Systematic Biology Advance Access published March 11, 2016

Does gene tree discordance explain the mismatch between macroevolutionary models and empirical patterns of tree shape and branching times?

Tanja Stadler[1,2,*], James H. Degnan[3], Noah A. Rosenberg[4]

February 22, 2016

[1] ETH Zürich, Department of Biosystems Science and Engineering, Mattenstrasse 26, 4058 Basel, Switzerland
[2] Swiss Institute of Bioinformatics (SIB), Switzerland
[3] Department of Mathematics and Statistics, University of New Mexico, 311 Terrace NE, Albuquerque, NM, 87131, USA
[4] Department of Biology, Stanford University, 371 Serra Mall, Stanford, CA 94305, USA
∗ Corresponding author: tanja.stadler@bsse.ethz.ch

Keywords: Birth–death process, genealogy, multispecies coalescent, phylogeny

Abstract

Classic null models for speciation and extinction give rise to phylogenies that differ in distribution from empirical phylogenies. In particular, empirical phylogenies are less balanced and have branching times closer to the root compared to phylogenies predicted by common null models. This difference might be due to null models of the speciation and extinction process being too simplistic, or due to the empirical datasets not being representative of random phylogenies. A third possibility arises because phylogenetic reconstruction methods often infer gene trees rather than species trees, producing an incongruity between models that predict species tree patterns and empirical analyses that consider gene trees. We investigate the extent to which the difference between gene trees and species trees under a combined birth–death and multispecies coalescent model can explain the difference in empirical trees and birth–death species trees. We simulate gene trees embedded in simulated species trees and investigate their difference with respect to tree balance and branching times. We observe that the gene trees are less balanced and typically have branching times closer to the root than the species trees. Empirical trees from TreeBase are also less balanced than our simulated species trees, and model gene trees can explain an imbalance increase of up to 8% compared to species trees. However, we see a much larger imbalance increase in empirical trees, about 100%, meaning that additional features must also be causing imbalance in empirical trees. This simulation study highlights the necessity of revisiting the assumptions made in phylogenetic analyses, as these assumptions, such as equating the gene tree with the species tree, might lead to a biased conclusion.

Which macroevolutionary processes give rise to empirical phylogenies? This question has puzzled biologists for almost as long as empirical phylogenies have been inferred. It can be argued that neither the discrete tree shapes nor the numerical branching times of empirical trees are explained well by current null models of

macroevolution (Blum and François, 2006; Etienne and Rosindell, 2012).

For the discrete tree shape, approaches to testing macroevolutionary null models typically rely on tree-balance statistics, measuring the extent to which sizes of sister clades differ at internal nodes of phylogenies (Sackin, 1972; Colless, 1982; Mooers and Heard, 1997; Aldous, 2001; Felsenstein, 2004). In balanced trees, sister clades have similar numbers of taxa, whereas in unbalanced trees, their numbers of taxa differ substantially. Tests of a macroevolutionary model compare theoretical or simulation-based predictions of the model about tree balance to observations from empirical trees (Heard, 1996; Agapow and Purvis, 2002; Heard and Mooers, 2002; Blum and François, 2006; Bortolussi et al., 2006). Tests of predictions about branching times proceed similarly, examining representations of the number of lineages through time (Harvey et al., 1994) and evaluating the extent to which lineages accumulate nearer the present rather than early in the phylogeny.

Perhaps the simplest model describing the shapes of phylogenies is the constant-rate birth–death model, in which speciations are represented by birth events and extinctions by death events (Kendall, 1948, 1949; Nee et al., 1994). Under this model, each species at each point in time has the same rate λ for speciation and the same rate μ for extinction. When examining theoretical phylogenies under the model and empirical phylogenies constructed primarily from molecular data, studies typically observe that empirical phylogenies are much less balanced than is predicted by the constant-rate birth–death model (Aldous and Pemantle, 1996; Blum and François, 2006; Hagen et al., 2015). As all the so-called species-speciation-exchangeable models (Stadler, 2013)—including the Yule pure-birth model, diversity-dependent models, and environment-dependent models—predict the same tree-shape distribution as the constant-rate birth–death process, a large class of models predicts phylogenies to be more balanced than those that have been reported. Furthermore, branching times in empirical phylogenies are generally closer to the root of the tree than is predicted by the constant-rate birth–death model (Etienne and Rosindell, 2012).

The mismatch of a simple null model such as the constant-rate birth–death process with empirical phylogenies built from molecular sequences has been described with two classes of explanations: the null model might be a poor description of the macroevolutionary process (Aldous and Pemantle, 1996; Heard, 1996; Heard and Mooers, 2002), or alternatively, it might be a reasonable model that fails because it is applied to non-representative sets of empirical phylogenies that possess various forms of bias, including selection bias and taxon sampling bias (Mooers and Heard, 1997; Heath et al., 2008). We investigate yet a third possibility: the model and data are both reasonable, but *the species evolution process that the models describe is not the same as the gene lineage evolution process that the molecular sequences represent.*

When testing macroevolutionary hypotheses on empirical phylogenies, are the models and data commensurable? In typical models for macroevolution, species trees are considered, representing the branching order of species. Frequently, however, empirical species trees are inferred from one or a small number of concatenated sequence alignments, and the inferred *gene tree*—the tree of genetic lineages at a particular region of the genome—is implicitly treated as an estimated *species tree*. Gene trees can be highly discordant with their underlying species tree (Degnan, 2013, Table 2), even when gene trees are estimated with high accuracy. Therefore it is not clear that models of species evolution correctly describe properties of accurately inferred gene trees.

Here, using a hierarchical model, we investigate the difference in tree balance and branching times between gene trees and species trees. In our model, the process of species evolution—speciation and extinction—employs the simple birth–death process. Gene trees for a particular species tree, however, are described by the multispecies coalescent model of gene lineage evolution conditional on the species tree (Rannala and Yang, 2003; Degnan and Rosenberg, 2009). The hierarchical model enables us to investigate the extent to which tree balance differs in gene trees—the data source of empirical phylogenies—and species trees, the source of predictions

about the data. Under our model, we find that gene trees typically have greater imbalance compared to species trees. We investigate if the imbalance in empirical phylogenies—which exceeds that of birth–death species trees—can be explained with the hierarchical model under the assumption that empirical phylogenies are, in fact, gene trees.

The multispecies coalescent null model assumes that gene lineages merge within the species tree branches according to a coalescent process (Degnan and Rosenberg, 2009). Typical analyses of gene trees under the multispecies coalescent treat a fixed species tree as a parameter (Degnan and Salter, 2005; Degnan et al., 2012; Wu, 2012); here, we permit the species tree to vary as in empirical macroevolutionary studies, examining the distribution of gene trees given a birth–death distribution of species trees. We perform a simulation study over a range of parameter combinations.

The discrete tree shape, the discrete temporal ordering of the branching events, and the continuous branching times uniquely describe a phylogenetic tree. We study the gene-tree and species-tree distributions under the nested model, focusing on tree shape and branching times. As these quantities are high-dimensional objects, we calculate summary statistics.

For tree shape, we examine the well-known Colless statistic (Colless, 1982); we also consider the Sackin statistic (Sackin, 1972) and a statistic recording the number of cherries in a tree (McKenzie and Steel, 2000). These statistics measure the imbalance of tree shapes, the Colless and Sackin statistics increasing with increasing imbalance, and the cherry statistic decreasing with increasing imbalance.

For the branching times, we consider the γ statistic (Pybus and Harvey, 2000), measuring the temporal locations of branching events. Increasing γ corresponds to moving branching times in a tree closer to the tips. A constant-rate pure-birth tree has an expected γ of 0, and γ increases with an increasing amount of extinction.

Under the hierarchical model, our simulation poses three questions. (1) How different are the shapes of gene trees compared to species trees? (2) How different

are the branching times of gene trees and species trees? (3) How different are the model gene trees from empirical gene trees? We first formally define the species tree and gene tree models. We then discuss our simulation results and compare the simulated gene trees to a database of empirical phylogenies.

The Hierarchical Model

The Birth–Death Model of Speciation and Extinction

The constant-rate birth–death model of speciation and extinction begins at time T in the past with a single species. Each species has a birth rate $\lambda > 0$ and a death rate μ with $0 \leq \mu \leq \lambda$. The values of λ and μ apply to all species. At the present, extant species lineages are independently sampled, each with probability ρ, $0 < \rho \leq 1$, for inclusion in the final species phylogeny. We assume an improper uniform-$(0, \infty)$ distribution on T and condition on the final phylogeny having n sampled tips. In other words, the resulting simulated tree set is analogous to the following procedure: we draw a time T from the uniform-$(0, \infty)$ distribution; we simulate for time T starting with a single species; we keep the tree if we obtain n extant sampled present-day species; we repeat the procedure until we obtain the required number of trees. However, we employ mathematical theory to make the simulations efficient (Aldous and Popovic, 2005; Gernhard, 2008a). Our simulations vary three parameters: the speciation rate λ, "turnover" μ/λ, and sampling probability ρ.

To facilitate interpretations, we note that different parameter values for λ, μ/λ, and ρ can give rise to exactly the same species tree distribution. When decreasing the sampling probability ρ while increasing the speciation rate λ and turnover μ/λ, we can obtain the same distribution of phylogenetic trees (Stadler, 2009).

We recall the parameter transformations that generate identical phylogenetic tree distributions. Consider arbitrary $\lambda > 0$ and μ/λ with $0 \leq \mu/\lambda \leq 1$, and let

$\rho = 1$. Choose a sampling probability ρ', with $0 < \rho' < 1$. The increased values of λ' and μ'/λ' producing the same distribution as $(\lambda, \mu/\lambda, 1)$ are (Stadler, 2009)

$$\lambda' = \frac{\lambda}{\rho'} \quad (1)$$

$$\frac{\mu'}{\lambda'} = 1 + \rho'\left(\frac{\mu}{\lambda} - 1\right). \quad (2)$$

Note that by eq. 1, λ' increases with decreasing ρ'. For turnover, noting that $\rho' \leq 1$, it follows that $1 + \rho'(\frac{\mu}{\lambda} - 1) \geq \frac{\mu}{\lambda}$, so that $\frac{\mu'}{\lambda'} \geq \frac{\mu}{\lambda}$. Furthermore, eq. 2 reveals that turnover μ'/λ' increases with decreasing ρ'. In the case of $\mu/\lambda = 1$, decreasing ρ' increases the speciation rate λ', while turnover μ'/λ' is fixed at 1.

Beginning from choices for $(\lambda', \mu'/\lambda', \rho')$ with $\lambda' > 0$, $0 \leq \mu'/\lambda' \leq 1$, and $\rho' < 1$, the parameter values $(\lambda', \mu'/\lambda', \rho')$ of a partially sampled speciation–extinction process give rise to the same phylogenetic tree distribution as a process with complete sampling $(\lambda, \mu/\lambda, 1)$ if and only if $\frac{\mu'/\lambda' - 1}{\rho'} + 1 = \frac{\mu}{\lambda} \geq 0$; if $\frac{\mu'/\lambda' - 1}{\rho'} + 1 < 0$, then no birth-death process producing the identical phylogenetic tree distribution with complete sampling exists (the second requirement on $\frac{\mu}{\lambda}$, namely $\frac{\mu}{\lambda} \leq 1$, is satisfied for all permissible λ', μ', ρ', following from $\frac{\mu'}{\lambda'} \leq 1$).

The Coalescent Model for Gene Lineages

Within a species lineage, we assume that gene lineages coalesce backward in time according to Kingman's coalescent (Kingman, 1982a,b). Under Kingman's coalescent, the waiting time in calendar units for two gene lineages to find their common ancestor is exponentially distributed with rate $1/(Ng)$, where N is the haploid effective size of the population along the species lineage and g is the length of a generation in calendar units (Hudson, 1990; Drummond et al., 2005). Following the assumptions of the multispecies coalescent, gene lineages that do not coalesce along a species tree branch persist into ancestral species branches, where they also have the opportunity to coalesce with other gene lineages entering the ancestral species

from other descendant species (Degnan and Rosenberg, 2009).

Simulation Design

We simulated species phylogenies under a constant-rate birth–death model with speciation rate λ, extinction rate μ, and sampling probability ρ for each extant species. We simulated 100,000 species trees on n tips for each parameter combination (λ, μ, ρ), for $n = 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100$.

Next, conditional on species trees, we simulated one gene tree per species tree, assuming a sample of one gene lineage per extant species. We assumed a constant effective population size N and a constant generation time g for a species, with $Ng = 1$ for each species (meaning N and g may differ across species, but with a constant product). One coalescent time unit—the expected time to coalescence of two lineages—is N generations, or Ng calendar time units. A speciation rate of λ events per coalescent time unit means that in expectation, a species splits into two species after $1/\lambda$ coalescent time units, or equivalently, after Ng/λ calendar time units (in our setting, $Ng/\lambda = 1/\lambda$).

We compared the distributions of tree shape and branching times of the gene trees to those of the species trees. We summarized the gene-tree and species-tree distributions using three summary statistics of tree shape, applied separately to both gene trees and species trees: the Colless index C, the Sackin index S, and the number of cherries H. We denote the gene tree statistics by C_g, S_g, and H_g, and the species tree statistics by C_s, S_s, and H_s. For these statistics, we report ratios, $\overline{C}_g/\overline{C}_s$, $\overline{S}_g/\overline{S}_s$, and $\overline{H}_g/\overline{H}_s$, where the numerator represents the mean value of the statistic computed across gene trees and the denominator is the corresponding mean across species trees. The higher the ratios $\overline{C}_g/\overline{C}_s$ and $\overline{S}_g/\overline{S}_s$, and the lower the ratio $\overline{H}_g/\overline{H}_s$, the more imbalanced the gene trees are in relation to the species trees. Because these statistics are correlated, we present only the Colless statistic in the main text and provide the other two statistics in the supplement. The statistic we

report is equivalent to the average across simulations of $1 + (C_g - C_s)/\overline{C}_s$, where $C_g - C_s$ is the difference in the Colless statistic for species tree–gene tree pairs. The value of $C_g - C_s$ depends on both the birth–death parameters and the sample size, so that dividing by $\overline{C}_s$ helps to standardize it.

For branching times, we summarized the gene-tree and species-tree distributions using the branching-time statistic γ. As γ is already normalized for tree size and in fact has expectation 0 for a range of species tree models, we reported the average of the difference $\gamma_g - \gamma_s$ between γ values computed on gene trees and on species trees. We denote the average difference by $\overline{\gamma}_g - \overline{\gamma}_s$. The smaller the difference $\overline{\gamma}_g - \overline{\gamma}_s$, the closer the branching times of the gene trees are to the root compared to the corresponding branching times of the species tree.

The simulations and analyses were performed in R unless otherwise indicated. The code was added to the R package TreeSim v2.2 (Stadler, 2011).

Simulation Results: Tree Shape

Figure 1 and SI Figures 1 and 2 present the ratios $\overline{C}_g/\overline{C}_s$, $\overline{S}_g/\overline{S}_s$, and $\overline{H}_g/\overline{H}_s$, respectively, of the summary statistics for simulated gene trees and species trees. We briefly summarize the results for shapes of gene trees compared to species trees.

Both for very small and very large λ, the gene trees and species trees have approximately the same average tree shape. For intermediate λ, however, in the biologically plausible range, gene trees evolving on species trees have a different shape distribution from the species trees themselves. For high turnover μ/λ, the imbalance was greatest in our simulations for $\lambda = 5$, representing an average of five speciation events in each N-generation unit of coalescent time. For low turnover, the maximal imbalance was observed for $\lambda = 2$, two speciation events per N generations. The effect was larger for trees with many taxa, producing an increase of $\sim$8% for the Colless statistic (Figure 1) and $\sim$1.8% for Sackin (SI Figure 1), and a $\sim$1.3% decrease for the cherry statistic (SI Figure 2). Thus, we might expect to overestimate

the tree imbalance from empirical data when using gene trees instead of species trees.

We next discuss differences in gene tree and species tree properties in detail, as a function of speciation rate λ, turnover μ/λ, sampling probability ρ, and the number of species n used in the simulations. First, we examine the limits of very small and very large λ, and we then consider the roles of the parameters in the biologically relevant intermediate cases.

Extreme Values of λ

The extreme cases of $\lambda \to 0$ and $\lambda \to \infty$ illustrate the limiting behavior of the statistics. We use $\lambda = 10^{-3}$ to represent $\lambda \to 0$, and $\lambda = 10^{7}$ for $\lambda \to \infty$.

$\lambda \to 0$.—For small λ, speciation is rare, and therefore, species tree branches are very long. Consequently, sufficient time exists for each gene lineage coalescence to occur on the most recent species tree branch for which the coalescence is possible. Each gene tree then has the same shape as the species tree on which it has evolved. Thus, the ratios of the mean Colless, Sackin, and cherry statistics for simulated gene trees and for the underlying simulated species trees all approach 1.

$\lambda \to \infty$.—For large λ, speciation is frequent, and species tree branches are infinitesimally short. Thus, all gene lineage coalescences occur prior to the root of the species tree. Gene-tree shapes then follow the shapes of gene trees under the Kingman coalescent. It can be shown that Kingman's coalescent and constant-rate birth–death trees produce the same distribution of tree shapes (Aldous and Pemantle, 1996). Thus, as in the $\lambda \to 0$ case, but for a different reason, the ratios of the mean Colless, Sackin, and cherry statistics for gene trees and species trees equal 1.

Intermediate λ

For intermediate values of λ, we observed in our simulations that gene trees were less balanced than species trees, as the Colless and Sackin ratios exceeded 1, and the cherry ratio was less than 1 (Figure 1 and SI Figures 1 and 2). Further, these ratios move farther from 1 for larger trees.

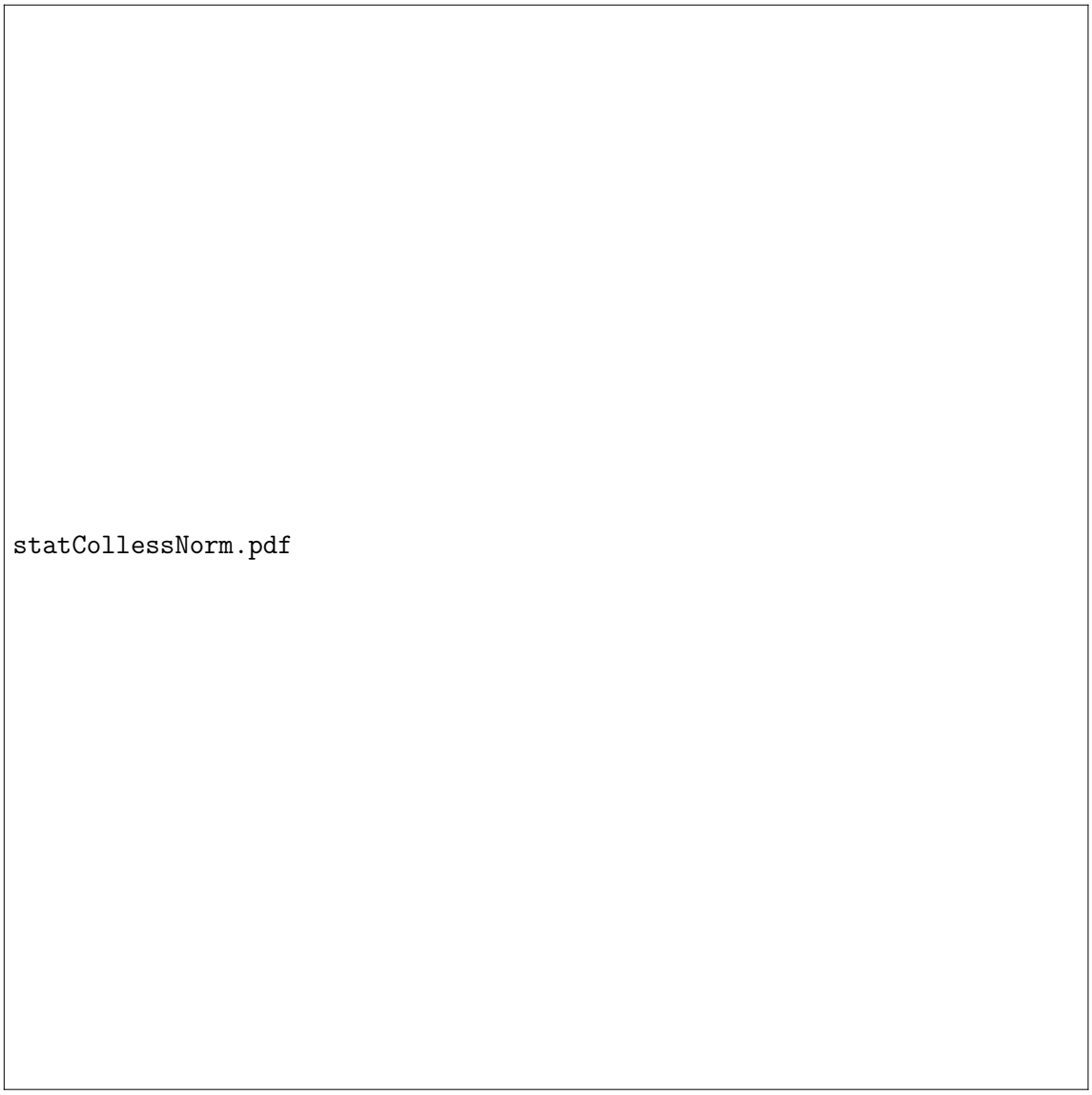

Figure 1: Mean Colless statistic of gene trees divided by mean Colless statistic of species trees ($\overline{C}_g/\overline{C}_s$). Solid lines correspond to complete species sampling $\rho = 1$, dashed lines to sampling probability $\rho = 0.75$, and dot-dashed lines to sampling probability $\rho = 0.5$. Plots are obtained based on 100,000 simulated species tree–gene tree pairs at each choice of parameter values, taking means separately for the gene trees and the species trees.

Small $\lambda \leq 2$

Varying $\lambda \leq 2$*, fixed turnover* μ/λ*, and complete sampling* $\rho = 1$.—In our simulations, the difference between gene trees and species trees in tree balance increases with λ for these parameter values. As species tree branches become shorter with increasing λ, gene coalescences might not happen on the first allowed branch, and therefore, they might not follow the same pattern as speciation events.

Fixed $\lambda \leq 2$*, varying turnover* μ/λ*, and complete sampling* $\rho = 1$.—Here, the difference between gene trees and species tree is larger for small turnover compared to large turnover. For $\lambda \leq 2$, species tree branches are relatively long, so that most, though not all, gene coalescences happen on the first branch allowed. Trees with small μ/λ have younger root ages and therefore shorter branches compared to trees with large μ/λ (Figures 3 and 4 of Stadler (2008)). Thus, the probability that gene coalescences do not happen on the first species tree branch—so that they might not follow the same pattern as speciation events—increases with decreasing turnover.

Fixed $\lambda \leq 2$*, fixed turnover* μ/λ*, and varying sampling probability* ρ.—Sparser sampling, as represented by smaller ρ, decreases the difference in balance between gene trees and species trees. Recall that a process with sampling probability ρ', speciation rate λ' and extinction rate μ' is equivalent to a process with complete sampling $\rho = 1$ and a smaller speciation rate $\lambda \leq \lambda'$ and smaller turnover $\mu/\lambda \leq \mu'/\lambda'$, provided $\frac{\mu'/\lambda'-1}{\rho'}+1 \geq 0$. The smaller speciation rate λ produces longer species-tree branch lengths compared to a process with parameters λ', μ', and $\rho = 1$, and thus decreases tree-shape differences between gene trees and species trees. On the other hand, the smaller turnover μ/λ produces shorter trees compared to a process with parameters λ', μ', and $\rho = 1$, and thus increases the difference of gene trees and species trees. We observe from the figures that the effect of a smaller speciation rate—meaning longer branches and thus less difference between gene trees and species trees—dominates, so that for fixed λ and μ/λ, decreasing the sampling

fraction increases the agreement between gene-tree and species-tree shape.

Note that for a turnover $\mu/\lambda = 1$, we have $\mu/\lambda = \mu'/\lambda'$. Thus, arbitrary λ and $\rho = 1$ produces the same tree balance ratio as $\lambda' = 2\lambda$ and $\rho' = 0.5$. This property can be verified in our figures by comparing $\lambda = 0.5$ and $\lambda' = 1$, $\lambda = 1$ and $\lambda' = 2$, $\lambda = 10$ and $\lambda' = 20$, or $\lambda = 50$ and $\lambda' = 100$.

Large $\lambda \geq 5$

Varying $\lambda \geq 5$*, fixed turnover* μ/λ*, and complete sampling* $\rho = 1$.—The difference in balance between gene trees and species trees decreases with increasing λ, particularly for the larger λ values ($\lambda \geq 50$). As λ increases, species tree branches become so short that most coalescences happen prior to the species tree root. Such coalescences occur according to the Kingman coalescent, inducing the same tree shapes as the constant-rate birth–death process. Thus, as λ increases, the gene-tree shape distribution approaches the same distribution as that of species trees.

Fixed $\lambda \geq 5$*, varying turnover* μ/λ*, and complete sampling* $\rho = 1$.—We observe a larger difference between gene trees and species trees for high turnover compared to low turnover; for $\mu/\lambda = 1$, this result begins at $\lambda \geq 50$. Species trees become very short for large λ, and for fixed λ, low-turnover species trees are shorter than high-turnover trees. Thus, more gene coalescences happen above the root for low-turnover trees, so that the approach of the distribution of gene-tree shapes to the same distribution seen for species trees is faster at low turnover.

Fixed $\lambda \geq 5$*, fixed turnover* $\mu/\lambda < 1$*, and varying sampling probability* ρ.—For these values, incomplete sampling minimally changes tree balance: the effects of incomplete sampling, amounting to a process with complete sampling and both a decrease in λ that produces a greater difference between gene trees and species trees as well as a decrease in turnover that produces a smaller difference, cancel.

Fixed $\lambda \geq 5$*, fixed turnover* $\mu/\lambda = 1$*, and varying sampling probability* ρ.—In this case, incomplete sampling can be seen as a process with the same turnover and

complete sampling, but a decreased speciation rate—meaning that species trees are longer for smaller ρ. Consequently, decreasing ρ increases the difference between gene trees and species trees. Note again that statistics in the different plots are the same for different (λ, ρ) pairs with the same value of $\lambda\rho$.

Simulation Results: Branching Times

Figure 2 presents the difference $\overline{\gamma}_g - \overline{\gamma}_s$ of the γ statistics for simulated gene trees and species trees. Briefly, gene trees tend to have a smaller γ statistic than species trees for low to medium values of λ, and a larger γ for large $\lambda \gtrsim 50$, depending on the turnover. As was observed for tree shapes, all effects increased in magnitude with the number of taxa n. A value of $\lambda = 50$ means that a speciation occurs on average after $Ng/50$ calendar time units, which seems very high. Because γ is smaller for gene trees than for species trees for realistic values of λ, we expect to underestimate γ_s from empirical data when using gene trees instead of species trees.

We discuss below differences in branching times between gene trees and species trees in detail, as a function of λ, μ/λ, and ρ (Figure 2).

Extreme Values of λ

$\lambda \to 0$.—For small λ and hence long species tree branch lengths compared to the coalescent rate for gene lineages, each gene tree coalescence occurs immediately prior to its associated speciation event. Thus, the branching times are nearly identical for gene trees and species trees, and $\overline{\gamma}_g - \overline{\gamma}_s$ is close to 0.

$\lambda \to \infty$.—For large λ and hence short branch lengths compared to the coalescent rate, gene tree coalescences happen prior to the root of the species tree, so that the gene trees are Kingman-coalescent trees. Kingman-coalescent branch lengths are in expectation, up to a scaling constant, equal to constant-rate birth–death branch lengths with $\lambda = \mu$ (Gernhard, 2008b). Thus, γ_g is in expectation equal to γ_s for

constant-rate birth–death trees with $\lambda = \mu$. The value of γ_s depends on μ/λ and ρ, so that for large λ, the behavior of $\overline{\gamma}_g - \overline{\gamma}_s$ depends on the other parameters.

$\lambda \to \infty$, varying turnover μ/λ, and complete sampling $\rho = 1$.—For these parameter values, $\overline{\gamma}_s$ decreases as μ/λ decreases (Pybus and Harvey, 2000). Thus, $\overline{\gamma}_g - \overline{\gamma}_s$ is increasingly positive with decreasing turnover. As for turnover $\mu/\lambda = 1$, the constant-rate birth–death trees equal in expectation Kingman-coalescent trees up to a scaling constant; thus, we obtain $\overline{\gamma}_g - \overline{\gamma}_s \approx 0$ (Figure 2, $\lambda = 10^7$).

$\lambda \to \infty$, fixed turnover $\mu/\lambda < 1$, varying sampling probability $\rho < 1$.—In this case, species trees can be interpreted to arise from a process with complete sampling $\rho = 1$ and decreased turnover. Thus, $\overline{\gamma}_s$ decreases for decreasing sampling probability ρ, so that $\overline{\gamma}_g - \overline{\gamma}_s$ increases.

$\lambda \to \infty$, fixed turnover $\mu/\lambda = 1$, varying sampling probability $\rho < 1$.—At $\mu/\lambda = 1$, incomplete sampling does not change relative branch lengths (Stadler (2008), Figure 3d). Incomplete sampling can be interpreted as a process with decreased speciation rate λ, turnover $\mu/\lambda = 1$, and complete sampling $\rho = 1$. Thus, with $\mu/\lambda = 1$, $\overline{\gamma}_s$ is the same for all sampling probabilities.

Intermediate λ

Varying λ, fixed turnover μ/λ, and complete sampling $\rho = 1$.—As λ increases, $\overline{\gamma}_g - \overline{\gamma}_s$ first becomes more negative, then switches ($\lambda \approx 5 - 20$) and becomes more positive.

Fixed λ, varying turnover μ/λ, and complete sampling $\rho = 1$.—We observe a decrease of $\overline{\gamma}_g - \overline{\gamma}_s$ with increasing turnover, meaning gene trees have branching events closer to the root compared to species trees for increasing turnover. Note that $\mu/\lambda = 1$ and small $\lambda < 5$ is an exception; these trees are very long, and gene trees are almost equal to species trees. Because $\overline{\gamma}_g - \overline{\gamma}_s$ changes from negative to positive for increasing λ, for small λ, gene trees and species trees are most similar in γ for small turnover, whereas for large λ, they are most similar for large turnover.

By contrast, recall that for shape statistics, for increasing turnover, a switch occurred from decreasing to increasing $\overline{C}_g/\overline{C}_s$ values for λ in $[2, 20]$. $\overline{C}_g/\overline{C}_s$ exceeded 1 for all λ. Thus, for small λ, gene trees and species trees were most similar in shape for large turnover, whereas for large λ, they were most similar for small turnover.

Fixed λ, fixed turnover $\mu/\lambda < 1$, and varying sampling probability $\rho < 1$.—The value $\overline{\gamma}_g - \overline{\gamma}_s$ increases with decreasing sampling, meaning gene trees had branching events closer to the tips compared to species trees. Recall that a process with decreased sampling is equivalent to a complete sampling process and decreased birth rate and turnover. A decrease in λ leads to an increase in $\overline{\gamma}_g - \overline{\gamma}_s$ for small λ and a decrease for large λ (see paragraph *"Varying λ, fixed turnover μ/λ, and complete sampling $\rho = 1$"* in this section). A decrease in turnover leads to an increase in $\overline{\gamma}_g - \overline{\gamma}_s$ (see paragraph *"Fixed λ, varying turnover μ/λ, and complete sampling $\rho = 1$"* in this section). The effect of turnover dominates.

Fixed λ, fixed turnover $\mu/\lambda = 1$, and varying sampling probability $\rho < 1$.—Incomplete sampling increases $\overline{\gamma}_g - \overline{\gamma}_s$ for small λ and decreases $\overline{\gamma}_g - \overline{\gamma}_s$ for large λ. The reason is that for $\mu/\lambda = 1$, a process with decreased sampling is equivalent to a complete sampling process with decreased birth rate and turnover 1. Recall that a decrease in λ increases $\overline{\gamma}_g - \overline{\gamma}_s$ for small λ and decreases it for large λ.

Simulation Results: Comparing Gene Trees to Their Species Trees

We have reported average gene tree balance compared to average species tree balance (Figure 1). This approach does not give an indication of the joint distribution of shape statistics for gene trees and species trees and therefore of the extent to which the shape can differ for a gene tree and its underlying species tree. To illustrate this joint variability, we simulate distributions of $1 + (C_g - C_s)/\overline{C}_s$ and C_g/C_s and the joint distribution of C_g and C_s for $\lambda = 0.1$, 2, 20, and 1000, with $\mu = 0$, for $n = 100$ taxa.

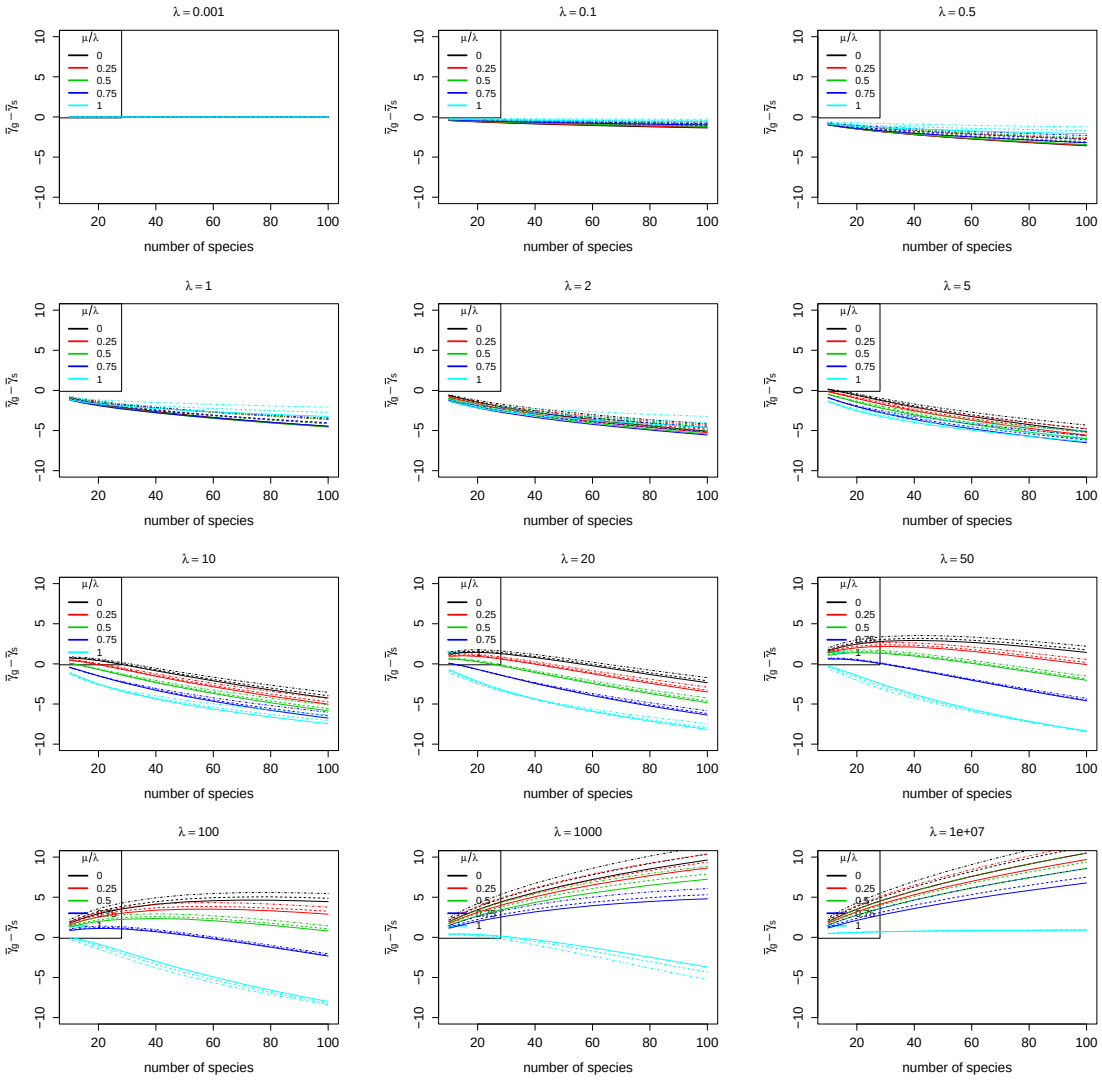

Figure 2: Mean γ statistic of gene trees minus mean γ statistic of species trees ($\overline{\gamma}_g - \overline{\gamma}_s$). Solid lines correspond to complete species sampling $\rho = 1$, dashed lines to sampling probability $\rho = 0.75$, and dot-dashed lines to sampling probability $\rho = 0.5$. Plots are obtained based on 100,000 simulated species tree–gene tree pairs at each choice of parameter values, taking means separately for the gene trees and the species trees.

As discussed above, for small λ, gene tree balance closely accords with species tree balance ($\overline{C}_g/\overline{C}_s \approx 1$), as species tree branches are very long and the gene tree and species tree are hence highly correlated (Figure 3). For increasing λ, the correlation decreases as the species tree branches become shorter, and in the $\lambda \to \infty$ limit, gene tree balance is independent of species tree balance. Because of this independence, the gene trees give rise to the same shape distribution as the species trees, and thus for large λ, again $\overline{C}_g/\overline{C}_s \approx 1$—but now, with a low correlation coefficient between C_g and C_s.

For $\lambda = 0.1$, 49.8% of gene trees have a higher Colless statistic than the underlying species trees and 48.1% have a lower value, the remaining cases having identical values for the gene tree and species tree (Figure 3). For $\lambda = 2$, 62% of gene trees have a higher Colless statistic than the underlying species tree. The percentage drops to 53% for $\lambda = 20$ and is again nearly 50% for $\lambda = 1000$. For $\lambda = 2$ and $n = 100$, the average value of C_g/C_s is 1.12, somewhat larger than the corresponding value $\overline{C}_g/\overline{C}_s = 1.08$ for $\lambda = 2$ and $n = 100$ (Figure 1).

Empirical Trees

To determine whether the difference in tree balance between gene trees and species trees under the model can explain the excess imbalance in empirical trees, we reanalyzed a set of empirical phylogenies from TreeBASE (Hagen et al., 2015; Sanderson et al., 1994). This set of phylogenies included 2759 fully resolved trees, 156 of which possessed calendar-time branch-length information. We hypothesize that many of these phylogenies are not species trees, but are either gene trees or trees that result from concatenation of genes.

Recall that the species-tree Colless value for each tree size is independent of speciation rate, turnover, and sampling probability, as all constant-rate birth–death processes induce the same distribution on tree shapes (Aldous and Pemantle, 1996). We calculated the average Colless statistics C_d for all empirical phylogenies for all

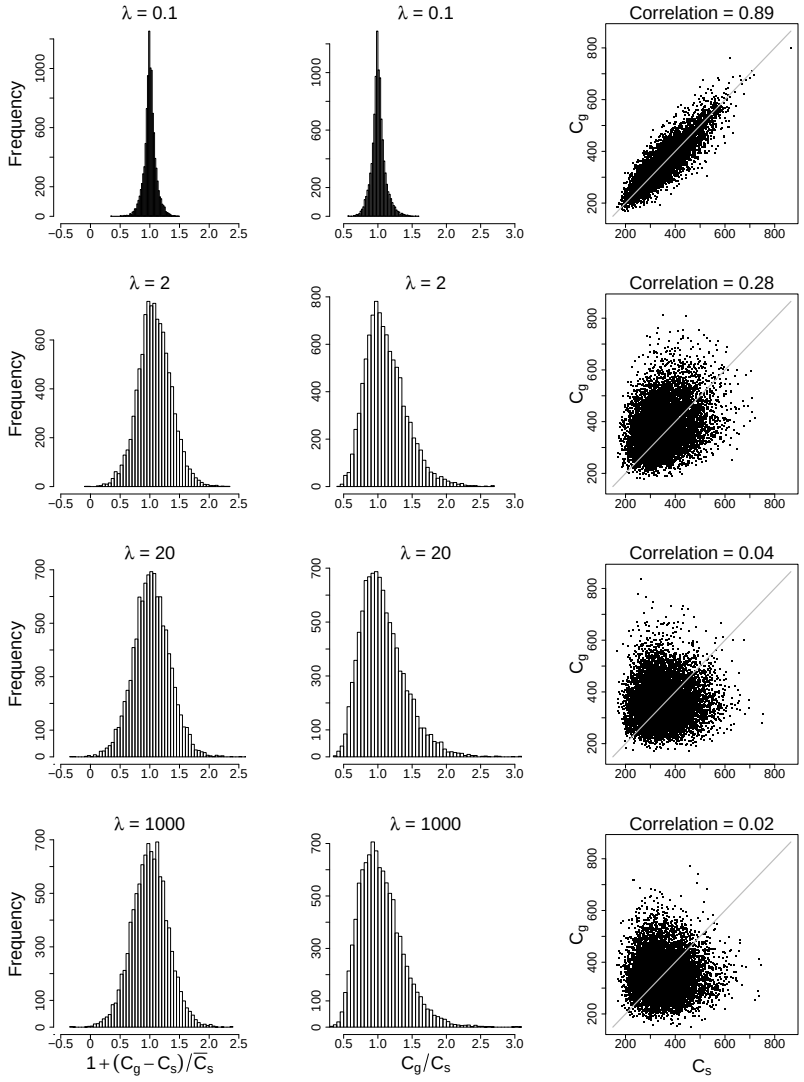

Figure 3: Distributions of $1+(C_g-C_s)/\overline{C}_s$ and C_g/C_s, and the joint distribution of C_g and C_s. All plots are for the birth process only with no extinction and are based on 10,000 independent gene tree–species tree pairs simulated in Hybrid-Lambda (Zhu et al., 2015). Grey lines in the scatterplots represent the line $C_g = C_s$; above the line, based on the Colless statistic, the gene tree has less balance than the species tree.

sizes up to $n = 100$, and we report $C_d/\overline{C}_s$ for each tree. This ratio is on average about 2 (Figure 4), so that empirical phylogenies have about twice the Colless value as constant-rate birth–death species trees. Although our simulations detected the correct direction for the deviation from the baseline value of 1, they also revealed that the multispecies coalescent with the constant-rate birth–death model can only explain an increase of the Colless statistic in gene trees compared to species trees by a factor of 1.08.

For the empirical phylogenies that reported branch lengths scaled in calendar time, although relatively few data points were available, we further calculated the γ statistic for completeness, plotting the empirical γ values together with simulated mean γ_s values for different μ/λ and ρ (SI Figure 3). We did not plot $\overline{\gamma}_g - \overline{\gamma}_s$, as such a calculation would yield an excessive 15 points (5 turnover values, 3 sampling probabilities) for each empirical data point. Because relatively few trees with branch length information are available for each value of the number of species, it is not feasible to take an expectation of empirical γ for each tree size, as we did for the simulations and empirical Colless statistic. The relationship between the empirical and simulated trends in $\overline{\gamma}_g - \overline{\gamma}_s$ is therefore difficult to discern.

Summary

Using simulations, we have quantified the difference in tree shape and branching times between gene trees and species trees under a simple hierarchical model, incorporating a constant-rate birth–death process for species trees, and a multispecies coalescent for gene trees conditional on species trees. The results suggest that although in limiting cases of very low and very high speciation rate, gene trees and species trees have the same distribution of shapes, for a variety of intermediate parameter values, gene trees are in expectation less balanced than the species trees. Branching times in gene trees and species trees differ except in the limiting case of

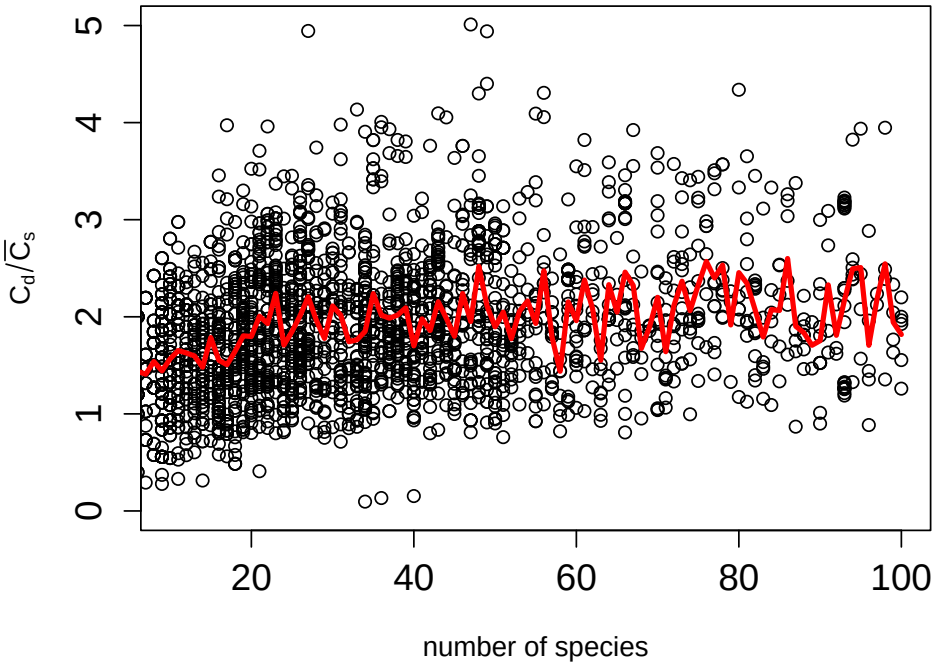

Figure 4: The Colless statistic for empirical trees from TreeBASE. Each black dot represents a tree. We normalized each empirical Colless value by dividing by the expected species-tree Colless value. The expected species-tree Colless value is independent of speciation rate λ, turnover μ/λ and species sampling ρ. The red line represents the mean of the normalized Colless statistic for each fixed tree size.

very low speciation rate.

Depending on the question of interest, either of two effect sizes could be reported for the balance ratio for gene trees and species trees: 1.12 obtained from the average of the ratios, we which denote $\overline{C_g/C_s}$ (Figure 3), or 1.08 obtained from $\overline{C}_g/\overline{C}_s$ (Figure 1). If we compare a species tree to its embedded gene tree, the effect size based on $\overline{C_g/C_s}$ is appropriate; for our data application, however, we compared a set of empirical trees to a set of model species trees. Thus, we do not consider pairs, but averages of two distributions, which calls for the latter effect size, $\overline{C}_g/\overline{C}_s$. If gene trees and species trees follow the same shape distribution, then the ratio $\overline{C}_g/\overline{C}_s$ of

the expected shape statistics is equal to 1; however, the mean value of the ratio, $\overline{C_g/C_s}$, does not generally equal 1 under the null hypothesis that C_g and C_s have the same distribution. In particular, for two random variables X and Y, both the expectations $E(X/Y)$ and $E(Y/X)$ can exceed 1. Thus, we suggest that $\overline{C_g/C_s}$ is less appropriate than $\overline{C}_g/\overline{C}_s$ as a measure of the difference in shape distributions.

The observed difference between gene trees and species trees highlights a problem in tests of species-tree models that make use of empirical phylogenies, demonstrating that empirical phylogenies obtained by taking gene trees as estimates of species trees follow a different tree-shape distribution than that predicted for species trees themselves. It is thus problematic to equate an inferred gene tree to the species tree when testing for the most appropriate species tree model.

Gene trees are expected to be less balanced compared to the underlying species tree, with branching events closer to the root for most biologically relevant parameter regions that do not involve implausibly large speciation rates. It is noteworthy that our comparison of model gene trees to model species trees yields qualitatively similar patterns to the comparison of empirical trees to model species trees: empirical phylogenies are less balanced than predicted by birth–death models (Blum and François, 2006), and they have branching events closer to the root compared to birth–death trees (Etienne and Rosindell, 2012).

Under the model, the differences in tree shape and branching times between gene trees and species trees depend on a speciation rate λ, a turnover rate μ/λ, and a sampling rate ρ. In particular, the relative timing of branching events in gene trees compared to species trees depends mainly on the speciation rate λ: gene-tree branching events are closer to the root than in species trees for small λ, and closer to the tips for large λ. This result reflects the fact that for higher speciation rates, species-tree branches are short, and thus, coalescences occur in more ancestral populations, making gene trees more like Kingman-coalescent trees.

We emphasize that our model is a neutral model: speciation rates, extinction

rates, and coalescent rates are assumed to be the same through time and across lineages. However, relaxing this assumption to allow for rate heterogeneity will not eliminate incomplete lineage sorting and thus, as in the constant-rate case, we expect that gene trees will continue to differ in balance from species trees.

Are our parameter settings in the range of empirically observed parameter values? We can use the great ape tree to examine if our model parameters are sensible in light of empirical observations. Recent estimates of the branch lengths in the great ape tree, for which there is considerable evidence of incomplete lineage sorting (Ebersberger et al., 2007; Burgess and Yang, 2008; Hobolth et al., 2011), lie between 0.7 and 3.7 coalescent time units (Schrago, 2014). Consider a birth–death model for a species tree. The pure-birth model has the property that the mean branch length in the species tree is $1/(2\lambda)$ coalescent units (Stadler and Steel, 2012), meaning $\lambda = 0.5$ induces a mean branch length of 1. Thus, with $\mu = 0$, setting λ to 0.5—a value among those on which our analysis has focused—places branch lengths within the range observed in the great ape tree.

For $\mu > 0$, a mean branch length of 1 suggests higher λ; for $\lambda = \mu$, the expected pendant branch length under a birth–death process is $1/\lambda$ (Mooers et al., 2012), so that the expected pendant branch length is 1 at $\lambda = \mu = 1$. Bokma et al. (2012) estimated the mean λ for the hominoid primate tree to be 0.46 per myr (95% confidence interval 0.12-1.37). Assuming $N = 30,000$ and $g = 25$ years—approximate values from Schrago (2014) for the ancestor of humans and chimpanzees—produces $\lambda = 0.46 \times 30,000 \times 25 \times 10^{-6} = 0.276$ speciations per coalescent unit (95% confidence interval 0.072–0.822). Turnover was estimated close to 1, as the mean μ was 0.43 myr (95% confidence interval 0.01–1.44). These similarities of empirical trees to a model with λ and μ on the order of 0.1 to 1 indicate that our approach of centering parameter choices around such values is reasonable.

Obtaining unbiased empirical species trees requires using appropriate methods for inferring species trees. Recent developments in estimation methods permit joint

inference of species trees and gene trees, or inference of species trees from multiple gene trees (Degnan and Rosenberg, 2009; Edwards, 2009; Liu et al., 2015; Szöllősi et al., 2015; Ogilvie et al., 2016). Species trees estimated by such methods take into account the hierarchical production of gene trees from species trees, and they do not rely on an implicit or explicit identification of species trees with gene trees. Thus, the shapes of species trees obtained by these methods would be expected to follow a distribution appropriate to species trees. In our empirical analysis, however, the set of previously published empirical phylogenies that we used to determine the difference between empirical and model species trees dates as far back as 1994—prior to the widespread use of phylogenetic tools that distinguish between gene trees and species trees. The hypothesis that many of the empirical trees are in fact gene trees rather than species trees explains some of the excess imbalance observed in empirical tree-shape distributions; however, because our inflation of the Colless statistic is only $\sim$1.08 for gene trees compared to species trees and the empirical inflation of the statistic is $\sim$2, other factors are required for explaining the imbalance in empirical trees. Because our number of time-calibrated empirical trees is low, our temporal computations have been less exhaustive compared to those we performed for tree shape; unlike for shape, at present, the empirical γ values—of which there are fewer—are explained reasonably well by species-tree γ values.

We comment on two of the many factors that could influence the difference between empirical trees and gene trees and species trees under our model. First, we assumed in our analyses that the gene trees and species trees are known without error. It is possible that reconstruction biases in tree estimation (Mooers and Heard, 1997; Holton et al., 2014) could contribute to a difference between empirical and theoretical distributions for trees. Second, even when species tree inference is informed by gene tree discordance, species tree inference methods might generate shape biases. For example, the minimize deep coalescence criterion (Maddison and Knowles, 2006; Than and Nakhleh, 2009) is expected to produce highly balanced

tree estimates (Than and Rosenberg, 2014) and indeed its empirical estimates are more balanced than those obtained by other methods from the same data (DeGiorgio et al., 2014).

We hope that this paper stimulates analytic and simulation-based investigations of more complex nested species tree–gene tree models, thereby linking extensive traditions modeling species trees (Nee et al., 1994; Stadler, 2013) and modeling gene trees conditional on fixed species trees (Degnan and Rosenberg, 2009). Only if we understand the predictions produced by plausible null models—and the relationships between those models and the assumptions underlying empirical trees—can we produce a proper account of the macroevolutionary phenomena that give rise to species tree patterns.

Acknowledgments. TS is supported in part by the European Research Council under the Seventh Framework Programme of the European Commission (PhyPD: grant agreement number 335529). JHD and NAR acknowledge support from NIH grant R01 GM117590.

References

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

Aldous, D. and R. Pemantle, eds. 1996. Random discrete structures vol. 76 of *The IMA Volumes in Mathematics and its Applications*. Springer-Verlag, New York papers from the workshop held in Minneapolis, Minnesota, November 15–19, 1993.

Aldous, D. and L. Popovic. 2005. A critical branching process model for biodiversity. Adv. Appl. Prob. 37:1094–1115.

Aldous, D. J. 2001. Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. Statist. Sci. 16:23–34.

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

Bokma, F., V. van den Brink, and T. Stadler. 2012. Unexpectedly many extinct hominins. Evolution 66:2969–2974.

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

Burgess, R. and Z. Yang. 2008. Estimation of hominoid ancestral population sizes under Bayesian coalescent models incorporating mutation rate variation and sequencing errors. Mol. Biol. Evol. 25:1979–1994.

Colless, D. H. 1982. Phylogenetics: the theory and practice of phylogenetic systematics. Syst. Zool. 31:100–104.

DeGiorgio, M., J. Syring, A. J. Eckert, A. Liston, R. Cronn, D. B. Neale, and N. A. Rosenberg. 2014. An empirical evaluation of two-stage species tree inference strategies using a multilocus dataset from North American pines. BMC Evol. Biol. 14:67.

Degnan, J. H. 2013. Anomalous unrooted gene trees. Syst. Biol. 62:574–590.

Degnan, J. H. and N. A. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24:332–340.

Degnan, J. H., N. A. Rosenberg, and T. Stadler. 2012. The probability distribution of ranked gene trees on a species tree. Math. Biosci. 235:245–255.

Degnan, J. H. and L. A. Salter. 2005. Gene tree distributions under the coalescent process. Evolution 59:24–37.

Drummond, A. J., A. Rambaut, B. Shapiro, and O. G. Pybus. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. 22:1185–1192.

Ebersberger, I., P. Galgoczy, S. Taudien, S. Taenzer, M. Platzer, and A. von Haeseler. 2007. Mapping human genetic ancestry. Mol. Biol. Evol. 24:2266–2277.

Edwards, S. V. 2009. Is a new and general theory of molecular systematics emerging? Evolution 63:1–19.

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

Felsenstein, J. 2004. Inferring Phylogenies. Sinauer, Sunderland, MA.

Gernhard, T. 2008a. The conditioned reconstructed process. J. Theor. Biol. 253:769–778.

Gernhard, T. 2008b. New analytic results for speciation times in neutral models. Bull. Math. Biol. 70:1082–1097.

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

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

Heard, S. B. 1996. Patterns in phylogenetic tree balance with variable and evolving speciation rates. Evolution 50:2141–2148.

Heard, S. B. and A. Ø. Mooers. 2002. Signatures of random and selective mass extinctions in phylogenetic tree balance. Syst. Biol. 51:889–897.

Heath, T. A., D. J. Zwickl, J. Kim, and D. M. Hillis. 2008. Taxon sampling affects inferences of macroevolutionary processes from phylogenetic trees. Syst. Biol. 57:160–166.

Hobolth, A., J. Y. Dutheil, J. Hawks, M. H. Schierup, and T. Mailund. 2011. Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciiation and widespread selection. Genome Res. 21:349–356.

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

Hudson, R. R. 1990. Gene genealogies and the coalescent process. Oxford Surv. Evol. Biol. 7:1–44.

Kendall, D. G. 1948. On some modes of population growth leading to R. A. Fisher's logarithmic series distribution. Biometrika 35:6–15.

Kendall, D. G. 1949. Stochastic processes and population growth. J. Roy. Statist. Soc. Ser. B. 11:230–264.

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

Kingman, J. F. C. 1982b. On the genealogy of large populations. J. Appl. Prob. 19A:27–43.

Liu, L., S. Wu, and L. Yu. 2015. Coalescent methods for estimating species trees from phylogenomic data. J. Syst. Evol. 53:380–390.

Maddison, W. P. and L. L. Knowles. 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55:21–30.

McKenzie, A. and M. Steel. 2000. Distributions of cherries for two models of trees. Math. Biosci. 164:81–92.

Mooers, A., O. Gascuel, T. Stadler, H. Li, and M. Steel. 2012. Branch lengths on birth–death trees and the expected loss of phylogenetic diversity. Syst. Biol. 61:195–203.

Mooers, A. Ø. and S. B. Heard. 1997. Inferring evolutionary process from phylogenetic tree shape. Q. Rev. Biol. 72:31–54.

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

Ogilvie, H. A., J. Heled, D. Xie, and A. J. Drummond. 2016. Computational performance and statistical accuracy of *beast and comparisons with other methods. Syst. Biol. in press.

Pybus, O. G. and P. H. Harvey. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. Proc. R. Soc. Lond. B 267:2267–2272.

Rannala, B. and Z. Yang. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics 164:1645–1656.

Sackin, M. J. 1972. "Good" and "bad" phenograms. Syst. Zool. 21:225–226.

Sanderson, M. J., M. J. Donoghue, W. Piel, and T. Eriksson. 1994. TreeBASE: a prototype database of phylogenetic analyses and an interactive tool for browsing the phylogeny of life. Am. J. Bot. 81:183.

Schrago, C. G. 2014. The effective population sizes of the anthropoid ancestors of the human–chimpanzee lineage provide insights on the historical biogeography of the great apes. Mol. Biol. Evol. 31:37–47.

Stadler, T. 2008. Lineages-through-time plots of neutral models for speciation. Math. Biosci. 216:163–171.

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. Simulating trees with a fixed number of extant species. Syst. Biol. 60:676–684.

Stadler, T. 2013. Recovering speciation and extinction dynamics based on phylogenies. J. Evol. Biol. 26:1203–1219.

Stadler, T. and M. Steel. 2012. Distribution of branch lengths and phylogenetic diversity under homogeneous speciation models. J. Theor. Biol. 297:33–40.

Szöllősi, G. J., E. Tannier, V. Daubin, and B. Boussau. 2015. The inference of gene trees with species trees. Syst. Biol. 64:e42–e62.

Than, C. and L. Nakhleh. 2009. Species tree inference by minimizing deep coalescences. PLoS Comput. Biol. 5:e1000501.

Than, C. V. and N. A. Rosenberg. 2014. Mean deep coalescence cost under exchangeable probability distributions. Discrete Appl. Math. 174:11–26.

Wu, Y. 2012. Coalescent-based species tree inference from gene tree topologies under incomplete lineage sorting by maximum likelihood. Evolution 66:763–775.

Zhu, S., J. H. Degnan, S. J. Goldstien, and B. Eldon. 2015. Hybrid-lambda: simulation of multiple merger and kingman gene genealogies in species networks and species trees. BMC Bioinformatics 16.