Heritability of Extinction Rates Links Diversification Patterns in Molecular Phylogenies and Fossils

DANIEL L. RABOSKY1,2,3,*

1Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA; 2Fuller Evolutionary Biology Program, Cornell Laboratory of Ornithology, 159 Sapsucker Woods Road, Ithaca, NY 14850, USA; and 3Department of Integrative Biology, University of California, Berkeley, CA 94720, USA;

*Correspondence to be sent to: Department of Integrative Biology, University of California, Berkeley, CA 94720 USA; E-mail: drabosky@berkeley.edu.

Abstract.—Time-calibrated molecular phylogenies provide a valuable window into the tempo and mode of species diversification, especially for the large number of groups that lack adequate fossil records. Molecular phylogenetic data frequently suggest an initial “explosive speciation” phase, leading to widespread speculation that ecological niche-filling processes might govern the dynamics of species diversification during evolutionary radiations. However, these patterns are difficult to reconcile with the fossil record. The fossil record strongly suggests that extinction rates have been high relative to speciation rates, but such elevated background extinction should erase the signal of early, rapid speciation from molecular phylogenies. For this reason, extinction rates in molecular phylogenies are frequently estimated as zero under the widely used birth–death model. Here, I construct a simple model that combines phylogenetically patterned extinction with pulsed turnover dynamics and constant diversity through time. Using approximate Bayesian methods, I show that heritable extinction can easily explain the phenomenon of explosive early diversification, even when net diversification rates are equal to zero. Several assumptions of the model are more consistent with both the fossil record and neontological data than the standard birth–death model and it may thus represent a viable alternative interpretation of phylogenetic diversification patterns. These results suggest that variation in the absolute rate of lineage turnover through time, in conjunction with phylogenetically nonrandom extinction, may underlie the apparent diversity-dependent speciation observed in molecular phylogenies. [Adaptive radiation; approximate Bayesian computation; birth-death process; macroevolution; speciation.]

Molecular phylogenetic studies can potentially complement fossil-based analyses of evolutionary radiations because time-calibrated molecular phylogenies provide information about the timing of speciation events in groups for which minimal fossil data are available. The ease of obtaining DNA sequence data has led to a rapid increase in the availability of time-calibrated phylogenetic trees, and many studies have used these data to infer patterns of species diversification through time (e.g., Nee et al. 1992; Barraclough and Vogler 2002; Harmon et al. 2003). Perhaps the most surprising finding from these studies is that, many, if not most, suggest an initial burst of lineage accumulation early in the history of evolutionary radiations (Ruber and Zardoya 2005; McPeek 2008; Phillimore and Price 2008). This pattern has frequently been interpreted as diversity-dependent speciation mediated by ecological opportunity, whereby rapid diversification is facilitated by an abundance of resources and paucity of competing species (Nee et al. 1992; Phillimore and Price 2008; Price 2008; Rabosky and Lovette 2008a). Under this model, speciation rates are high initially but subsequently decline in conjunction with the rise of species diversity within a particular ecological or biogeographic theater.

However, extinction shapes phylogenetic trees by “pulling” nodes close to the present (Fig. 1) and thus obscures the signal of early, rapid diversification (Rabosky and Lovette 2008b,c). This occurs because high background extinction changes the age structure of lineages that survive to the present to be observed. If the speciation rate $\lambda$ is high relative to the extinction rate $\mu$, then many lineages will have been around for a comparatively long period of time. However, if $\mu$ approaches $\lambda$, the turnover rate of lineages will be much higher, and most lineages will have been in existence for a relatively brief period of time. A corollary of this is that early rapid diversification is more difficult to detect in molecular phylogenies if extinction rates have been high (Rabosky and Lovette 2008b). This underlies the observation that extinction rates estimated from molecular phylogenies are typically close to zero (Nee 2006; Weir 2006; Purvis 2008; Rabosky and Lovette 2008b). Indeed, under the widely used birth–death model (Nee et al. 1994), phylogenies showing an excess of early speciation events almost invariably appear to have trivial extinction (Purvis 2008; Rabosky and Lovette 2008b).

These results directly conflict with the fossil record (Bokma 2009), which overwhelmingly suggests that extinction rates have been high relative to speciation rates (Ricklefs 2007, 2009), in every group for which appropriate data are available (Stanley 1979; Gilinsky 1994; Alroy 1996, 2008, 2009). Studies of paleodiversity through time lend further support to this idea because a consistent and substantial excess of speciation relative to extinction would lead to exponentially increasing diversity for many groups. Yet, this is clearly not the case: studies that have investigated sampling-standardized diversity through time have found no evidence for recent exponential increases in diversity (Alroy et al. 2008; Rabosky and Sorhannus 2009), and many groups appear to be characterized by approximate constancy of diversity through time (Alroy 2000; Jaramillo et al. 2006; Ricklefs 2007, 2009). Moreover, clade age is unrelated to species richness in many higher taxa, a result that implies ecological limits on clade diversity through time (Ricklefs 2006a; Ricklefs et al. 2007; Rabosky 2009a, 2009b).

A more subtle problem with putative “diversity-dependent” interpretations of these patterns is that,
for most systems, there is little historical evidence for conditions of high ecological opportunity and empty ecological space that could have triggered rapid speciation. It is certainly the case that diversification on isolated islands may be triggered by ecological opportunity (Schluter 2000) and that conditions favoring explosive ecological and species diversification might follow the profound biotic upheaval of mass extinctions (Foote 1996; Sepkoski 1998). But it is much more difficult to see how so many radiations of taxa in ecologically complex continental systems might have experienced much greater niche availability early in their history. There is little evidence for massive clearing of ecological space by extinction over timescales consistent with molecular phylogenetic analyses of diversification (Ricklefs 2007; Alroy 2009). Even if Pleistocene climate oscillations resulted in recent extinction pulses in the temperate zone, many groups from tropical regions that would have been minimally affected by these events have often undergone explosive early diversification (Weir 2006).

Although biases associated with taxon sampling (Pybus and Harvey 2000; Purvis et al. 2008) and tree reconstruction method (Revell et al. 2005) might partially account for apparent conflict between molecular and fossil data, it is also possible that the conflict is an artifact of the standard birth–death model used to draw inferences about species diversification rates from molecular data. Here, I construct a simple alternative to the birth–death model that assumes 1) constant diversity through time, 2) heritable extinction tendencies, and 3) a balanced speciation–extinction process with zero net diversification. The effects of heritable extinction rates on lineage accumulation curves have not been studied, although a number of studies have documented both heritability of diversification rates (Savolainen et al. 2002; Davies et al. 2004) and phylogenetic clustering of extinction and extinction risk (Purvis et al. 2005; Vamosi and Wilson 2008). In the new model, each extinction event results in an immediate replacement, as in the neutral community model of Hubbell (2001).

I apply the model to a recent phylogeny of Dendroica wood warblers, a well-studied group of North American songbirds that appear to have undergone a diversity-dependent decline in the rate of speciation with trivial background extinction (Rabosky and Lovette 2008a, 2008b). Because no likelihood function can presently be specified for the model, I use approximate Bayesian computation to infer posterior distributions of parameters and to evaluate absolute model fit. I show that a simple turnover pulse with phylogenetically patterned extinction can account for speciation–extinction dynamics observed in molecular phylogenies, with ecological implications that contrast sharply with inferences based on the birth–death model.
Heritable Extinction with Pulsed Turnover Model

The birth–death model as typically used for inference on species diversification rates assumes per-lineage rates of speciation ($\lambda$) and extinction ($\mu$) in a clade that grows from an initial diversity of $n = 2$ lineages (Nee et al. 1994). Here, I develop an alternative model specifying heritable extinction with pulsed turnover (HEPT) that does not allow diversity to increase or decrease through time. The HEPT model is closely related to the Moran (1962) process, which is central to many aspects of coalescence theory in population genetics (Hudson 1991; Ewens 2004). Under the simple Moran process, lineages may go extinct, but clade size remains constant. When one lineage goes extinct, another lineage immediately undergoes a speciation event, and there is thus no change in diversity. In the context of speciation–extinction dynamics, the model assumes that diversity is regulated at an equilibrium value through time, perhaps due to saturation of ecological niches, and there is no lag time for species replacement after extinction. This model has not been widely incorporated into macroevolutionary studies (Nee 2001, 2006), although Hey (1992) assessed the fit of the simple Moran process to lineage accumulation curves and Ricklefs (2006b, 2007) considered the implications of the process for species turnover within metacommunities and biogeographic regions.

The model I consider contains 5 parameters ($x_c$, $\kappa$, $r_1$, $r_2$, $\omega$) and assumes that diversity is regulated at a constant level, that rates of species turnover vary through time, and that extinction is “heritable” or phylogenetically patterned. At some time ($x_c + \kappa$) before present, per-lineage speciation and extinction rates $r_1$ shift to a new rate $r_2$ (the “turnover pulse” phase), which is maintained for $2\kappa$ time units, until time $x_c - \kappa$, whereupon $r_2$ shifts back to $r_1$ (Fig. 2). The parameter $x_c$ thus equals the elapsed time between the present and the midpoint of the period of high lineage turnover. A crucial feature of this model is the heritability of extinction, which is specified by $\omega$. If a particular species goes extinct, the next extinction event will eliminate a randomly drawn member of its sister taxon with probability $\omega$. Otherwise, a taxon goes extinct at random from the full set of $N$ taxa with probability $1 - \omega$. When $\omega = 0$, the HEPT model reduces to the simple Moran process, at least when $r_1 = r_2$: all lineages have equivalent probabilities of extinction, and turnover rates are constant through time.

Fitting the Model to Data with Approximate Bayesian Computation

I used approximate Markov chain Monte Carlo (MCMC) (Marjoram et al. 2003; Plagnol and Tavare 2004) to fit the HEPT model to a complete, species-level phylogeny for the *Dendroica* subgroup of North American wood warblers (Rabosky and Lovette 2008a). The tree contains all 25 species in the *Dendroica* group that occur within North America and was constructed under a Bayesian relaxed clock approach from approximately 9500 bp of combined nuclear and mitochondrial DNA. Analyses were conducted in BEAST (Drummond et al. 2006), recognizing 4 data partitions; the final tree set was pooled from a total of 341 million sampled generations across 11 independent MCMC runs. Several *Dendroica* species from Caribbean islands are not included in this tree because it seems unlikely that they would have been involved in ecological interactions with continental wood warbler species.

Rabosky and Lovette (2008a) reported that *Dendroica* has undergone a severe decline in the rate of diversification through time, consistent with several previous studies that have also reported declining diversification through time in this group (Lovette and Bermingham 1999; Phillimore and Price 2008; Rabosky and Lovette 2008b). I scaled the basal divergence of the *Dendroica* tree to 5 million years before present (Lovette and Bermingham 1999). I conducted all analyses described below on the maximum clade credibility tree from the Bayesian analysis because all trees in the credible set show similar patterns of early rapid diversification (e.g., figure 2b in Rabosky and Lovette 2008a); results presented here should be robust to phylogenetic uncertainty. This tree is available as supplementary material at http://www.sysbio.oxfordjournals.org.
To obtain independent samples from the joint posterior distribution of HEPT model parameters, I used an approximate MCMC approach (Marjoram et al. 2003; Plagnol and Tavare 2004) that does not require calculation of likelihoods (Marjoram et al. 2003). The approach requires only that we have one or more sufficient summary statistics that can be used to compare the observed data with data simulated under the model given a set of parameters. The approach is essentially a variant of rejection sampling: parameters are sampled from proposal distributions, and data are simulated under the HEPT model with the sampled parameters. We then compute summary statistics that describe the match between the observed data and the simulated data. Parameter sets that can reproduce phylogenetic patterns consistent with the observed data are retained, whereas those that cannot are rejected.

In this case, the depth (or crown age) of simulated phylogenetic trees ($L$) and the $\gamma$-statistic (Pybus and Harvey 2000) provide convenient summary statistics that describe the match between the observed wood warbler tree and simulated trees. The $\gamma$-statistic is a measure of the extent to which speciation events in a molecular phylogeny are clustered at the base or tips of a tree; $\gamma < 0$ implies clustering near the base of the tree and decelerating speciation through time (Pybus and Harvey 2000; Rabosky and Lovette 2008b). Although alternative summary statistics are certainly possible, $\gamma$ is a simple metric that has been widely adopted by researchers in this field (e.g., McPeek 2008; Phillimore and Price 2008). Moreover, this metric captures precisely the pattern we wish to explain, namely, the rapid burst of lineage accumulation frequently observed at the base of phylogenetic trees.

The rejection criterion in an approximate Bayesian framework depends on the match between the observed data ($D$) and the simulated data ($D'$) as determined by the difference in summary statistics. We thus generate a candidate set of parameters $\theta_0$ and simulate data $D'$ under the model. The parameters $\theta_0$ are retained if the distance $\rho(D', D)$ is less than some threshold $\epsilon$. Parameter proposals could only be accepted if simulated trees approximately matched the observed tree in 2 summary statistics: the crown age or tree depth (5 million years), and the value of the $\gamma$ statistic ($-3.48$).

To generate phylogenies under the HEPT model, a candidate tree was first generated under a pure birth model with $\lambda = 0.1$. This tree was then evolved under the HEPT model for 20 million years. The rejection criteria (discussed below) ensured that simulated trees retaining any signal of this initial pure birth tree were not accepted; increasing the simulation duration to 100 million years did not change the results, but dramatically increased computational time.

Initial parameters for each independent chain were sampled randomly from prior distributions $\pi(\cdot)$. Priors were uniform for all parameters ($\gamma_1$ and $\gamma_2$ (0, 30); $\kappa$ (0, 5), and $\omega$ (0, 1]) except $\kappa$, which received a normal (mean = 5, standard deviation = 2) prior. The MCMC sampler then iterated through the following steps until a target number of generations had been reached (Marjoram et al., 2003):

S1: If at state $\theta$, propose a move to state $\theta'$ according to a transition kernel $q(\theta \rightarrow \theta')$.
S2: Generate data $D'$ under the HEPT model with parameters $\theta'$.
S3: Compute $\rho(D', D)$. If $\rho(D', D) \leq \epsilon$, go to S4. Otherwise, reject the proposal and return to S1.
S4: Compute an approximation of the Hastings ratio as

$$h = \min \left( 1, \frac{\pi(\theta')q(\theta \rightarrow \theta)}{\pi(\theta)q(\theta' \rightarrow \theta')} \right)$$

where $\pi(\theta)/\pi(\theta')$ is the prior ratio and $q(\theta' \rightarrow \theta)/q(\theta \rightarrow \theta')$ is the ratio of transition kernels between states.
S5: Accept the proposal $\theta'$ with probability $h$, otherwise stay at $\theta$. Return to S1.

Here, the ratio of transition kernels is always equal to one. The criterion $\rho(D', D)$ is simply the absolute difference in $\gamma$ and $L$ between the simulated and observed data, scaled by the absolute value of the observed $\gamma$ and $L$. Thus, for gamma, this was simply $|\gamma_{SIM} - \gamma_{OBS}|/|\gamma_{OBS}|$. After experimenting with a range of $\epsilon$ between 0.05 and 0.5, I found that $\epsilon = 0.2$ performed well for both $\gamma$ and $L$. A proposed state could thus only be accepted if the scaled differences in both $\gamma$ and $L$ were less than $\epsilon$.

I ran 10 chains for 100,000 generations each, sampling parameters every 100 generations.

I repeated these analyses assuming fixed $\omega$ values. Multiple chains were run for scenarios corresponding to low ($\omega = 0.1$), moderate ($\omega = 0.5$), and high ($\omega = 1.0$) extinction heritability. Together with the simulations in which $\omega$ was treated as a free parameter, this gave a total of 4 evolutionary scenarios analyzed with approximate MCMC. All simulation and analyses were conducted in the R programming environment, with some code borrowed from the Ape (Paradis et al. 2004), Geiger (Harmon et al. 2008), and Laser (Rabosky 2006) packages.

Convergence Analysis and Diagnostics

Each MCMC chain was initiated with a randomly drawn set of parameters that satisfied $\rho(D', D) \leq \epsilon$. There was thus no need to include a burn-in phase as is typical with standard MCMC because the chain was already sampling from the target distribution. To assess convergence, I calculated the scale reduction factor of Gelman and Rubin (1992). This statistic essentially compares within- and between-chain variances for each parameter and estimates the factor by which the scale parameter of the estimated posterior densities for each parameter would shrink were the chains to run to infinity. There are no absolute guidelines for interpreting this statistic, but values less than 1.05 are typically taken to reflect convergence (Gelman et al. 2003). Prior to computing the Gelman–Rubin statistic, data were transformed to improve normality. In addition, I estimated the effective sample size, adjusted for
autocorrelation, for both summary statistics and parameters sampled with approximate MCMC. This is computed by using autoregressive models to estimate the mean number of generations until the autocorrelation drops to zero; this quantity is then used to approximate the number of effectively independent draws obtained from the posterior distribution. Effective sample sizes were computed using the MCMC diagnostic package “coda” for R.

**Lineage Accumulation Curves**

I assessed whether lineage accumulation curves generated under the HEPT model showed evidence for high or low extinction when analyzed with several variants of the birth–death model. In this case, the simulation model specifies \( \lambda = \mu \), so a failure to recover high extinction relative to speciation implies a failure of the birth–death model to reconstruct a basic feature of the underlying evolutionary process. For each value of \( \omega \), mean log-lineage accumulation curves were tabulated from the sample of simulated phylogenies obtained during MCMC. These phylogenies necessarily satisfied \( \rho(D', D) \leq 0.2 \) for both \( \gamma \) and \( L \). The first model fitted to the mean lineage accumulation curves was simply a constant-rate birth–death process (Nee et al. 1994). The second was the BOTHVAR model from Rabosky and Lovette (2008b) that allows simultaneous, independent changes in both speciation and extinction rates, such that the net diversification is modeled as

\[
\rho(t) = \lambda(t) - \mu(t),
\]

where

\[
\lambda(t) = \lambda_0 \exp(-kt)
\]

and

\[
\mu(t) = \mu_0 (1 - \exp(-kt)).
\]

I estimated the time-integrated extinction fraction \( \mu/\lambda \) as

\[
\frac{\mu}{\lambda} = \frac{\int \mu(t) \, dt}{\int \lambda(t) \, dt}
\]

integrated from the time of the basal divergence to the present. The final model specified exponentially declining speciation and extinction rates, where the extinction rate is a constant fraction \( \upsilon \) of the speciation rate, leading to

\[
\lambda(t) = \lambda_0 \exp(-kt)
\]

and

\[
\mu(t) = \upsilon \lambda(t).
\]

**Posterior Predictive Simulations**

The preceding analyses tell us little about how well the HEPT model fits the wood warbler data. To determine whether the model is capable of recovering the major features of wood warbler diversification, I assessed absolute model fit by simulating phylogenies under parameters sampled from the joint posterior distribution of \( \theta \) conditional on several values of \( \omega \). For each simulated data set, I then calculated the summary statistics \( \gamma \) and \( L \), which indicate 1) whether the simulated data set shows evidence for an apparent slowdown in the rate of diversification through time, and 2) whether the simulated tree is of the same age as the wood warbler radiation. Here, I was simply using parametric simulation to determine the adequacy of the HEPT model, given the joint posterior distribution of parameters \( f(\theta|D, \omega, M) \). Formally, we are interested in the distribution of one or more summary statistics \( Z' \) given the summary statistic \( Z \) observed for the real data,

\[
\Pr(Z'|Z, \omega, M) = \int \Pr(Z'|\theta, M, \omega) \Pr(\theta|D, M, \omega) \, d\theta.
\]

In practice, parameters were sampled at random and with replacement from the posterior distribution of parameters \( f(\theta|D, \omega, M) \). Each random draw consisted of selecting \( x_c, \kappa, r_1, r_2 \) from a single generation of the pooled MCMC chains, thus ensuring that parameters were sampled in proportion to their joint posterior probability. A single simulation of the HEPT model was performed for each sampled set of parameters, and the distribution of \( Z' \) was tabulated from 50,000 such random draws per \( \omega \) model. I evaluated absolute model fit under scenarios of high (\( \omega = 1.0 \)), moderate (\( \omega = 0.5 \)), and low (\( \omega = 0.1 \)) extinction heritability. I found the posterior distributions \( f(\theta|D, \omega, M) \) of \( x_c, \kappa, r_1, \) and \( r_2 \) conditional on these values of \( \omega \) by running 5 MCMC chains per \( \omega \).

**RESULTS**

**HEPT Model**

Analysis of 10 independent MCMC chains (Table S1) with the Gelman–Rubin statistic (Gelman et al. 2003) suggested that all chains converged on the same target distribution (Table S2). Mean parameter values and standard deviations for each chain are given in Table S3 (\( \omega \) variable model) and Table S4 (\( \omega = 0.1, 0.5, \) and 1.0 models); values are generally similar among independent runs. Effective sample sizes for summary statistics and parameters further indicated that a large number of effectively independent draws from the posterior distribution were obtained for all \( \omega \) models considered (Table S5). Figure 3 shows posterior distributions for \( \omega, \kappa, \) and rate parameters inferred using approximate MCMC. The posterior distribution of \( \omega \) suggests that patterns of wood warbler diversification through time are most consistent with phylogenetically structured extinction under the HEPT model (Fig. 3a). The distribution of \( \kappa \) and the ratio of turnover rates (\( r_2/r_1 \)) implies that a broad range of turnover conditions, both in terms of rates and duration, may also be consistent with the observed data (Fig. 3b,c). However, there is a strong negative relationship between these parameters: turnover pulses of short duration require much greater turnover rates, and viceversa (Fig. 3d).

Per-lineage estimates of speciation and extinction were summarized from the results for each \( \omega \) category.
(\(\omega\) variable; \(\omega = 1.0\); \(\omega = 0.5\); \(\omega = 0.1\)); median values of each posterior distribution, as well as 2.5% and 97.5% quantiles, are shown in Table 1. The magnitude of the turnover pulse (\(r_2\)) is negatively correlated with \(\omega\). When extinction events are not phylogenetically clustered (\(\omega = 0.1\)), increased speciation and extinction rates are required to generate phylogenies consistent with the observed data.

**Lineage Accumulation Curves**

Visual comparison of mean log-lineage accumulation curves obtained for different \(\omega\) values clearly indicate that high extinction heritability provides the best match to the observed pattern of lineage accumulation through time for the wood warbler data (Fig. 4). However, even low extinction heritability yielded a “concave down” lineage accumulation curve consistent with declining diversification through time. For all 3 variants of the birth–death model, extinction was inferred to be a trivial fraction of the speciation rate (Table 2), in spite of the fact that the simulation model specified exact equivalence between these parameters.

**Posterior Predictive Simulations**

When extinction shows high phylogenetic structure (\(\omega = 1.0\)), posterior predictive distributions of both \(\gamma\) and \(L\) are centered on the observed values for the wood warbler tree (Fig. 5a,b; arrows indicate observed values), indicating that this model fits the data well. The model does not perform as well with intermediate extinction heritability (\(\omega = 0.5\)) and appears to perform poorly with low extinction heritability. For \(\omega = 0.1\) and 0.5, both \(\gamma\) and \(L\) show bimodal distributions (Fig. 5a, b), and these values are correlated (Fig. 5c). Trees that are young (3–8 Ma) show lineage accumulation patterns consistent with declining diversification rates (\(\gamma < 0\)). However, trees that are far older than the wood warbler tree show \(\gamma > 0\), which is inconsistent with temporally declining diversification.

This bimodality in tree depth and \(\gamma\) is due solely to the fact that some trees have a long waiting time between the initial bifurcation in the phylogenies and the next speciation event in the reconstructed tree. This was immediately obvious from visual inspection of both simulated trees and corresponding lineage-through-time plots (Fig. 6). To demonstrate this, I computed \(\gamma\) for each tree from the \(\omega = 0.1\) posterior predictive simulations but ignored the basal divergence in the tree (Fig. 6). When this basal bifurcation is ignored, the distribution of \(\gamma\) strongly suggests temporally declining diversification of similar magnitude to that observed in wood warblers (Fig. 5d), even though these trees were
simulated under $\omega = 0.1$. Thus, even limited diversity models with phylogenetically unstructured extinction can generate lineage-through-time patterns consistent with temporally declining or diversity-dependent diversification. However, this pattern is more readily apparent with heritable extinction, which tends to eliminate the signal of diversification that occurred prior to the turnover pulse. This leaves the impression that a single clade has undergone rapid diversification followed by a temporal decline in rates.

**Discussion**

These results have broad implications for the interpretation of species diversification patterns as inferred from molecular phylogenies. I demonstrated that a simple limited diversity model, with equivalent speciation and extinction rates and zero net diversification through time, can generate patterns of lineage accumulation in molecular phylogenies that suggest temporally declining diversity-dependent speciation rates. This apparent slowdown in the rate of speciation is driven solely by variation in the rate of lineage turnover through time in conjunction with phylogenetically patterned extinction. Such a pulsed turnover model can account for patterns of lineage accumulation observed in North American wood warblers, a group believed to have undergone explosive speciation early in its history (Rabosky and Lovette 2008a, 2008b). Posterior predictive simulations under the HEPT model consistently recovered major features of wood warbler diversification, including the rapid accumulation of lineages during the early stage of the radiation.

As with any study that incorporates prior information and MCMC sampling, 2 major questions include 1) whether results are robust to potential convergence issues and 2) the sensitivity of results to the choice of priors. I considered 3 pieces of information when assessing convergence: the behavior of the Gelman–Rubin convergence diagnostic, the means of parameter values and summary statistics among independent chains, and the autocorrelation-adjusted effective size of parameters sampled with MCMC. All these indicators suggest that convergence has occurred (Tables S2–S5). However, the central result in this paper is robust to potential convergence problems. Failure to achieve convergence should yield parameter estimates that perform poorly at predictive simulation relative to samples from the true posterior distribution. Hence, it is rather unlikely that convergence problems underlie the ability of the HEPT model to recover major features of the wood warbler lineage accumulation curve as demonstrated through posterior predictive simulation (Fig. 5). This likewise applies to choice of priors: if priors were poorly specified or exerted undue influence on final parameter estimates, this should have led to worse performance of posterior predictive simulation.

The HEPT model is implicitly ecological, in that there are limits on clade diversity and that extinction events are immediately followed by speciation events. However, this differs fundamentally from the ecological explanations that have previously been advanced to explain the apparent slowdown in diversification through time observed in many molecular phylogenies (Ruber and Zardoya 2005; McPeek 2008; Phillimore and Price 2008; Rabosky and Lovette 2008a). Under a model of diversification mediated by ecological opportunity, diversification rates decline through time as the number of species rises within a particular ecological or biogeographic theater. This may be attributable to a higher frequency of ecological speciation during the early phases of radiations (Rice and Hostert 1993; Schluter 2000; Gavrilets and Vose 2005). Alternatively, if new populations persist better in depauperate environments, then the formation of geographically isolated populations and the effective rate of speciation will decline through time as the number of species increases (Price 2008). The HEPT model lacks these features and specifies only
that speciation events occur immediately after extinction events. Indeed, phylogenies generated under the HEPT model do not even correspond to “radiations” in any meaningful sense because rapid lineage accumulation that occurs at the base of phylogenetic trees yet is unaccompanied by an increase in species diversity or expansion of ecological space.

A central challenge posed by molecular phylogenetic analyses of diversification has been to explain why estimates of extinction are so often near zero under the birth–death model (Nee 2006; Bokma 2009; Rabosky and Lovette 2008c). This observation is striking in light of the fossil record, which typically suggests that speciation and extinction rates are roughly equivalent when considering sufficiently long time intervals (Pearson 1996; Sepkoski 1998; Alroy 2000, 2008). I have shown that, in contrast to the birth–death model, a limited diversity model with zero net diversification can account for patterns of lineage accumulation observed in real phylogenies.

An alternative to the birth–death model that may also explain the apparent pattern of rapid diversification early is McPeek’s (2008) metacommunity model, where he showed that manipulation of ecological similarity between parent and progeny species could generate phylogenetic patterns consistent with the appearance of both temporal declines and increases in diversification through time. In contrast to the model presented here, McPeek’s (2008) model predicts that clades showing apparent bursts of lineage diversification early in their history should be characterized by ecological speciation or at least a tendency for speciation events to be associated with ecological differentiation. Both these models make predictions that can be tested with additional data, from either the fossil record or patterns of ecological divergence among species within radiations. The HEPT model predicts 1) that diversity should be roughly constant through time, 2) that pulses of phylogenetically patterned extinction should be seen in the fossil record, and 3) that these turnover pulses should correspond in time to the apparent rapid speciation at the base of molecular phylogenetic trees. Clearly there is a need to better integrate perspectives on diversification from molecular phylogenies and fossils.

The HEPT model is best able to explain the wood warbler data when extinction rates show considerable phylogenetic structure (Fig. 5a,b); this is primarily due to the reduced waiting time for coalescence from \( N = 3 \) to \( N = 2 \) species. When extinction is minimally heritable (\( \omega = 0.1 \)), this waiting time may be extremely long relative to the total depth of the tree (Fig. 7), and including this “long stem” leads to old trees showing \( \gamma > 0 \) (Fig. 5c). However, when this initial waiting time is omitted, the waiting times between the remaining speciation events are consistent with a major decline in the diversification rate through time (Figs. 5d and 6). Few studies have explicitly quantified the extent to which extinction in the fossil record is phylogenetically structured (Purvis 2008) although the sudden disappearance of many groups in their entirety strongly suggests that phylogeny is a strong predictor of extinction (Bininda-Emonds et al. 2007). More evidence emerges from recent work on extinction risk in extant taxa, where numerous studies have shown that threatened and endangered species are frequently clustered with respect to phylogeny (Purvis et al. 2000; Koh et al. 2004; Purvis et al.
Figure 6. Many trees simulated under low $\omega$ were characterized by extremely long waiting times between the initial bifurcation and the next speciation event in the phylogenetic tree consisting of extant taxa only. a) Representative tree from posterior predictive simulations with low extinction heritability ($\omega = 0.1$) showing long waiting time following initial bifurcation. Note apparent burst of diversification at approximately 5 Ma, which corresponds to the timing of wood warbler diversification. b) Lineage accumulation curve for tree shown in (a), with $\gamma$ much greater than that observed for the wood warbler data ($\gamma = -3.48$). c) Lineage accumulation curve after omitting the initial bifurcation in the tree; $\gamma$ for the remaining speciation times is consistent with the wood warbler data set. A great number of trees in simulations under $\omega = 0.1$ and $\omega = 0.5$ were characterized by similar long waiting times at the base of the tree, accounting for the bimodal distribution of both $\gamma$ and tree depth (Fig. 5b). However, when these basal bifurcations are ignored, the distribution of $\gamma$ is strikingly similar to the wood warbler data (Fig. 5d).

Figure 7. Ratio of the initial waiting time to total tree depth for phylogenies simulated under model with low extinction heritability ($\omega = 0.1$). Many phylogenies failed to eliminate this basal divergence, leading to exceptionally long waiting times until the next speciation event preserved in the reconstructed tree. Trees with high extinction heritability almost always eliminated this long waiting period, such that most trees simulated under $\omega = 1.0$ showed tree depth values consistent with the wood warbler tree.

2005; Vamosi J.C. and Vamosi S.M., 2005; Jablonski 2008; Vamosi and Wilson 2008). Although more work in this area is clearly needed, the HEPT model assumption of phylogenetically structured extinction is likely to be no more problematic than the standard birth–death model, which typically assumes identical extinction rates across all lineages.

The HEPT model is closely related to the simple Moran process as studied by previous researchers (Hey 1992; Ricklefs 2006b) but predicts dramatically different patterns of lineage accumulation through time. The simple Moran process predicts extremely long waiting times for the final coalescence events in a phylogenetic tree. That few phylogenies show evidence for such long waiting times underlies Hey’s (1992) finding that a pure birth model provided a better fit to phylogenies than a Moran process. Likewise, Ricklefs (2006b, 2009) has pointed out that a Moran process requires timescales for species turnover that are inconsistent with the known history of life on Earth. For example, he estimates that, under the Moran process, the complete turnover of South American suboscine birds could require billions of years. However, phylogenetically patterned extinction can lead to lineage turnover in much shorter periods of time and eliminate the long waiting times for coalescence observed under the simple Moran process (Figs. 6 and 7). It may prove fruitful to revisit these and other results obtained for the simple Moran process after relaxing the assumption of identical extinction rates among contemporaneous species.

One potential weakness of the model is that it appears to require very high speciation and extinction rates
to explain the wood warbler data. The rates given in Table 1 \( (r_2) \) appear to be at the upper end of the spectrum of previous estimates of speciation (Coyne and Orr 2004). However, it is somewhat misleading to compare my estimates with previous estimates for several reasons. First and foremost, previous compilations of speciation rates from phylogenetic data (Coyne and Orr 2004) have often estimated rates assuming \( \mu = 0 \). As higher background extinction rates are assumed, the estimated speciation rate will necessarily increase. For example, a simple estimate of the speciation rate based on clade age and species richness (Magallon and Sanderson 2001) for the wood warbler tree gives dramatically different results depending on the assumed level of background extinction. If we assume \( \mu/\lambda = 0 \), the estimated speciation rate is 0.51 lineages/million year, but if \( \mu/\lambda = 0.99 \), the estimate spikes to 4.70 lineages/million year. This latter rate is higher than median rates estimated under the HEPT model even when extinction is not phylogenetically structured.

A second point is that many previous estimates of speciation assume homogeneous diversification through time (Magallon and Sanderson 2001; McCune 2004; Nee 2001). Because the HEPT model explicitly estimates rates during pulsed turnover phases, it is only natural that these high rates should be higher than rates averaged across the entirety of evolutionary radiations. The evidence for rate variation through time is substantial (Purvis et al. 2008; Rabosky 2009b), and it is not appropriate to compare estimates that assume rate constancy through time with those that do not.

It is important to note that the HEPT model, as formulated here, does not incorporate a mechanism by which phylogenetically patterned extinction might arise. A broad range of biologically relevant phenomena might lead to apparent heritability of extinction. For example, a key innovation might arise in some particular species that leads to increased species proliferation at the expense of other species. The result might be a turnover process driven by speciation and competitive displacement, where the descendants of the species in which the innovation arose come to dominate the biota. Such a mechanism would likely give patterns consistent with HEPT model predictions. Likewise, if species with particular trait values are more likely to go extinct, and if those trait values show phylogenetic patterning, then we might also expect to observe patterns consistent with those reported here. Another issue that remains to be resolved is the comparability of the heritability parameter, \( \omega \), across phylogenetic trees of differing diversities. As one reviewer of this manuscript pointed out, formulation of a metric of heritable extinction that is comparable across groups would be useful. This may enable us to quantify the magnitude of heritability required to produce a given lineage diversification pattern and should facilitate comparisons with patterns in the fossil record.

Limited diversity models with pulsed turnover and phylogenetically structured extinction represent a radically different interpretation of phylogenetic diversification patterns. In contrast to the standard birth–death process, these models can recover apparent bursts of lineage diversification early in a clade’s history with high background extinction and zero net diversification. Moreover, because these lineage accumulation patterns can arise despite constant diversity through time, apparent “bursts” of diversification do not necessarily require the depauperate environments specified by the ecological opportunity model. These results further imply that early rapid bursts of lineage accumulation in molecular phylogenies need not correspond to an adaptive radiation model of diversification but can arise through temporal variation in turnover rates in conjunction with ecological limits on clade growth. It is too early to determine the extent to which limited diversity models might replace or augment the standard birth–death process, but they have the potential to dramatically change our perspective.

A major limitation of diversification studies from phylogenetic data is that these competing classes of models cannot be directly compared. How well does the HEPT model fit relative to a birth–death model with diversity-dependent speciation rates? While the HEPT model can generate lineage accumulation curves that are similar to observed data, so too can the birth–death model (e.g., Rabosky and Lovette 2008a). There is no straightforward way to compare these models because birth–death models (as typically implemented) do not treat tree depth or crown clade age as a random variable; it is taken as a fixed quantity. This is not the case for the HEPT model, which is why tree depth is an appropriate summary statistic. Although these results suggest limits to inference based on molecular phylogenies alone, consideration of additional data may be highly informative. For example, some diversity-dependent models—but not HEPT models—predict that rapid lineage diversification in phylogenetic trees should be associated with rapid divergence of ecological traits (Price 1997). This can be tested with trait data from extant species (e.g., Harmon et al. 2003; Freckleton and Harvey 2006), leading to richer inference than would be possible with phylogenies alone.

The question is not whether clades radiated: at some point they clearly did. However, this study raises questions about whether lineage accumulation patterns observed in molecular phylogenies correspond in any way to this radiation. We have traditionally focused on identifying key innovations and ecological conditions that promote rapid species accumulation in growing clades, but it may be equally valid to ask whether time-varying turnover rates and constant diversity can account for patterns of speciation in molecular phylogenies.

**Funding**

This research was supported by U.S. National Science Foundation grants OSIE-0612855 and DEB-0814277 and by the Fuller Evolutionary Biology Program at the Cornell Laboratory of Ornithology.
ACKNOWLEDGEMENTS
I thank A. McCune, I. Lovette, A. Phillimore, R. Ricklefs, K. Wagner, J. Koslow, V. Savolainen, and J. Sullivan for comments on the manuscript and/or discussion of these topics.

SUPPLEMENTARY MATERIAL
Supplementary material can be found at http://www.sysbio.oxfordjournals.org/.

REFERENCES

Received 22 February 2009; reviews returned 20 April 2009; accepted 8 September 2009
Associate Editor: Vincent Savolainen