102.

Barnosky, A. D. 2001. Distinguishing the effects of the red queen and court jester on miocene mammal evolution in the northern rocky mountains. Journal of Vertebrate Paleontology 21:172–185.

Barraclough, T. G. 1998. Sister-group analysis in identifying correlates of diversification. Evol Ecol 12:751–754.

Barraclough, T. G., A. P. Vogler, and P. H. Harvey. 1998. Revealing the factors that promote speciation. Philosophical Transactions of the Royal Society B: Biological Sciences 353:241–249.

Bascompte, J., P. Jordano, C. J. Melián, and J. M. Olesen. 2003. The nested assembly of plant–animal mutualistic networks. Proceedings of the National Academy of Sciences 100:9383–9387.

Beaulieu, J. M. and B. C. O'Meara. 2015. Extinction can be estimated from moderately sized molecular phylogenies. Evolution 69:1036–1043.

Beaulieu, J. M. and B. C. O'Meara. 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. Systematic biology 65:583–601.

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.

Benton, M. J. and P. N. Pearson. 2001. Speciation in the fossil record. Trends in Ecology & Evolution 16:405–411.

Bersier, L.-F. and P. Kehrli. 2008. The signature of phylogenetic constraints on food-web structure. Ecological Complexity 5:132–139.

Bertoin, J. 2002. Self-similar fragmentations. Pages 319–340 *in* Annales de l'Institut Henri Poincare (B) Probability and Statistics vol. 38 No longer published by Elsevier.

Bertoin, J. 2006. Random fragmentation and coagulation processes vol. 102. Cambridge University Press.

Blum, M. G. and O. François. 2006. Which random processes describe the tree of life? a large-scale study of phylogenetic tree imbalance. Systematic Biology 55:685–691.

Blüthgen, N., J. Fründ, D. P. Vázquez, and F. Menzel. 2008. What do interaction network metrics tell us about specialization and biological traits. Ecology 89:3387–3399.

Bortolussi, N., E. Durand, M. Blum, and O. François. 2006. apTreeshape: statistical analysis of phylogenetic tree shape. Bioinformatics 22:363–364.

Briand, F. 1983. Environmental control of food web structure. Ecology 64:253–263.

Brock, C. D., L. J. Harmon, and M. E. Alfaro. 2011. Testing for temporal variation in diversification rates when sampling is incomplete and nonrandom. Systematic Biology 60:410–419.

Bromham, L., X. Hua, and M. Cardillo. 2015. Detecting macroevolutionary self-destruction from phylogenies. Systematic biology 65:109–127.

Cagnolo, L., A. Salvo, and G. Valladares. 2011. Network topology: patterns and mechanisms in plant-herbivore and host-parasitoid food webs. Journal of Animal Ecology 80:342–351.

Canard, E., N. Mouquet, L. Marescot, K. J. Gaston, D. Gravel, and D. Mouillot. 2012. Emergence of structural patterns in neutral trophic networks. PLoS One 7:e38295.

Canard, E., N. Mouquet, D. Mouillot, M. Stanko, D. Miklisova, and D. Gravel. 2014. Empirical evaluation of neutral interactions in host-parasite networks. The American Naturalist 183:468–479.

Chan, K. M. and B. R. Moore. 2004. Symmetree: whole-tree analysis of differential diversification rates. Bioinformatics 21:1709–1710.

Chen, J. 2012. Gunifrac: generalized unifrac distances. R package version 1:2012.

Claramunt, S. and J. Cracraft. 2015. A new time tree reveals earth history's imprint on the evolution of modern birds. Science Advances 1:e1501005.

Coelho, M. T. P. and T. F. Rangel. 2018. Neutral community dynamics and the evolution of species interactions. The American Naturalist 191:421–434.

Colless, D. H. 1982. Phylogenetics: The theory and practice of phylogenetic systematics.

Condamine, F. L., J. Rolland, and H. Morlon. 2013. Macroevolutionary perspectives to environmental change. Ecology letters 16:72–85.

Condamine, F. L., E. F. Toussaint, A.-L. Clamens, G. Genson, F. A. Sperling, and G. J. Kergoat. 2015. Deciphering the evolution of birdwing butterflies 150 years after alfred russel wallace. Scientific Reports 5:11860.

Crisp, M. D. and L. G. Cook. 2009. Explosive radiation or cryptic mass extinction? interpreting signatures in molecular phylogenies. Evolution 63:2257–2265.

Cruaud, A., N. Rønsted, B. Chantarasuwan, L. S. Chou, W. L. Clement, A. Couloux, B. Cousins, G. Genson, R. D. Harrison, P. E. Hanson, et al. 2012. An extreme case of plant–insect codiversification: figs and fig-pollinating wasps. Systematic Biology 61:1029–1047.

Cusimano, N. and S. S. Renner. 2010. Slowdowns in diversification rates from real phylogenies may not be real. Systematic biology 59:458–464.

Cusimano, N., T. Stadler, and S. S. Renner. 2012. A new method for handling missing species in diversification analysis applicable to randomly or nonrandomly sampled phylogenies. Systematic biology 61:785–792.

Davies, T. J., A. P. Allen, L. Borda-de Água, J. Regetz, and C. J. Melián. 2011. Neutral biodiversity theory can explain the imbalance of phylogenetic trees but not the tempo of their diversification. Evolution: International Journal of Organic Evolution 65:1841–1850.

Dormann, C. F., B. Gruber, and J. Fründ. 2008. Introducing the bipartite package: analysing ecological networks. interaction 1:0–2413793.

Dormann, C. F. and R. Strauss. 2014. A method for detecting modules in quantitative bipartite networks. Methods in Ecology and Evolution 5:90–98.

Drury, J., J. Clavel, M. Manceau, and H. Morlon. 2016. Estimating the effect of competition on trait evolution using maximum likelihood inference. Systematic biology 65:700–710.

Duchêne, D., S. Duchêne, and S. Y. Ho. 2015. Tree imbalance causes a bias in phylogenetic estimation of evolutionary timescales using heterochronous sequences. Molecular ecology resources 15:785–794.

Egas, M., U. Dieckmann, and M. W. Sabelis. 2004. Evolution restricts the coexistence of specialists and generalists: the role of trade-off structure. The American Naturalist 163:518–531.

Ehrlich, P. R. and P. H. Raven. 1964. Butterflies and plants: a study in coevolution. Evolution 18:586–608.

Eklöf, A., U. Jacob, J. Kopp, J. Bosch, R. Castro-Urgal, N. P. Chacoff, B. Dalsgaard, C. Sassi, M. Galetti, P. R. Guimaraes, et al. 2013. The dimensionality of ecological networks. Ecology letters 16:577–583.

Elias, M., C. Fontaine, and F. F. van Veen. 2013. Evolutionary history and ecological processes shape a local multilevel antagonistic network. Current Biology 23:1355–1359.

Emerson, B. C. and N. Kolm. 2005. Species diversity can drive speciation. Nature 434:1015.

Etienne, R. S. and B. Haegeman. 2012. A conceptual and statistical framework for adaptive radiations with a key role for diversity dependence. The American Naturalist 180:E75–E89.

Etienne, R. S., B. Haegeman, T. Stadler, T. Aze, P. N. Pearson, A. Purvis, and A. B. Phillimore. 2011. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. Proc. R. Soc. B Page rspb20111439.

Etienne, R. S., H. Morlon, and A. Lambert. 2014. Estimating the duration of speciation from phylogenies. Evolution 68:2430–2440.

Etienne, R. S., A. L. Pigot, and A. B. Phillimore. 2016. How reliably can we infer diversity-dependent diversification from phylogenies? Methods in Ecology and Evolution 7:1092–1099.

Etienne, R. S. and J. Rosindell. 2012. Prolonging the past counteracts the pull of the present: protracted speciation can explain observed slowdowns in diversification. Systematic Biology 61:204–213.

FitzJohn, R. G. 2010. Quantitative traits and diversification. Systematic biology 59:619–633.

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

FitzJohn, R. G., W. P. Maddison, and S. P. Otto. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. Systematic biology 58:595–611.

Fontaine, C. 2013. Ecology: abundant equals nested. Nature 500:411.

Fontaine, C., P. R. Guimarães, S. Kéfi, N. Loeuille, J. Memmott, W. H. van Der Putten, F. J. van Veen, and E. Thébault. 2011. The ecological and evolutionary implications of merging different types of networks. Ecology letters 14:1170–1181.

Fontaine, C. and E. Thébault. 2015. Comparing the conservatism of ecological interactions in plant–pollinator and plant–herbivore networks. Population Ecology 57:29–36.

Fordyce, J. A. 2010. Host shifts and evolutionary radiations of butterflies. Proceedings of the Royal Society of London B: Biological Sciences 277:3735–3743.

Fort, H., D. P. Vázquez, and B. L. Lan. 2016. Abundance and generalisation in mutualistic networks: solving the chicken-and-egg dilemma. Ecology letters 19:4–11.

Fortuna, M. A., D. B. Stouffer, J. M. Olesen, P. Jordano, D. Mouillot, B. R. Krasnov, R. Poulin, and J. Bascompte. 2010. Nestedness versus modularity in ecological networks: two sides of the same coin? Journal of Animal Ecology 79:811–817.

Freckleton, R. P., A. B. Phillimore, and M. Pagel. 2008. Relating traits to diversification: a simple test. The American Naturalist 172:102–115.

Futuyma, D. J. and A. A. Agrawal. 2009. Macroevolution and the biological diversity of plants and herbivores. Proceedings of the National Academy of Sciences 106:18054–18061.

Gandon, S. 2002. Local adaptation and the geometry of host–parasite coevolution. Ecology Letters 5:246–256.

Gascuel, F., R. Ferrière, R. Aguilée, and A. Lambert. 2015. How ecology and landscape dynamics shape phylogenetic trees. Systematic biology 64:590–607.

Gavrilets, S., H. Li, and M. D. Vose. 2000. Patterns of parapatric speciation. Evolution 54:1126–1134.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2014. Bayesian data analysis vol. 2. CRC press Boca Raton, FL.

Gelman, A., X.-L. Meng, and H. Stern. 1996. Posterior predictive assessment of model fitness via realized discrepancies. Statistica sinica Pages 733–760.

Gibson, R. H., B. Knott, T. Eberlein, and J. Memmott. 2011. Sampling method influences the structure of plant–pollinator networks. Oikos 120:822–831.

Gilbert, G. S. and C. O. Webb. 2007. Phylogenetic signal in plant pathogen–host range. Proceedings of the National Academy of Sciences 104:4979–4983.

Gilman, R. T., S. L. Nuismer, and D.-C. Jhwueng. 2012. Coevolution in multidimensional trait space favours escape from parasites and pathogens. Nature 483:328.

Goldberg, E. E. and B. Igić. 2012. Tempo and mode in plant breeding system evolution. Evolution 66:3701–3709.

Goldberg, E. E., J. R. Kohn, R. Lande, K. A. Robertson, S. A. Smith, and B. Igić. 2010. Species selection maintains self-incompatibility. Science 330:493–495.

Goldberg, E. E., L. T. Lancaster, and R. H. Ree. 2011. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. Systematic Biology 60:451–465.

Gómez, J. M., M. Verdú, and F. Perfectti. 2010. Ecological interactions are evolutionarily conserved across the entire tree of life. Nature 465:918.

Gotelli, N. J. and W. Ulrich. 2012. Statistical challenges in null model analysis. Oikos 121:171–180.

Grafen, A. 1989. The phylogenetic regression. Phil. Trans. R. Soc. Lond. B 326:119–157.

Green, P. J. 1995. Reversible jump markov chain monte carlo computation and bayesian model determination. Biometrika 82:711–732.

Grilli, J., M. Adorisio, S. Suweis, G. Barabás, J. R. Banavar, S. Allesina, and A. Maritan. 2017. Feasibility and coexistence of large ecological communities. Nature communications 8:0.

Guill, C. and B. Drossel. 2008. Emergence of complexity in evolving niche-model food webs. Journal of theoretical biology 251:108–120.

Hagen, O., K. Hartmann, M. Steel, and T. Stadler. 2015. Age-dependent speciation can explain the shape of empirical phylogenies. Systematic biology 64:432–440.

Hallinan, N. 2012. The generalized time variable reconstructed birth–death process. Journal of theoretical biology 300:265–276.

Harvey, M. G. and D. L. Rabosky. 2017. Continuous traits and speciation rates: Alternatives to state-dependent diversification models. Methods in Ecology and Evolution .

Harvey, P. H., R. M. May, and S. Nee. 1994. Phylogenies without fossils. Evolution 48:523–529.

Heard, S. B. 1992. Patterns in tree balance among cladistic, phenetic, and randomly generated phylogenetic trees. Evolution 46:1818–1826.

Hedges, S. B., J. Marin, M. Suleski, M. Paymer, and S. Kumar. 2015. Tree of life reveals clock-like speciation and diversification. Molecular biology and evolution 32:835–845.

Heled, J. and R. R. Bouckaert. 2013. Looking for trees in the forest: summary tree from posterior samples. BMC evolutionary biology 13:221.

Hembry, D. H., J. B. Yoder, and K. R. Goodman. 2014. Coevolution and the diversification of life. The American Naturalist 184:425–438.

Ho, S. Y., M. J. Phillips, A. Cooper, and A. J. Drummond. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Molecular biology and evolution 22:1561–1568.

Höhna, S. 2015. The time-dependent reconstructed evolutionary process with a key-role for mass-extinction events. Journal of theoretical biology 380:321–331.

Höhna, S., T. Stadler, F. Ronquist, and T. Britton. 2011. Inferring speciation and extinction rates under different sampling schemes. Molecular biology and evolution 28:2577–2589.

Holton, T. A., M. Wilkinson, and D. Pisani. 2014. The shape of modern tree reconstruction methods. Systematic biology 63:436–441.

Hua, X. and L. Bromham. 2016. Phylometrics: an r package for detecting macroevolutionary patterns, using phylogenetic metrics and backward tree simulation. Methods in Ecology and Evolution 7:806–810.

Hubbell, S. P. 2001. The unified neutral theory of biodiversity and biogeography (mpb-32)(monographs in population biology) .

Huelsenbeck, J. P. and M. Kirkpatrick. 1996. Do phylogenetic methods produce trees with biased shapes? Evolution 50:1418–1424.

Huelsenbeck, J. P., B. Larget, and D. Swofford. 2000. A compound poisson process for relaxing the molecular clock. Genetics 154:1879–1892.

Hugall, A. F. and M. S. Lee. 2007. The likelihood node density effect and consequences for evolutionary studies of molecular rates. Evolution 61:2293–2307.

Hunter, J. P. 1998. Key innovations and the ecology of macroevolution. Trends in ecology & evolution 13:31–36.

Ispolatov, I., V. Madhok, and M. Doebeli. 2016. Individual-based models for adaptive diversification in high-dimensional phenotype spaces. Journal of theoretical biology 390:97–105.

Ives, A. and H. Godfray. 2006. Phylogenetic analysis of trophic associations. The American Naturalist 168:E1–E14.

Jablonski, D. 1987. Heritability at the species level: analysis of geographic ranges of cretaceous mollusks. Science 238:360–363.

Jabot, F. and J. Bascompte. 2012. Bitrophic interactions shape biodiversity in space. Proceedings of the National Academy of Sciences 109:4521–4526.

Janz, N. 2011. Ehrlich and raven revisited: mechanisms underlying codiversification of plants and enemies. Annual review of ecology, evolution, and systematics 42.

Jetz, W., G. Thomas, J. Joy, K. Hartmann, and A. Mooers. 2012. The global diversity of birds in space and time. Nature 491:444.

Jordano, P., J. Bascompte, and J. M. Olesen. 2003. Invariant properties in coevolutionary networks of plant–animal interactions. Ecology letters 6:69–81.

Joy, J. B. 2013. Symbiosis catalyses niche expansion and diversification. Proceedings of the Royal Society of London B: Biological Sciences 280:20122820.

Kass, R. E. and A. E. Raftery. 1995. Bayes factors. Journal of the american statistical association 90:773–795.

Kay, K. M. 2006. Reproductive isolation between two closely related hummingbird-pollinated neotropical gingers. Evolution 60:538–552.

Kay, K. M. and R. D. Sargent. 2009. The role of animal pollination in plant speciation: integrating ecology, geography, and genetics. Annual Review of Ecology, Evolution, and Systematics 40:637–656.

Keitt, T. H. and H. E. Stanley. 1998. Dynamics of north american breeding bird populations. Nature 393:257.

Kingman, J. F. C. 1982. The coalescent. Stoch. Proc. Appl. 13:235–248.

Kondoh, M. 2003. Foraging adaptation and the relationship between food-web complexity and stability. Science 299:1388–1391.

Kopp, M. and S. Gavrilets. 2006. Multilocus genetics and the coevolution of quantitative traits. Evolution 60:1321–1336.

Kozak, K. H., D. W. Weisrock, and A. Larson. 2006. Rapid lineage accumulation in a non-adaptive radiation: phylogenetic analysis of diversification rates in eastern north american woodland salamanders (plethodontidae: Plethodon). Proceedings of the Royal Society of London B: Biological Sciences 273:539–546.

Krasnov, B. R., M. A. Fortuna, D. Mouillot, I. S. Khokhlova, G. I. Shenbrot, and R. Poulin. 2012. Phylogenetic signal in module composition and species connectivity in compartmentalized host-parasite networks. The American Naturalist 179:501–511.

Krause, A. E., K. A. Frank, D. M. Mason, R. E. Ulanowicz, and W. W. Taylor. 2003. Compartments revealed in food-web structure. Nature 426:282.

Krishna, A., P. R. Guimaraes Jr, P. Jordano, and J. Bascompte. 2008. A neutral-niche theory of nestedness in mutualistic networks. Oikos 117:1609–1618.

Kubo, T. and Y. Iwasa. 1995. Inferring the rates of branching and extinction from molecular phylogenies. Evolution 49:694–704.

Lambert, A., H. Morlon, and R. S. Etienne. 2015. The reconstructed tree in the lineage-based model of protracted speciation. Journal of mathematical biology 70:367–397.

Lambert, A. and T. Stadler. 2013. Birth–death models and coalescent point processes: The shape and probability of reconstructed phylogenies. Theoretical population biology 90:113–128.

Lambert, A. and M. Steel. 2013. Predicting the loss of phylogenetic diversity under non-stationary diversification models. Journal of theoretical biology 337:111–124.

Lambert, A. et al. 2017. Probabilistic models for the (sub) tree (s) of life. Brazilian Journal of Probability and Statistics 31:415–475.

Lartillot, N., M. J. Phillips, and F. Ronquist. 2016. A mixed relaxed clock model. Phil. Trans. R. Soc. B 371:20150132.

Lewinsohn, T. M., P. Inácio Prado, P. Jordano, J. Bascompte, and J. M Olesen. 2006. Structure in plant–animal interaction assemblages. Oikos 113:174–184.

Lewitus, E. and H. Morlon. 2017. Detecting environment-dependent diversification from phylogenies: a simulation study and some empirical illustrations. Systematic biology .

Liow, L. H., T. B. Quental, and C. R. Marshall. 2010. When can decreasing diversification rates be detected with molecular phylogenies and the fossil record? Systematic Biology 59:646–659.

Loeuille, N. and M. Loreau. 2005. Evolutionary emergence of size-structured food webs. Proceedings of the National Academy of Sciences 102:5761–5766.

Lovette, I. J. and E. Bermingham. 1999. Explosive speciation in the new world dendroica warblers. Proceedings of the Royal Society of London B: Biological Sciences 266:1629–1636.

Lozupone, C. A., M. Hamady, S. T. Kelley, and R. Knight. 2007. Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. Applied and environmental microbiology 73:1576–1585.

Maddison, W. P. 2006. Confounding asymmetries in evolutionary diversification and character change. Evolution 60:1743–1746.

Maddison, W. P., P. E. Midford, and S. P. Otto. 2007. Estimating a binary character's effect on speciation and extinction. Systematic biology 56:701–710.

Maddison, W. P. and M. Slatkin. 1991. Null models for the number of evolutionary steps in a character on a phylogenetic tree. Evolution 45:1184–1197.

Magallón, S. and M. J. Sanderson. 2001. Absolute diversification rates in angiosperm clades. Evolution 55:1762–1780.

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: International Journal of Organic Evolution 64:2731–2745.

Maliet, O., F. Gascuel, and A. Lambert. 2018. Ranked tree shapes, non-random extinctions and the loss of phylogenetic diversity. Systematic biology Page syy030.

Manceau, M. and A. Lambert. 2017. The species problem from the modeler's point of view. bioRxiv Page 075580.

Manceau, M., A. Lambert, and H. Morlon. 2015. Phylogenies support out-of-equilibrium models of biodiversity. Ecology letters 18:347–356.

Manceau, M., A. Lambert, and H. Morlon. 2016. A unifying comparative phylogenetic framework including traits coevolving across interacting lineages. Systematic biology 66:551–568.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer research 27:209–220.

May, M. R., S. Höhna, and B. R. Moore. 2016. A bayesian approach for detecting the impact of mass-extinction events on molecular phylogenies when rates of lineage diversification may vary. Methods in Ecology and Evolution 7:947–959.

May, M. R. and B. R. Moore. 2016. How well can we detect lineage-specific diversification-rate shifts? a simulation study of sequential aic methods. Systematic biology 65:1076–1084.

May, R. M. 1972. Will a large complex system be stable? Nature 238:413.

Maynard Smith, J. 1976. A comment on the red queen. The American Naturalist 110:325–330.

McGill, B. 2003. Strong and weak tests of macroecological theory. Oikos 102:679–685.

McPeek, M. A. 2008. The ecological dynamics of clade diversification and community assembly. The American Naturalist 172:E270–E284.

McPeek, M. A. and J. M. Brown. 2007. Clade age and not diversification rate explains species richness among animal taxa. The American Naturalist 169:E97–E106.

Meyer, A. L. and J. J. Wiens. 2018. Estimating diversification rates for higher taxa: Bamm can give problematic estimates of rates and rate shifts. Evolution 72:39–53.

Miller, A. H. 1949. Some ecologic and morphologic considerations in the evolution of higher taxonomic categories. Ornithologie als biologische Wissenschaft Pages 84–88.

Missa, O., C. Dytham, and H. Morlon. 2016. Understanding how biodiversity unfolds through time under neutral theory. Phil. Trans. R. Soc. B 371:20150226.

Mitchell, J. S. and D. L. Rabosky. 2017. Bayesian model selection with bamm: effects of the model prior on the inferred number of diversification shifts. Methods in Ecology and Evolution 8:37–46.

Mitter, C., B. Farrell, and B. Wiegmann. 1988. The phylogenetic study of adaptive zones: has phytophagy promoted insect diversification? The American Naturalist 132:107–128.

Moen, D. and H. Morlon. 2014. Why does diversification slow down? Trends in Ecology & Evolution 29:190–197.

Montoya, J. M., S. L. Pimm, and R. V. Solé. 2006. Ecological networks and their fragility. Nature 442:259.

Mooers, A., O. Gascuel, T. Stadler, H. Li, and M. Steel. 2011. Branch lengths on birth–death trees and the expected loss of phylogenetic diversity. Systematic biology 61:195–203.

Mooers, A. O. and S. B. Heard. 1997. Inferring evolutionary process from phylogenetic tree shape. The quarterly review of Biology 72:31–54.

Moore, B. R., S. Höhna, M. R. May, B. Rannala, and J. P. Huelsenbeck. 2016. Critically evaluating the theory and performance of bayesian analysis of macroevolutionary mixtures. Proceedings of the National Academy of Sciences 113:9569–9574.

Moore, J. E. and R. K. Swihart. 2007. Toward ecologically explicit null models of nestedness. Oecologia 152:763–777.

Morlon, H. 2014. Phylogenetic approaches for studying diversification. Ecology letters 17:508–525.

Morlon, H., E. Lewitus, F. L. Condamine, M. Manceau, J. Clavel, and J. Drury. 2016. Rpanda: an r package for macroevolutionary analyses on phylogenetic trees. Methods in Ecology and Evolution 7:589–597.

Morlon, H., T. L. Parsons, and J. B. Plotkin. 2011. Reconciling molecular phylogenies with the fossil record. Proceedings of the National Academy of Sciences 108:16327–16332.

Nee, S., E. C. Holmes, R. M. May, and P. H. Harvey. 1994a. Extinction rates can be estimated from molecular phylogenies. Phil. Trans. R. Soc. Lond. B 344:77–82.

Nee, S. and R. M. May. 1997. Extinction and the loss of evolutionary history. Science 278:692–694.

Nee, S., R. M. May, and P. H. Harvey. 1994b. The reconstructed evolutionary process. Phil. Trans. R. Soc. Lond. B 344:305–311.

Nee, S., A. O. Mooers, and P. H. Harvey. 1992. Tempo and mode of evolution revealed from molecular phylogenies. Proceedings of the National Academy of Sciences 89:8322–8326.

Newman, M. E. 2004. Analysis of weighted networks. Physical review E 70:056131.

Nielsen, A. and J. Bascompte. 2007. Ecological networks, nestedness and sampling effort. Journal of Ecology 95:1134–1141.

Norris, R. D. and P. M. Hull. 2012. The temporal dimension of marine speciation. Evolutionary Ecology 26:393–415.

Nuismer, S. L., P. Jordano, and J. Bascompte. 2013. Coevolution and the architecture of mutualistic networks. Evolution 67:338–354.

Olesen, J. M., J. Bascompte, Y. L. Dupont, and P. Jordano. 2007. The modularity of pollination networks. Proceedings of the National Academy of Sciences 104:19891–19896.

Onstein, R. E., W. J. Baker, T. L. Couvreur, S. Faurby, J.-C. Svenning, and W. D. Kissling. 2017. Frugivory-related traits promote speciation of tropical palms. Nature ecology & evolution 1:1903.

Pennell, M. W., B. A. Sarver, and L. J. Harmon. 2012. Trees of unusual size: biased inference of early bursts from large molecular phylogenies. PloS one 7:e43348.

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

Pigot, A. L., A. B. Phillimore, I. P. Owens, and C. D. L. Orme. 2010. The shape and temporal dynamics of phylogenetic trees arising from geographic speciation. Systematic biology 59:660–673.

Pillai, P., A. Gonzalez, and M. Loreau. 2011. Metacommunity theory explains the emergence of food web complexity. Proceedings of the National Academy of Sciences 108:19293–19298.

Poisot, T. and D. Stouffer. 2016. How ecological networks evolve. bioRxiv Page 071993.

Pybus, O. G. and P. H. Harvey. 2000. Testing macro–evolutionary models using incomplete molecular phylogenies. Proceedings of the Royal Society of London B: Biological Sciences 267:2267–2272.

R Development Core Team. 2012. R: A Language and Environment for Statistical Computing. {R Foundation for Statistical Computing}, Vienna, Austria.

Rabosky, D. L. 2009a. Ecological limits and diversification rate: alternative paradigms to explain the variation in species richness among clades and regions. Ecology letters 12:735–743.

Rabosky, D. L. 2009b. Ecological limits on clade diversification in higher taxa. The American Naturalist 173:662–674.

Rabosky, D. L. 2010. Extinction rates should not be estimated from molecular phylogenies. Evolution 64:1816–1824.

Rabosky, D. L. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. PloS one 9:e89543.

Rabosky, D. L. 2016a. Challenges in the estimation of extinction from molecular phylogenies: a response to beaulieu and o'meara. Evolution 70:218–228.

Rabosky, D. L. 2016b. Reproductive isolation and the causes of speciation rate variation in nature. Biological Journal of the Linnean Society 118:13–25.

Rabosky, D. L. 2017. How to make any method" fail": Bamm at the kangaroo court of false equivalency. arXiv preprint arXiv:1711.03253 .

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

Rabosky, D. L. and E. E. Goldberg. 2017. Fisse: A simple nonparametric test for the effects of a binary character on lineage diversification rates. Evolution 71:1432–1442.

Rabosky, D. L. and H. Huang. 2015. A robust semi-parametric test for detecting trait-dependent diversification. Systematic Biology 65:181–193.

Rabosky, D. L. and I. J. Lovette. 2008a. Density-dependent diversification in north american wood warblers. Proceedings of the Royal Society of London B: Biological Sciences 275:2363–2371.

Rabosky, D. L. and I. J. Lovette. 2008b. Explosive evolutionary radiations: decreasing speciation or increasing extinction through time? Evolution 62:1866–1875.

Rabosky, D. L., J. S. Mitchell, and J. Chang. 2017. Is bamm flawed? theoretical and practical concerns in the analysis of multi-rate diversification models. Systematic biology 66:477–498.

Rabosky, D. L., G. J. Slater, and M. E. Alfaro. 2012. Clade age and species richness are decoupled across the eukaryotic tree of life. PLoS biology 10:e1001381.

Ramsey, J., H. Bradshaw Jr, and D. W. Schemske. 2003. Components of reproductive isolation between the monkeyflowers mimulus lewisii and m. cardinalis (phrymaceae). Evolution 57:1520–1534.

Redding, D. W. and A. Ø. Mooers. 2006. Incorporating evolutionary measures into conservation prioritization. Conservation Biology 20:1670–1678.

Revell, L. J., L. J. Harmon, and R. E. Glor. 2005. Under-parameterized model of sequence evolution leads to bias in the estimation of diversification rates from molecular phylogenies. Systematic Biology 54:973–983.

Rezende, E. L., P. Jordano, and J. Bascompte. 2007. Effects of phenotypic complementarity and phylogeny on the nested structure of mutualistic networks. Oikos 116:1919–1929.

Ribera, I., T. Barraclough, and A. Vogler. 2001. The effect of habitat type on speciation rates and range movements in aquatic beetles: inferences from species-level phylogenies. Molecular Ecology 10:721–735.

Ricklefs, R. E. 2007a. Estimating diversification rates from phylogenetic information. Trends in Ecology & Evolution 22:601–610.

Ricklefs, R. E. 2007b. History and diversity: explorations at the intersection of ecology and evolution. The American Naturalist 170:S56–S70.

Rivera-Hutinel, A., R. Bustamante, V. Marín, and R. Medel. 2012. Effects of sampling completeness on the structure of plant–pollinator networks. Ecology 93:1593–1603.

Rohr, R. P., S. Saavedra, and J. Bascompte. 2014. On the structural stability of mutualistic systems. Science 345:1253497.

Rosenzweig, M. L. 1992. Species diversity gradients: we know more and less than we thought. Journal of mammalogy 73:715–730.

Rosindell, J., S. J. Cornell, S. P. Hubbell, and R. S. Etienne. 2010. Protracted speciation revitalizes the neutral theory of biodiversity. Ecology Letters 13:716–727.

Rosindell, J., L. J. Harmon, and R. S. Etienne. 2015. Unifying ecology and macroevolution with individual-based theory. Ecology letters 18:472–482.

Rueffler, C., T. J. Van Dooren, and J. A. Metz. 2006. The interplay between behavior and morphology in the evolutionary dynamics of resource specialization. The American Naturalist 169:E34–E52.

Saavedra, S., R. P. Rohr, J. M. Olesen, and J. Bascompte. 2016. Nested species interactions promote feasibility over stability during the assembly of a pollinator community. Ecology and evolution 6:997–1007.

Sackin, M. 1972. "good" and "bad" phenograms. Systematic Biology 21:225–226.

Sánchez-Reyes, L. L., H. Morlon, and S. Magallón. 2016. Uncovering higher-taxon diversification dynamics from clade age and species-richness data. Systematic biology 66:367–378.

Sanderson, M. J. 1990. Estimating rates of speciation and evolution: a bias due to homoplasy. Cladistics 6:387–391.

Santamaría, L. and M. A. Rodríguez-Gironés. 2007. Linkage rules for plant–pollinator networks: trait complementarity or exploitation barriers? PLoS biology 5:e31.

Schoenly, K., R. Beaver, and T. Heumier. 1991. On the trophic relations of insects: a food-web approach. The American Naturalist 137:597–638.

Shao, K.-T. and R. R. Sokal. 1990. Tree balance. Systematic Zoology 39:266–276.

Slowinski, J. B. and C. Guyer. 1993. Testing whether certain traits have caused amplified diversification: an improved method based on a model of random speciation and extinction. The American Naturalist 142:1019–1024.

Smith, C. I., O. Pellmyr, D. M. Althoff, M. Balcazar-Lara, J. Leebens-Mack, and K. A. Segraves. 2008. Pattern and timing of diversification in yucca (agavaceae): specialized pollination does not escalate rates of diversification. Proceedings of the Royal Society of London B: Biological Sciences 275:249–258.

Stadler, T. 2011. Mammalian phylogeny reveals recent diversification rate shifts. Proceedings of the National Academy of Sciences 108:6187–6192.

Stadler, T., D. L. Rabosky, R. E. Ricklefs, and F. Bokma. 2014. On age and species richness of higher taxa. The American Naturalist 184:447–455.

Staniczenko, P. P., J. C. Kopp, and S. Allesina. 2013. The ghost of nestedness in ecological networks. Nature communications 4:1391.

Stenseth, N. C. and J. Maynard Smith. 1984. Coevolution in ecosystems: Red queen evolution or stasis? Evolution 38:870–880.

Takahashi, D., Å. Brännström, R. Mazzucco, A. Yamauchi, and U. Dieckmann. 2013. Abrupt community transitions and cyclic evolutionary dynamics in complex food webs. Journal of theoretical biology 337:181–189.

Ter Braak, C. J. 2006. A markov chain monte carlo version of the genetic algorithm differential evolution: easy bayesian computing for real parameter spaces. Statistics and Computing 16:239–249.

Thébault, E. and C. Fontaine. 2010. Stability of ecological communities and the architecture of mutualistic and trophic networks. Science 329:853–856.

Thompson, J. N. 1997. Evaluating the dynamics of coevolution among geographically structured populations. Ecology 78:1619–1623.

Thompson, J. N. and B. M. Cunningham. 2002. Geographic structure and dynamics of coevolutionary selection. Nature 417:735.

Thorne, J. L., H. Kishino, and I. S. Painter. 1998. Estimating the rate of evolution of the rate of molecular evolution. Molecular biology and evolution 15:1647–1657.

Ulrich, W., M. Almeida-Neto, and N. J. Gotelli. 2009. A consumer's guide to nestedness analysis. Oikos 118:3–17.

Valdovinos, F. S., B. J. Brosi, H. M. Briggs, P. Moisset de Espanés, R. Ramos-Jiliberto, and N. D. Martinez. 2016. Niche partitioning due to adaptive foraging reverses effects of nestedness and connectance on pollination network stability. Ecology letters 19:1277–1286.

van der Niet, T. and S. D. Johnson. 2012. Phylogenetic evidence for pollinator-driven diversification of angiosperms. Trends in Ecology & Evolution 27:353–361.

Van Valen, L. 1973. A new evolutionary law. Evol Theory 1:1–30.

Vázquez, D. P. 2005. Degree distribution in plant–animal mutualistic networks: forbidden links or random interactions? Oikos 108:421–426.

Vázquez, D. P., N. P. Chacoff, and L. Cagnolo. 2009. Evaluating multiple determinants of the structure of plant–animal mutualistic networks. Ecology 90:2039–2046.

Vázquez, D. P., R. Poulin, B. R. Krasnov, and G. I. Shenbrot. 2005. Species abundance and the distribution of specialization in host–parasite interaction networks. Journal of Animal Ecology 74:946–955.

Venditti, C., A. Meade, and M. Pagel. 2010. Phylogenies reveal new interpretation of speciation and the red queen. Nature 463:349.

Voje, K. L., Ø. H. Holen, L. H. Liow, and N. C. Stenseth. 2015. The role of biotic forces in driving macroevolution: beyond the red queen. Proc. R. Soc. B 282:20150186.

Weir, J. T. and S. Mursleen. 2013. Diversity-dependent cladogenesis and trait evolution in the adaptive radiation of the auks (aves: Alcidae). Evolution: International Journal of Organic Evolution 67:403–416.

Yang, Z. and C. E. Rodríguez. 2013. Searching for efficient markov chain monte carlo proposal kernels. Proceedings of the National Academy of Sciences 110:19307–19312.

Yoder, J. B. and S. L. Nuismer. 2010. When does coevolution promote diversification? The American Naturalist 176:802–817.

Yule, G. 1925. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. Philos. Trans. R. Soc. Lond. B 213:402–410.

Zink, R. M. and J. B. Slowinski. 1995. Evidence from molecular systematics for decreased avian diversification in the pleistocene epoch. Proceedings of the National Academy of Sciences 92:5832–5835.