

Heterogeneous Rates of Molecular Evolution and Diversification Could Explain the Triassic Age Estimate for Angiosperms

JEREMY M. BEAULIEU^{1,*}, BRIAN C. O'MEARA², PETER CRANE³, AND MICHAEL J. DONOGHUE⁴

¹National Institute for Mathematical and Biological Synthesis (NIMBioS); ²Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996-1610, USA; ³School of Forestry and Environmental Studies, Yale University, 195 Prospect Street, New Haven, CT 06511, USA; and ⁴Department of Ecology and Evolutionary Biology, Yale University, PO Box 208106, New Haven, CT 10620, USA

*Correspondence to be sent to: National Institute for Mathematical & Biological Synthesis (NIMBioS), University of Tennessee, 1122 Volunteer Boulevard, Knoxville, TN 37996, USA; E-mail: jbeaulieu@nimbios.org.

Received 10 March 2015; reviews returned 21 April 2015; accepted 28 April 2015
Associate Editor: Mark Fishbein

Abstract.—Dating analyses based on molecular data imply that crown angiosperms existed in the Triassic, long before their undisputed appearance in the fossil record in the Early Cretaceous. Following a re-analysis of the age of angiosperms using updated sequences and fossil calibrations, we use a series of simulations to explore the possibility that the older age estimates are a consequence of (i) major shifts in the rate of sequence evolution near the base of the angiosperms and/or (ii) the representative taxon sampling strategy employed in such studies. We show that both of these factors do tend to yield substantially older age estimates. These analyses do not prove that younger age estimates based on the fossil record are correct, but they do suggest caution in accepting the older age estimates obtained using current relaxed-clock methods. Although we have focused here on the angiosperms, we suspect that these results will shed light on dating discrepancies in other major clades. [Angiosperms; birth–death; divergence times; land plants; rates of molecular evolution.]

Controversy surrounds the sometimes major differences between age estimates for clades based on the fossil record versus molecular clock methods (e.g., metazoans: Peterson et al. 2004; eukaryotes, Parfrey et al. 2011; mammals: Meredith et al. 2011; dos Reis et al. 2012; O'Leary et al. 2013). Flowering plants (angiosperms) provide a classic example. The generally accepted fossil evidence for the existence of the angiosperm crown clade dates to ca. 140 Ma (Valanginian to Hauterivian stages of the Early Cretaceous; Hughes 1994; Brenner 1996; see also reviews reviewed by Friis et al. 2006; 2011), whereas, unless the age of the crown is fixed to reflect the fossil age (Magallón et al. 2015), recent molecular phylogenetic estimates mostly imply that the angiosperm crown existed in the Triassic, some 200 Ma, or even earlier (e.g., Bell et al. 2010; Magallón; Smith et al. 2010; Clark et al. 2011; Zeng et al. 2014). Even though molecular dating methods have steadily increased in complexity—further relaxing the assumption of substitution rate inheritance (e.g., uncorrelated relaxed-clocks; Drummond et al. 2006; Drummond and Rambaut 2007) and treating fossil calibrations as probabilistic priors—the gap between the stratigraphic record and molecular age estimates for angiosperms has remained stubbornly persistent.

Consistency among the several molecular age estimates of a Triassic origin of flowering plants requires explaining a gap in the fossil record of at least 60 myr. One possibility is that crown angiosperms existed

during that time interval, but were not ecologically dominant and/or were living in environments where fossilization was unlikely (c.f., Feild et al. 2004; Smith et al. 2010). Given repeated claims of angiosperm fossils from the Triassic (e.g., Cornet 1986; Seward 1904; Hochuli and Feist-Burkhardt 2004; 2013) such an explanation might seem plausible, but all such claims are disputed (although some could represent lineages along the angiosperm stem; Doyle 2012). Of course, if any fossil from the Triassic proved to be a legitimate member of the angiosperm crown, the debate would instantly be settled in favor of the older molecular dates.

On the other hand, it is possible that the truth is more in line with the much younger fossil-based dates, in which case there must be serious methodological issues with current molecular dating methods that have not yet been addressed. This is the topic that we investigate here. Specifically, we focus on the possibility that clade-specific heterogeneity in rates of molecular evolution and the nature of taxon sampling could cause a systematic bias in age estimation using certain relaxed-clock methods. We reanalyze the age of angiosperms using updated sequences and fossil calibrations and then explore potential sources of error in a series of simulations designed to test whether methodological biases might partially explain why molecular-clock studies have consistently yielded much older ages than those suggested by the fossil record.

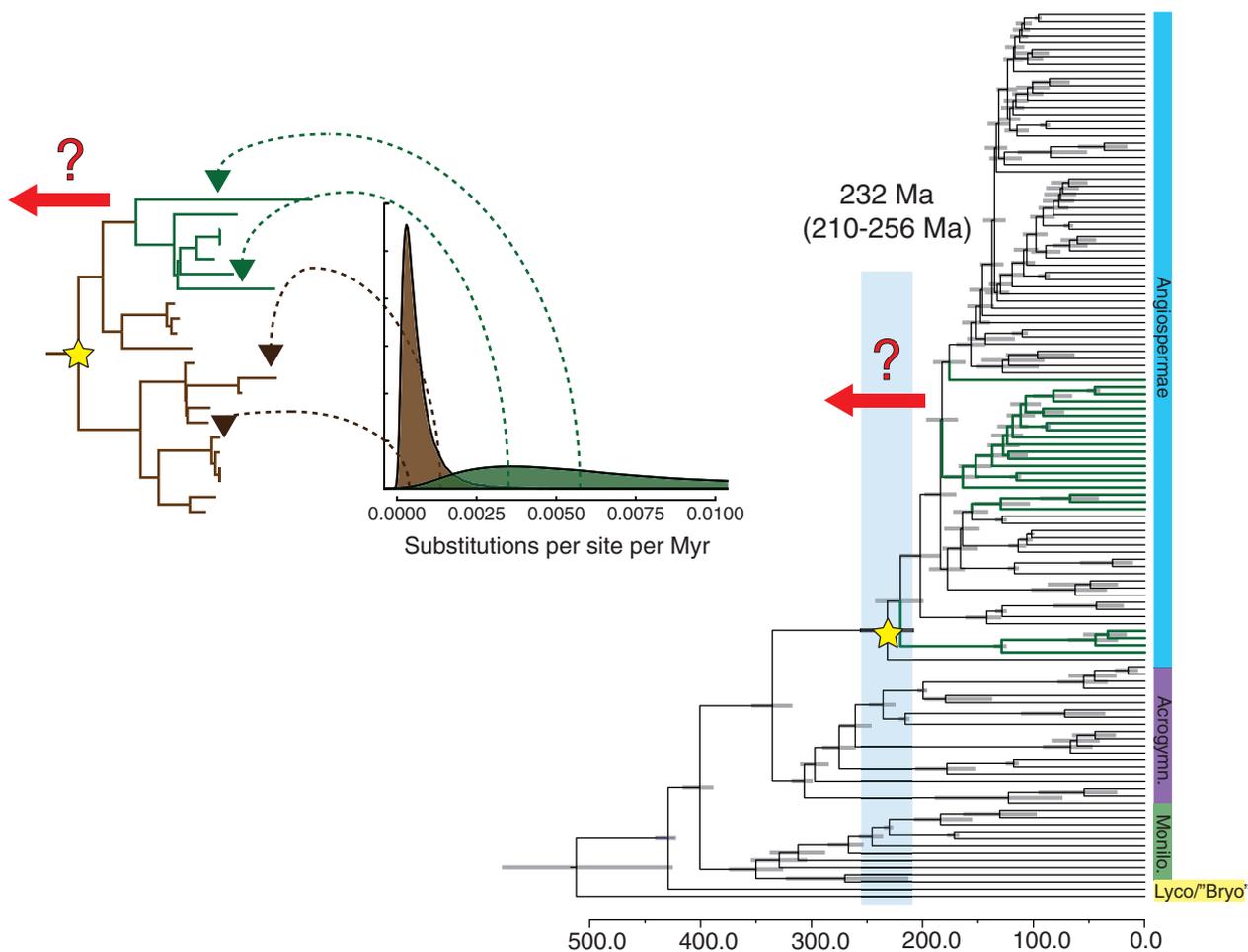


FIGURE 1. Divergence time estimates for land plants from a four-gene analysis (*atpB*, *rbcL*, *psbA*, and 18S), with the inferred Triassic age for angiosperms highlighted (star and light blue box). See Supplementary information for information on fossil calibrations. On the left, we illustrate our concern that herbaceous clades (green branches; often with elevated substitution rates relative to woody plants), positioned not far from the origin of angiosperm, may push back the age of the angiosperm crown.

A REANALYSIS OF THE AGE OF ANGIOSPERMS

We developed a set of 24 fossil calibrations mainly selected from those used by Smith et al. (2010), but supplemented with several additional calibrations proposed by Doyle and Endress (2010). Fifteen of these fossils calibrate nodes within flowering plants, with the remaining fossils providing temporal information in the other land plant clades (see Supplementary Fig. S1, available on Dryad at <http://dx.doi.org/10.5061/dryad.629sc>). We also assembled a molecular dataset of published sequences of nuclear 18S and chloroplast *atpB*, *psbB*, and *rbcL* across land plants. We estimated divergence times using the uncorrelated lognormal (UCLN) clock model implemented in BEAST (Drummond et al. 2006; Drummond and Rambaut 2007). Our data set includes representatives of every angiosperm order (*sensu* Angiosperm Phylogeny Group, 2009) and an expanded sample for Nymphaeales, Austrobaileyales, and Magnoliidae to permit a more precise placement of several fossils.

It also includes representatives of every major clade of acrogymnosperms (the clade containing the four major extant lineages of non-angiosperm seed plants: conifers, gnetophytes, cycads, and ginkgos), seven taxa representing monilophytes (ferns and allies), and a single lycophyte (club moss) to root the tree.

Analytical details, including justification for the placement of each fossil calibration point, and our rationale for assigning prior probabilities, are provided in the Supplementary Materials (available on Dryad at <http://dx.doi.org/10.5061/dryad.629sc>). However, our aim is not to refine previous studies. For present purposes, the important point is that our age estimate for crown angiosperms is consistent with previous results, with a median age of 232 Ma and a 95% HPD of 210–256 Ma (Fig. 1). This suggests that the angiosperm crown clade existed in the Triassic, and implies a gap in the fossil record of around 90 myr. Age estimates within several subgroups of angiosperms (e.g., campanulids—a clade that contains the sunflowers and their relatives [Asterales], carrots and their relatives

[Apiales], and honeysuckles and relatives [Dipsacales] are also consistent with previous estimates (Beaulieu et al. 2013).

THE POTENTIAL IMPACT OF CLADE-SPECIFIC RATE HETEROGENEITY

In angiosperms, significant differences in the rate of molecular evolution are correlated with growth habit, with herbaceous clades exhibiting higher and more variable molecular rates than related woody clades (Gaut et al. 1992, 1996; Laroche et al. 1997; Kay et al. 2006; Smith and Donoghue 2008; Lanfear et al. 2013). There is some evidence that relaxed clock methods can perform poorly in the presence of significant among-lineage rate variation (Wertheim et al. 2012), or when sister clades differ substantially in substitution rates (Dornburg et al. 2012), both of which can lead to overestimation of ages, although in the latter case the effects can be alleviated (at a cost of precision) when there are many calibration points spanning both clades. In the case of angiosperms our worry is that rapidly evolving herbaceous lineages will tend to appear older than they really are, and that this has the potential to “trickle-down” and bias age estimates for nodes below and some distance from the inferred shift in growth habit (also see Smith et al. 2010). Specifically, we wonder whether multiple shifts to the herbaceous habit nested not too far within the angiosperms might have the effect of pushing back the estimated age of the angiosperm crown (Fig. 1).

To explore this possibility, we conducted a set of simulations to test the impact of shifts in molecular substitution rates not far from a node of interest. Using the tree in Figure 1, we fixed the crown age of angiosperms at 140 Ma, based on the fossil record, and then asked how well we recovered this age as we varied the difference in rate of evolution between herbaceous clades and woody ones. Specifically, we supposed for this purpose that four herbaceous angiosperm clades whose stem connect nearer to the root—Nymphaeales, Piperales, Monocotyledonae, and *Ceratophyllum*—had variously elevated rates of molecular evolution. The “true” ages of nodes in the seed plant portion of our tree for which we have calibrations were set by treating as fixed ages the median value of the prior distribution applied to each fossil calibration (see Supplementary Materials, available on Dryad at <http://dx.doi.org/10.5061/dryad.629sc>); the ages of all other nodes were obtained by smoothing the molecular rates using r8s, which assumes an autocorrelated model of rate variation (Sanderson 2002).

Shifts in molecular rate due to life-history differences were simulated by independently drawing rates for branches within Nymphaeales, Piperales, monocotyledons, and *Ceratophyllum* from a lognormal distribution that differed from all other branches in the tree (Fig. 2). For this purpose we used the inferred parametric shape of the lognormal distribution of rates from our reanalysis of land plants (mean = $5e-4$,

sd = 0.75) as a baseline for increasing the rate of the herbaceous clades to a mean that was initially three times ($3\times$) higher than all other branches in the tree. The $3\times$ rate difference roughly corresponds to the average difference between the woody and herbaceous clades examined by Smith and Donoghue (2008).

Using the fixed tree topology shown in Figure 1, and the associated ages (see above), we generated a set of 100 molecular trees (differing in branch lengths) that were used to simulate gene alignments of 1000 sites in length using SeqGen (Rambaut and Grassly 1997). We assumed a general time reversible (GTR) model of nucleotide substitution using the inferred parameters from our land plant study. Each simulated data set was then analyzed using the UCLN model implemented in BEAST with 7,500,000 generations following the removal of the first 2,500,000 steps as burn-in. We also conducted a complimentary set of analyses that incorporated a calibration prior for the age of crown seed plants (age = 317.0 Ma) from our r8s analysis. We focus our discussion on simulations that exclude this calibration point, but the results were consistent between the two sets of analyses (see Supplementary Materials, available on Dryad at <http://dx.doi.org/10.5061/dryad.629sc>). All trees were summarized with TreeAnnotator with the consensus ages representing the median estimate of the posterior distribution.

Our primary finding is that when we simulated significant clade-specific rate heterogeneity located not far from the origin of angiosperms, we obtained age estimates for the crown node that are much older than the age assumed in the simulation, despite use of an uncorrelated rate model. As shown in Figure 3, across 100 randomly generated data sets the age of angiosperms was estimated to be, on average, 209.5 Ma, a difference of about 70 myr from our fixed age of 140 Ma. It is important to note that the impact of rate heterogeneity does not appear to be widely scattered throughout our seed plant tree. Instead, it appears to strongly affect only a few nodes that occur in the general vicinity of the simulated rate shifts (Fig. 3), but this includes the angiosperm crown node. Increasing the rate difference between woody and herbaceous lineages overestimates the age of angiosperms even further. When we repeated the simulations assuming a $6\times$ rate difference (Fig. 4), we estimated the crown age of angiosperms to be 244.3 Ma, on average, with even the youngest ages never falling outside of the Triassic (2.5% quantile = 213.9 Ma).

Of course, the situation is likely to be far more complicated than our simple two-rate scenario. For instance, there is evidence that the substitution rate in woody angiosperms is nearly $4\times$ higher than that in woody acrogymnosperms (Buschiazzi et al. 2012), due, at least in part, to the generally shorter times to first reproduction in woody angiosperms (Verdú 2002). Simulated data sets under such a three-rate scenario resulted in a distribution of the inferred age of angiosperms very similar (Fig. 4) to that for the $6\times$ rate difference (Fig. 4) (average at 244.9 Ma [95% CI = 208.7–275.2]).

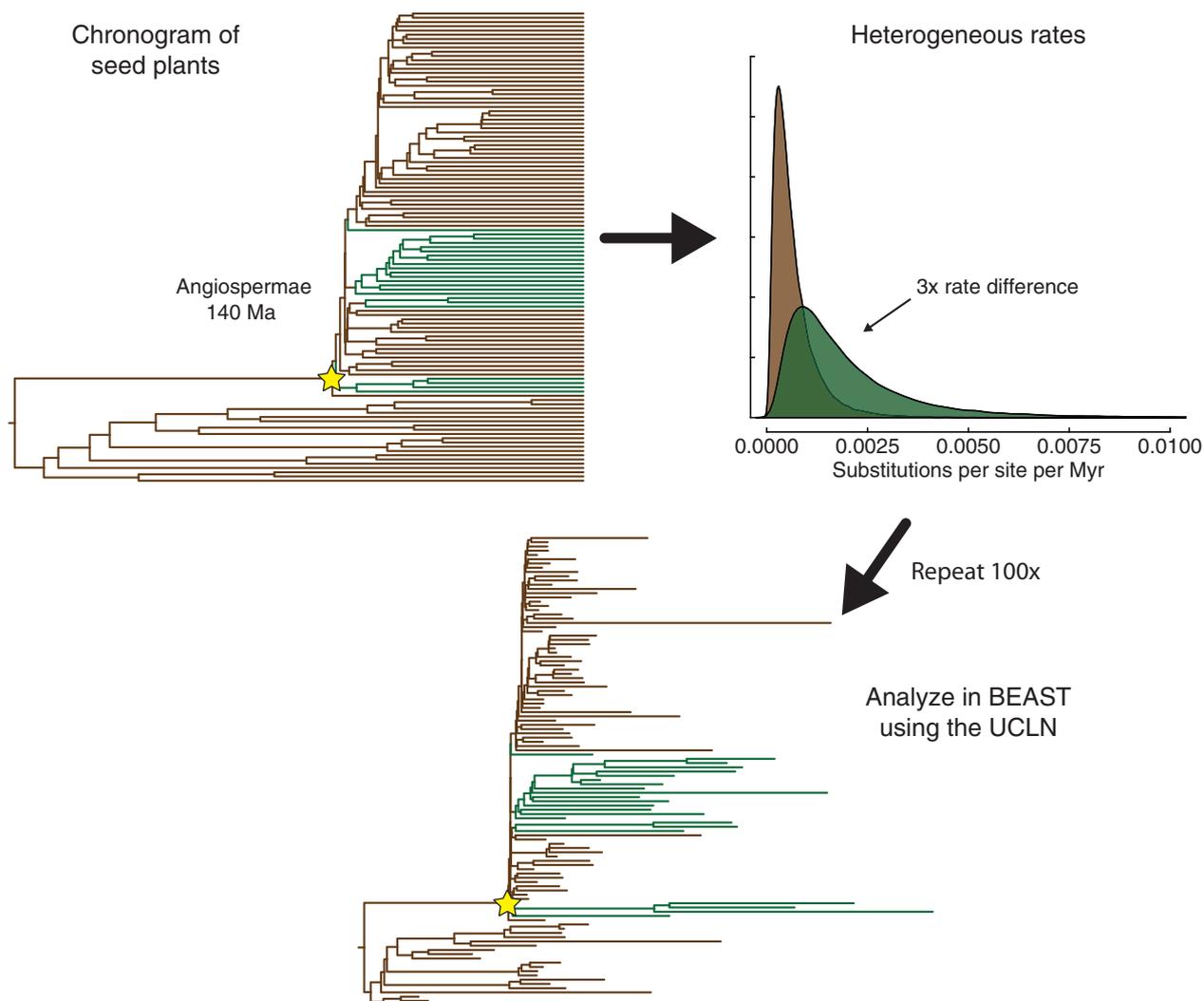


FIGURE 2. Basic steps in our simulation of clade-specific rate heterogeneity. We isolated the seed plant portion of our land plant tree and generated a set of “true” ages by treating as fixed ages the median value of the prior distribution applied to each fossil calibration from our reanalysis of land plants, with all other node ages obtained using *r8s* (see main text). We fixed the age of crown angiosperms (star) to be 140 Ma to reflect the fossil record. Faster molecular rates within the angiosperms (due to the evolution of the herbaceous habit) were simulated by independently drawing rates for the green branches (Nymphaeales, Piperales, Monocotyledonae, and *Ceratophyllum*) from a lognormal distribution with a mean initially 3× higher than all other (woody) branches in the tree. This process was repeated 100 times, and we analyzed each data set using the UCLN model implemented in BEAST.

We performed additional simulations that applied an angiosperm molecular rate distribution to the Gnetales, a small clade of acrogymnosperms that consistently exhibits elevated substitution rates (Chaw et al. 2000; Donoghue and Doyle 2000; Burleigh and Mathews 2004; Mathews 2009). Whether we draw rates for Gnetales from the rate distribution for woody or herbaceous angiosperms the age estimated for crown angiosperms is still much older than 140 myr (see Supplementary Materials, available on Dryad at <http://dx.doi.org/10.5061/dryad.629sc>).

THE POTENTIAL IMPACT OF REPRESENTATIVE SAMPLING

The simulations described above implicitly assumed that if there were no clade-specific rate variation the

relaxed-clock method would perform well. In other words, if all branches were randomly drawn from a single lognormal distribution of molecular rates, then the age estimate for crown angiosperms would be centered on the “true” age of 140–130 Ma. As a post-hoc check, we ran a set of simulations that assumed the same baseline lognormal distribution of rates for all branches in our seed plant tree. We were surprised to still find a considerable bias, with the age of crown angiosperms estimated to be 188 Ma on average. This is 20–60 Ma younger than estimates incorporating rate heterogeneity (see above), but still nearly 50 myr older than the fixed age used in the simulation. In other words, rate heterogeneity may explain part of the discrepancy between fossil-based and molecule-based estimates, but by no means all of it.

2 Rate scenario

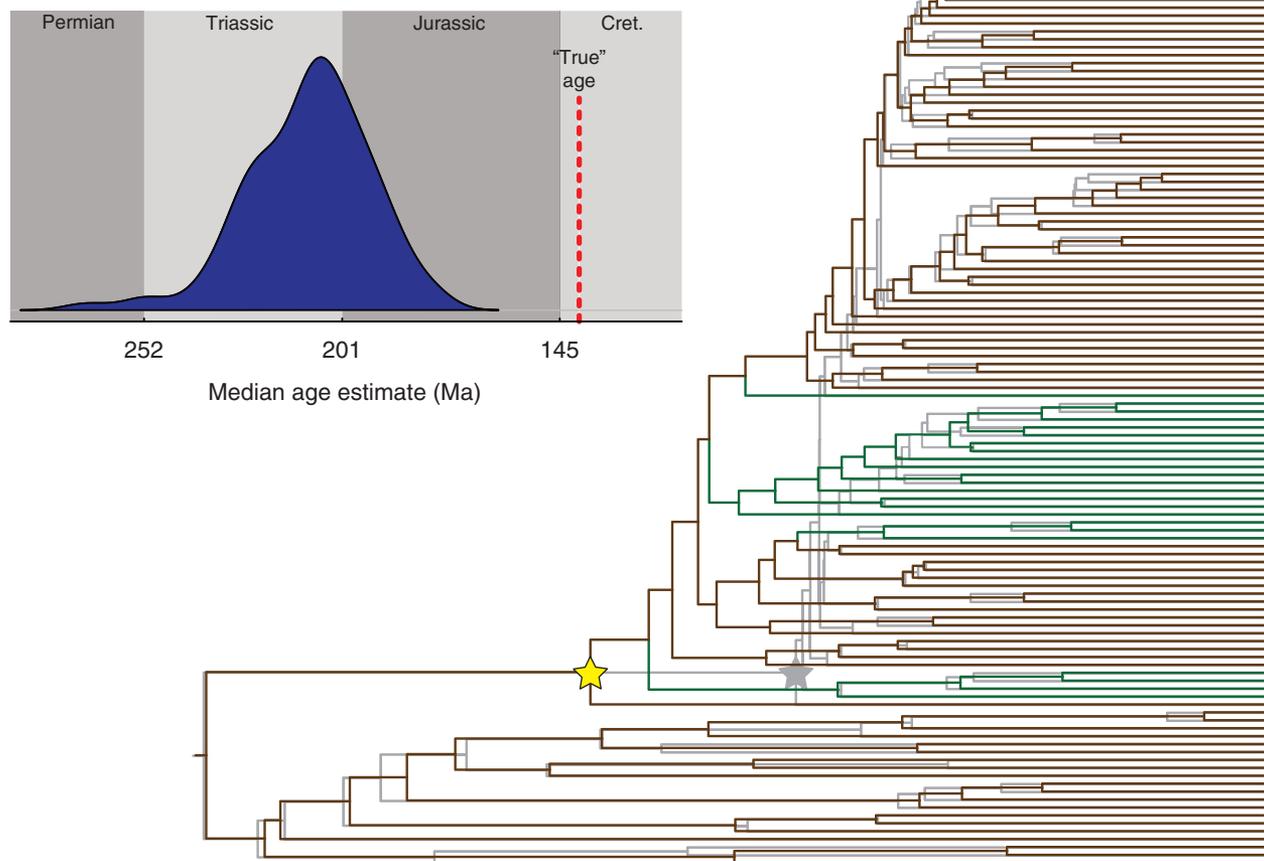


FIGURE 3. Results from the two-rate simulation (Fig. 2). Across 100 randomly generated data sets, the age of angiosperms was overestimated on average by roughly 70 myr (compare yellow vs. gray star). On the right, the chronogram of a randomly selected simulated tree is overlain on the “true” tree (shown in gray).

To explore this further we first compared the lengths of each branch in the “true” tree with the inferred branch lengths across the 100 randomly generated data sets that contain no lineage-specific rate heterogeneity (Fig. 5). This revealed that internal branch lengths were consistently inferred to be longer than the “true” branch lengths, while terminal branches were generally shorter. This overestimation and underestimation of internal and terminal branches, respectively, could be viewed as largely compensatory. However, in terms of percent difference from the true branch length, the bias for internal branches is much larger and can be upward of nearly 1000% of the true length (Fig. 5). Cumulatively, this has the effect of increasing the total length of the tree, even in the absence of rate heterogeneity.

We suspect that this tendency to push dates back is caused by the way in which extant diversity is sampled in such studies. In our representative sampling of the major land plant lineages, each terminal branch is generally a placeholder from anywhere from a handful to many thousands of species (e.g., Asterales, with ca. 27,000 species, represented here by *Helianthus annuus*; Asparagales, with ca. 26,000 species, represented by *Apostasia styphelioides*, and Lamiales, with ca. 24,000

species, represented by *Antirrhinum majus*). Perhaps this sort of sampling creates a problem for a dating method that requires a robust estimate of the underlying birth–death process as a means of calibrating rates.

With BEAST, there is a natural tension between whether differences in branch length reflect variation in the rate of molecular evolution or differences in time, with differences in time being inferred from a combination of calibration (from the fossil record) and an estimate of the diversification process, which for the purposes of the current discussion we treat as a birth–death model. By contrast, most other dating methods, such as penalized likelihood (Sanderson 2002), do not rely upon a diversification model that provides a separate signal for the true branch lengths. Under normal circumstances, the estimates of the birth–death parameters should provide an adequate measure of the expected wait times (i.e., $\frac{1}{b+d}$) between successive nodes. BEAST should be able to discern whether to shorten relatively long branches to accommodate a birth–death process that expects generally shorter wait times, or to lengthen relatively short branches to account for longer expected wait times. However, in the case of our tree, and many other studies that employ representative

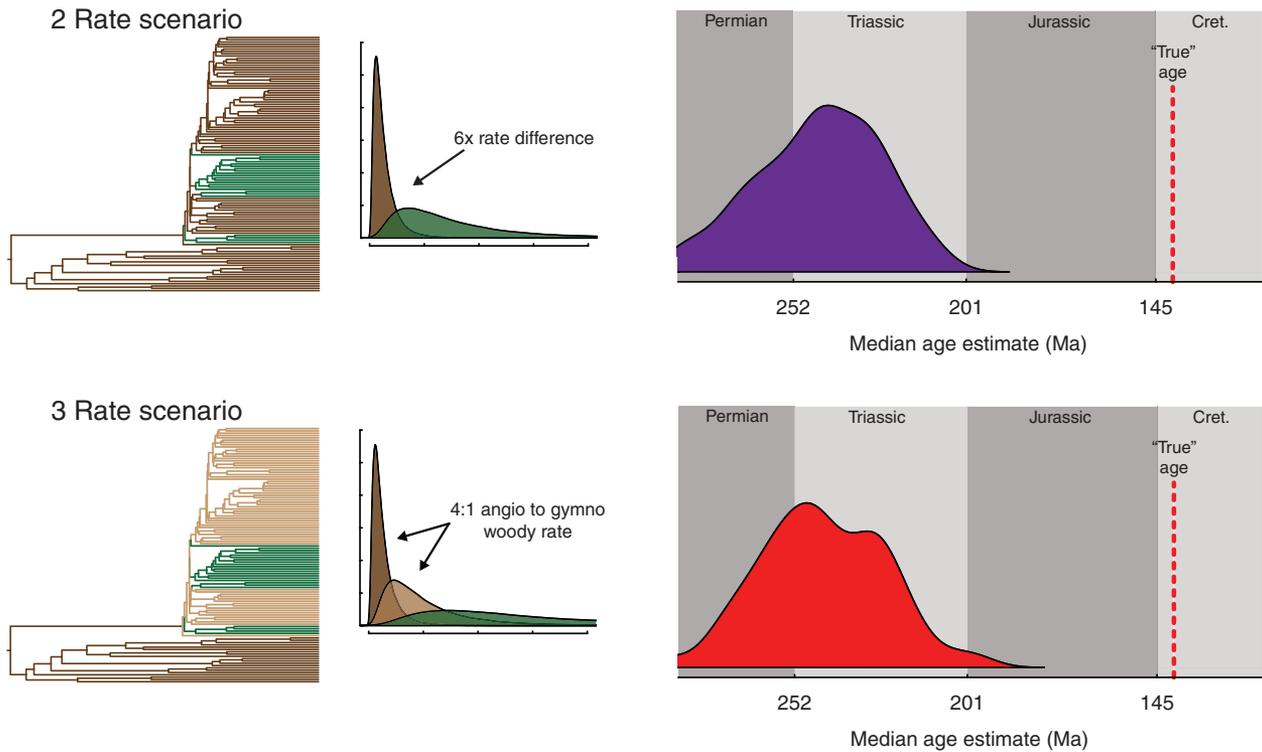


FIGURE 4. Biases in additional simulations. Above: a two-rate scenario, where rates for herbaceous clades were drawn from a lognormal distribution $6\times$ higher than woody clades. Below: a three-rate scenario in which acrogymnosperms (the clade containing the four major clades of extant “gymnosperms”), woody angiosperms, and herbaceous angiosperms have different rates of molecular evolution.

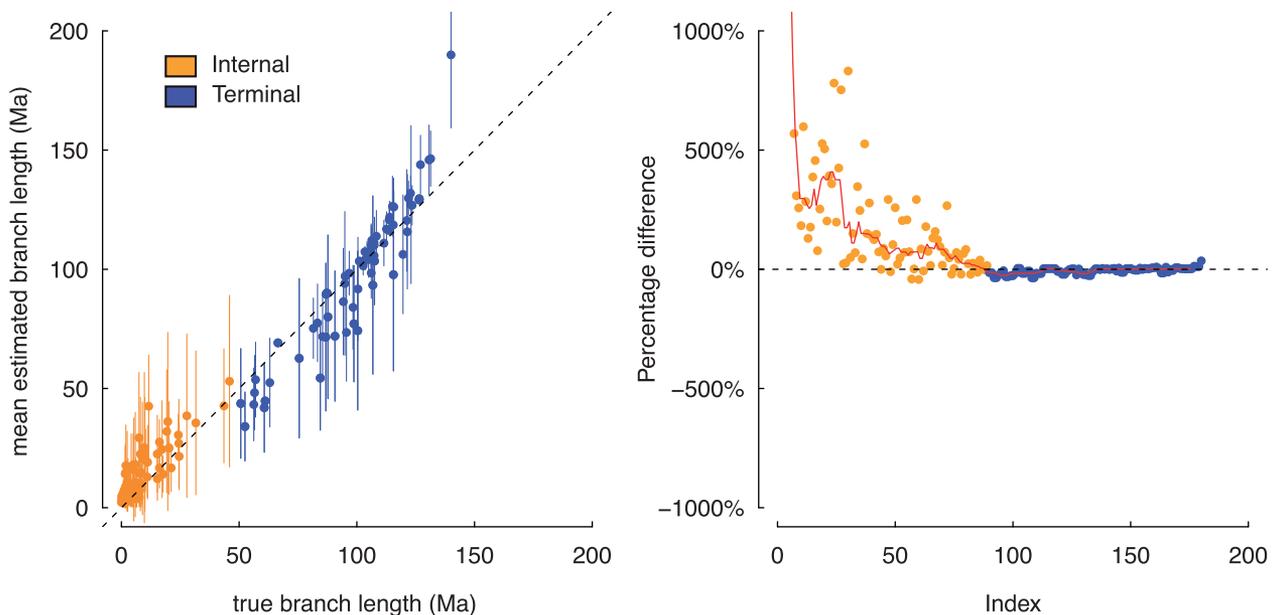


FIGURE 5. Distribution of bias between internal and terminal branches assuming that all branches are randomly drawn from a single lognormal distribution (i.e., “no-variation” scenario). Left: Relationship between “true” branch lengths and the mean of the estimated branch lengths from 100 randomly generated data sets (line shows the 95% CI). Right: When viewed as percentage difference, the bias is seen to be in internal branches (line shows the trend in the rolling mean).

sampling, the wait times in the tree cannot be reflecting some common underlying birth–death process. In fact, such representative sampling insures a particular kind of heterogeneity: one process that generated the internal

branch lengths, which should generally better reflect the true expected wait times (assuming, of course, extinction has been random and that all major land lineages have been sampled), and another that would be consistent

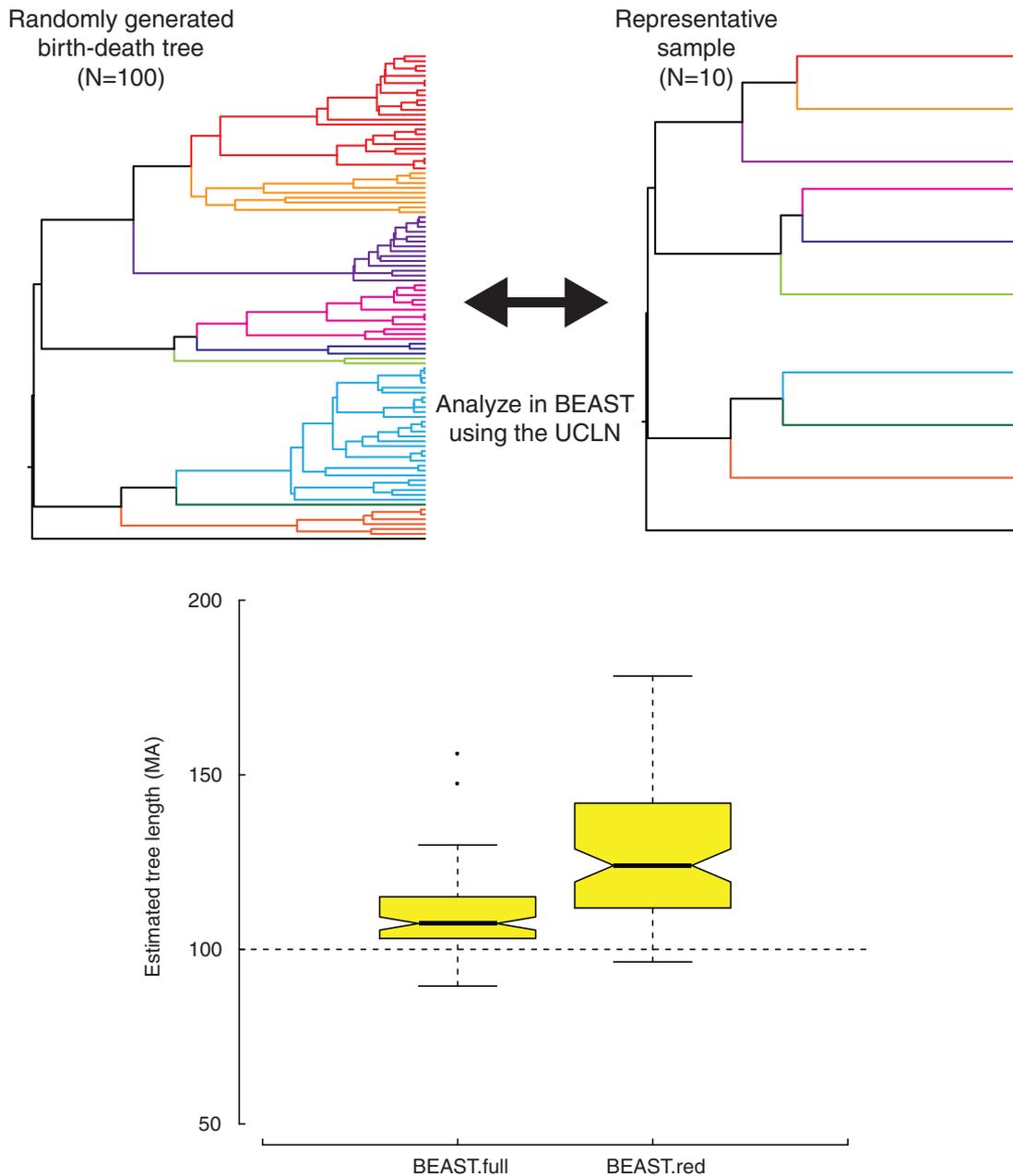


FIGURE 6. Basic steps in our simulation to test the influence of the birth–death prior. 100 sequence data sets were simulated on a random birth–death tree with 100 taxa. Each simulated data set was then pruned to ensure a single representative of the 10 major lineages (taxonomic groups shown with different colors). The “complete” and “sampled” data sets were analyzed in BEAST. Yellow boxplots below show an average bias of ca. 7 myr in the median age estimate for the complete trees, and an average bias of ca. 24 myr for the sampled trees.

with the long durations in the terminal branches where, ostensibly, nothing happened (see Fig. 1). Because terminal branches make up more than half the tree, the “average” should be biased toward the longer wait times of the terminal branches. The worry is that, as a result, the internal branches in our seed plant tree are being systematically lengthened.

To explore this issue we conducted a set of simulations where we compared age estimates obtained from a complete tree with those of a representative sample of the same tree (Fig. 6). We generated a single random

birth–death tree (birth = 0.08, death = 0.04) that contained 100 species; the total length of the tree was rescaled to reflect 100 time units. We followed the same procedures described above: we generated a set of 100 molecular trees under our baseline UCLN distribution, brought them into SeqGen (Rambaut and Grassly 1997), and simulated gene alignments of 1000 sites assuming a GTR model of nucleotide substitution. Each simulated data set was also then selectively pruned down to 10 tips to ensure a single representative of 10 major lineages (shown with different colors in Fig. 6). Both the

“complete” and the “sampled” data sets were analyzed in BEAST assuming a UCLN model of substitution, a fixed topology, and calibration priors that were set such that the true ages were the median values in a lognormal distribution (mean = 1.5, sd = 0.75).

When the “complete” trees were analyzed there was an increase of around 7 myr in the median age estimate across the 100 data sets. When the “sampled” trees were analyzed this increased to around 24 myr. This, we think, reflects the great disparity in the lengths of internal and terminal branch lengths in the sampled tree. The impact of this disparity on the birth–death process can be seen in the estimates of net diversification, which were more than three times lower in the sampled (mean = 0.009) than in the complete data set (mean = 0.032). When the parameters of the birth–death process are converted to an estimate of the expected wait times, the sampled trees show wait times that are, on average (mean = 70.0 myr), nearly an order of magnitude longer than in the complete trees (mean = 9.82 myr).

CONCLUDING REMARKS

What can we conclude about Triassic age estimates for the flowering plants based on molecular data? Unfortunately, it appears that these could largely reflect the methodological issues highlighted here. As our simulations imply, such an age could be largely explained by the potential additive effects of not properly accounting for (1) clade-specific heterogeneity in molecular substitution rates, and (2) the use of a representative sample of the major land plant groups. It is possible, in other words, that crown angiosperms are ca. 140–130 myr, and that we are obtaining much older age estimates using BEAST owing to these two sources of bias. We are not asserting here that the crown angiosperms are actually only ca. 140 myr (or perhaps slightly older), only that the molecular dating analyses conducted to date do not strongly rule out this possibility.

The shifts in plant habit and molecular rate that actually occurred in the course of angiosperm evolution will have been far more complicated than we have represented in our simulations. For example, within the species-rich eudicot clade there are many transitions in habit, but for purposes of our simulations we treated this entire clade as having only evolved under a woody rate. There is also the vexing problem of how lineages that are missing as a result of extinction might complicate the issue. It is unclear what impact having many more nested rate shifts would have on our results. Of course, the impact of rate heterogeneity may also vary depending on the number and the identity of the genes that are analyzed, and more detailed comparisons along these lines would be useful. In any event, what is clear is that even simple scenarios of rate heterogeneity can have a pronounced impact on age estimation.

Regarding representative sampling (both of extant and extinct lineages), we note that BEAST does implement

a birth–death prior that assumes incomplete sampling, which could potentially alleviate the problem. With this approach, the normal birth–death parameters are estimated with the addition of a third variable, ρ , that scales estimates of birth and death to compensate for incomplete sampling (Gernhard 2008). We applied this prior to our land plant data set (see Supplementary Materials, available on Dryad at <http://dx.doi.org/10.5061/dryad.629sc>), but found little change in age estimates for angiosperms. Furthermore, the mean of the posterior distribution for the sampling frequency was $\rho = 0.059$ (95% HPD = 0.001–0.117), which is nowhere near the dismal sampling frequency in our tree (ca. 0.0003% assuming 375,000 species of land plants and no extinction). This is consistent with recent theoretical work demonstrating that ρ must be considered a known quantity and cannot be estimated from a phylogeny while also estimating speciation and extinction rates (Stadler 2013), as BEAST does. Otherwise, there is a tendency to overestimate the sampling frequency (ρ) at the cost of underestimating both speciation and extinction rates.

Even if we could treat sampling frequency as fixed (either through direct specification or by restricting the prior on ρ), the assumption of a constant birth–death process would still misrepresent the heterogeneity in diversification across land plants. For example, the success of angiosperms is generally linked to higher overall rates of diversification (e.g., Sanderson and Donoghue 1994; Smith et al. 2011), which would translate into shorter expected wait times within angiosperms relative to other land plants. With the assumption of a single distribution from which waiting times are drawn, rapidly diversifying clades like the angiosperms will appear to be older than they really are when analyzed as a part of a larger clade.

Under the circumstances, it behooves us to remain humble and to honestly assess potential biases not only in the fossil record, but also in our methods for analyzing molecular data. Our simulations highlight two potentially general systematic biases introduced by (1) the phylogenetic position and the magnitude of shifts in rates of molecular evolution, and (2) a representative sampling scheme that can create difficulties for methods that rely on birth–death processes. Although we have focused here on the angiosperms, we suspect that these results will shed light on dating discrepancies in other major clades.

SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.629sc>.

FUNDING

This work was supported by the National Institute for Mathematical and Biological Synthesis [to J.M.B.], an Institute sponsored by the National Science Foundation,

the U.S. Department of Homeland Security, and the U.S. Department of Agriculture through NSF Award No. EF-0832858, with additional support from The University of Tennessee, Knoxville.

ACKNOWLEDGMENTS

The authors thank Mark Fishbein, Susana Magallón, and an anonymous reviewer for thoughtful comments and suggestions for improving the manuscript. They also thank Jim Doyle, Tom Near, Andrew Leslie, Stephen Smith, Alex Dornburg, and Nick Matzke for helpful discussions.

REFERENCES

- Angiosperm Phylogeny Group. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161: 105–121.
- Beaulieu J.M., Tank D.C., Donoghue M.J. 2013. A Southern Hemisphere origin for campanulid angiosperms, with traces of the break-up of Gondwana. *BMC Evol. Biol.* 13:80.
- Bell C.D., Soltis D.E., Soltis P.S. 2010. The age and diversification of the angiosperms revisited. *Am. J. Bot.* 97:1296–303.
- Brenner G.J. 1996. Evidence for the earliest stage of angiosperm pollen evolution: A paleoequatorial section from Israel. In: Taylor D.W., Hickey L.J., editor. *Flowering plant origin, evolution and phylogeny*. New York: Chapman & Hall. p. 91–115.
- Burleigh J.G., Mathews S. 2004. Phylogenetic signal in nucleotide data from seed plants: Implications for resolving the seed plant tree of life. *Am. J. Bot.* 91:1599–1613.
- Buschiazio E., Ritland C., Bohlmann J., Ritland K. 2012. Slow but not low: Genomic comparisons reveal slower evolutionary rate and higher dN/dS in conifers compared to angiosperms. *BMC Evol. Biol.* 12:8.
- Chaw S.-M., Parkinson C.L., Cheng Y., Vincent T.M., Palmer J.D. 2000. Seed plant phylogeny inferred from all three plant genomes: Monophyly of extant gymnosperms and origin of Gnetales from conifers. *Proc. Natl Acad. Sci. U. S. A.* 97:4086–4091.
- Clark J.T., Warnock R.C.M., Donoghue P.J. 2011. Establishing a time-scale for plant evolution. *New Phytol.* 192:266–301.
- Cornet B. 1986. The leaf venation and reproductive structures of a Late Triassic angiosperm, *Sanmiguelia lewisii*. *Evol. Theory* 7: 231–309.
- Donoghue M.J., Doyle J.A. 2000. Seed plant phylogeny: Demise of the anthophyte hypothesis? *Curr. Biol.* 10:R106–R109.
- Dornburg A., Brandley M.C., McGowen M.R., Near T.J. 2012. Relaxed clocks and inferences of heterogeneous patterns of nucleotide substitution and divergence time estimates across whales and dolphins (Mammalia: Cetacea). *Mol. Biol. Evol.* 29: 721–736.
- dos Reis M., Inoue J., Hasegawa M., Asher R.J., Donoghue P.C.J., Yang Z. 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc. R. Soc. B* 279:3491–3500.
- Doyle J.A. 2012. Molecular and fossil evidence on the origin of angiosperms. *Annu. Rev. Earth Planet. Sci.* 40:301–326.
- Doyle J.A., Endress P.K. 2010. Integrating Early Cretaceous fossils into the phylogeny of living angiosperms: Magnoliidae and eudicots. *J. Syst. Evol.* 48:1–35.
- Drummond A.J., Ho S.Y.W., Phillips M.J., Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4:e88.
- Drummond A.J., Rambaut A. 2007. Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214.
- Feild T.S., Arens N.C., Doyle J.A., Dawson T.E., Donoghue M.J. 2004. Dark and disturbed: A new image of early angiosperm ecology. *Paleobiology* 30:82–107.
- Friis E.M., Pedersen K.R., Crane P.R. 2006. Cretaceous angiosperm flowers: Innovation and evolution in plant reproduction. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 232:251–293.
- Friis E.M., Crane P.R., Pedersen R. 2011. The early flowers and angiosperm evolution. Cambridge: Cambridge University Press. p. 1–596.
- Gaut B.S., Muse S.V., Clark W.D., Clegg M.T. 1992. Relative rates of nucleotide substitution at the *rbcL* locus of monocotyledonous plants. *J. Mol. Evol.* 35:292–303.
- Gaut B.S., Morton B.R., McCraig B.C., Clegg M.T. 1996. Substitution rate comparisons between grasses and palms: Synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. *Proc. Natl Acad. Sci. U. S. A.* 93: 10274–10279.
- Gernhard T. 2008. The conditioned reconstructed process. *J. Theor. Biol.* 253:769–778.
- Hochuli P.A., Feist-Burkhardt S. 2004. A boreal early cradle of Angiosperms? Angiosperm-like pollen from the Middle Triassic of the Barents Sea (Norway). *J. Micropalaeontol.* 23:97–104.
- Hochuli P.A., Feist-Burkhardt S. 2013. Angiosperm-like pollen and Afropollis from the Middle Triassic (Anisian) of the Germanic Basin (Northern Switzerland). *Front. Plant Sci.* 4:1–14.
- Hughes N.F. 1994. The enigma of angiosperm origins. Cambridge: Cambridge University Press. p. 1–303.
- Kay K.K., Whittall J.B., Hodges S.A. 2006. A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: An approximate molecular clock with life history effects. *BMC Evol. Biol.* 6:36.
- Lanfear R., Ho S.Y.W., Davies T.J., Moles A.T., Aarssen L., Swenson N.G., Warman L., Zanne A.E., Allen A.P. 2013. Taller plants have lower rates of molecular evolution: The rate of mitosis hypothesis. *Nat. Commun.* 4:1879.
- Laroche J., Li P., Maggia L., Bousquet J. 1997. Molecular evolution of angiosperm mitochondrial introns and exons. *Proc. Natl Acad. Sci. U. S. A.* 94:5722–5727.
- Magallón S. 2010. Using fossils to break long branches in the molecular dating: A comparison of relaxed clocks applied to the origin of angiosperms. *Syst. Biol.* 59:384–399.
- Magallón S., Gómez-Acevedo S., Sánchez-Reyes L.L., Hernández-Hernández T. 2015. A meta-calibrated timetree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* 207:437–453.
- Mathews S. 2009. Phylogenetic relationships among seed plants: Persistent questions and the limits of molecular data. *Am. J. Bot.* 96:228–236.
- Meredith R.W., Janecka J.E., Gatesy J., Ryder O.A., Fisher C.A., Teeling E.C., Goodbla A., Eizirik E., Simao T.L.L., Stadler T., Rabosky D.L., Honeycutt R.L., Flynn J.J., Ingram C.M., Steiner C., Williams T.L., Robinson T.J., Burk-Herrick A., Westerman M., Ayoub N.A., Springer M.S., Murphy W.J. 2011. Impacts of the Cretaceous terrestrial revolution and KPg Extinction on mammal diversification. *Science* 334:521–524.
- O’Leary M.A., Bloch J.I., Flynn J.J., Gaudin T.J., Giallombardo A., Giannini N.P., Goldberg S.L., Kraatz B.P., Luo Z.-X., Meng J., Ni X., Novacek M.J., Perini F.A., Randall Z.S., Rougier G.W., Sargis E.J., Silcox M.T., Simmons N.B., Spaulding M., Velazco P.M., Weksler M., Wible J.R., Cirranello A.L. 2013. The placental mammal ancestor and the post-K-Pg radiation of placentals. *Science* 339: 662–667.
- Parfrey L.W., Lahr D.J., Knoll A.H., Katz L.A. 2011. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc. Natl Acad. Sci. U. S. A.* 108:13624–13629.
- Peterson K.J., Lyons J.B., Nowak K.S., Takacs C.M., Wargo M.J., McPeck M.A. 2004. Estimating metazoan divergence times with a molecular clock. *Proc. Natl Acad. Sci. U. S. A.* 101:6536–6541.
- Rambaut A., Grassly N.C. 1997. Seq-Gen: An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput. Appl. Biosci.* 13:235–238.
- Sanderson M.J. 2002. Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. *Mol. Biol. Evol.* 14:1218–1231.
- Sanderson M.J., Donoghue M.J. 1994. Shifts in diversification rate with the origin of angiosperms. *Science* 264:1590–1593.

- Seward A.C. 1904. Catalogue of the Mesozoic Plants in the Department of Geology, British Museum (Natural History). The Jurassic Flora. II. London: British Museum of Natural History.
- Smith S.A., Beaulieu J.M., Donoghue M.J. 2010. An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proc. Natl Acad. Sci. U. S. A.* 107:5897–5902.
- Smith S.A., Beaulieu J.M., Stamatakis A., Donoghue M.J. 2011. Understanding angiosperm diversification using small and large phylogenetic trees. *Am. J. Bot.* 98:404–414.
- Smith S.A., Donoghue M.J. 2008. Rates of molecular evolution are linked to life history in flowering plants. *Science* 322:86–89.
- Stadler T. 2013. How can we improve accuracy of macroevolutionary rate estimates? *Syst. Biol.* 62:321–329.
- Verdú M. 2002. Age at maturity and diversification in woody angiosperms. *Evolution* 56:1352–1361.
- Wertheim J.O., Fourment M., Pond S.L.K. 2012. Inconsistencies in estimating the age of HIV-1 subtypes due to heterotachy. *Mol. Biol. Evol.* 29:451–456.
- Zeng L., Zhang Q., Sun R., Kong H., Zhang N., Ma H. 2014. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nat. Commun.* 5: 4956.