Calibrated Tree Priors for Relaxed Phylogenetics and Divergence Time Estimation

JOSEPH HELED[1,*] AND ALEXEI J. DRUMMOND[1,2,3]

[1]*Department of Computer Science, University of Auckland, Auckland, New Zealand;* [2]*Bioinformatics Institute, University of Auckland, Auckland, New Zealand; and* [3]*Allan Wilson Centre for Molecular Ecology and Evolution, University of Auckland, Auckland, New Zealand;*
Correspondence to be sent to: Department of Computer Science, University of Auckland, Auckland, New Zealand;
E-mail: jheled@gmail.com

Abstract.—The use of fossil evidence to calibrate divergence time estimation has a long history. More recently, Bayesian Markov chain Monte Carlo has become the dominant method of divergence time estimation, and fossil evidence has been reinterpreted as the specification of prior distributions on the divergence times of calibration nodes. These so-called "soft calibrations" have become widely used but the statistical properties of calibrated tree priors in a Bayesian setting have not been carefully investigated. Here, we clarify that calibration densities, such as those defined in BEAST 1.5, do not represent the marginal prior distribution of the calibration node. We illustrate this with a number of analytical results on small trees. We also describe an alternative construction for a calibrated Yule prior on trees that allows direct specification of the marginal prior distribution of the calibrated divergence time, with or without the restriction of monophyly. This method requires the computation of the Yule prior conditional on the height of the divergence being calibrated. Unfortunately, a practical solution for multiple calibrations remains elusive. Our results suggest that direct estimation of the prior induced by specifying multiple calibration densities should be a prerequisite of any divergence time dating analysis. [Calibration; divergence time estimation; tree priors.]

In addition to observed sequence data, a Bayesian phylogenetic analysis can incorporate other sources of knowledge through the application of informative priors. The use of so-called "soft calibrations" (Rannala and Yang 2005) in the form of informative prior distributions on the divergence times of internal nodes during a phylogenetic analysis, has become increasingly common. This is especially the case in Bayesian phylogenetic models that support "relaxed phylogenetics," in which genetic distances are partitioned into divergence times and lineage-specific substitution rates using a relaxed molecular clock (Drummond et al. 2006).

Although these methods are now quite widely used, the statistical properties of prior distributions subject to calibration densities have not been carefully investigated, *when the ranked tree topology is a random variable*. In the relaxed phylogenetic models implemented in BEAST (Drummond and Rambaut 2007), calibration is achieved by one of three means: (i) calibration of the rate of evolution through an informative prior on the substitution rate, (ii) calibration through heterochronous data (Drummond et al. 2002; Drummond et al. 2003) (iii) calibration by specification of an informative prior distribution on the divergence time(s) of one or more internal nodes. Whereas the first two methods are relatively straightforward and have been well studied, the statistical properties of the third option in a Bayesian setting have not been well studied when the tree is a random variable.

Here we aim to highlight some of the statistical properties of calibration on internal nodes in BEAST 1.5, a commonly used Bayesian MCMC implementation, and give a new method of constructing a calibration prior that has more intuitive statistical properties when the tree is a random variable. Below we give two examples illustrating how the current implementation of calibration in BEAST may induce non-uniform prior distribution on the tree topology, and how the marginal prior distribution of the calibrated nodes may differ from the calibration density used to construct the tree prior. While the form of the calibrated tree prior can be computed directly for simple cases, in general, the precise relationship between the calibration densities used to construct the tree prior and the actual marginal priors on the calibrated nodes can only be investigated by direct simulation of the tree prior using Markov chain Monte Carlo (MCMC). When there is only a single calibration density, we introduce an alternative method of specifying the marginal prior distribution of the calibrated node directly. However, when using the existing calibration method, we recommend direction simulation of the tree prior as a standard precursor to all relaxed phylogenetic analyses involving internal node calibration densities.

The Construction of a Calibrated Tree Prior in BEAST

When calibrating the divergence times of some internal nodes, the tree prior is constructed in BEAST using three main ingredients:

1. One or more "calibration densities," each applied to the divergence time of the most recent common ancestor of a subset of the taxa.
2. A parametric "tree prior" and associated hyperparameters and hyperpriors that specify a density on the topology and all the divergence times of the tree.
3. Zero or more additional constraints on the topology in the form of subsets of taxa that are constrained to be monophyletic.

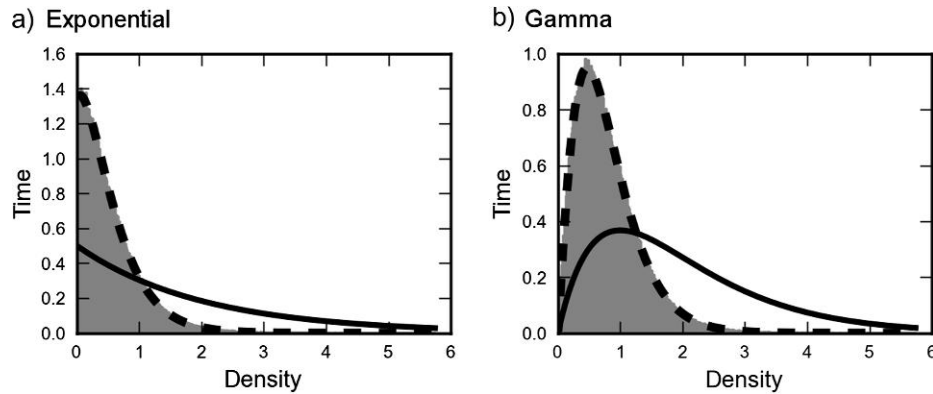In BEAST, these ingredients are combined in a particular way to construct a prior distribution on time

138

FIGURE 1.   a) Density of T_{AB} from a BEAST run with a Yule prior ($\lambda = 1$) and exponential calibration density with mean 2 (black line). The induced density (gray) matches the theoretical density (black dashed line). b) is the same as a) with a gamma ($\Gamma(k=2, \theta=1)$) calibration density.

trees. The combination of the latter two ingredients is quite unproblematic from the point of view of interpretation. The resulting distribution is simply the relevant parametric "tree prior" *conditional* on the topological constraints. Although this interpretation is simple, it is worth noting that the resulting distribution of both the divergence times and (obviously) the tree topology will differ from the unconstrained distribution.

However, the first of these ingredients can be incorporated into the model in a number of ways. A general method for computing a conditional birth–death sampling prior for a tree with a fixed ranked topology has been described (Rannala and Yang 2005), but this is not suitable when we wish to infer the topology. For the birth–death model, the special case where the node in question is the root is given in Gernhard (2008, theorem 4.1 with k = 1). In BEAST, the calibration density is combined with the tree prior by simply taking their product. This ignores the overlap in state space of the two densities and we will call this the *multiplicative construction*. In papers applying BEAST, the calibration density is often known as the "calibration prior" or the "prior on the divergence times," but we will avoid using the term prior, and use "calibration density" instead, since in the multiplicative construction this distribution does not correspond to the marginal prior distribution of the associated divergence time. If the birth rate and the calibration density are really independent sources of information about the phylogeny, then this may not be a bad method to construct the calibrated tree prior, although this construction certainly does not follow the rules of probability calculus. Specifically, the multiplicative construction is problematic in situations where the researcher expects the calibration prior to represent the marginal distribution of the calibrated node, and can lead to unexpected results.

For example, consider associating a calibration density on T_{AB}, the time of the most recent common ancestor (MRCA) of A and B in a three-taxon tree A, B, C. A Yule prior with a birth rate $\lambda = 1$ is used for the tree and an exponential density with mean 2 is used to calibrate T_{AB}. The density of T_{AB} obtained by running BEAST

using only the prior is shown in Figure 1a. The same setup with a gamma (Γ) calibration density is shown in Figure 1b.

Note that those are the expected outputs. Although difficult in general, here we can analytically compute the marginal prior density (shown as a dashed line), which exactly matches the distribution sampled by MCMC with BEAST. Also note that in case (a) the tree $((A, B), C)$ is preferred over the other two possible trees for any value of μ. This may seem counterintuitive at first because one might expect the pairing of A and B to be less probable when the mean of T_{AB} is larger than the expected height of the tree. See Appendix 1 for more details and other examples.

A Multiple Calibration Data Set of Marsupials

As the number of calibration densities used to construct the tree prior grows, it becomes hard to describe the joint prior on the calibrated node heights or their individual marginal priors analytically. However, it is always possible to examine the mismatch between the specified calibration densities and marginal densities that result from the multiplicative construction. In Phillips et al. (2009), sequences of seven nuclear genes and the complete mitochondrial (mt) genome protein-coding and RNA-coding DNA sequences for seven placental mammals, three marsupials, two monotremes, and two sauropsids were analyzed, aimed at dating the echidna–platypus divergence. Here, we have rerun the MCMC analysis without the sequence data to show the marginal distributions that result from the multiplicative construction (Figs 2 and 3) on the eight calibrated nodes, alongside the calibration densities specified. We follow the authors by constraining mammals and sauropsids to be monophyletic. Because no specific prior was placed on the birth rate (implied improper prior between zero and infinity), it can be seen that the root height distribution almost matches the calibration density, but most of the others show strong modality, due to the interaction between the calibration densities and the topological constraints. When all
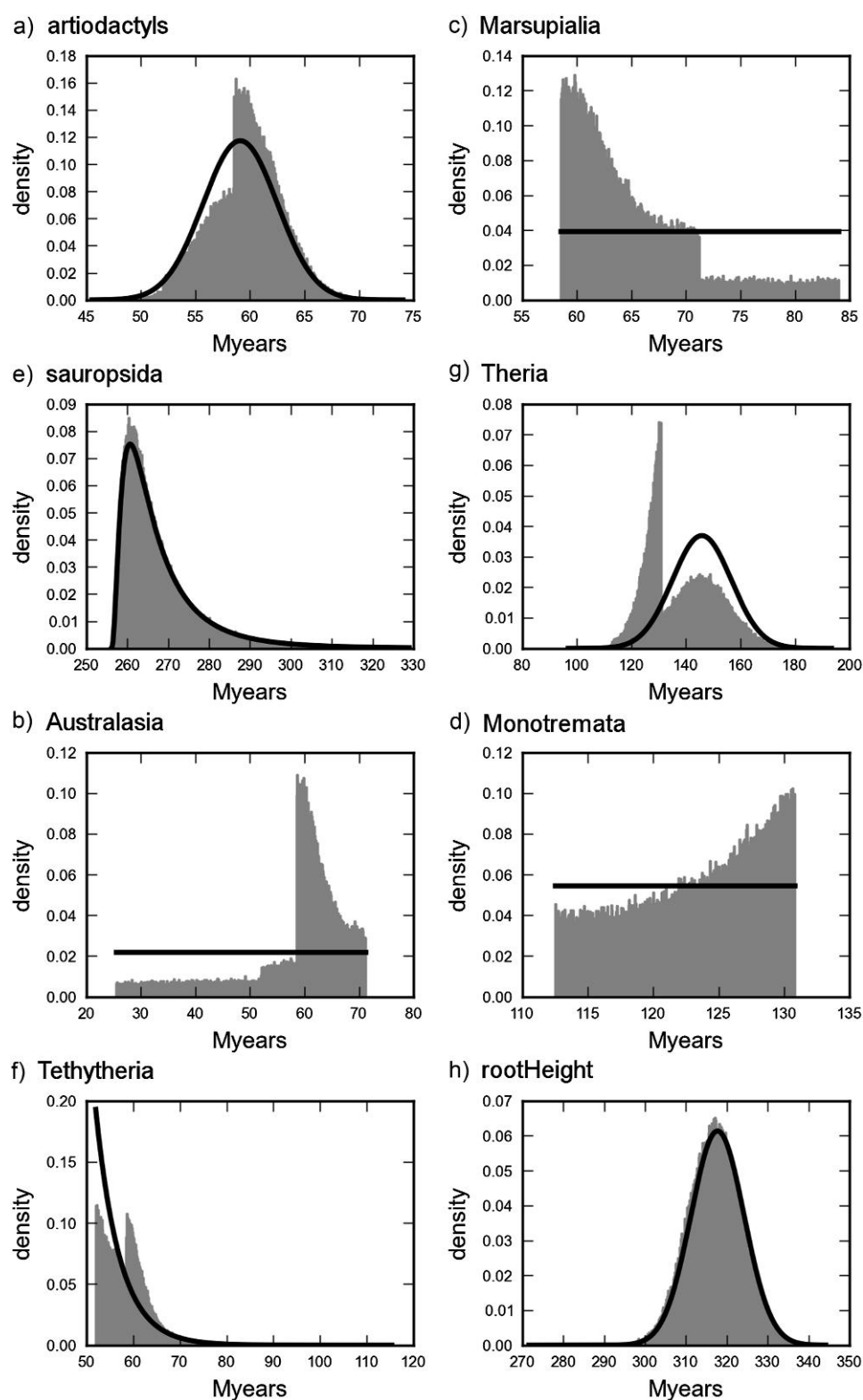
FIGURE 2. The marginal prior distributions that result from the multiplicative construction (gray) versus calibration densities (black line) specified for the calibrated nodes from Phillips et al. (2009). The marginal prior distributions were obtained from an MCMC run using the prior only. The calibration densities are as defined by the authors.
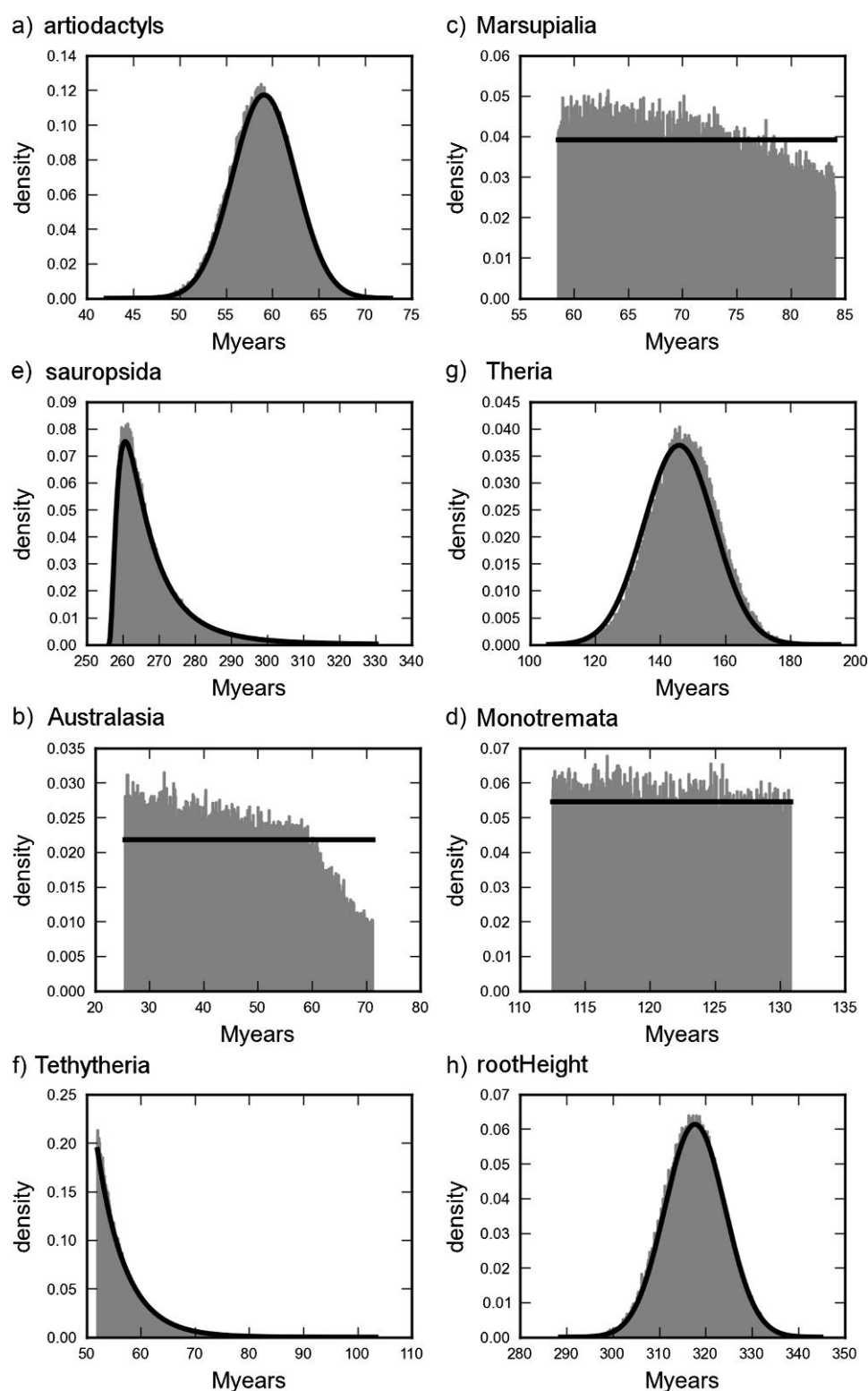
FIGURE 3.   The marginal prior distributions that result from the multiplicative construction (gray) versus calibration densities (black line) specified for the calibrated nodes from Phillips et al. (2009), when monophyly is enforced for all calibrated nodes. The marginal prior distributions were obtained from an MCMC run using the prior only. The calibration densities are as defined by the authors.

of the calibrated nodes have monophyly enforced, the marginal prior distributions are much better matches to the calibration densities, but there are still small discrepancies evident (Fig 3).

<div align="center">CONSTRUCTING A TREE PRIOR WITH AN ARBITRARY
MARGINAL DISTRIBUTION ON THE TIME OF AN
INTERNAL NODE</div>

What are the desired properties of a calibrated tree prior? First, we would like the marginal density of the calibrated node to match the calibration density; and second, conditional on the calibrated node height, we want two trees to have relative prior densities proportional to some sensible generative process like the Yule, birth–death sampling (Stadler 2009, 2010) or coalescent (Kingman 1982; Griffiths and Tavare 1994) tree prior.

Let $\tau(g)$ be the *time of the most recent common ancestor* (TMRCA) for calibrated taxa on genealogy g from the space of all genealogies G. Consider the function $\rho_G(g)$, a candidate for a calibrated tree prior density on the space of genealogies and $\rho_T(\cdot)$, the desired marginal calibration density on $\tau(g)$. The following properties are desired:

I The marginal density on the calibrated node is equal to the calibration density:

$$\rho_T(x) = \int_{\substack{g \in G \\ \tau(g)=x}} \rho_G(g) = \int_{g \in G} 1(\tau(g) = x)\rho_G(g)dg \quad (1)$$

In words, the total density of all trees for which the TMRCA of calibrated taxa is x equals the calibration density at x. The integral is written first informally, then formally using the indicator function $1(\cdot)$, which is equal to 1 when the argument is true, 0 otherwise. Note that this is an integral over all genealogies, or time trees. When the genealogy is represented as $g = \{\psi, h\}$, where ψ is the ranked topology and h is a vector of internal node heights in order, the integral can be written as

$$\rho_T(x) = \sum_{\psi \in \Psi} \int 1(\tau(\{\psi, h\}) = x)\rho_G(\{\psi, h\}) \, dh. \quad (2)$$

II When restricted to a subset of trees with equal calibrated node height, the density is proportional to the uncalibrated target density f_G:

$$\tau(g_1) = \tau(g_2) \implies \frac{\rho_G(g_1)}{\rho_G(g_2)} = \frac{f_G(g_1)}{f_G(g_2)}. \quad (3)$$

Constructing $\rho_G : G \to \mathbb{R}$, which satisfies (I) and (II) is quite easy:

$$\rho_G(g) = f_G(g) \frac{\rho_T(\tau(g))}{f_T(\tau(g))}, \quad (4)$$

where $f_T(\cdot)$ is the marginal distribution of τ under f_G. We call this the *conditional construction*. Informally, equation 4 can be written as

$$\text{new joint prior} = \text{old joint prior} \times \frac{\text{new marginal}}{\text{old marginal}}.$$

It is easy to see that $\rho_G(\cdot)$ satisfies (I) and (II), and integrates to 1. For genealogies with equal calibration, the calibration and marginal are equal, so their ratio is f_G as the other two terms cancel (II). And when integrating over trees with equal calibration, the calibration and marginal can be moved out of the integral, which leaves only the f_G term inside, which then cancels with the marginal, leaving the calibration (I).

The conditional construction is useful in practice only if the marginal density $f_T(\cdot)$ can be computed efficiently

$$f_T(x) = \int_{g \in G} 1(\tau(g) = x)f_G(g)dg. \quad (5)$$

Note that the domain of $f_G(\cdot)$ may depend on the conditions imposed on g. If taxa ϕ are not required to be monophyletic, the domain is all genealogies (G). When ϕ is required to be monophyletic, the domain is all genealogies that have ϕ as a monophyletic clade (G_ϕ).

<div align="center">*Yule Tree Prior on Four Taxa with One
Monophyletic Calibration*</div>

We now describe in detail how to compute the marginal prior (equation 5) for the Yule tree prior with calibration on a 2-taxon monophyletic clade (A,B) in a 4-taxon tree. We then show how the same can be done in the general case for Yule tree prior on n taxa and one calibration. In the following sections we shall use f_Y for the uncalibrated density instead of f_G to make it clear that the results involving f_Y are specific to the Yule prior.

There are four ranked trees (Fig. 4). One, in which T_{CD} is lower than T_{AB} (Case 1), and three ranked trees where T_{AB} has the most recent divergence time (Case 2).

Let $T = (T_1, T_2, T_3)$ be the intracoalescent time intervals. For example, in Case 1, the interval between the leaves and $T_3 = T_{CD}$, $T_2 = T_{AB} - T_{CD}$, and so on. Under the Yule prior, each interval is distributed exponentially, $T_{i-1} \sim \text{Exp}(i\lambda)$, and the joint density for T is

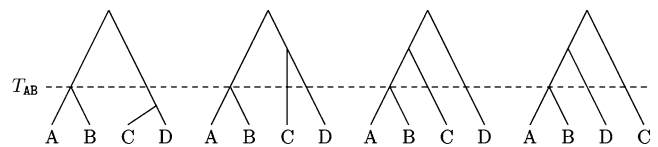$$f_Y(t) = 24\lambda^3 e^{-\lambda(4t_3 + 3t_2 + 2t_1)}.$$

FIGURE 4. The four ranked trees with four taxa and (A,B) monophyletic. τ_{AB} is the same in all cases.

Because for Case 1 we have $T_{AB} = T_2 + T_3$, the marginal density is given by

$$u^{(1)}(t_{AB}) = \int_0^\infty \int_0^{t_{AB}} f_Y(t_1, t_{AB} - t_3, t_3) \, dt_3 \, dt_1$$

$$= \int_0^\infty \int_0^{t_{AB}} 24\lambda^3 e^{-\lambda(4t_3 + 3(t_{AB} - t_3) + 2t_1)} \, dt_3 \, dt_1$$

$$= 12\lambda(e^{-3\lambda t_{AB}} - e^{-4\lambda t_{AB}}). \tag{6}$$

Note the range $[0, t_{AB}]$ in the integral of t_3, which keeps the branch length positive. For Case 2, we obtain

$$u^{(2)}(t_{AB}) = \int_0^\infty \int_0^\infty f_Y(t_1, t_2, t_{AB}) \, dt_2 \, dt_1 = 4\lambda e^{-4\lambda t_{AB}}. \tag{7}$$

Because there are three ranked trees with density $u^{(2)}$ and one with $u^{(1)}$, the marginal Yule distribution is given by

$$f(t_{AB}) = \frac{1}{4}u^{(1)}(t_{AB}) + \frac{3}{4}u^{(2)}(t_{AB}) = 3\lambda e^{-3\lambda t_{AB}}. \tag{8}$$

Yule tree on four taxa with one calibration prior, no monophyly.—The construction for the monophyletic clade can be adapted to placing a calibration on T_{AB} without enforcing monophyly. Instead of two cases we have three: A, B is monophyletic (Case I), the common ancestor of A, B has three descendants (Case II), and the common ancestor of A, B is the root (Case III). We already have the densities when A, B is monophyletic, and the density for Case II is given by equation (6). We still need a density for Case III:

$$u^{(3)}(t_{AB}) = \int_0^{t_{AB}} \int_0^{t_{AB} - t_3} f_Y(t_{AB} - t_3 - t_2, t_2, t_3) \, dt_2 \, dt_3$$

$$= 12\lambda e^{-2\lambda t_{AB}}(1 - e^{-\lambda t_{AB}})^2.$$

The three densities are combined by weighting them according to the number of ranked topologies to which they apply. For Case I we have, as before, 1 and 3 ranked topologies with densities $u^{(1)}$ and $u^{(2)}$. For Case II, there are 4 ranked topologies with density $u^{(1)}$, and for Case III there are 10 with density $u^{(3)}$. Together we get

$$f(t_{AB}) = \frac{10u^{(3)} + 3u^{(2)} + 5u^{(1)}}{10 + 3 + 5}$$

$$= \frac{\lambda e^{-2\lambda t_{AB}}}{3}(12e^{-2\lambda t_{AB}} - 30e^{-\lambda t_{AB}} + 20).$$

Yule tree prior with one monophyletic calibration prior

The four taxa case can be generalized to any monophyletic clade ϕ of size n_c in an $n = n_c + n_o$ taxa tree. Formally, the genealogy g is a pair $g = \{\psi, h\}$, where ψ is the

ranked topology and h is a vector of the internal node heights in reverse order, $h = (h_1, h_2, \ldots, h_{n-1})$. Because ϕ is monophyletic, $\psi \in \Psi_\phi$, where Ψ_ϕ is the set of ranked topologies that contain the monophyletic clade ϕ.

Now, the Yule density for the heights (Gernhard (2008) section 6.1) is equally divided between all ranked trees having those heights. Because there are $|\Psi_\phi|$ of them, the density for the genealogy g is,

$$f_Y(g) = \frac{1}{|\Psi_\phi|} n! \, e^{-\lambda h_1} \prod_{i=1}^{n-1} \lambda e^{-\lambda h_i}. \tag{9}$$

Define $i(\psi)$ as the rank of the MRCA of ϕ. The marginal Yule density is given by

$$f_{T_\phi}(x) = \sum_{\psi \in \Psi_\phi} \int_{h_1=x}^\infty \cdots \int_{h_{i(\psi)-1}=x}^{h_{i(\psi)-2}} \int_{h_{i(\psi)+1}=0}^x \tag{10}$$

$$\cdots \int_{h_{n-2}=0}^{h_{n-3}} \int_{h_{n-1}=0}^{h_{n-2}} f_Y(\{\psi, h\})$$

Surprisingly, this multiple integral evaluates to a simple expression that depends only on the size of ϕ (n_c) and does not depend on n (Appendix 3)

$$f_{T_\phi}(x) = \begin{cases} \frac{1}{2}(n_c^3 - n_c)\lambda e^{-3\lambda x}(1 - e^{-\lambda x})^{n_c-2} & \text{if } n_c < n, \\ n_c(n_c - 1)\lambda e^{-2\lambda x}(1 - e^{-\lambda x})^{n_c-2} & \text{if } n_c = n. \end{cases} \tag{11}$$

Yule tree prior with one calibration prior, no monophyly.— Deriving the conditional density for the age of the most recent common ancestor of a subset of taxa ϕ, without the constraint of monophyly of ϕ is more involved. Again, assume a subset ϕ of size n_c in an $n = n_c + n_o$ taxa tree. The conditional density is broken into $n_o + 1$ cases:
Case 0: taxa set ϕ is monophyletic,
Case 1: the most recent common ancestor of taxa ϕ has $n_c + 1$ descendants,
Case 2: the most recent common ancestor of taxa ϕ has $n_c + 2$ descendants,
$\vdots$
Case n_o: the most recent common ancestor of taxa ϕ is the root.

Because in Case k we have a monophyletic clade of size $k + n_c$, the density for that case is given by $f_{T_{\phi_k}}(x)$ (11), where ϕ_k is of size $k + n_c$. Note that in (11) only the size of the clade matters. To combine the densities from all cases into an overall density we need the number of ranked topologies for each case. Those counts, when scaled to add to 1, act as the coefficients w_k in the final equation

$$\tilde{f}_{T_\phi}(x) = \sum_{k=0}^{n_o} w_k f_{T_{\phi_k}}(x). \tag{12}$$

For the derivation of w_k, see (14) in Appendix 2.

Note that the formula works for the special case $n_c = 1$, that is when we wish to condition on the time a particular taxon "attaches" to the tree. In that case, the marginal density has the simple form $2\lambda e^{-2\lambda x}$.

REVISITING AN ANALYSIS OF THE CHACMA BABOON (PAPIO URSINUS)

Sithaldeen et al. (2009) provide a phylogenetic analysis of the Chacma Baboon sequences sampled across the entire range of the species. The authors analyze 52 mtDNA samples, using a Yule prior coupled with calibration densities on the root and two nested monophyletic clades. Although this calibrated tree prior has multiple calibration densities and therefore does not fall under the cases previously described, we can derive the marginal density in this particular case using the same methods (Appendix 4).

Before applying the new prior, we run both the original analysis and a prior-only version. Figure 5a shows the calibration density as specified in the BEAST analysis for the three nodes, together with the induced density from the prior-only run and the posterior values from the full analysis. We can clearly see that the posterior values match the induced prior almost
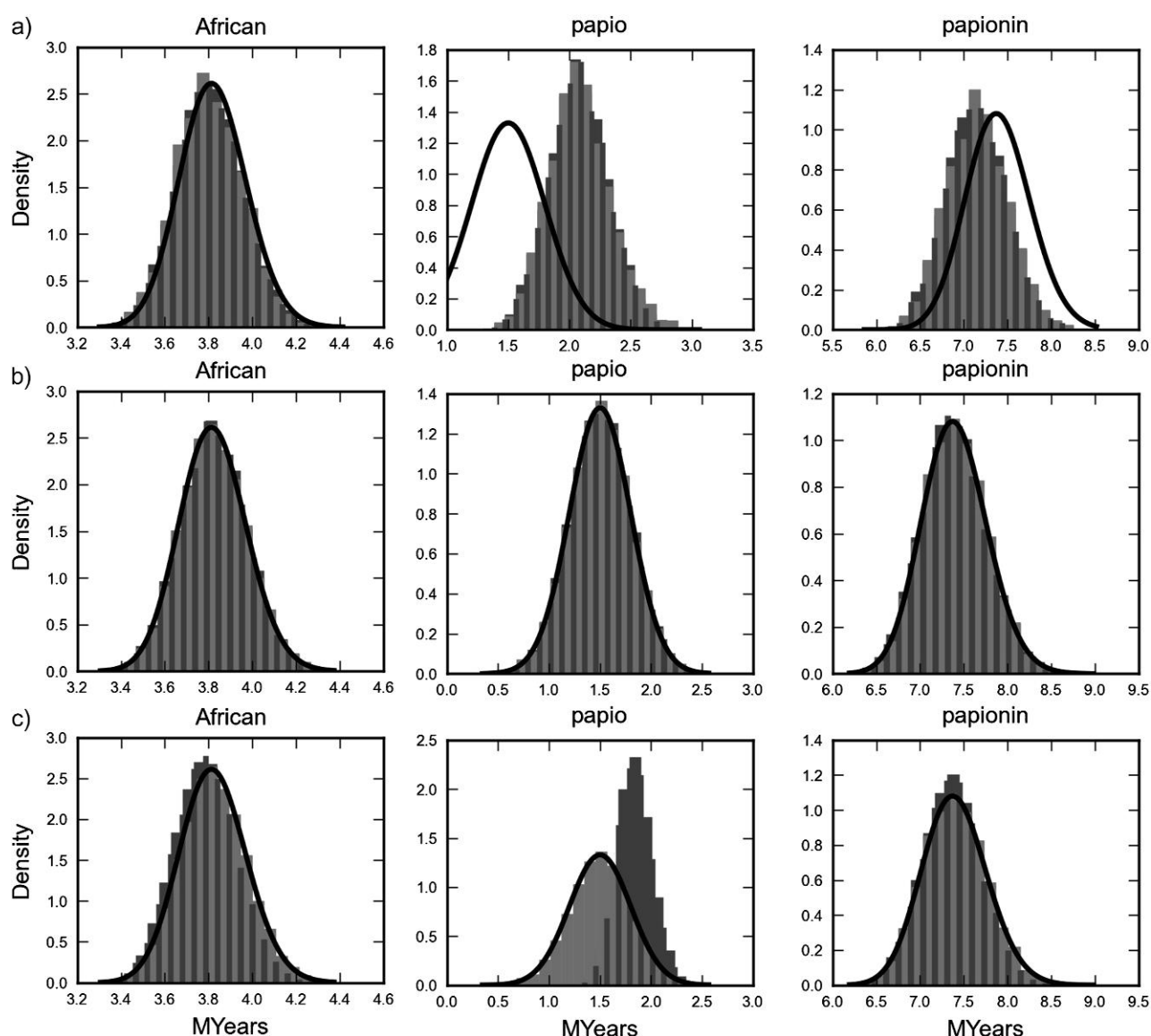

FIGURE 5. a) Induced marginal prior distributions (gray) versus calibration densities (black line) versus marginal posterior distributions (dark gray) for the calibrated nodes from (Sithaldeen et al. 2009) using the multiplicative construction. The induced marginal prior distributions were obtained from an MCMC run using the prior only. The calibration densities are as defined by the authors. b) Similar to (a) but using the conditional-construction. The marginal distributions match the calibration densities as expected. c) Similar to (b), using the conditional construction and a strict molecular clock.

to perfection, and that the induced prior is shifted in varying degrees from the calibration priors due to the interaction with the Yule prior. It is not really surprising that the analysis was able to "match" all three marginal divergence time priors because it used a relaxed clock with a wide and flat prior on the rate mean and variance, accommodating branch rate/time combinations whose products satisfy the desired branch length in substitutions while also producing branch times that closely match the marginal tree prior on the calibrated nodes.

When using the conditional construction, the calibrated density matches the MCMC-sampled prior as expected, and the posterior from the full analysis is almost identical to the prior, presumably because of the relaxed clock (Fig. 5b). However, with the use of a strict clock and the conditional construction we can see that, whereas the sampled prior matches the specified marginal calibration densities, the posterior distribution shows how this prior knowledge has been updated by the data (Fig 5c). That we did not see this with the relaxed clock suggests we may have overparameterized the model in that case.

Given that the data had no visible effect on the posterior distribution of the calibrated divergence times, it is reasonable to assume that in the relaxed clock analysis the prior plays a significant role in the divergence times of the noncalibrated nodes as well. This is indeed the case. Figure 6 shows the trees with the key divergence times from the original, multiplicative construction run (Fig. 6a), and the tree from the conditional construction run (Fig. 6b). We can see that the lowest calibration node in the conditional construction run matches the expected mean of the calibration prior, and as a result all the divergence times in the subtree below have earlier times as well.

DISCUSSION AND CONCLUSIONS

It is sometimes possible to construct a calibrated tree prior that factorizes precisely into a standard process-based tree prior conditional on the divergence times of the calibrated nodes and an independent marginal prior on those divergence times. We have demonstrated this for one calibrated node in the Yule prior. In order to produce such a conditional construction, one simply needs to be able to efficiently calculate the marginal distribution of the calibrated nodes under the uncalibrated tree prior of choice.

Other conditions on tree priors are also possible. For example, conditioning on the root height of the tree is fairly straightforward for both the Yule model and the more general birth–death model of speciation (Stadler, T. *personal communication*). In fact, the original description of both the Yule and birth–death models in a phylogenetic context were in the form of a conditioning on the root height. However, those formulations did not condition on the number of taxa, which is also required. Nevertheless, arriving at a Yule probability density conditional on both the root height and the number of taxa is straightforward from that earlier work.
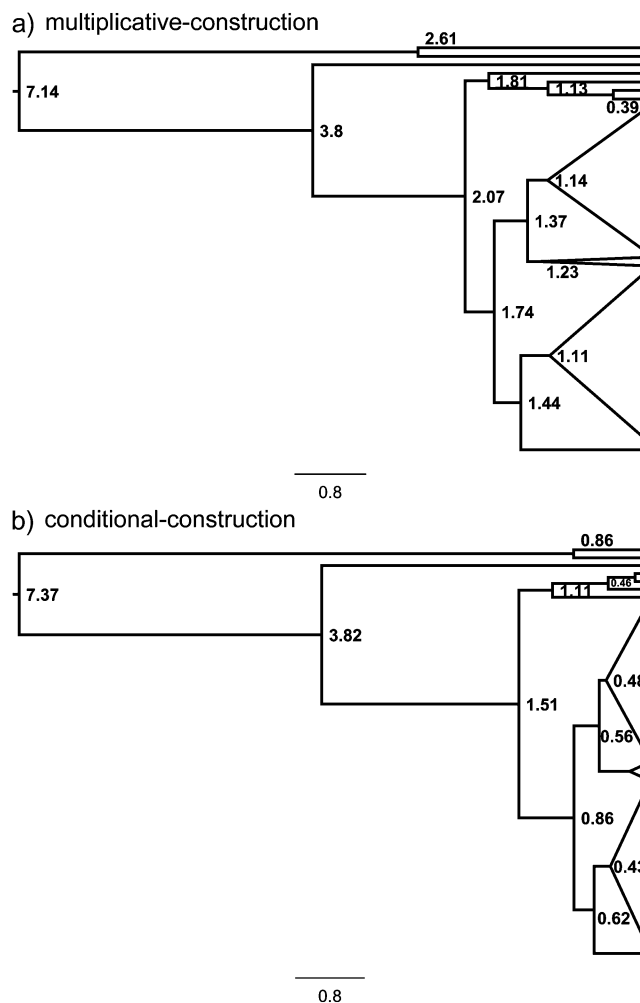

FIGURE 6. a) Posterior estimates for divergence times using the data from Sithaldeen et al. (2009) and the multiplicative construction. b) Posterior estimates for the same data using the conditional construction and relaxed molecular clock as for (a).

We are fairly confident that the methods presented here can be extended to handle multiple marginal prior distributions on internal nodes. However, the formulas for the conditional densities grow rapidly in size as a function of the number of conditions and taxa. As a result, the determination and evaluation of those conditional priors may become a practical problem.

The method that BEAST implements for constructing calibrated tree priors can lead to marginal distributions on calibrated nodes that are very different than the calibration densities chosen, as seen in Figure 2. In practice, any multiple calibration analysis should always involve direct computation of the calibrated tree prior (by MCMC), and preferably report the actual marginal calibration prior for nodes of interest. Finally, in general, both multiplicative construction and the conditional construction produce nonuniform distributions on the (ranked) topology.

REFERENCES

Drummond A., Ho S., Phillips M., Rambaut A. 2006. Relaxed phylo-genetics and dating with confidence. PLoS Biol. 4:e88.
Drummond A., Rambaut A. 2007. Beast: Bayesian evolutionary analy-sis by sampling trees. BMC Evol. Biol. 7:214.
Drummond A.J., Nicholls G.K., Rodrigo A.G., Solomon W. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics. 161:1307–1320.
Drummond A.J., Nicholls G.K., Rodrigo A.G., Solomon W. 2003. Genealogies from time-stamped sequence data. In: Buck, C.E.Millard A.R., editors. Tools for constructing chronologies: crossing disciplinary boundaries, Lecture Notes in Statistics. Vol. 177, Chapter 7, p. 149–174. London: Springer.
Gernhard T. 2008. The conditioned reconstructed process. J. Theor. Biol. 253:769–778.
Griffiths R., Tavare S. 1994. Sampling theory for neutral alleles in a varying environment. R. Soc. Lond. Philos. Trans. B. 344:403–410.
Kingman J. 1982. The coalescent. Stoch. Process.Appl. 13:235–248.
Phillips M., Bennett T., Lee M. 2009. Molecules, morphology, and ecol-ogy indicate a recent, amphibious ancestry for echidnas. Proc. Natl. Acad. Sci. U.S.A. 106:17089–17094.
Rannala B., Yang Z. 2005. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol. Biol. Evol. 23:212–226.
Sithaldeen R., Bishop J., Ackermann R. 2009. Mitochondrial DNA analysis reveals Plio-Pleistocene diversification within the chacma baboon. Mol. Phylogenet. Evol. 53:1042–1048.
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. 2010. Sampling-through-time in birth-death trees. J. Theor. Biol. 267:396–404.

APPENDIX 1

Examples of Calibrated Tree Priors Using the multiplicative Construction

In the multiplicative construction used by BEAST 1.5, the tree topology and divergence times are influenced both by the calibration density and by the birth rate (λ) of the Yule model of tree branching. These two sources of information are combined to construct a prior density on the tree.

For our first example, we consider associating a cal-ibration density with T_{AB}, the time of the MRCA of A and B in a three-taxon tree A, B, C. A Yule prior with a birth rate λ is used for the tree and an exponential den-sity with mean $1/\mu$ is used to calibrate T_{AB}. Label with T_2 is the time of the youngest internal node and with T_1, the time between the root and the youngest inter-nal node. Obviously, because T_{AB} is a function of T_1, T_2, the Yule model already implies a marginal density on

T_{AB}, so the Yule density and the calibration density share state space and are not independent.

In BEAST 1.5, this is ignored and the density of the Yule model is simply multiplied by the calibration den-sity to form the multiplicative construction. This prod-uct gives the following density on the spaces of trees:

$$f(t_1, t_2, \psi) = \begin{cases} \frac{1}{z}2\lambda e^{-2\lambda t_1}3\lambda e^{-3\lambda t_2}\mu e^{-\mu t_2} & \text{when } \psi = ((A,B),C) \\ \frac{1}{z}2\lambda e^{-2\lambda t_1}3\lambda e^{-3\lambda t_2}\mu e^{-\mu(t_1+t_2)} & \text{when } \psi = ((A,C),B) \\ & \text{or } ((B,C),A), \end{cases}$$

where the terms involving λ come from the Yule prior and the terms involving μ from the calibration. Since the product of two densities is not in general a proper den-sity, the expressions have to be normalized by the con-stant factor $1/z$. The resulting probabilities of the three possible topologies are,

$$f((\text{A},\text{B}),\text{C}) = \frac{1}{z}\int_0^\infty\int_0^\infty 2\lambda e^{-2\lambda t_1}3\lambda e^{-3\lambda t_2}\mu e^{-\mu t_2}\, dt_1 dt_2$$

$$= \frac{1}{z}\frac{3\lambda\mu}{3\lambda+\mu}$$

$$f((\text{A},\text{C}),\text{B}) = f((\text{B},\text{C}),\text{A})$$

$$= \frac{1}{z}\int_0^\infty\int_0^\infty 2\lambda e^{-2\lambda t_1}3\lambda e^{-3\lambda t_2}\mu e^{-\mu(t_1+t_2)}\, dt_1 dt_2$$

$$= \frac{1}{z}\frac{6\lambda^2\mu}{(3\lambda+\mu)(2\lambda+\mu)}$$

The constant factor z is easily computed because the three integrals have to sum to 1, giving $z = \frac{3\lambda\mu(6\lambda+\mu)}{(2\lambda+\mu)(3\lambda+\mu)}$. So, the relative ratio of the three topologies is $2\lambda+\mu:2\lambda:2\lambda$ and tree ((A,B),C) is preferred for any value of μ. Fur-thermore, it can be shown that under the multiplica-tive construction $E[T_2] = 1/{\mu+3\lambda}$ instead of $1/{3\lambda}$ under the Yule, whereas $E[T_1] = 1/{2\lambda}(1 - 4\mu\lambda/{(\mu+2\lambda)(\mu+6\lambda)})$ instead of $1/{2\lambda}$.

One may think that placing a calibration density on the nonmonophyletic clade is the cause of the problem. However, we can repeat the calculation for a four-taxon tree while enforcing monophyly of A, B. In one of the four possible topologies (first left in (Fig. 4), the TMRCA of (A, B) is the larger of the two internal nodes ($T_{AB} = T_2 + T_3$) and is the smaller of the two in the other three cases ($T_{AB} = T_3$). The total densities for those two cases are

$$f_3(t_{AB}) = \int_0^\infty\int_0^\infty\int_0^\infty 2\lambda e^{-2\lambda t_1}3\lambda e^{-3\lambda t_2}\lambda e^{-4\lambda t_3}\mu e^{-\mu t_3}\, dt_1 dt_2 dt_3$$

$$= \frac{4\lambda\mu}{\mu+4\lambda}$$

$$f_{23}(t_{AB}) = \int_0^\infty \int_0^\infty \int_0^\infty 2\lambda e^{-2\lambda t_1} 3\lambda e^{-3\lambda t_2} \lambda e^{-4\lambda t_3} \mu e^{-\mu(t_2+t_3)} \, dt_1 dt_2 dt_3$$

$$= \frac{12\lambda^2\mu}{(\mu+3\lambda)(\mu+4\lambda)}.$$

Now, because there are two ranked topologies with the unranked topology $((A,B),(C,D))$, the ratio is

$$f\big(\big(\big((A,B),C\big),D\big)\big) : f\big(\big((A,B),(C,D)\big)\big)$$

$$= \frac{f_3(t_{AB})}{f_3(T_{AB}) + f_{23}(T_{AB})}$$

$$= \frac{\mu+3\lambda}{\mu+6\lambda}.$$

So, a ratio of $\mu + 6\lambda : \mu + 3\lambda : \mu + 3\lambda$ is obtained for the three topologies $((A,B),(C,D))$, $(((A,B),C),D)$ and $(((A,B),D),C)$. Again, the first topology is preferred regardless of μ.

Even when restricting the three-taxon tree to a single topology by enforcing monophyly, the induced prior on divergence times depends on the specific interaction between the tree prior and the calibration density. Consider a Yule prior with birth rate λ and a gamma prior with shape 2 and scale θ ($\Gamma(2,\theta)$) on T_{AB}. The expected divergence time under this combination can be shown to be $\frac{2\theta}{1+3\lambda\theta}$; which would always be less than the mean of the calibration density, 2θ. Finally, instead of fixing λ we can assume a hyperprior on λ—a very common practice in BEAST. This results in an increase in the dimensionality of the state space, and when deriving expectations or clade probabilities, we need to integrate over the divergence times and λ to obtain the constant normalizing factor. To compute the expected divergence time of (A,B), where λ has a uniform hyperprior of $[0,N]$, we first derive the constant normalization factor,

$$z = \int_0^N \int_0^\infty \int_0^\infty 2\lambda e^{-2\lambda t_1} 3\lambda e^{-3\lambda t_2} \frac{t_2 e^{-t_2/\theta}}{\theta^2} \frac{1}{N} dt_1 dt_2 d\lambda$$

$$= \int_0^N \frac{3\lambda}{(3\lambda\theta+1)^2 N} d\lambda$$

$$= \frac{(3N\theta+1)\log(3N\theta+1) - 3N\theta}{3(3N\theta+1)N\theta^2}.$$

The expectation under the multiplicative construction is

$$E[T_2] = \frac{1}{z} \int_0^N \int_0^\infty \int_0^\infty t_2 \, 2\lambda e^{-2\lambda t_1} 3\lambda e^{-3\lambda t_2} \frac{t_2 e^{-t_2/\theta}}{\theta^2} \frac{1}{N} dt_1 dt_2 d\lambda$$

$$= \frac{1}{z} \int_0^N \frac{6\lambda\theta}{(3\lambda\theta+1)^3 N} d\lambda$$

$$= \frac{9N^2\theta^3}{(3N\theta+1)((3N\theta+1)\log(3N\theta+1) - 3N\theta)}$$

$$\approx \frac{\theta}{\log(3N\theta+1) - 1}, \text{ for large } N,$$

with $N = 100$, the average divergence time is approximately $\frac{\theta}{\log(300\theta+1)-1}$, which is less than θ for any $\theta > 0.006$.

APPENDIX 2
Number of Ranked Topologies for a Nonmonophyletic Clade

Here we derive the coefficients w_k used in the formula for calculating the conditional density for the time of the MRCA of the nonmonophyletic taxa set ϕ. The coefficient w_k is the ratio of r_k, the number of ranked topologies for case k, to the total number $\mathcal{R}_n$ of ranked topologies for an $n = n_c + n_o$ taxa tree.

$$w_k = \frac{r_k}{\mathcal{R}_n} \qquad\qquad \left(\sum_k w_k = 1\right)$$

$$\mathcal{R}_n = \prod_{i=2}^n \binom{i}{2}$$

Here, r_k is the number of ranked topologies, where n_c taxa are part of a $n_c + k$ taxa subtree. The number is the product of

(i) the number of ways to choose k taxa from n_o to be part of the clade ϕ;
(ii) $\mathcal{C}_{k,n_c}$, the number of ranked trees with $n_c + k$ taxa whose common ancestor of ϕ is the root; and
(iii) $\mathcal{D}_{n_o-k, n_c+k}$, the number of ways to combine the remaining $n_o - k$ taxa with the clade in (i).

(i) is simply $\binom{n_o}{k}$. For (ii), we start with the $\mathcal{R}_{n_c}$ ways to coalesce n_c lineages. For each of those, we can add the remaining k in some fixed order. The first lineage can attach itself to $2 + 3 + \cdots + n_c = \binom{n_c+1}{2} - 1$ places to create a different ranked topology, the second to $2 + 3 + \cdots + n_c + (n_c + 1) = \binom{n_c+2}{2} - 1$, and so on, giving

$$\mathcal{C}_{k,n_c} = \mathcal{R}_{n_c} \prod_{i=n_c+1}^{k+n_c} \left[\binom{i}{2} - 1\right].$$

Let $\mathcal{D}_{l,m}$ be the number of ways to combine l lineages and a fixed subtree with m lineages. Examine the possible choices for the first coalescent: Either two of the l lineages are joined ($\binom{l}{2}$ ways), or this is a coalescent in the subtree. This observation leads to the following recursive formula:

$$\mathcal{D}_{l,m} = \mathcal{D}_{l,m-1} + \binom{l}{2}\mathcal{D}_{l-1,m}.$$

With the initial condition $\mathcal{D}_{0,m} = 1$ and $\mathcal{D}_{l,1} = \mathcal{R}_{l+1}$. It is easy to show the above has the solution

$$\mathcal{D}_{l,m} = \binom{l+m}{l-1} \mathcal{R}_l. \tag{13}$$

Substituting $n_o - k$ for l and $n_c + k$ for m gives the required count for (iii). All three put together give

$$r_k = \binom{n_o}{k} \prod_{i=3}^{n_c} \binom{i}{2} \prod_{i=n_c+1}^{k+n_c} \left[\binom{i}{2} - 1 \right] \prod_{i=2}^{n_o+1-k} \frac{i(i+n_c+k)}{2}. \tag{14}$$

Computational Note

The counts r_k are large and we need to evaluate w_k directly. Some tedious manipulations result in an expression that does not involve large numbers

$$w_k = \binom{n_o+n_c}{2}^{-1} \prod_{i=0}^{k-1} \frac{n_o - i}{i+1} \left(1 - \frac{2}{(i+n_c)(i+n_c+1)} \right)$$

$$\prod_{i=k}^{n_o-2} \frac{(i-k+2)\left(1+\frac{2}{i+n_c}\right)}{i+n_c+1}.$$

APPENDIX 3

Conditional Density of the TMRCA for a Monophyletic Clade ϕ

Here we derive the simple form of the marginal Yule density when the genealogy has a single monophyletic clade ϕ of size n_c in a tree with n taxa.

First note that the total number of those genealogies can be obtained from equation (13)

$$|\Psi_\phi| = \mathcal{R}_n^{(n_c)} = \mathcal{R}_{n_c} \mathcal{D}_{n-n_c,n_c}.$$

Partition all ranked topologies according to $i_\phi(\psi) = k + 1$, that is group together topologies having k heights above the root of ϕ.

$$f_{T_\phi}(x) = \sum_{k=1}^{n_o} \sum_{\substack{\psi \in \Psi_\phi \\ i_\phi(\psi)=k}} \int_{h_1=x}^{\infty} \cdots \int_{h_{k-1}=x}^{h_{k-2}} \int_{h_{k+1}=0}^{x}$$

$$\cdots \int_{h_{n-2}=0}^{h_{n-3}} \int_{h_{n-1}=0}^{h_{n-2}} f_Y(\{\psi, h\}).$$

Under both conditions, the multiintegral has the same value in each case. The integrals can be separated into two independent groups, the $n - k - 2$

heights below x ($n_c - 2$ from ϕ, $n_o - k$ from outside), and the k heights above x. The first group integrates to $\frac{1}{(n-k-2)!}(1 - e^{-\lambda x})^{n-k-2}$, the second to $\frac{1}{(k+1)!}(e^{-\lambda x})^{k+1}$. Both from the simple observation that the integral on k unrestricted heights is $k!$ times the integral on the order statistic. The root of ϕ contributes $\lambda e^{-\lambda x}$, giving

$$f_{T_\phi}(x) =$$

$$\lambda \sum_{k=1}^{n_o} \sum_{\substack{\psi \in \Psi_\phi \\ i_\phi(\psi)=k}} \frac{1}{\mathcal{R}_n^{(n_c)}} \frac{n!}{(n-k-2)!\,(k+1)!}$$

$$\times (1 - e^{-\lambda x})^{n-(k+2)} (e^{-\lambda x})^{k+2}$$

$$= \lambda \sum_{k=1}^{n_o} \left(\frac{1}{\mathcal{R}_n^{(n_c)}} \frac{n!}{(n-k-2)!\,(k+1)!} \right.$$

$$\left. \times (1 - e^{-\lambda x})^{n-(k+2)} (e^{-\lambda x})^{k+2} \right) \sum_{\substack{\psi \in \Psi_\phi \\ i_\phi(\psi)=k}} 1.$$

The last step is possible because none of the terms depend on the specific topology. The number of ranked topologies under our criteria is,

$$\sum_{\substack{\psi \in \Psi_\phi \\ i_\phi(\psi)=k}} 1 = \binom{n-k-2}{n_c-2} \mathcal{R}_{k+1} \mathcal{R}_{n_c} \mathcal{R}_{n_o,k},$$

where $\mathcal{R}_{n,k} = \prod_{i=k+1}^{n} \binom{i}{2}$, the number of ranked ways to reduce n lineages to k.

Now it is straightforward (but tedious) to show that

$$\frac{n!}{(n-k-2)!\,(k+1)!} \binom{n-k-2}{n_c-2} \mathcal{R}_{k+1} \mathcal{R}_{n_c} \mathcal{R}_{N_o,k}$$

$$= \frac{n_c^3 - n_c}{2} \binom{n_o-1}{k-1} \mathcal{R}_n^{(n_c)}.$$

After replacing the above and factoring out,

$$f_{T_\phi}(x) = \frac{n_c^3 - n_c}{2} \lambda e^{-3\lambda x} (1 - e^{-\lambda x})^{n_c-2} \sum_{k=1}^{n_o} \binom{n_o-1}{k-1}$$

$$\times (1 - e^{-\lambda x})^{n_o-k} (e^{-\lambda x})^{k-1}$$

$$= \frac{n_c^3 - n_c}{2} \lambda e^{-3\lambda x} (1 - e^{-\lambda x})^{n_c-2} (1 - e^{-\lambda x} + e^{-\lambda x})^{n_o-1}$$

$$= \frac{n_c^3 - n_c}{2} \lambda e^{-3\lambda x} (1 - e^{-\lambda x})^{n_c-2}$$

APPENDIX 4

Conditional Density of Three Nested Clades with Three Taxa Outside the Main Clade

Here we derive the marginal density for three nested calibration points in a $n + 3$ taxon tree. The first

(lowest) calibration point is on the root of an n-taxon monophyletic clade ϕ, the second on the $n + 1$ clade containing ϕ and one additional taxon, and the third is on the root of the tree, which includes the remaining two taxa. Let the heights of the calibration points be x_0, x_1 and x_2, where x_0 is the height of the root.

Of the $n + 2$ heights in the tree, three are calibrated and $n - 2$ are below x_2, which leaves just one height, x, outside ϕ. This gives us three cases, the first where the two outside taxa coalesce before x_2 ($x < x_2$), the second where they coalesce between x_1 and x_2 ($x_1 \leq x < x_2$), and the third where $x_1 \leq x < x_0$. The marginal densities for the three cases are as follows:

$$f_1(x_0, x_1, x_2) = (n + 3)\,!\,\lambda^3 e^{-2\lambda x_0} e^{-\lambda x_1} e^{-\lambda x_2} \frac{(1 - e^{-\lambda x_2})^{n-1}}{(n - 1)!}$$

$$f_2(x_0, x_1, x_2) = (n + 3)\,!\,\lambda^3 e^{-2\lambda x_0} e^{-\lambda x_1} e^{-\lambda x_2}$$
$$\times \frac{(1 - e^{-\lambda x_2})^{n-2}}{(n - 2)!}(e^{-\lambda x_2} - e^{-\lambda x_1})$$

$$f_3(x_0, x_1, x_2) = (n + 3)\,!\,\lambda^3 e^{-2\lambda x_0} e^{-\lambda x_1} e^{-\lambda x_2}$$
$$\times \frac{(1 - e^{-\lambda x_2})^{n-2}}{(n - 2)!}(e^{-\lambda x_1} - e^{-\lambda x_0}).$$

For each of the possible $\mathcal{R}_n$ ways to coalesce ϕ, there are $n - 1$ ways to place h between the $n - 2$ heights of ϕ (Case 1), only one way in Case 2 (no other heights between x_2 and x_1), and again only one way in Case 3, but here there are $\mathcal{R}_3$ ways to coalesce the three lineages between x_1 and the root. So, the ratios of the three cases are $n - 1 : 1 : 3$, which let us combine the f_1, f_2 and f_3 into the required density:

$$f(x_0, x_1, x_2) = \frac{1}{n + 3}((n - 1)f_1(x_0, x_1, x_2) + f_2(x_0, x_1, x_2)$$
$$+ 3f_3(x_0, x_1, x_2))$$
$$= (n - 1)n(n + 1)(n + 2)\lambda^3 e^{-\lambda(2x_0 + x_1 + x_2)}$$
$$\times (1 - 3e^{-\lambda x_0} + 2e^{-\lambda x_1})(1 - e^{-\lambda x_2})^{(n-2)}. \tag{15}$$