Syst. Biol. 65(1):35–50, 2016

DOI:10.1093/sysbio/syv065
Advance Access publication September 24, 2015

Quantifying Age-dependent Extinction from Species Phylogenies

HELEN K. ALEXANDER[1,*], AMAURY LAMBERT[2,3], AND TANJA STADLER[4]

[1]*Institute for Integrative Biology, ETH Zürich, 8092 Zürich, Switzerland;* [2]*Laboratoire de Probabilités et Modèles Aléatoires CNRS UMR 7599, UPMC Univ Paris 06, Paris, France;* [3] *Center for Interdisciplinary Research in Biology CNRS UMR 7241, Collège de France, Paris, France; and* [4]*Department of Biosystems Science and Engineering, ETH Zürich, 4058 Basel, Switzerland;*
**Correspondence to be sent to: Institute for Integrative Biology, ETH Zürich, CHN H.74, Universitätsstrasse 16, 8092 Zürich, Switzerland; E-mail: helen.alexander@env.ethz.ch.*

Received 16 April 2014; reviews returned 21 August 2015; accepted 1 September 2015
Associate Editor: Edward Susko

Abstract.—Several ecological factors that could play into species extinction are expected to correlate with species age, i.e., time elapsed since the species arose by speciation. To date, however, statistical tools to incorporate species age into likelihood-based phylogenetic inference have been lacking. We present here a computational framework to quantify age-dependent extinction through maximum likelihood parameter estimation based on phylogenetic trees, assuming species lifetimes are gamma distributed. Testing on simulated trees shows that neglecting age dependence can lead to biased estimates of key macroevolutionary parameters. We then apply this method to two real data sets, namely a complete phylogeny of birds (class Aves) and a clade of self-compatible and -incompatible nightshades (Solanaceae), gaining initial insights into the extent to which age-dependent extinction may help explain macroevolutionary patterns. Our methods have been added to the R package TreePar. [Aves, coalescent point process, diversification, macroevolution, maximum likelihood estimation, phylogenetics, Solanaceae]

The proposition that extinction of species could depend on their age, i.e., time since arising by speciation, has a long history of investigation and debate. In a seminal 1973 paper, Leigh Van Valen proposed that the rate of extinction of taxa is independent of their age, presenting apparently log-linear survivorship curves of taxa within defined groups as evidence of this effect (Van Valen 1973). The methodology behind this result was subsequently criticized on several grounds (Raup 1975; McCune 1982; Pearson 1995) and application of more sophisticated statistical methods later suggested that species extinction is indeed age dependent in some groups (Pearson 1995; Doran et al. 2006). Nonetheless, the sum of evidence remains ambiguous (Liow et al. 2011), and various patterns seem plausible. For instance, an increased risk of extinction with age has specifically been hypothesized for asexual or selfing plant species, due to the accumulation of deleterious mutations not broken up by recombination, i.e., Muller's ratchet (Muller 1964; Johnson et al. 2011). More generally, an age-dependent signal could arise through various processes, not necessarily requiring "genetically inbuilt 'senescence' " (Pearson 1995, p. 134). Biotic interactions, i.e., competition and co-evolution among species, have been invoked to explain age-independent extinction (Van Valen 1973; Liow et al. 2011), or alternatively to argue that older species are more prone to extinction. For instance, Pearson (1995) speculated that in evolutionarily static taxa, old species may be less fit than new species, while in gradually evolving taxa, old species may become too specialized and susceptible to environmental changes. On the other hand, we speculate that new species also face distinct challenges that could result in higher extinction risk, particularly when speciation is triggered by a founder event (initiated by a small population prone to demographic stochasticity) or by environmental stress (cf. Doran et al. 2006). Species lifetime has also been associated with ecological variables, such as range and dispersal characteristics (reviewed by McCune 1982); age dependence then arises insofar as age correlates with these ecological variables.

Quantitative tests of hypotheses regarding age-dependent extinction have so far been based on paleontological data, thus on a limited set of species, and have suffered from problems associated with defining durations of species existence from the fossil record (Pearson 1995; Liow et al. 2011). On the other hand, phylogenetic trees based on genetic data of extant species have become increasingly available in recent years, with a corresponding development of statistical inference methods to quantify macroevolutionary processes from these trees. Birth–death models have been extensively used to model the distribution of phylogenetic trees. In reverse, given a reconstructed tree, its structure (topology and branch lengths) can be used to infer rates of birth and death (Thompson 1975; Nee et al. 1994b). In the context of species trees, these rates are interpreted as speciation and extinction rates. Recent advances have allowed the incorporation of greater biological realism by modeling dependence of these rates on factors such as time, number of species in a clade, or a species trait (recently reviewed by Pyron and Burbrink 2013; Stadler 2013b; Morlon 2014). Very few phylogenetic models have considered age dependence (reviewed in Morlon 2014), and to the best of our knowledge, no likelihood-based method is yet available to infer age-dependent rates. Likelihood-based approaches have the advantages of maximal statistical power, because they use all information about the tree, and readily comparable output across different models (Morlon 2014). In this article, we will present an inference method under a model that incorporates age-dependent death

and a constant birth rate (age-dependent birth cannot yet be treated; see "Discussion" section), where all sampling occurs at the present. This model applies to trees of extant species, which we consider here, but could also be applied to data on the individual organism or cell level. The necessary mathematical results have recently been developed (Lambert 2010; Lambert and Stadler 2013), and our aim here is to implement and test a computational framework for application to data.

The manuscript is organized as follows. We first describe the model, present the mathematical likelihood expressions, and explain our computational approach. Though already-existing mathematical results are more general (Lambert 2010), we confine ourselves here to lifetimes described by gamma distributions, a family of unimodal distributions where both mean and variance can be controlled independently. This family contains as a special case the exponential distribution, corresponding to a constant (age-independent) death/extinction rate, to which we compare our results. We apply the inference method to simulated trees to rigorously confirm the method's accuracy, investigate power, and identify biases to be expected when age dependency in the death rate is neglected. Finally, we apply our method to two real data sets: (1) a complete phylogeny of birds (class Aves) (Jetz et al. 2012), where we investigate robustness of our model fit to phylogenetic uncertainty and compare it to the fits of previously investigated models; and (2) a phylogeny of self-incompatible (SI) and -compatible (SC) nightshade species (Solanaceae), where we shed new light on the question of whether breeding system affects species survival (Goldberg et al. 2010; Goldberg and Igić 2012).

Methods

Model

The derivation of the likelihood is based on the assumption that a sampled phylogenetic tree, $\mathcal{T}$, is generated by the following model (Lambert 2010; Lambert and Stadler 2013):

- The process is initiated by one individual at time 0 and observed after a fixed time T_s.
- Each individual gives birth at constant rate λ. The old individual keeps its current age, while the newly born individual is assigned age zero.
- The lifetime of each individual is random, with an arbitrary distribution characterized by probability density g, and realized lifetimes are independent of one another.
- $\mathcal{T}$ is the tree spanned by all individuals sampled at time T_s, where each extant individual at time T_s is independently sampled with probability p (i.e., Bernoulli sampling). The expected sampling fraction is thus equal to p.

Table 1. Definitions of model parameters and functions

Symbol	Definition
λ	per-lineage speciation rate
$g(a)$	probability density of the species lifetime distribution as a function of age, a
$\mu(a)$	extinction rate as a function of age, a, related to $g(a)$ by Equation (2)
k	shape parameter of gamma lifetime distribution
θ	scale parameter of gamma lifetime distribution
ℓ	mean lifetime, equal to $k\theta$ under the gamma distribution
η	net diversification rate, i.e., asymptotic exponential growth rate of number of extant lineages, given by Equation (3)
ϵ	turnover, i.e., relative extinction fraction, given by $\epsilon=1-\eta/\lambda$
n	number of sampled tips (extant species)
p	expected sampling fraction
T_s	stem age of tree, i.e., time since progenitor arose
T_c	crown age of tree, i.e., time since first speciation event giving rise to sampled tips

We consider the context where each "individual" is a species, with "birth" corresponding to a speciation event and "death" to extinction of a species, and will thus refer to speciation and extinction rates, rather than birth and death rates, throughout the article. In this macroevolutionary context, T_s is called the stem age of the tree. All definitions of model parameters and functions are collected in Table 1.

Note that speciation is asymmetric, i.e., there is a "mother" species whose age is not reset upon speciation. This asymmetry could be captured in the tree-generating process, for instance by orienting trees such that the new species is always placed branching off to the right. However, it turns out (Lambert and Stadler 2013; see also the section "Mathematical Likelihood Formulae" below) that all tree orientations are equally likely under this model. That is, distinguishing the mother and daughter species is irrelevant to the likelihood of the tree and thereby the parameter estimates obtained. (Note however that orientation does matter for an age-dependent extinction model in which sampling can occur continually through time; Lambert et al. 2014.)

For computational implementation, we suppose here that lifetimes are gamma distributed. The properties of this distribution and justification for its selection are clarified in the following sub-section. Furthermore, for the purposes of inference, we will assume throughout that p is a fixed, known quantity, while λ and the parameters of the lifetime distribution g are to be estimated. Often in data sets there is indeed a good independent estimate of the proportion of species sampled in a clade.

Lifetime Distribution and Extinction Rate

The gamma lifetime distribution with shape parameter k and scale parameter θ is characterized by the following density:

$$g(a)=\frac{a^{k-1}\exp(-a/\theta)}{\Gamma(k)\theta^k} \quad (1)$$

The gamma density is unimodal, and its mean ($k\theta$) and variance ($k\theta^2$) can be varied independently. It contains as a special case (when $k=1$) the exponential distribution.

Macroevolutionary models of the "birth–death" type have generally been formulated in terms of an extinction rate, μ, which may depend on various factors (Stadler 2013b). In our model, μ is a function of age, a. The relationship between lifetime distribution and extinction rate is given by (Lambert and Stadler 2013):

$$\mu(a)=\frac{g(a)}{\int_a^\infty g(s)ds}. \quad (2)$$

Taking g to be the gamma density, we have a constant extinction rate $\mu(a)=1/\theta$ if and only if $k=1$, i.e., in the special case of an exponentially distributed lifetime. When $k\neq 1$, there is no explicit formula for $\mu(a)$, but it can be shown (see Supplementary Text available on Dryad at http://dx.doi.org/10.5061/dryad.7894h) that for $k<1$ the extinction rate decreases with age, while for $k>1$ the extinction rate increases with age. In the limiting case where $k\to\infty$ and $\theta\to 0$ simultaneously, with $k\theta\to\ell$, the distribution approaches a Dirac delta, i.e., lifetimes have a fixed duration of ℓ. The choice of a gamma lifetime distribution therefore captures qualitatively different extinction patterns, while containing the previously studied case of constant extinction rate (exponentially distributed lifetime), over which it introduces just one additional parameter.

Composite Parameters

While we have described the model in terms of the parameters λ, k, and θ, composite parameters can provide further insight into the diversification process. Furthermore, as we will see later, composite parameters are sometimes more precisely estimated than individual model parameters.

First, we will often refer to the *mean lifetime*, ℓ, equal to $k\theta$ for a gamma-distributed lifetime. This composite parameter allows direct comparison of estimates under the gamma and the exponential models.

Second, in diversity-independent diversification processes such as those considered here, the number of lineages grows asymptotically exponentially at a rate η called the *net diversification rate*, a widely used measure in macroevolutionary analyses (Pyron and Burbrink 2013; Moen and Morlon 2014). More generally, this exponential growth rate of a population is classically known as the *Malthusian parameter* (Rice 2004, p. 16). For a given lifetime density g and speciation rate λ, η is a solution (Lambert 2010) to

$$\eta=\lambda\int_0^\infty g(x)\left(1-e^{-\eta x}\right)dx. \quad (3)$$

If $\lambda\ell>1$, η is the unique positive solution to Equation (3). If $\lambda\ell<1$, η is negative, and when g is the gamma density, η is the unique negative solution to Equation (3). In the critical case where $\lambda\ell=1$, $\eta=0$.

When g is the gamma density, $\eta(\lambda,k,\theta)$ must be found numerically, except when lifetimes are distributed exponentially, where $\eta(\lambda,1,\theta)=\lambda-1/\theta$. However, it can be proven (see Supplementary Text) that: (i) for given k and θ, $\eta(\lambda,k,\theta)$ increases (asymptotically linearly) with speciation rate λ, and (ii) for given λ and mean lifetime ℓ, $\eta(\lambda,k,\ell/k)$ increases with k, approaching an asymptotic value corresponding to the case when all lifetimes are fixed equal to ℓ. This last observation shows that the net diversification rate $\eta(\lambda,k,\theta)$, and correspondingly the expected number of lineages after a given time, varies with k even when λ and ℓ are fixed. That is, $\eta(\lambda,k,\ell/k)$ depends not only on speciation rate λ and mean lifetime ℓ, but also on the particular distribution of lifetimes, as characterized by k.

The final composite parameter used here is *turnover*, or relative extinction fraction, which we denote ϵ. In a constant-rate model, this parameter has been defined as μ/λ where μ is the extinction rate and λ is the speciation rate (Pyron and Burbrink 2013). Similarly to the net diversification rate, this definition should be adjusted when extinction is age dependent. We propose to define $\epsilon:=(\lambda-\eta)/\lambda$, which is the asymptotic population-level extinction rate divided by the (still constant) speciation rate.

Mathematical Likelihood Formulae

According to the coalescent point process (CPP) representation (Popovic 2004; Lambert 2010; Lambert and Stadler 2013), a tree $\mathcal{T}$ generated by our model, with stem age T_s and n sampled tips, can be fully characterized by the list of its node depths, $\{x_i\}_{i=1}^{n-1}$. The node depths are the times *since present* of speciation events in the tree (cf. Figure 1 in Lambert and Stadler 2013), not to be confused with the times *between* speciation events. Under our model, for a given lifetime density g, speciation rate λ, and sampling fraction p, there is a random variable H with density f, such that the node depths form a sequence of independent copies of H, stopped at its first value larger than T_s (Lambert and Stadler 2013).

If f is known, then the tree likelihood is given by the formulae in Section 3.2 of Lambert and Stadler (2013), along with their Proposition 2 to account for sampling. We use the likelihood conditioned on sampling at least one tip (event S) and on either the stem age T_s or the crown age T_c, the latter being the time since the first speciation event giving rise to sampled descendants in $\mathcal{T}$. Conditioning on stem age we have:

$$\mathcal{L}_s(\mathcal{T}|T_s,S)=\frac{1}{F(T_s)}\prod_{i=1}^{n-1}f(x_i), \quad (4)$$

while conditioning on crown age (assuming $T_c\equiv x_{n-1}$) we have:

$$\mathcal{L}_c(\mathcal{T}|T_c,S)=\frac{1}{F(T_c)^2}\prod_{i=1}^{n-2}f(x_i). \quad (5)$$

where $F(t) := 1/\Pr(H > t) = 1/\int_t^\infty f(s)ds$. In general, $\mathcal{T}$ refers to the entire tree as characterized by its topology and branch lengths. However, in accordance with the CPP representation, these likelihood expressions make it clear that any ordering of node depths is equally likely under this model (see Section 4 in Lambert and Stadler 2013). Thus, the likelihood $\mathcal{L}_s$ (resp. $\mathcal{L}_c$) is simply proportional to the probability density of observing the node depths $\{x_i\}_{i=1}^{n-1}$ (resp. $\{x_i\}_{i=1}^{n-2}$), which implicitly also specify the number of tips, conditioned on stem (resp. crown) age and on having sampled descendants. We will typically consider the likelihood conditioned on the crown age, since this quantity tends to be better known from data than the stem age. We will, however, use conditioning on stem age applied to individual subclades of SC nightshades (see section "Nightshades Data Analysis" below). Computationally there is no difference in the ease of use of either formula.

It remains to characterize f from the model ingredients. For a given lifetime density g, speciation rate λ, and sampling probability p, Proposition 6 in Lambert and Stadler (2013) yields

$$F(t) = 1/\Pr(H > t) = 1 - p + pW(t)$$

and

$$f(t) = -\frac{\mathrm{d}}{\mathrm{d}t}(1/F(t)) = \frac{pW'(t)}{\left(1-p+pW(t)\right)^2}$$

where W is the so-called *scale function* associated with g and λ. W is a differentiable function defined by its Laplace transform:

$$L(W)(y) := \int_0^\infty \exp(-ys)W(s)\mathrm{d}s = 1/\psi(y) \qquad \text{for } y > \eta$$

where

$$\psi(y) := y - \lambda \int_0^\infty g(x)(1-\exp(-yx))\mathrm{d}x \qquad \text{for } y \geq 0$$

is called the *Laplace exponent*. Incidentally, the net diversification rate η is a root of ψ (see Equation (3)) and $W(t)$ is the expected number of species extant at time t conditional on there being at least one (Lambert 2010). Taking g to be the gamma density given by Equation (1), we can write ψ in closed form as:

$$\psi(y) = y - \lambda\left(1-(1+\theta y)^{-k}\right)$$

Numerical Evaluation of the Likelihood

Analytical expressions for the scale function W, and in turn the likelihood, are available in the special case $k = 1$ (exponentially distributed lifetimes; Lambert and Stadler 2013), but not for general k. We thus implement a numerical computational method in Matlab. A numerical inverse Laplace transform method has previously been developed for stability in calculating scale functions of spectrally negative Lévy processes (Surya 2008), which includes the scale function W required here (Lambert 2010). As a basis we use the Matlab code provided in Surya (2008), with parameters of the inverse Laplace transform itself (controlling numerical error in W) set as chosen there. The derivative of W is then approximated with a central difference, where numerical error is controlled through the spacing between points at which W is evaluated. We take 500 grid points evenly spaced on the interval $[0, T]$ (see Supplementary Text), where T is either the stem age, T_s, or the crown age, T_c, according to the choice of conditioning in the likelihood formula. Node depths in a tree (simulated or constructed from real data) are rounded to the nearest grid point value, thus rounding errors in W are also controlled by grid spacing.

Notice that the gamma distribution has the advantage of presenting a simple, closed-form expression for its Laplace transform. Lifetime distributions for which such a closed-form expression is not available would require a significantly more computationally intensive implementation, either calculating ψ numerically or evaluating the likelihood instead via Equation 15 in Lambert and Stadler (2013).

Maximum Likelihood Inference and Model Selection

Likelihood optimization is performed using the built-in Matlab function 'fmincon'. The likelihood is optimized over the parameters (λ, k, θ) under the gamma model, or over (λ, θ) with k fixed to one under the exponential model. In both cases, we specify a fixed value of p. In the case of exponentially distributed lifetimes, we know from analytical expressions that the likelihood is fully characterized by two composite parameters ($\lambda - 1/\theta$ and λp), and thus p cannot be estimated together with λ and θ (Stadler 2009). However, in the case of gamma-distributed lifetimes, all four parameters λ, θ, k, and p are identifiable and thus could be estimated given a large enough tree, as we show in the Supplementary Text. Nonetheless, fixing p is expected to make the inference faster and more precise for the remaining parameters.

Optimization is run from multiple initial points, accepting the result that yields the highest maximized likelihood, to increase confidence that a global peak is located (see Supplementary Text). Maximum likelihood estimates (MLEs) of parameters for a given phylogeny are denoted with a hat (e.g. $\hat{\lambda}$). We compare the gamma and exponential lifetime distribution models using the likelihood ratio test with one degree of freedom at a 5% significance level.

Availability of Code in Matlab and R

Our numerical method of likelihood evaluation and optimization is implemented in Matlab. We additionally compile the necessary functions (using 'mcc' in Matlab) into stand-alone applications, which can be used on compatible systems without requiring a

Matlab license, thanks to the Matlab Compiler Runtime (MCR) freely available from Mathworks at http://www.mathworks.com/products/compiler/mcr/. Our Matlab source code and compiled application files are available to download from https://github.com/cevo-public. We further provide an interface to call these functions in R, included in the package TreePar v3.2 (Stadler 2011) available on CRAN. The steps to set up MCR and apply our method in R are described in the R help pages of the relevant functions ('create.mat', 'LikAge', and 'bd.age.optim.matlab').

Simulated Trees

To analyze performance of the inference method, we simulate trees under constant speciation rate and gamma-distributed species lifetimes, the key assumptions of the model described above. We could study performance as a function of either tree size (number of tips, i.e., sampled extant species) or tree age. We argue that the former is a more informative analysis for empirical systematists, because the number of sampled species is unequivocally known in real data.

Sets of simulated trees with a given number of tips are generated using the R package TreeSimGM (Hagen and Stadler 2013). The package allows two options for obtaining trees with a fixed number of tips; we typically use the more computationally efficient option (denoted 'gsa=FALSE'), which is expected to introduce a slight bias in parameter estimates that decreases with tree size. This expectation was confirmed by a limited set of simulations repeated with the computationally intensive 'gsa=TRUE' option (see Supplementary Text for details).

The TreeSimGM package currently only allows complete sampling ($p=1$). To generate an incompletely sampled tree with n sampled tips and given sampling fraction $p<1$, we first simulate a tree with n/p tips, then select n tips uniformly at random. This sampling procedure is much more computationally efficient than Bernoulli sampling (where each tip is sampled independently with probability p) when a particular number of sampled tips is desired, and is not expected to yield substantially different patterns in the tree structure.

For each final sampled tree, the stem age and list of node depths (including the crown age) are recorded and used for maximum likelihood inference as described above. That is, we maximize the probability of the observed tree (out of all trees with the same age) given the parameters, as given by the likelihood in Equations (4) and (5). Thus, the conditions applied in the likelihood formula (fixed tree age, random number of tips, Bernoulli sampling) are not identical to the conditions under which the tree was generated (fixed number of tips, random tree age, fixed sampling fraction). We take this approach because, mathematically, only likelihood formulae that condition on tree age are available for our model (see Supplementary Text for further discussion). Nonetheless, we expect that conditioning on different features (age versus number of tips) of a given tree will result in similar parameter estimates for sufficiently large trees, in accordance with an investigation of this issue for the constant-rate birth–death model (Stadler 2013a). As we will see, it turns out that the discrepancy in conditioning between the simulations and the likelihood formula also does not appear to be problematic for the estimates we achieve under the present model. In conclusion, we emphasize that the aim of the simulation study is to assess the performance of the inference method on a tree of a given size, providing empirical biologists with relevant information when dealing with data sets consisting of a known number of samples.

Confidence Intervals

We use two types of confidence intervals here. First, from a set of simulated trees, the "95% (bootstrap) confidence interval" for a parameter estimate refers to the smallest interval containing the MLEs of 95/100 simulated trees. Second, for any individual tree (applied here only to single trees reconstructed from real data), we can find the profile likelihood (95%) confidence interval for a parameter estimate. The boundaries of this interval are the minimum and maximum values of the focal parameter for which the likelihood, optimized over the remaining parameters with the focal parameter value fixed, would not be rejected by a likelihood ratio test (at 5% significance level) in comparison to the likelihood optimized over all parameters.

Simulation Study

To test the performance of the inference method, we estimate parameters from simulated trees. Parameter values used for simulation are set as follows. We select a range of values of the lifetime distribution shape parameter: $k=1$ (exponential distribution), $k=0.5$ (more variable than exponential, with mode at zero), $k=5$ and $k=100$ (bell shaped and progressively less variable). The scale parameter, θ, is set to $1/k$ such that the mean lifetime ℓ is always equal to one; that is, time is scaled in units of average lifespan. Supplementary Fig. S1 illustrates the chosen lifetime distributions. We also select a range of values for the net diversification rate η (0.25, 0.5, 1, 1.5), and tune speciation rate λ to yield the chosen η. Our main findings (using all η values) are derived from simulated trees with $n=1000$ tips and complete sampling ($p=1$). We further consider the influence of tree size by comparing the performance on fully sampled, 100-tip trees (only for $\eta=0.5$), and the influence of incomplete sampling by comparing the performance on trees containing $n=1000$ sampled tips representing sampling fraction $p=0.5$ (only for $\eta=0.25$).

Aves Data Analysis

Phylogenetic trees of all 9993 extant species in class Aves have been previously constructed in a Bayesian

framework (Jetz et al. 2012), yielding a posterior distribution of trees available at http://birdtree.org. The construction was undertaken with four approaches: based on either of two backbones (Hackett or Ericson) and either the full set of 9993 species or only those 6670 species for which genetic sequence data are available (see Jetz et al. 2012 for details).

To test robustness of our results to phylogenetic uncertainty, we analyze a set of 100 trees randomly drawn from the posterior distribution in each of the four approaches. That is, we apply our inference approach separately to each tree to obtain a maximum likelihood point estimate, and then look at the range of point estimates across sets of 100 trees. We assume sampling is complete ($p=1$) for the set of all known bird species, and account for incomplete sampling ($p=6670/9993\approx 0.67$) in the set of species with available genetic data. The likelihood is conditioned on the crown age of each tree, which ranges from 89.0 to 149 million years (myr).

In addition, to obtain confidence intervals on parameter estimates (reflecting uncertainty in the model fit to a fixed tree), we conduct a more detailed analysis on one tree from the set constructed under the Hackett backbone including all species (H-full). This tree has a crown age (107.9 myr) and MLE point estimates (see Supplementary Table S11) close to the median values in the set of 100 trees. The magnitudes of confidence intervals on parameter estimates in this tree are thus expected to be reasonably representative of the majority of trees.

The size of this data set makes it a promising test case for picking up signals left in the tree structure by particular patterns of diversification. Jetz et al. (2012) compared the fit of nine models of diversification (see their Supplementary Discussion Table 1), including constant-rate models with or without extinction, several models with time-varying parameters, and a "clade shift" model incorporating heterogeneity in rates across lineages. To additionally compare the fit of our model, we use the Akaike Information Criterion (Akaike 1974), $\mathrm{AIC}=2(-\log(\mathcal{L})+m)$, where m is the number of model parameters and $\mathcal{L}$ is the likelihood under the given model. Since the previous model comparison was based on a set of trees containing all 9993 species constructed with the Hackett backbone, we also take trees constructed with this approach, but the particular trees drawn from the posterior distribution are not the same.

Nightshades Data Analysis

We analyze a previously constructed phylogeny of 356 extant nightshade species, including 135 SI species and 221 SC species (Goldberg et al. 2010). This phylogeny represents sampling fractions of $p=0.162$ among SI species and 0.150 among SC species (see Supplementary Material of Goldberg et al. 2010; we follow these authors in assuming independent sampling). The previous study used a model that assumes speciation and extinction rates depend only on the character state (SI/SC) and incorporates transitions between states. Since our model does not incorporate transitions, we separate SI and SC species in the maximum likelihood tree for our analysis. While transitions from self-incompatibility to self-compatibility are common, there is strong evidence that reverse transitions have not occurred within the family (Goldberg and Igić 2012 and references therein). We therefore consider the subtree of SI species obtained simply by pruning all SC clades. The end of a species lifetime on this subtree should thus be interpreted as either extinction *or* a transition to self-compatibility. We treat each clade consisting of only SC species as an independent realization of a fixed model (i.e., with the same parameter values for all SC clades) arising from one event of losing self-incompatibility. The overall likelihood of the parameters given all SC data is thus taken as the product of the likelihood of the 73 individual clades. SC clade size ranges from 1 to 21 species, and since many (35) clades are singletons, we conduct two analyses: one including all clades with likelihood conditioned on stem age, and one including only non-singleton clades with likelihood conditioned on crown age as usual.

We apply maximum likelihood inference as described above to both the SI subtree and the SC clades. Clear peaks in the likelihood surface were located under both the gamma and the exponential models for the SI subtree, and the exponential model for the SC clades. However, under the gamma model in the SC case, likelihood increased very gradually and indefinitely with dramatic increases in λ and concurrent decreases in k and $\ell=k\theta$ (with θ itself remaining at a similar value). We stopped the optimization at the parameter values reported in Supplementary Table S13, which should thus be interpreted as conservative estimates for those parameters that continued to change in the direction of increasing likelihood (λ, k, and ℓ) or approximate estimates for those that remained roughly constant (θ, η, and ϵ). Similarly, the confidence interval lower bound on λ and upper bounds on k and ℓ are conservative estimates determined at the cut-off point. If the likelihood under the gamma model indeed continues to increase with more extreme parameter values, the reported confidence interval bounds would become more stringent and any improvement over the exponential model would only be strengthened.

Profile likelihood confidence intervals were constructed as described above for parameters of particular interest (λ, k, ℓ) for both SI and SC data. For the SI species, we additionally conducted a bootstrap analysis by using the MLE parameter values of the gamma model to simulate 100 trees with $n/p=135/0.162=833$ tips, followed by sampling $n=135$ tips, as described under "Simulated Trees." Twenty-six simulations failed because they exceeded computational capacity; this appears to be due to the accumulation of too many extinct lineages, attributable to the high turnover in this parameter set. Our analysis is based on the remaining 74 successfully simulated trees. The

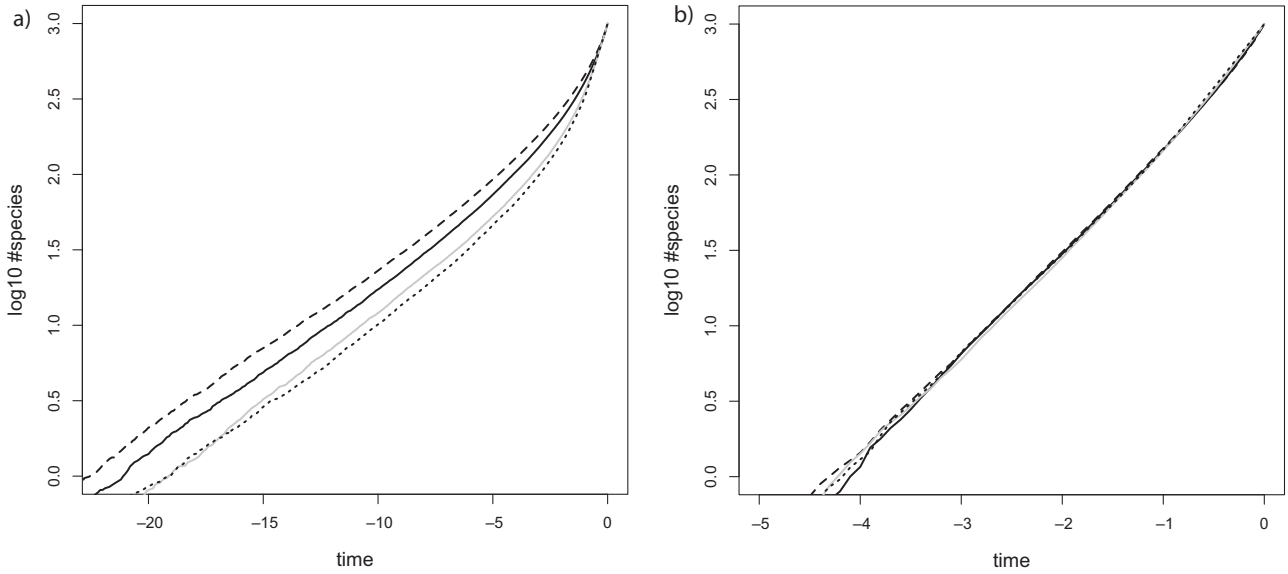

FIGURE 1. LTT plots for trees simulated under gamma-distributed lifetimes. Each curve plots the number of lineages (on a base-10 log scale), averaged over 100 simulated trees per parameter set, versus time since present. All trees are completely sampled and have $n = 1000$ species at present. Net diversification rate η used in simulations is fixed to 0.25 in panel (a) or 1.5 in panel (b), and lifetime shape parameter k varies (black dashed: $k = 0.5$; black solid: $k = 1$; gray solid: $k = 5$; black dotted: $k = 100$). $k = 1$ corresponds to exponentially distributed species lifetimes (constant extinction rate), $k < 1$ corresponds to more variable lifetimes (extinction rate decreases with age), and $k > 1$ corresponds to less variable, bell-shaped lifetime distributions (extinction rate increases with age).

median MLEs from these trees (not shown) are close to the true parameter values used for simulation, indicating that discarding the failed simulations did not introduce any obvious bias. For the SC species, we did not conduct a bootstrap analysis, since it is not clear what would be an appropriate set of clades to simulate.

To test whether SI and SC species exhibit differences in lifetime distribution, we compare a model where all parameters are allowed to differ between SI and SC species (thus characterized by the six parameters λ_{SI}, k_{SI}, θ_{SI}, λ_{SC}, k_{SC}, θ_{SC}) with two restricted models: (1) where the shape parameter of the lifetime distribution does not differ ($k_{SI} = k_{SC}$), or (2) where the mean lifetime does not differ ($\ell_{SI} = \ell_{SC}$), but individual values of both k and θ may differ. The total log likelihood of the full (six-parameter) model is obtained, under the assumption of independence, simply by summing the log likelihoods of the SI and SC data as previously determined individually. For each restricted model, we optimize the likelihood jointly over both the SC and SI data sets with the focal parameter restricted to be the same in both. We again obtain results conditioning likelihood on either stem age or crown age of SC clades. We compare each restricted model to the full model using the likelihood ratio test with one degree of freedom.

RESULTS: SIMULATION STUDY

We test the computational method on simulated trees under chosen parameter sets, as described in the "Methods" section. Below we outline general observations on the performance of the inference method. The median and 95% confidence intervals of the MLEs for all parameter sets are reported in Supplementary Tables S1–S9. Lineages-through-time (LTT) plots (Fig. 1) provide an additional way to visualize the influence of the parameters on the trees (Harvey et al. 1994). Note that since all topologies are equally likely under our model (Lambert and Stadler 2013), all information relevant to the likelihood of the tree is contained in the set of branching times, or equivalently the accumulation of LTT.

Inference under the Gamma Lifetime Distribution

Accuracy of inference.—First, and most importantly, inference assuming the gamma model is essentially effective in recovering parameter values accurately. For all 16 sets of parameter values tested for completely sampled 1000-tip trees, the median MLE (across 100 simulated trees per parameter set) is close to the true value used for simulations, and the true value is well within the 95% confidence interval (Supplementary Tables S1–S2). Similar quantitative results are obtained if we consider 1000-tip trees obtained by 50% sampling from 2000-tip trees (Supplementary Tables S4–S5). Thus, incomplete sampling in itself does not appear to compromise inference performance, if we compare trees with the same number of sampled tips, provided the sampling fraction (p) is known. When sampling is complete but tree size is reduced to 100 tips, median MLEs generally remain close to the true values, but unsurprisingly, confidence intervals are much larger

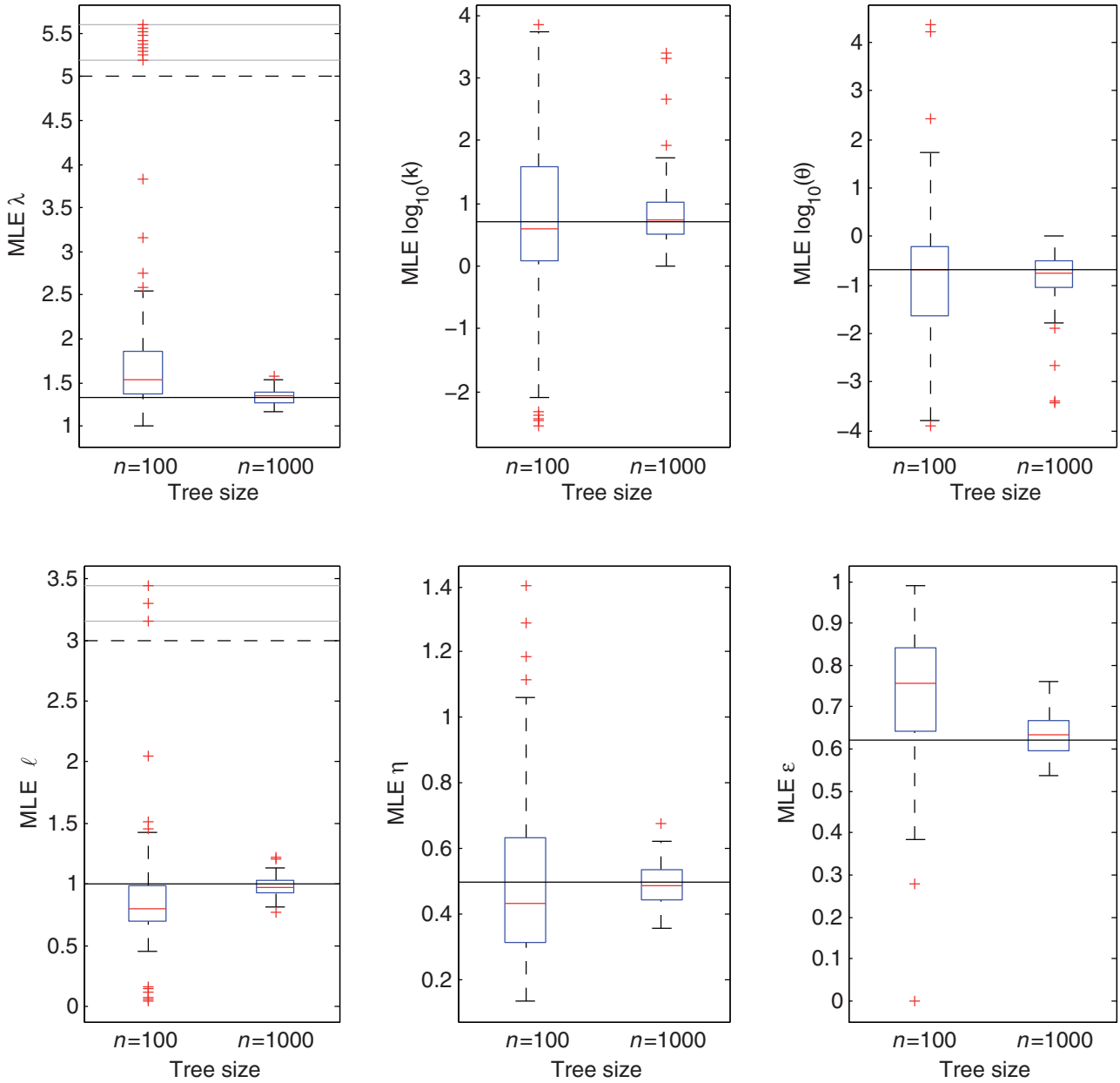

FIGURE 2. Dependence of inference quality under the gamma model on tree size. Parameters inferred under the gamma lifetime distribution model are compared for completely sampled trees with $n=100$ versus $n=1000$ extant species, simulated under true values of $\eta=0.5$ and $k=5$. Illustrated is the distribution of MLEs for each parameter (defined in Table 1) across 100 simulated trees per parameter set. Note that k and θ are plotted on base-10 log scales. The box contains the interquartile range with the median marked as a line. The solid black line indicates the true parameter value. Where MLE ranges are large, the dashed black line indicates an arbitrary upper data limit, beyond which outliers are drawn between the gray lines. For the 100-tip trees, 10 estimates of λ fall above a value of 5 (actual values ranging from 7.19 to 99.3) and three estimates of ℓ fall above a value of 3 (actual values ranging from 61.7 to 561).

(Fig. 2 and Supplementary Tables S7–S8). A slight bias in parameter estimates, which decreases with tree size, can be attributed to a slight bias in the simulated trees themselves, due to the use of a faster but less accurate option for simulating trees of a desired size. Using the more accurate tree simulation option reduces the bias; see "Methods" section and Supplementary Text for details. That is, this bias does not appear to be introduced by the inference method.

Limited precision in estimating individual lifetime parameters.—The variability in the MLEs of the lifetime distribution parameters, k and θ, is much greater than that of the speciation rate, λ. However, the mean lifetime, $\ell=k\theta$, is much more tightly estimated than k and θ individually, reflected by a strong negative correlation between these estimates (detailed in the Supplementary Text). A large k value is particularly difficult to infer precisely; this is unsurprising, since the

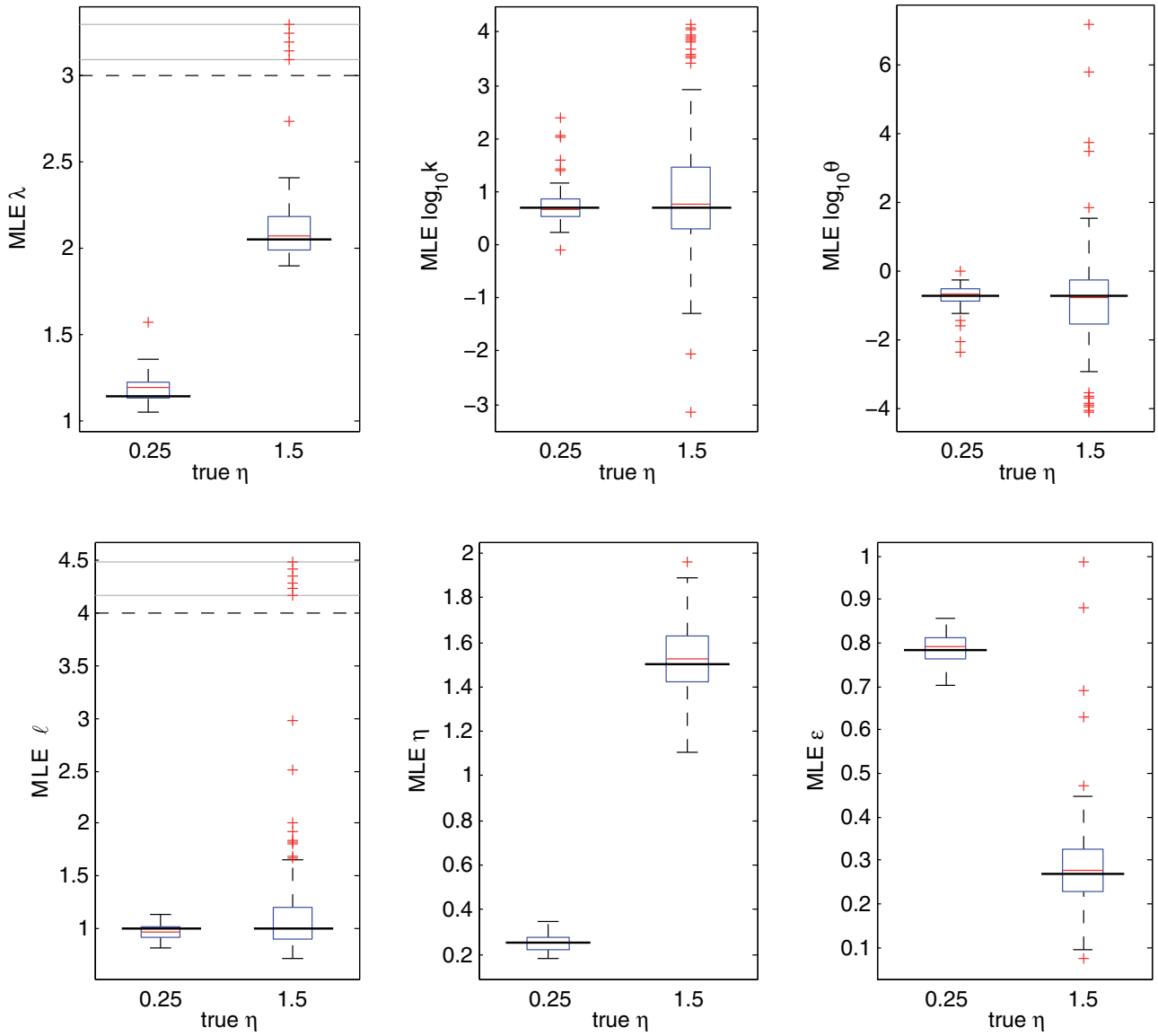

FIGURE 3. Dependence of inference quality under the gamma model on net diversification rate, η. Parameters inferred under the gamma lifetime distribution model are compared for completely sampled trees with $n=1000$ extant species, simulated under true values of $k=5$ and $\eta=0.25$ versus $\eta=1.5$. Illustrated is the distribution of MLEs for each parameter (defined in Table 1) across 100 simulated trees per parameter set. Plotting conventions are as in Figure 2. For true η of 1.5, outliers consist of five estimates of λ falling above a value of 3 (with actual values ranging from 3.3 to 141) and six estimates of ℓ falling above a value of 4 (with actual values ranging from 5.4 to 1.07e4).

gamma distribution converges to a limiting distribution as $k \to \infty$, such that large k values result in very similar distributions. For fixed η, the normalized sizes of the confidence intervals of k and θ indeed tend to increase with true k (those of λ, however, decrease).

Decreasing precision in lifetime parameter estimates with increasing net diversification rate.—The variability across trees in estimated k and θ tends to increase with the true value of the net diversification rate η (see Fig. 3 and the normalized size of the confidence intervals reported in Supplementary Tables S1–S2). There is an intuitive reason for this trend. For fixed k and θ, larger net diversification rate η corresponds to lower turnover, i.e., fewer species have arisen and gone extinct by the time the phylogeny reaches a given size and is observed. In turn, the tree contains less information about the lifetime distribution. The difficulty in inferring parameters at large η is visually indicated in the LTT plots: for smaller η (Fig. 1a), there is a clear separation between curves for different k values, while this separation virtually disappears for larger η (Fig. 1b).

Inference under the Exponential Lifetime Distribution

We now turn to the results when the exponential model (i.e., constant, age-independent extinction

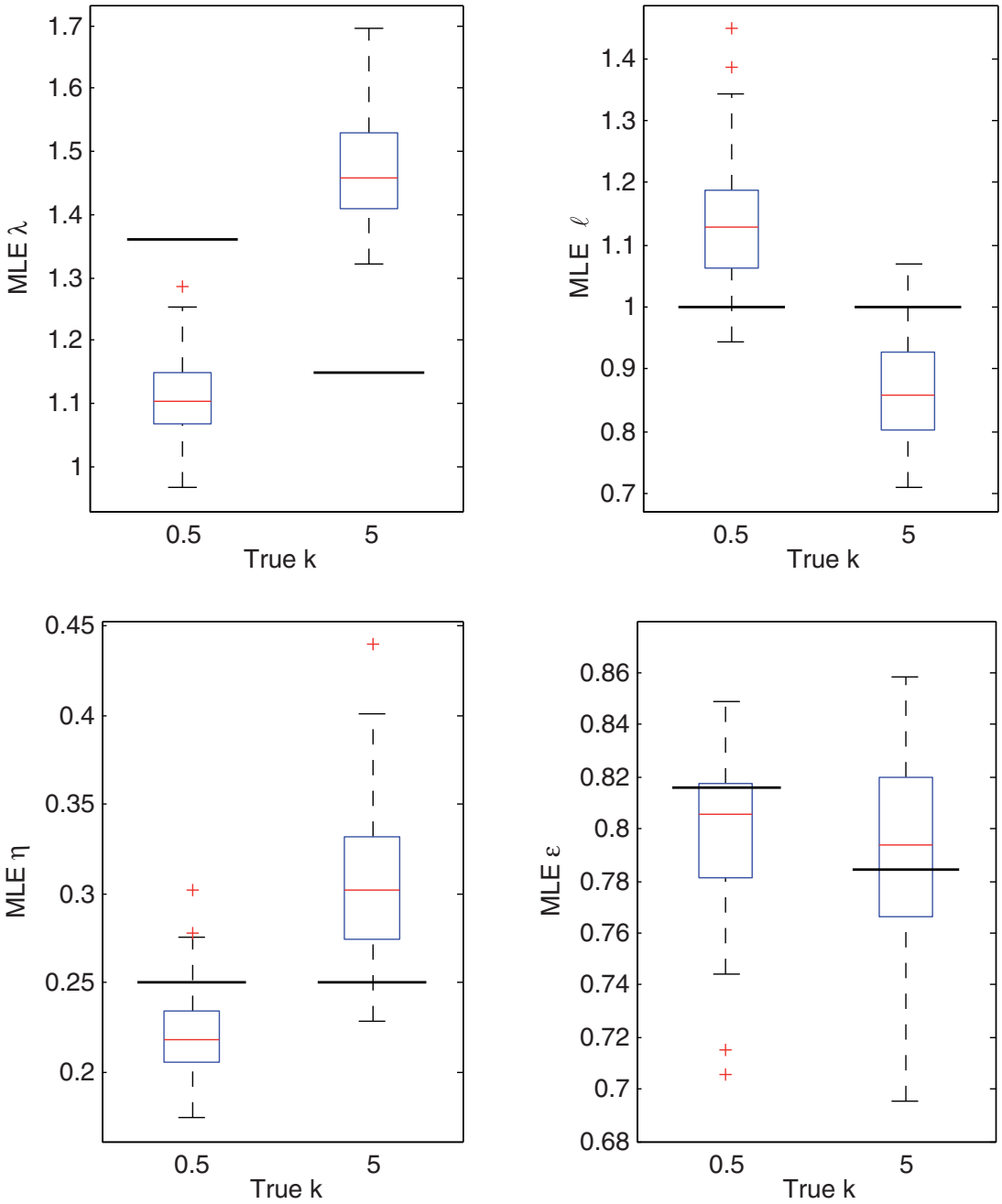

FIGURE 4. Biases in parameter estimates obtained with the exponential (age-independent) model when extinction actually is age dependent. Completely sampled trees with $n=1000$ extant species are simulated using true values of $\eta=0.25$ and lifetime distribution shape parameter $k=0.5$ (lifetime more variable than exponential) or $k=5$ (lifetime less variable than exponential). Illustrated is the distribution of MLEs for each parameter, inferred under the exponential model, across 100 simulated trees per parameter set. Plotting conventions are as in Figure 2.

rate) is assumed for parameter inference. When simulations are indeed conducted under an exponential lifetime distribution (i.e., true $k=1$), inference either under the general gamma distribution or constrained to the exponential return similar median MLEs for corresponding parameters. However, biases in parameter estimates using the exponential model arise when true k deviates from one. When simulations use a lifetime distribution that is more variable than exponential ($k=0.5$), speciation rate (λ) and net diversification rate (η) tend to be underestimated; conversely, when simulations use a less variable lifetime distribution ($k=5$ or 100), λ and η tend to be overestimated (Fig. 4). This result can be explained by recalling that less variable lifetimes (larger k) result in larger η for given λ and ℓ (see "Composite Parameters" section). When k is actually larger than one, using an exponential distribution supposes that lifetimes are

more variable than they actually are. Speciation rate (λ) and net diversification rate (η) are then over-estimated to compensate in explaining the observed growth of the tree, and vice versa when k is actually less than one. A bias in estimated mean lifetime (ℓ) is also apparent, but shows a more complex pattern. Mean lifetime is overestimated at small η and underestimated at large η when $k<1$, with the reverse for $k>1$. The effect on turnover (ϵ) is less clear. For $k=0.5$, ϵ is consistently but only slightly underestimated by the median MLE, while for $k>1$, the results are inconsistent. In all cases, the true value of ϵ still falls within the 95% confidence interval of the estimates. It is possible that the slight bias in the simulated trees themselves (see "Methods" section), which turns out to inflate the estimates of ϵ (Supplementary Text), masks any bias introduced in inference by assuming the exponential model.

These qualitative patterns remain consistent when sampling is incomplete ($p=0.5$, tested only for $\eta=0.25$; see Supplementary Table S6). Interestingly, however, the magnitude of bias appears to be exacerbated for λ, ℓ, and ϵ, but reduced for η.

Further insight into these biases can be gained from LTT plots for the completely sampled trees (Fig. 1). A "pull of the present" effect has been described by Nee et al. (1994a) for the exponential model when turnover is large: in the very recent past, the LTT plot has a steeper slope, indicating faster lineage accumulation, because species that arose recently have not yet gone extinct. Specifically, the slope is initially η and increases to λ near the present. For the gamma model, we find that this pull of the present effect is intensified when k is large. This can be understood by considering the extreme case as $k\to\infty$, such that all lifetimes last exactly ℓ time units: then no lineages arising within the past ℓ time units have yet gone extinct. In contrast, a smaller k value (more variable lifetime distribution) tempers the pull of the present by the early extinction of some lineages. This effect is not explained by different speciation rates: for fixed η, larger k actually corresponds to smaller λ, which would be expected to produce a shallower recent slope under the exponential model. Indeed, this effect clarifies the observed biases in inference under the exponential model. If k is large, the LTT curve bends upward earlier than expected, and fitting a line through the initial portion of the curve yields an overestimate of η. Furthermore, the curve for large k is steeper than expected in the recent past, resulting in overestimation of λ.

Power to Distinguish between Gamma and Exponential Lifetime Distributions

Table 2 reports the proportion of simulated trees in each parameter set for which the likelihood ratio test rejects the exponential model. When simulations actually are conducted under an exponential distribution ($k=1$), this proportion is close to the expected Type I error of 5% (the significance level used for the test), ranging from 3 to 10/100 simulations. When true $k\neq1$, the proportion rejected represents the power of the statistical test, and as expected, the power increases as k deviates further from one. The power is lower for larger η; this reflects the greater difficulty in precise lifetime parameter identification noted above for larger η. Incomplete sampling (tested only with $\eta=0.25$) appears to result in lower power for $k<1$, but similar power for $k>1$. Smaller trees (tested only with $\eta=0.5$) yield substantially lower power to reject the exponential distribution.

TABLE 2. Power to reject exponential model[a]

True η	Sampled tree size (n)	Sampling fraction (p)	True k			
			$k=0.5$	$k=1$	$k=5$	$k=100$
$\eta=0.25$	1000	1	0.41	0.10	0.89	1.00
		0.5	0.24	0.03	0.91	0.97
$\eta=0.5$	1000	1	0.31	0.05	0.81	0.95
	100		0.12	0.03	0.13	0.21
$\eta=1$	1000	1	0.28	0.03	0.42	0.76
$\eta=1.5$	1000	1	0.22	0.07	0.28	0.38

[a]Proportion of simulated trees (out of 100) rejecting the exponential model ($k=1$) using the likelihood ratio test at a significance level of 5%, depending on true values of the model parameters η and k; number of sampled tips, n; and sampling fraction, p.

RESULTS: APPLICATION TO DATA

Birds (Class Aves)

We apply our inference method to a published phylogeny of 9993 bird species (Jetz et al. 2012). First, we analyze robustness of our model parameter estimates to phylogenetic uncertainty by conducting estimation on sets of 100 trees drawn from the posterior distribution under each of four tree construction approaches. The range of MLEs obtained for each set (Fig. 5) shows patterns comparable to those seen in the simulation study: speciation rate (λ) and composite parameter (ℓ, η and ϵ) estimates vary across much less than an order of magnitude among trees, while estimated individual parameters of the lifetime distribution (k and θ) each vary across a few orders of magnitude but are highly negatively correlated (not shown). Thus, phylogenetic uncertainty appears to induce similar patterns of variability in parameter estimates as stochasticity in realizations of trees generated under fixed parameter values. Furthermore, the range of parameter estimates is similar, regardless of the tree construction method. Important qualitative results are extremely consistent across trees. All 100 trees from each of the four sets yield an estimated lifetime shape parameter (k) significantly larger than one (corresponding to a more bell-shaped distribution); that is, the exponential distribution is rejected by the likelihood ratio test, indeed with extremely high significance (p-value $<10^{-15}$ in every case). Furthermore, consistent tendencies appear when the exponential model is used for parameter

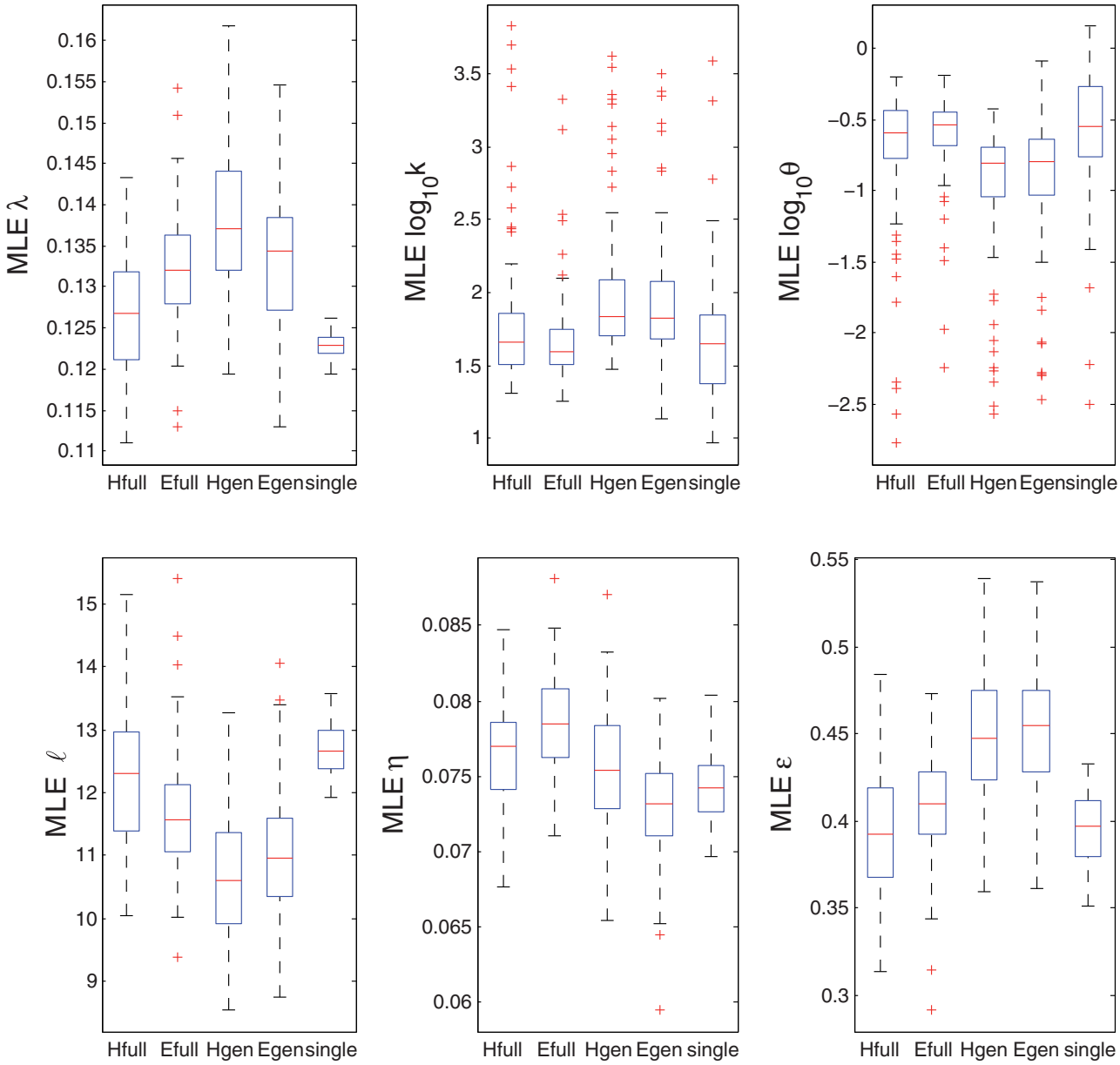

FIGURE 5. Range of parameter estimates across sets of Aves phylogenies. Illustrated is the distribution of MLEs for each parameter (defined in Table 1) across sets of 100 Aves phylogenies. Time units are millions of years The first four columns represent sets drawn from the posterior distribution of trees reconstructed by Jetz et al. (2012) using either a Hackett (H) or Ericson (E) backbone and either including all species (full) or only those with genetic data (gen). The far right column represents a set of trees generated by bootstrap simulations using the maximum likelihood parameters estimated for a single "typical" tree in the 'Hfull' set (see "Methods" section). The box contains the interquartile range with the median marked as a line.

estimation: every single tree yields a point estimate of higher speciation rate (λ) and net diversification rate (η), larger mean lifetime (ℓ), and lower turnover (ϵ) under the exponential model as compared to the gamma model. Quantitative results, specifically the median MLE for each parameter in each set of 100 trees, estimated under both the gamma and exponential models, are reported in Supplementary Table S10. Note that the time scale here is millions of years; when rescaled to average lifetime, parameter estimates are comparable to those used in the simulation study.

Further analysis of one "typical" tree (see "Methods" section) suggests that uncertainty in our model fit to a particular tree is comparable in magnitude to the phylogenetic uncertainty under any particular reconstruction method, and smaller than the phylogenetic uncertainty across reconstruction methods (Fig. 5 and Supplementary Table S11). The estimated lifetime distribution shape parameter, $\hat{k}$, has a large confidence interval, but due to the asymptotic convergence of the gamma distribution for large k, the extremes of the confidence interval

still correspond to reasonably similar, bell-shaped distributions (Supplementary Fig. S2). All 100 trees simulated under the MLEs of the gamma model reject the exponential model with very high significance using the likelihood ratio test (p-value $< 10^{-9}$ in all cases), and the exponential model yields higher $\hat{\lambda}$, $\hat{\eta}$, and $\hat{\ell}$ and lower $\hat{\epsilon}$ than the gamma model.

Despite its significant improvement of fit over the age-independent model, there are clear indications that age-dependent extinction fails to capture all features of the data. LTT plots of reconstructed Aves trees show a qualitatively distinct pattern from bootstrap simulated trees under the best-fitting gamma lifetime distribution model (Supplementary Fig. S3; see also Figure 1c in Jetz et al. 2012). In such a large and heterogeneous taxon, one would indeed expect multiple biological factors to play a role in diversification. An analysis based on clade size of passerine birds suggested that net diversification rate tends to be higher in clades with broader geographic distribution and in tropical as opposed to temperate regions, and that net diversification rate decreases over time, possibly due to saturation of ecological niches (Ricklefs 2006). A previous analysis of the complete Aves phylogenies compared the fit of nine birth–death models, each incorporating at most one dependency in the rates (Jetz et al. 2012). To additionally compare the fit of our age-dependent extinction model, we apply the AIC. Taking the median across our 100 trees we obtain $\mathrm{AIC}_{\mathrm{gam}} - \mathrm{AIC}_{\mathrm{exp}} = -136.0$ for the gamma model relative to the exponential model. That is, consistent with the likelihood ratio test results, we conclude that the gamma model fits better. For comparison, Jetz et al. (2012) found that various models of temporal variation yielded median difference in AIC (ΔAIC) ranging from $+2.1$ to -178.2 relative to the constant-rate birth–death model (equivalent to our exponential model). Our results thus suggest that allowing extinction rate to depend on age, assuming a gamma-distributed lifetime, yields an improvement in model fit comparable to the best previously tested models allowing speciation and/or extinction rate to depend on absolute time. On the other hand, with ΔAIC of -1893.4 relative to the constant-rate model, the clade-shift model tested by Jetz et al. (2012) still provides by far the best fit.

Nightshades (Solanaceae)

A previous phylogenetic analysis of 356 nightshade (Solanaceae) species found that self-compatible (SC) species have significantly higher rates of extinction than self-incompatible (SI) species, which outweigh a higher speciation rate to produce significantly lower (even negative) net diversification (Goldberg et al. 2010; Goldberg and Igić 2012). In light of the hypothesis that extinction rates of selfing plants may increase with species age (Johnson et al. 2011), we revisit the species-level differences between SI and SC with our age-dependent extinction model. For this purpose we separate the SI subtree from multiple SC clades thought to represent independent losses of self-incompatibility (see "Methods" section). We first confirmed that for both SI and SC species, our parameter estimates under the exponential model (Supplementary Tables S12 and S13) are in close agreement with previous work (Goldberg et al. 2010). Small deviations in the SC case may be explained by clades where the transition from SI to SC occurred later in the clade's history, rather than at the earliest split from the SI tree as our analysis assumes.

For the SI species subtree, under the gamma model we estimate a lifetime shape parameter of $\hat{k} = 4.53$, while applying the exponential model shows a tendency toward differences in $\hat{\lambda}$, $\hat{\ell}$, and $\hat{\eta}$ that are consistent with those seen in the simulation study when $k > 1$ (Supplementary Table S12). However, confidence intervals are wide and the exponential model is not rejected by the likelihood ratio test (p-value: 0.24). Only 14% (10/74) of trees simulated under the gamma model MLEs reject the exponential distribution, indicating that power is indeed low for this tree size and parameter set.

For the set of SC species clades, the exponential model is rejected by the likelihood ratio test (p-value of 0.020 or 0.025 for likelihood conditioning on stem or crown age, respectively) in favor of the gamma model with lifetime shape parameter $\hat{k} < 1$, i.e., an over-dispersed lifetime with highest extinction rate when young. Overall, the gamma model gives a clear signal for fast speciation during a lifetime that is on average short but relatively variable (Supplementary Table S13). A slightly negative estimated net diversification rate of SC clades is consistent with previous results (Goldberg et al. 2010) and with the observation that these clades are typically small. In line with our simulation study when $k < 1$ and η is small, the exponential model yields a lower estimate for λ and a higher estimate for ℓ compared to the age-dependent model.

Finally, we test whether SC and SI nightshade species show significantly different lifetime distributions. We find (Supplementary Table S14) that the full model, allowing all parameters to differ between SI and SC, fits significantly better than either a model fixing k (shape parameter) to be the same (likelihood ration test p-values: 0.023 or 0.021 for conditioning on stem or crown age of SC clades, respectively) or a model fixing ℓ (mean lifetime) to be the same (p-values: 0.020 or 0.019 for stem or crown age conditioning, respectively). We can thus support the previous conclusion (Goldberg et al. 2010) that SC species face a shorter mean time to extinction than SI species. (Note that by including transitions to self-compatibility in "extinction" of SI species, as described in the "Methods" section, if anything we over-estimate true extinction of SI species.) We uncover the additional factor that SC species appear to have especially high extinction rate at young age ($\hat{k}_{SC} < 1$) and higher lifetime variance-to-mean ratio than SI species ($\hat{\theta}_{SC} > \hat{\theta}_{SI}$), which could make an additional contribution to the demise of SC clades. Interestingly, in concordance with previous results (Goldberg et al. 2010), we also estimate a higher speciation rate for SC than for SI lineages.

Discussion

Here we have presented a significant methodological advance in inference from phylogenetic data, extending the arsenal of available birth–death-type models to allow the death/extinction rate to depend on the age of a lineage. In a macroevolutionary context, species age has been hypothesized to reflect extinction risk due to its correlation with characteristics such as population size, ecological traits, and nature of biotic interactions. While the role of species age in extinction has previously been investigated using the fossil record (Van Valen 1973; Pearson 1995; Doran et al. 2006), the development of statistical inference methods applicable to phylogenetic trees of extant species (possibly incompletely sampled) opens new avenues for exploring these hypotheses. Our results suggest that including age dependency in extinction rates may significantly improve model fits to real data sets, and that ignoring age dependency can bias parameter estimates.

To fit our age-dependent extinction model to a real data set of interest, the "user" needs a reconstructed phylogenetic tree along with an estimate of the fraction of extant species in the clade that have been sampled. The tree can be derived from any application (though we have focused here on trees of extant species), as long as it is ultrametric, i.e., all sampling occurs effectively at one time point. More precisely, the information about the tree required by our inference method is the stem or crown age, together with a list (in any order) of speciation times leading to the sampled extant species in the tree. Given this tree information, our freely available Matlab code or corresponding functions in the R package TreePar v3.2 (see "Methods" section) can be used to obtain MLEs of model parameters (speciation rate and lifetime distribution parameters) and the maximum likelihood value for use in model comparison.

In this first computational implementation, we have chosen the gamma distribution family to describe species lifetime. This family includes the exponential distribution, facilitating comparison to standard age-independent (constant-rate) models, but allows both mean and variance to be controlled independently. If the lifetime distribution is not exactly gamma, but qualitatively similar, we expect minimal effects on parameter estimates. For instance, we find (Supplementary Fig. S4) that lifetime mean and variance can be recovered with good accuracy assuming a gamma distribution, even when trees are simulated under a Weibull distribution (another two-parameter, unimodal family). Nonetheless, the gamma family may poorly capture other patterns, such as a bimodal lifetime distribution, where extinction risk is elevated for both very young and very old species. While likelihood formulae allow an arbitrary distribution of species lifetime (Lambert 2010), implementation of the current framework under other distribution families is expected to be more computationally intensive than the gamma, and, if described by more parameters, to require larger data sets for precise inference.

Our inference method applies to a given tree with known branching times. However, trees reconstructed from genetic sequence data obviously contain phylogenetic uncertainty, which should be taken into account in parameter estimates. We can get an idea of this impact by running maximum likelihood inference separately on multiple individual trees supported by the data, for instance drawn from the posterior distribution in a Bayesian approach to tree reconstruction (as applied here to Aves data). In the future, the age-dependent extinction model could be incorporated into a Bayesian framework (e.g., BEAST; Drummond and Rambaut 2007) for simultaneous inference of trees themselves and their macroevolutionary parameters.

A simulation study confirmed that our inference method is effective (on sufficiently large trees) in recovering true parameter values when trees are actually generated by a diversification process with constant speciation rate and gamma-distributed time until extinction. For fixed number of sampled species and fixed mean species lifetime, inference is more effective on slower-growing (older) trees. The individual parameters of the lifetime distribution (shape, k, and scale, θ) are the most difficult to infer precisely; however, their estimated values are highly correlated, such that mean lifetime is estimated much more precisely. Furthermore, uncertainty in k is moderated by the approach to a limiting distribution for large k. In practice, this means that large confidence intervals on k and θ estimated from a given tree are to be expected, but this does not preclude a precise estimate of mean lifetime, nor does it necessarily imply low power to detect deviations from a constant-rates model. We expect that estimation of age-dependent extinction/death rate will be more precise on trees that are sampled through time (Lambert et al. 2014), such as species trees supplemented with fossils or phylogenies of quickly evolving viral populations during an epidemic.

We also found that neglecting age dependence biases parameter estimates. Specifically, when younger species are more prone to extinction ($k<1$), applying an age-independent (exponential; $k=1$) model results in underestimation of speciation rate and net diversification rate, and vice versa when older species are more prone to extinction ($k>1$). Estimated mean lifetime is also biased according to a more complex pattern. Effects on estimated turnover were ambiguous from the simulation study, but interestingly, the exponential model consistently returned lower turnover than the gamma model in the Aves data. The possible biases in estimated mean species lifetime and turnover are particularly noteworthy, as it has been observed that trees derived from sequence data often result in inference of a lower extinction rate or turnover than that estimated from the fossil record (discussed by Morlon et al. 2011 and references therein). Depending on the typical parameter regime applicable to real species clades, neglect of age-dependent extinction in previous analyses could be one factor helping to resolve this discrepancy.

We illustrated the application of our method to real data with two test cases. We emphasize that we do not wish to put forward definitive biological explanations, but simply to test the performance of our method while raising interesting new hypotheses. First, in the Aves phylogeny, we found that a model where extinction rate increases with age provided a significantly better fit than age-independence. Parameter estimates under an age-independent model showed deviations from the best fit consistent with the biases seen in the simulation study. These signals were extremely robust to phylogenetic uncertainty, holding across all 100 reconstructed trees drawn from the posterior distribution under each of four reconstruction methods (Jetz et al. 2012). Phylogenetic uncertainty furthermore appears to be a larger source of uncertainty in parameter estimates than the fit of our model to a typical individual tree in this data set. Despite the significant improvement in fit gained by adding age-dependent extinction, there were signs that this model still did not provide a satisfactory description of the data (Supplementary Fig. S3). This result is not surprising, given the multitude of biological factors likely to play into the diversification of such a large and heterogeneous taxon and supported by previous studies (Ricklefs 2006; Jetz et al. 2012). A comparison of models modifying single factors was nonetheless enlightening: while a clade-shift model provides the best fit (Jetz et al. 2012), we found that age-dependent extinction yielded similar improvements over a constant-rates model as gained by time-dependent rates (Jetz et al. 2012). This suggests that age dependence should not be overlooked in future considerations of multi-factorial models.

Interestingly, it has been more generally observed that trees derived from real species data tend to be less balanced than those generated under a broad class of models, including constant speciation and extinction rates, diversity- or time-dependent speciation and/or extinction, or age-dependent extinction, that give rise to a uniform distribution on ranked oriented trees (Blum and François 2006; Lambert and Stadler 2013; Stadler 2013b). The Aves data considered here are no exception; comparing Colless statistics (Colless 1982) indicates that reconstructed Aves trees are less balanced than trees simulated under a constant-rates model (results not shown), which cannot be explained by any of the aforementioned models. Intriguingly, simulations have indicated that an age-dependent speciation rate can produce realistic levels of imbalance (Hagen et al. 2015). However, a statistical inference method accounting for age-dependent speciation is currently lacking, as trees generated under such a model can no longer be represented by a coalescent point process, and thus the likelihood cannot be derived with previously used mathematical approaches (Lambert and Stadler 2013).

Our second data analysis, involving a phylogeny of self-incompatible (SI) and self-compatible (SC) nightshades, extended previous results concerning the species-level disadvantage of self-compatibility. Our results suggested that the lifetime of SC species not only has a significantly lower mean than that of SI species, in agreement with previous results (Goldberg et al. 2010; Goldberg and Igić 2012), but also that it has a higher variance-to-mean ratio. Among SC species, the exponential lifetime distribution (constant rates) model was rejected in favor of a gamma lifetime distribution model where extinction rate decreases with age. This result is surprising in light of the hypothesis that selfing or asexual species face increasing extinction risk with age due to the accumulation of deleterious mutations (Johnson et al. 2011). One possible explanation is that the accumulation of mutations in a parent species, though not hindering speciation, continues to have deleterious effects on daughter species; that is, age of the entire clade and not only of the individual species is relevant. Our analysis may also be limited by the small size of the data set and our method of considering each SC clade independently; thus, these initial results should be interpreted with caution. We are not aware of any existing larger reconstructed trees containing both asexual/selfing and sexual/nonselfing species that are suitable for our analysis. Nonetheless, our framework is ready to be used as larger data sets become available.

Our focus here was on developing methodology to infer macroevolutionary rates under an age-dependent extinction model, and testing its accuracy when phylogenies are actually generated under this model. Future work should test the effects of model mis-specification, i.e., whether spurious signals of age dependence arise, or conversely true signals are obscured, when macroevolutionary rates depend on other factors. For instance, the importance of disentangling dependence of extinction rate on real time from dependence on species age has been emphasized in the paleontological literature (Pearson 1995; Doran et al. 2006). An extension of our model to allow rates of speciation and extinction additionally to vary with time would be mathematically fairly straightforward (Lambert and Stadler 2013). While elucidating the effects of model mis-specification is beyond the scope of the present study, we emphasize that our work represents an important addition to the set of models available for inference. Other factors potentially influencing speciation and extinction rates have also individually been incorporated into models generating phylogenetic trees (reviewed in Pyron and Burbrink 2013; Stadler 2013b; Morlon 2014). An important advantage of using a likelihood-based inference method is that alternative models can be compared side-by-side based on their likelihood given the data, as done for example with the Aves data. Assessing the best explanation for the data would, however, ideally involve integration of all these proposed factors in one model. We thus echo previous calls for the development of an integrated inference framework (Pyron and Burbrink 2013; Stadler 2013b; Morlon 2014), adding that age dependence now can and should be included in such a synthesis. With increasing model complexity, larger data sets will be required for inference. The growing availability of large reconstructed phylogenies thus provides an

unparalleled opportunity to gain insight into the biological factors shaping macroevolutionary dynamics.

Supplementary Materials

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.7894h.

Funding

This work was supported by the ETH Zürich (H.K.A., T.S.); the Swiss National Science Foundation (grant PZ00P3_136820 to T.S.); CNRS and Collège de France through the Center for Interdisciplinary Research in Biology (A.L.); and the French national research agency (ANR, grant MANEGE 'Modèles Aléatoires en Écologie, Génétique et Évolution' 09-BLAN-0215 to A.L.).

Acknowledgments

We thank Jeff Joy and Sally Otto for assistance obtaining and interpreting appropriate data sets, Louis Du Plessis for technical assistance in making our code distributable, and members of the Institute for Integrative Biology at ETH Zürich for helpful comments on this work. We also thank the Editor-in-Chief, Frank Anderson; the Associate Editor, Edward Susko; and several anonymous reviewers, for excellent suggestions that helped us to improve this manuscript.

References

Akaike H. 1974. A new look at the statistical model identification. IEEE Trans. Automat. Contr. AC-19:716–723.
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.
Colless D.H. 1982. Phylogenetics: the theory and practice of phylogenetic systematics. Syst. Zool. 31:100–104.
Doran N. A., Arnold A.J., Parker W.C., Huffer F.W. 2006. Is extinction age dependent? Palaios 21:571–579.
Drummond A., Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7:214.
Goldberg E.E., Igić B. 2012. Tempo and mode in plant breeding system evolution. Evolution 66:3701–3709.
Goldberg E.E., Kohn J.R., Lande R., Robertson K.A., Smith S.A., Igić B. 2010. Species selection maintains self-incompatibility. Science 330:493–495.
Hagen O., Hartmann K., Steel M., Stadler T. 2015. Age-dependent speciation can explain the shape of empirical phylogenies. Syst. Biol. 64:432–440.
Hagen O., Stadler T. 2013. TreeSimGM: simulating phylogenetic trees under a general model. Available from: URL http://cran.r-project.org/web/packages/TreeSimGM/index.html.
Hartmann K., Wong D., Stadler T. 2010. Sampling trees from evolutionary models. Syst. Biol. 59:465–476.
Harvey P.H., May R.M., Nee S. 1994. Phylogenies without fossils. Evolution 48:523–529.
Jetz W., Thomas G.H., Joy J.B., Hartmann K., Mooers A.O. 2012. The global diversity of birds in space and time. Nature 491:444–448.
Johnson M.T.J., FitzJohn R.G., Smith S.D., Rausher M.D., Otto S.P. 2011. Loss of sexual recombination and segregation is associated with increased diversification in evening primroses. Evolution 65: 3230–3240.
Lambert A., Alexander H.K., Stadler T. 2014. Phylogenetic analysis accounting for age-dependent death and sampling with applications to epidemics. J. Theor. Biol. 352:60–70.
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. 2010. The contour of splitting trees is a Lévy process. Ann. Probab. 38:348–395.
Liow L.H., Van Valen L., Stenseth N.C. 2011. Red Queen: from populations to taxa and communities. Trends Ecol. Evol. 26:349–358.
McCune A.R. 1982. On the fallacy of constant extinction rates. Evolution 36:610–614.
Moen D., Morlon H. 2014. Why does diversification slow down? Trends Ecol. Evol. 29:190–197.
Morlon H., Parsons T.L., Plotkin, J. 2011. Reconciling molecular phylogenies with the fossil record. Proc. Natl Acad. Sci. USA 108:16327–16332.
Morlon H. 2014. Phylogenetic approaches for studying diversification. Ecol. Lett. 17:508–525.
Muller H.J. 1964. The relation of recombination to mutational advance. Mutat. Res. 1:2–9.
Nee S., Holmes E.C., May R.M., Harvey P.H. 1994a. Extinction rates can be estimated from molecular phylogenies. Philos. Trans. R. Soc. Lond. B Biol. Sci. 344:77–82.
Nee S., May R.M., Harvey P.H. 1994b. The reconstructed evolutionary process. Philos. Trans. R. Soc. Lond. B Biol. Sci. 344:305–311.
Pearson P.N. 1995. Investigating age-dependency of species extinction rates using dynamic survivorship analysis. Histor. Biol. 10: 119–136.
Popovic L. 2004. Asymptotic genealogy of a critical branching process. Ann. Appl. Probab. 14:2120–2148.
Pyron R.A., Burbrink F.T. 2013. Phylogenetic estimates of speciation and extinction rates for testing ecological and evolutionary hypotheses. Trends Ecol. Evol. 28:729–736.
Raup D.M. 1975. Taxonomic survivorship curves and Van Valen's Law. Paleobiology 1:82–96.
Rice S.H. 2004. Evolutionary theory: mathematical and conceptual foundations. Sunderland, MA: Sinauer Associates, Inc.
Ricklefs R.E. 2006. Global variation in the diversification rate of passerine birds. Ecology 87:2468–2478.
Stadler T. 2009. On incomplete sampling under birth-death models and connections to the sampling-based coalescent. J. Theor. Biol. 261:58–66.
Stadler T. 2011. Mammalian phylogeny reveals recent diversification rate shifts. Proc. Natl Acad. Sci. 108:6187–6192.
Stadler T. 2013a. How can we improve accuracy of macroevolutionary rate estimates? Syst. Biol. 62:321–329.
Stadler T. 2013b. Recovering speciation and extinction dynamics based on phylogenies. J. Evol. Biol. 26:1203–1219.
Surya, B. A. 2008. Evaluating scale functions of spectrally negative Lévy processes. J. Appl. Prob. 45:135–149.
Thompson E.A. 1975. Human evolutionary trees. Cambridge, UK: Cambridge University Press.
Van Valen L. 1973. A new evolutionary law. Evol. Theory 1:1–30.