

The Impact of Fossils and Taxon Sampling on Ancient Molecular Dating Analyses

Laura A. Hug and Andrew J. Roger

Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada

The number and complexity of molecular dating studies has increased over the past decade. Along with a broadening acceptance of their utility has come significant controversy over the methods and models that are appropriate, as well as the accuracy of the estimates yielded by molecular clock analyses. Radically different age estimates have been published for the same divergences from analyses of different datasets with different fossil constraints obtained with different methods, and the underlying explanation for these differences is often unclear. Here we utilize two previously published datasets to examine the effect of fossil calibrations and taxon sampling on the age estimates for two deep eukaryote divergences in an attempt to discern the relative impact of these factors. Penalized likelihood, non-parametric rate smoothing, and Bayesian methods were utilized to generate age estimates for the origin of the Metazoa from a 7-gene dataset and for the divergence of Eukaryotes from a 129-gene dataset. From these analyses, it is clear that the fossil calibrations chosen and the method for applying constraints to these nodes have a large impact on age estimates, while the degree of taxon sampling within a dataset is less important in terms of the resulting age estimates. Concerns and recommendations for addressing these two factors when initiating a dating analysis are discussed.

Introduction

The concept of molecular dating has existed since Zuckerkandl and Pauling (1965) postulated that, if sequence changes occurred in a clock-like manner during evolution, it would be possible to determine the time elapsed since two species or groups diverged based on the differences present in their DNA sequences. This revolutionary idea sparked immediate interest with evolutionists, but the task of actually creating a model to properly describe the relationship between the divergence in molecular sequence and the time of divergence between groups has proven significantly more difficult than originally anticipated (Welch and Bromham 2005). Age estimates generated under molecular dating studies often deviate from the ages suggested by the paleontological record. In extreme cases, age estimates have been nearly double that suggested by the fossil record (Douzery et al. 2004; Hedges and Kumar 2004). For instance, Blair and Hedges (2005) reported divergence times for the arthropod-deuterostome divergence which are >500 million years (MY) older than corresponding fossil minimum dates from the Cambrian explosion. Such discrepancies are in part due to the incomplete fossil record, which, because of missing fossils, underestimates the divergence times, but is also due to statistical biases from the molecular dating studies, (Benton and Ayala 2003), and how fossil constraints are treated (Blair and Hedges 2005). Moreover, there are major discrepancies between estimates yielded by different methods of molecular dating (Aris-Brisou and Yang 2003; Blair and Hedges 2005; Roger and Hug 2006). In any case, the relative validity of fossil versus molecular divergence date estimation continues to be debated. Near & Sanderson (2004) have developed a fossil-based model cross-validation method in which the fossil calibrations are used to determine the phylogenetic model of best fit. Contrastingly, Kumar and Hedges (1998) postulated that the use of mo-

lecular dates could help direct the search for missing fossils. Ideally, advances in both areas will lead to a rapprochement of age estimates between the two forms of evidence.

For any molecular dating study, a basic requirement is at least one fossil constraint to anchor the estimated divergence times onto the timeline of the Earth. There has been significant discussion in recent literature as to the proper method of applying fossil constraints to a phylogenetic tree (Graur and Martin 2004; Hedges and Kumar 2004; Reisz and Muller 2004; Blair and Hedges 2005; Glazko, Koonin and Rogozin 2005; Drummond et al. 2006; Yang and Rannala 2006). Intuitively, a fossil date can function as nothing more than a minimum temporal boundary for the existence of the organism or node in question. In recent published studies, constraints have been set as minimum boundaries (Blair and Hedges 2005; Glazko, Koonin and Rogozin 2005; Near, Meylan and Shaffer 2005), fixed dates (Peterson et al. 2004; Peterson and Butterfield 2005), and, in some cases, maximum boundaries for a divergence point (Kishino, Thorne and Bruno 2001). These discrepancies, as well as differences in methods and models of dating, and the nature of the datasets, have led to considerable controversy over the age estimates recovered by these studies. There have been a few studies that actively examine the role of fossil constraints on age estimates (van Tuinen and Hadly 2004; Near, Meylan and Shaffer 2005; Roger and Hug 2006), but in general, the error introduced through the application of these constraints is ignored or unreported.

An additional current concern within the field is the use of a single fossil constraint as the sole primary calibration point within an analysis (Kumar and Hedges 1998; Kishino, Thorne and Bruno 2001; Hedges et al. 2004). Arguments in favor of this practice have highlighted the importance of the robustness of the constraints on the fossils chosen and the paucity of the fossil record, implying that there are very few fossils that can reliably be used in dating studies (Hedges and Kumar 2004). While this argument may have weight in terms of the scarcity of fossils for many of the organisms on the tree of life (particularly the protists and other organisms lacking hard structures for fossilization), it is challenged by the work of Benton and Donoghue

Key words: molecular dating, fossil constraints, Bayesian analysis, Penalized likelihood, taxon sampling.

E-mail: Andrew.Roger@dal.ca.

Mol. Biol. Evol. 24(8):1889–1897. 2007
doi:10.1093/molbev/msm115
Advance Access publication June 7, 2007

(2007), who outline 30 different well-constrained fossil calibrations for divergences along the tree of life. Furthermore, when the clock assumption of molecular divergence is violated, as is the case in these deep-level divergences, it is especially important that multiple fossil calibrations should be used, and that multiple genes be analyzed simultaneously (Thorne and Kishino 2002; Yang and Yoder 2003).

Several studies have also been published with multiple secondary time calibrations, all of which stem from earlier analyses that relied on a single calibration point (Hedges et al. 2004), a practice which seems assured to propagate error (Graur and Martin 2004). This practice of applying secondary calibrations from prior analyses requires that there be no known biases in the prior analysis (Hedges and Kumar 2004). Our earlier work (Roger and Hug 2006) showed that biases exist on multiple levels of molecular dating analyses, questioning the validity of such secondary calibrations. It has also been shown that certain calibration points can be inconsistent with the rest of the fossil calibrations used (Near, Meylan and Shaffer 2005). Near et al. (2005) suggest that these fossils should be removed from the analysis so that an internally consistent set of fossils are used to calibrate the phylogeny.

The positive impact of increased taxon sampling on phylogeny determination has often been posited (Delsuc, Brinkmann and Philippe 2005), although there has been considerable debate over the relative impacts of taxon versus site inclusion (Hedtke, Townsend and Hillis 2006). Improved taxonomic sampling can lead to the resolution of incorrect phylogenies through breaking of long branches (Hedtke, Townsend and Hillis 2006). On the other hand, it has been shown that inclusion of certain taxa, such as highly-divergent or fast-evolving species which contribute long branches to a tree, can prove detrimental in generating a phylogeny (Delsuc, Brinkmann and Philippe 2005). In these cases, it can be argued that it is better to omit certain sites or taxa in order to generate a better-resolved phylogeny (Gribaldo and Philippe 2002).

In the context of molecular dating studies, however, similar investigations into the impact of taxon sampling on date estimates have not been undertaken.

Here we investigate the impact of taxonomic sampling, fossil constraint sampling as well as different ways to apply such constraints on ancient molecular dating analyses. In a previous study (Roger and Hug 2006), we showed that different methods of constraining the fossil calibration points used had a pronounced and significant effect on age estimates. Here, we clarify and extend this observation by determining the impact of single versus multiple fossil constraints on age estimates, as well as examining the relative impact of taxon sampling versus fossil constraint application.

Materials and Methods

Datasets

Two recently published studies dealing with molecular clocks were chosen as test cases for this analysis. The first

(Peterson and Butterfield 2005) (abbreviated PB) contains 7 genes with 2,052 aligned amino acid sites and was kindly provided by Dr. Kevin Peterson at Dartmouth College. It comprises 30 Metazoan taxa and two outgroups. The authors used 10 fossil calibrations taken from the invertebrate fossil record in their dating analyses. The second dataset (Douzery et al. 2004) (abbreviated DZ) contains 129 genes with 30,399 amino acid sites, 36 taxa, and 6 fossil calibrations and was kindly provided by Dr. Hervé Philippe (Université de Montreal). Both studies were examining deep eukaryote divergences: PB the origin and timing of the Metazoan divergence, and DZ the age of the root of eukaryotes.

These datasets were chosen for their extensive taxon sampling, as well as their size: the smaller dataset (PB) is ideal for testing certain hypotheses that require computationally intensive algorithms, while the larger (DZ) allowed us to test whether increased sequence data would resolve the biases we observed with the PB dataset.

For all analyses, three nodes of interest of varying depth and proximity to fossil constraints were chosen from each dataset, in order to discern whether certain trends were consistent across the tree. In both cases, one of the nodes of interest was the deepest node on the tree, as it was expected that the greatest amount of variation, as well as systematic error in the estimates, would occur at the deeper nodes. The selected nodes are indicated with the numbers in bold on figure 1.

For both datasets, the published topology was taken as correct, and was held fixed in all subsequent analyses. For the PB dataset, phylogenetic analyses were undertaken to ascertain whether the presence of morphological characters in the original dataset contributed to the nature of the published topology. Using only the sequence data, an identical topology to that used in the original study was obtained using protein maximum likelihood analysis with the IQPNNI algorithm with a VT+ Γ model (Vinh le and Von Haeseler 2004).

Dating Software

Branch lengths for the PB dataset topology were estimated under a VT+ Γ model in Tree-Puzzle version 5.2 (Schmidt et al. 2002). The size of the DZ dataset precluded the use of Tree-Puzzle due to memory constraints inherent to the program, and thus branch lengths for the DZ dataset were estimated using PAML version 3.15 (Yang 1997) under a WAG+ Γ model, as the VT model is not available with this software.

Age estimates were determined using two different programs. For Non-Parametric Rate Smoothing (NPRS) (Sanderson 1997), and Penalized Likelihood (PL) (Sanderson 2002) analyses, the program r8s version 1.7 was used (Sanderson 2003).

All PL analyses used a logarithmic penalty, and the associated smoothing parameter was optimized via cross validation. Confidence intervals were generated for the nodes of interest only. In some cases, r8s failed to estimate confidence intervals for unknown reasons and in these cases only the age estimates were reported.

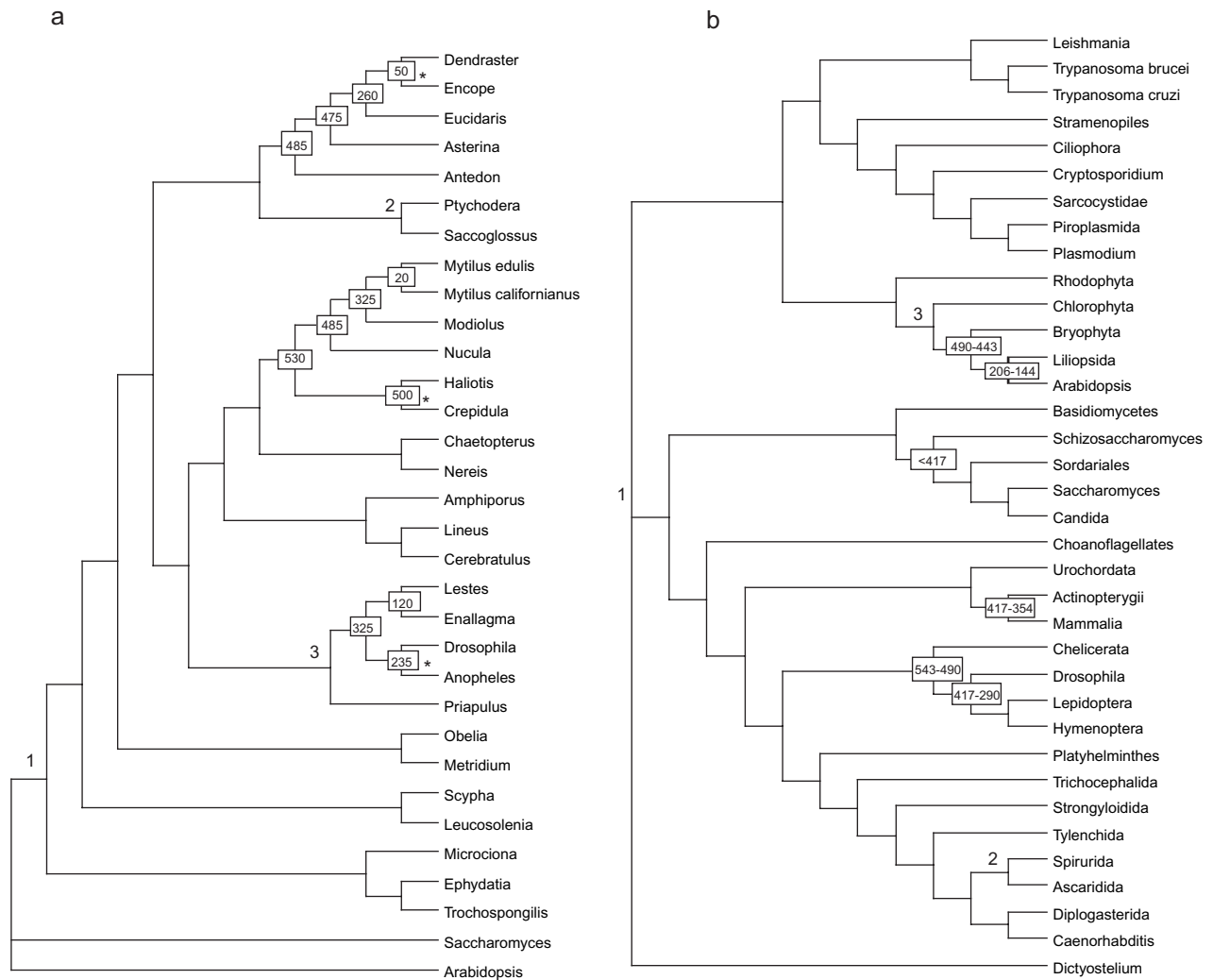


FIG. 1.—Assumed topologies for molecular clock studies. *a*) Topology used by Peterson-Butterfield (PB) in their analyses (Peterson and Butterfield 2005). *b*) Topology from the Douzery (DZ) dataset (Douzery et al. 2004). The nodes examined in the current study are labeled with the large numbers. For each dataset, one deep node and two shallow nodes (one affected by constraints, one independent) were chosen. Boxed numbers indicate fossil-dated (in millions of years) constrained nodes taken from the original studies. Fossil-dated nodes labeled with asterisks are those that were retained in the combined analysis of fossil calibrations and taxon sampling.

It should be noted here that Yang (2006) has pointed out a theoretical problem with identifiability of age estimates with the PL and NPRS methods when fossil dates with uncertainty are employed as calibrations. Nevertheless, we chose to examine the performance of these methods as they continue to be widely used.

All Bayesian relaxed clock analyses were conducted using the MULTIDISTRIBUTE software package (ESTBRANCHES & MULTIDIVTIME5b, (Thorne and Kishino 2002)). Using MULTIDIVTIME, it is possible to analyze multi-gene datasets as concatenated “super-genes” or as separate, component genes sharing the same underlying topology, but with different branch lengths. For all relevant analyses, both methods were used. All analyses discussed here were conducted using an adaptation to MULTIDIVTIME which allows for a discrete gamma rate heterogeneity among sites, implemented by Kumazawa, and available on the MULTIDISTRIBUTE website. Settings for MULTIDIVTIME priors were taken from Douzery et al. (2004).

Fossil Constraints

i) Method of constraining fossils

All fossil constraints were taken from the original studies (Douzery et al. 2004; Peterson and Butterfield 2005). For PB, nodes were constrained under three different methods: with the fossil age as the fixed age of the node (“all fixed” trial), with the fossil age as a minimum age boundary and the parent node’s fossil age (if applicable) as a maximum age boundary (“nearest-neighbour” trial), and finally with the fossil age as a minimum age boundary for the node, and a global maximum age limit for all constrained nodes of 1.5 billion years (“upper limit” trial) (figure 2a). Age estimates were determined using NPRS and PL in r8s for each of the three scenarios, and confidence intervals were generated for the three nodes of interest.

For DZ, five of the six fossil calibrations were originally published with minimum and maximum boundaries,

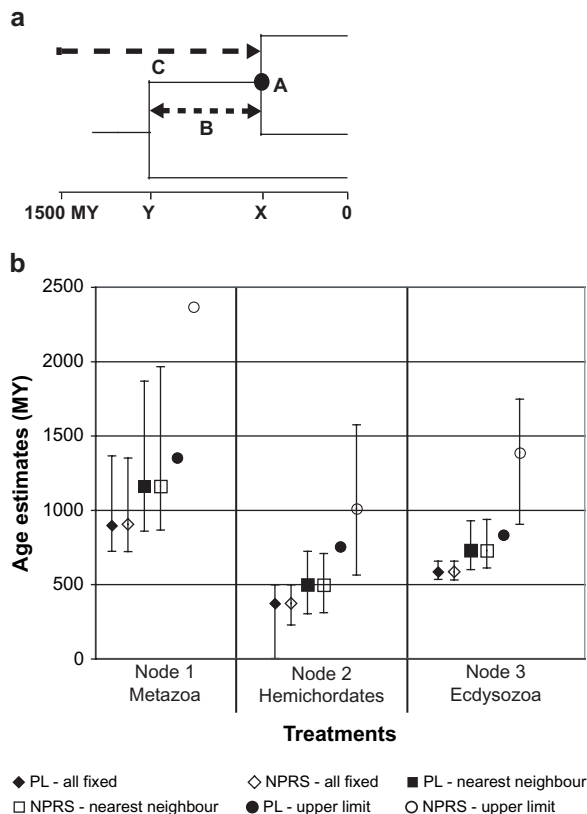


FIG. 2.—*a*) Depiction of constraint methods for nodes with a fossil calibration. A: fixed age constraint (age = X). B: “nearest-neighbour”-constraint ($Y \geq \text{age} \geq X$). C: upper limit constraint ($1500 \geq \text{age} \geq X$). *b*) Age estimates for the three nodes of interest in the PB dataset under all three (A-C) different methods of constraining fossil calibrations. Ages were generated in r8s under Penalized Likelihood (PL) and Non-Parametric Rate Smoothing (NPRS). Both models of rate evolution are presented here to highlight the relative effect of the fossil constraint methods. Confidence intervals were generated in r8s. Data points with missing confidence intervals are cases where the program failed to find a point of convergence.

while the sixth was treated as a maximum boundary only. These boundaries were based on the bounds of the geological periods in which the fossils associated with the nodes were found. We could not apply our “nearest-neighbour” constraint method in this case, as only two of the constrained nodes were associated with a constrained parent node. Therefore, for the DZ dataset, the constraints and associated minimum and maximum boundaries described in Douzery et al. (2004) were used instead of the “nearest-neighbour” method employed for the PB dataset. Fixed fossil constraints were equally impossible to apply, as the reported constraints were presented as ranges, not as fixed values as in the PB dataset. An “upper limit” trial was not assayed, based on the aberrant age estimates generated under such a trial for the PB dataset.

ii) Single fossil trials

To evaluate the impact of using only single fossil constraints on age estimates, both datasets were systematically analyzed using each of the provided fossil constraints singly. Age estimates were obtained using

Penalized Likelihood (PL) in r8s and Bayesian methods using MULTIDIVTIME under separate and concatenated conditions. For single fossil trials, the fossil date used was constrained using the “nearest-neighbour” method for PB and the minimum/maximum boundaries for the DZ dataset (as described above). The results from these ‘single fossil trials’ were compared to age estimates generated when all provided fossil dates were incorporated in the analysis.

Taxon Sampling

Four different levels of taxon sampling were tested. In all trials, taxa required to maintain a fossil calibration point were preserved. In the least exclusive “lineage” trial, one taxon from each terminal bifurcation (cherry) was deleted. The decision as to which taxa to remove was arbitrary. For PB, this resulted in 8 taxa of 32 being removed, and for DZ, 13 of 36. In the “nodes” trial, all taxa were removed that were not required to preserve the three nodes of interest on the tree and the fossil calibration nodes; otherwise choices of taxa to retain the nodes of interest were random. For PB, this led to 12 taxa eliminated, and 22 taxa excluded from DZ. The most extensive taxon deletion trial for DZ was the “node 1” trial, which preserved only those taxa required to maintain the deepest node on the tree and the fossil calibration nodes. Here, approximately half of the taxa were removed from each dataset: 15 of 32 from PB, and 25 of 36 from DZ. A fourth and final taxon sampling trial, “fossils kept”, was set up in conjunction with a fossil calibration trial. Here, three fossil calibrations were chosen from the PB dataset (marked with an asterisk in fig. 1a). All other fossil calibrations were disregarded, and taxa were preserved only if involved in maintaining the presence of either a node of interest, or one of the remaining fossil calibration points. This was the most extensive taxon deletion for this dataset, with 18 of 32 taxa excluded. The list of taxa preserved and deleted from each of these analyses, and the resultant trees are available in tables S1&2 and figures S1 & S2 of the supplementary material.

For all reduced datasets, age estimates were determined using Penalized Likelihood, as well as Bayesian relaxed clock methods under concatenated gene conditions. The PB dataset was also analyzed using Bayesian methods with unlinked branch lengths, as previous work had shown that this separate gene model provided a better fit to the data (Roger and Hug 2006). This analysis was not included for the DZ dataset due to its size and the prohibitive computational time required.

Results and Discussion

Impact of Fossil Calibration Constraints

The three different methods of constraining fossil calibrations: fixed, “nearest-neighbour,” and global upper bound of 1500 MY (fig. 2a), resulted in highly divergent age estimates. Estimates from using the fossil dates as fixed times of divergence were younger than those estimated under the “nearest-neighbour” constraints. Conversely, ages

estimated using the global upper limit of 1,500 MY for all fossil calibrations lead to age estimates skewed towards extremely ancient divergences. The results for the deepest node on the tree are depicted in figure 2*b*, with both NPRS and PL methods shown. The r8s program failed to calculate confidence intervals for a number of the older age estimates for unknown reasons.

While there is no objective method for determining which of the fossil constraint conventions, if any, is the most appropriate, some intuitive points can be discussed. First, it is apparent from figure 2 that the method of constraining fossil calibrations does have a significant effect on the age estimates produced, affecting the age of the deepest node by hundreds of millions of years for all relaxed clock methods we examined. Indeed the date estimates yielded by the “upper-limit” constraint method are in some cases absurd – for instance, using the NPRS method an age estimate for the origin of metazoa is pushed deep into the Proterozoic. This indicates that the debate surrounding the method of constraining fossils in molecular clock analysis (Graur and Martin 2004; Hedges and Kumar 2004; Reisz and Muller 2004; Blair and Hedges 2005; Glazko, Koonin and Rogozin 2005) is justified; molecular clock studies would benefit from greater caution taken when implementing these constraints. The real problem here and in other studies (Benton and Donoghue 2007) seems to be associated with how to define a maximum age constraint, especially when using software tools which do not permit ‘soft upper bounds’ (i.e. smoothly decreasing probability functions with older ages) as the r8s and MULTIDISTRIBUTE programs used here. Obviously, only fossils that are accepted by the paleontological community as belonging to a particular node and which have been dated with some confidence should be used as minimum bounds. Ideally, sequential fossil calibrations should be used along a lineage, in order to allow for the implementation of some kind of maximum age constraints, as seen in our “nearest-neighbour” calibrations (a similar procedure has been described by Benton and Donoghue as “phylogenetic bracketing”). In any other scenario, deriving a maximum age constraint, be it from a stratigraphic calibration, or an inference from indirect or external information, is much more difficult. Nevertheless, due to mathematical constraints inherent to relaxed clock dating, most available methods, including the two used here, require at least one maximum age constraint for accurate date estimation to give a unique solution.

The paucity of the fossil record, specifically for deeper divergences, may continue to be a limiting factor in the accuracy of dating studies. However, Berney and Pawlowski (2006) have recently made advances in this area through the use of the continuous microfossil record for deep eukaryote divergences, which provided the kind of multiple lineage-specific fossil calibrations we consider ideal for dating studies.

Impact of Using Single Fossil Calibrations

Considerable debate has surrounded the use of single versus multiple fossil constraints on molecular dating analyses (e.g. see Graur and Martin 2004; Hedges and Kumar

2004). We examined the impact of using a single fossil calibration compared to multiple fossil calibrations to estimate the age of nodes in both the PB and DZ analyses. Age estimates generated using only a single calibration varied widely depending on the fossil calibration used. For the PB dataset, a linear trend was observed between age estimates and the age of the fossil calibration used under the Bayesian relaxed clock method with unlinked branch lengths (fig. 3*a*) with ages for the deepest node ranging from 290-840 MY, a 550 MY spread. In PL analysis, a larger range of age estimates was produced, but without the linear trend (fig. 3*b*), where ages ranged from 760-1850 MY, a spread of more than a billion years. Interestingly, when the DZ dataset was analyzed the Bayesian ‘unlinked’ method, the linear trend seen with the PB dataset did not exist (fig. 3*c*). This may be due to the increased sequence information in this larger dataset: there is enough information in the alignment to over-ride a possible small sample bias affecting the single fossil calibration analysis. This trend may, alternatively, be due to the different fossil constraint methods used for the two datasets. Nevertheless, a wide range of age estimates was still generated for the deepest node (310-895 MY). The trend of older age estimates covering a larger range in PL compared to Bayesian analyses held true for the larger dataset, with ages for the deepest node in the DZ dataset estimated between 1060-3770 MY (fig. 3*d*). In cases where a fossil calibration did not have a maximum boundary, the r8s software was occasionally unable to generate age estimates under PL, and was consistently incapable of generating confidence intervals. It should be noted that this behaviour is not unexpected because a unique solution cannot be obtained from PL when using a single fossil calibration with an associated interval of uncertainty (Yang 2006); the optimal estimates themselves will form an interval. MULTIDIVTIME is not able to generate age estimates in the absence of at least one upper and one lower boundary, so for some fossil calibrations it was impossible to conduct this analysis.

These trials indicate that, regardless of dataset size, the use of only a single fossil calibration will result in the significant introduction of error into age estimates. It should be emphasized that it is not necessarily true that multiple fossil calibrations will result in a less biased age estimate, as the calibrations chosen must be consistent with each other (Near, Meylan and Shaffer 2005) and properly constrained (Benton and Donoghue 2007).

Taxon Sampling

The impact of taxon sampling on date estimates were, in general, less pronounced than fossil calibration sampling. In the “lineage” trial, one cherry from each terminal bifurcation was removed, while in the “nodes” and “node 1” trials, only those taxa required to maintain the fossil-constrained nodes, as well as the three nodes of interest or only the deepest node on the tree respectively, were kept. PL analysis of the PB and DZ datasets are shown in figure 4*a*. Not unexpectedly, the variation in age estimates with different taxonomic sampling schemes is most pronounced (i.e. a range of approximately 300 MY) in the smaller PB

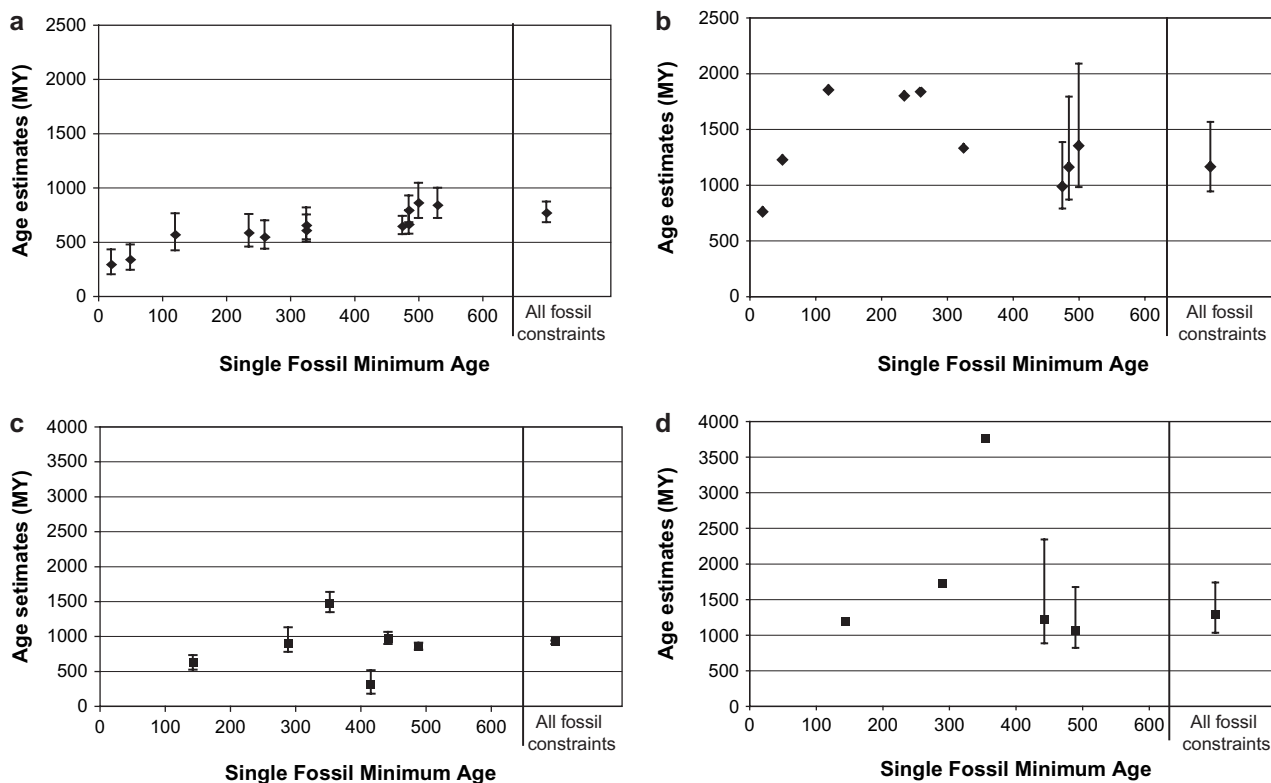


FIG. 3.—Results from the single fossil calibration trial. Age estimates for the deepest node from the PB and DZ datasets were generated by systematically utilizing only a single applicable fossil constraint in both MULTIDIVTIME and r8s. In MULTIDIVTIME, the PB dataset was analyzed with a VT+ Γ model and unlinked branch lengths (a) while Bayesian age estimates for the DZ dataset were determined under a WAG+ Γ model (c). Penalized likelihood was implemented using r8s for both the PB (b) and DZ (d) datasets. The “nearest-neighbour” fossil constraint method (Fig 2a) was used for the PB dataset. For the DZ dataset, fossil constraints were taken from the original paper as both a minimum and maximum bound were specified. Age estimates were plotted using the minimum age of the calibration point on the x-axis.

datasets and for the deepest node. Indeed there was almost no impact observed for node 2 in PB or node 1 and 2 for DZ, despite the removal of approximately half of the taxa from each dataset in the extreme case of the ‘node’ trials. A broadly similar situation occurred with the Bayesian analyses (fig. 4b). As observed previously (Roger and Hug 2006), the unlinked Bayesian analyses of the PB dataset generated significantly younger age estimates for node 1 than the linked analyses. However, it is also interesting to note that in these unlinked branch length analyses (fig. 4b, first column) the age estimates vary significantly less compared to the linked branch length analyses (fig. 4b, second column). We previously showed that unlinked branch lengths gave a significantly better fit to the data, even given the increased number of parameters optimized (Roger and Hug 2006), and the current analyses indicate that the unlinked method estimates are also more robust to taxonomic sampling. Overall, these trials suggest that taxonomic sampling has a relatively minor affect on age estimation for a given set of multiple fossil constraints.

Combined Trial

To discern the relative impact of taxon sampling compared to fossil sampling on age estimates, we carried out a trial where both factors were varied simultaneously.

Briefly, for these analyses the same taxonomic samples (i.e. ‘lineage,’ ‘all nodes’ and node 1) were used for the PB dataset as described above, but only a single fossil constraint was imposed on the analysis. Three such single fossil constraints were used, an ‘old,’ ‘medium’ and a ‘young’ constraint (see nodes marked with asterisks in fig. 1a). A fourth taxonomic sample that included only taxa to retain these three fossil constraints and node 1 (called “fossils kept”) was also used. Both PL and unlinked Bayesian analyses were carried for each of the ‘single’ fossil constraints and these taxonomic samples and the results are shown in figures 5a and 5b respectively. Once again, the PL analyses yielded much older node 1 age estimates, and showed greater variation in these estimates with taxonomic sampling and fossil sampling (fig. 5a) than the Bayesian analyses (fig. 5b). More importantly, however, it is clear that the impact of fossil calibrations on age estimates from either relaxed clock method is much greater than taxonomic sampling (fig. 5). Therefore, all other things being equal, it appears that the number and choice of fossils used to calibrate in relaxed molecular clock analyses is much more important than how well taxa are sampled in the analyses. Further, in designing molecular dating studies, if the phylogeny is known, then more effort should be directed at selecting key taxa to maximize the number of appropriate well-constrained fossils that can be used for analyses rather than generating extremely taxon-rich datasets.

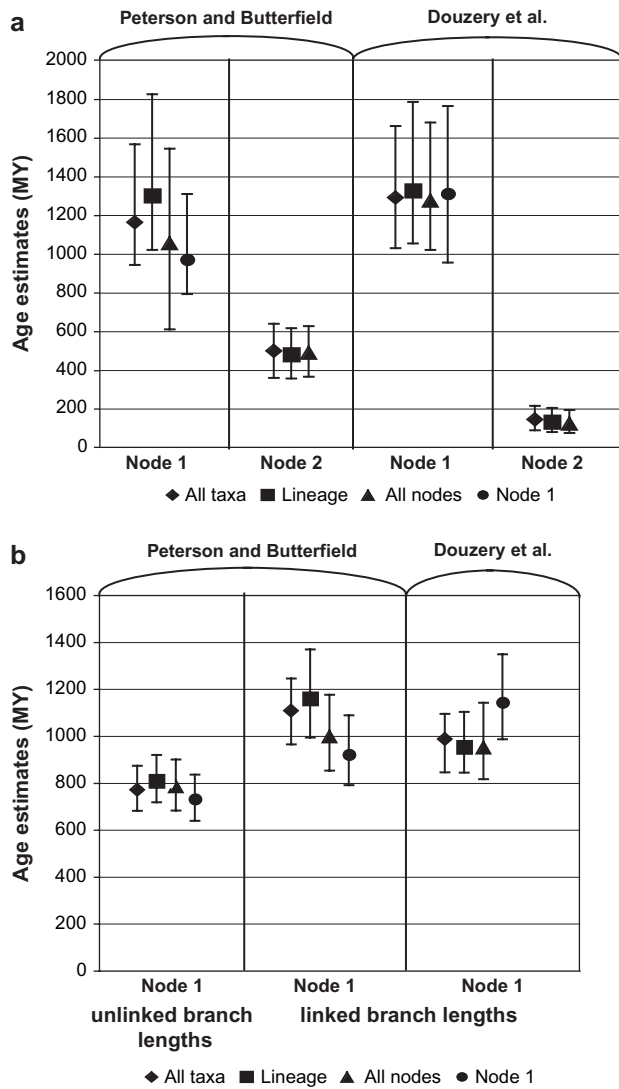


FIG. 4.—Age estimates from taxon sampling trials. Both datasets were analyzed under PL in r8s, and confidence intervals were generated (a). Bayesian analyses were conducted in MULTIDIVTIME with a VT+ Γ model for the PB dataset and a WAG+ Γ model for the DZ dataset. The PB dataset was analyzed with both linked and unlinked branch lengths, while the DZ dataset was analyzed only with linked branch lengths due to computational constraints (b). Only the results for the deepest nodes ('node 1' for each dataset) are depicted, as trends seen here held true across the trees. For both a & b, the three different taxon sampling trials are depicted: 'lineage' in which one cherry from each terminal bifurcation was deleted, 'all nodes,' in which only those taxa required for maintenance of the three node of interest and the constrained nodes were kept, and 'node 1,' in which only those taxa required for maintenance of the deepest node on the tree and the fossil-constrained nodes were kept for the analysis.

Conclusions

Our analyses confirm that methods of constraining fossil dates, sampling of fossil constraints, sampling of species and properties/assumptions of relaxed clock methods all contribute to variation in molecular clock estimates of deep divergences. Of these factors, several seem to be the most important. The method of constraining the calibrations has a crucial effect on the ages produced, and the use of single fossil constraints for an analysis generates highly unpredict-

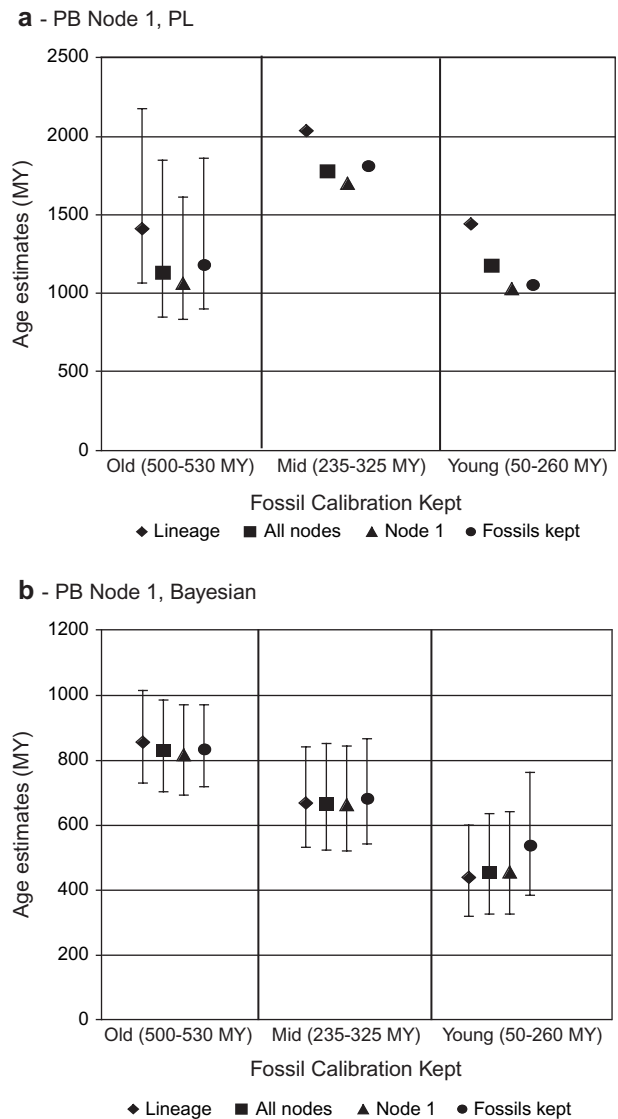


FIG. 5.—Age estimates from combined taxon sampling and fossil calibration trial for the PB dataset. Estimates were generated under PL in r8s (a) as well as under a Bayesian model with unlinked branch lengths in MULTIDIVTIME with a VT+ Γ model of rate evolution (b). For each trial, four taxon sampling reduced datasets ('lineage,' 'all nodes,' 'node 1,' and 'fossils kept') were run with a single fossil calibration. Three fossil calibrations of varying age ranges were chosen for use (highlighted by asterisks in Fig 1a), and were constrained under the "nearest-neighbour" method (Fig 2a). Age estimates varied significantly more based on the choice of the fossil constraint's age range compared to the number of taxa excluded from the analyses.

able and variable estimates. In general, greater care needs to be taken when assigning both minimal and, especially, maximal constraints on a node. The use of multiple, well-supported fossil dates is ideal, particularly if they are sequential along a lineage. Considering the newly available methods for determining the error due to fossil calibrations (Near and Sanderson 2004), and the steadily improving fossil record yielded by progress in paleontology, better-constrained and more applicable fossil dates are becoming available for many divergences reflected in molecular phylogenies (Benton and Donoghue 2007),

especially within the Phanerozoic. There is no longer any excuse for using a single calibration point in a study, or for ignoring the potential error introduced from an inconsistent calibration point.

The impact of taxon sampling on our age estimates was surprisingly low, with age estimates remaining generally consistent even when nearly 50% of the taxa in the analyses were excluded. This implies that, once a dataset has been developed that can be used to generate a resolved phylogeny, additional data may not be required to increase confidence in age estimates. Thus, priority should be given to obtaining as many fossil calibrations as possible and the taxonomic sampling of the sequence dataset be adjusted so as to optimize the chances of obtaining the correct tree that will contain the appropriate nodes corresponding to the fossil calibration dates. Attempts should be made to appropriately constrain fossil calibration dates (Benton and Donoghue 2007) and assess which, if any of the fossils are incompatible with the rest (Near and Sanderson 2004).

Supplementary Material

A list of taxa included and excluded from each taxon-sampling trial, as well as the representative topologies used in these analyses are available as supplementary tables S1 and S2, and figures S1 and S2.

Acknowledgments

We would like to thank Catharine Dunn, Jessica Leigh, and Jeffrey Thorne for technical assistance with the MULTIDISTRIBUTE package. This work was supported by Discovery grant #227085-2005 from the Natural Sciences and Engineering Research Council (NSERC) and grant MOP-62809 from the Canadian Institutes of Health Research (CIHR) awarded to A.J.R.. A.J.R. is supported by a fellowship from the Alfred P. Sloan Foundation, the Peter Lougheed/CIHR New Investigator fellowship, and the Canadian Institute for Advanced Research, Program in Evolutionary Biology. L.A.H. is supported by a NSERC PGS-M scholarship and a Killam Pre-Doctoral Fellowship.

Literature Cited

- Aris-Brisou S, Yang Z. 2003. Bayesian models of episodic evolution support a late Precambrian explosive diversification of the Metazoa. *Mol Biol Evol.* 20:1947–1954.
- Benton MJ, Ayala FJ. 2003. Dating the tree of life. *Science.* 300:1698–1700.
- Benton MJ, Donoghue PC. 2007. Paleontological evidence to date the tree of life. *Mol Biol Evol.* 24:26–53.
- Berney C, Pawlowski J. 2006. A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. *Proc Biol Sci.* 273:1867–1872.
- Blair JE, Hedges SB. 2005. Molecular clocks do not support the Cambrian explosion. *Mol Biol Evol.* 22:387–390.
- Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet.* 6: 361–375.
- Douzery EJ, Snell EA, Baptiste E, Delsuc F, Philippe H. 2004. The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? *Proc Natl Acad Sci USA.* 101:15386–15391.
- Drummond AJ, Ho SY, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4:e88.
- Glazko GV, Koonin EV, Rogozin IB. 2005. Molecular dating: ape bones agree with chicken entrails. *Trends Genet.* 21:89–92.
- Graur D, Martin W. 2004. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends Genet.* 20:80–86.
- Gribaldo S, Philippe H. 2002. Ancient phylogenetic relationships. *Theor Popul Biol.* 61:391–408.
- Hedges SB, Blair JE, Venturi ML, Shoe JL. 2004. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol Biol.* 4:2.
- Hedges SB, Kumar S. 2004. Precision of molecular time estimates. *Trends Genet.* 20:242–247.
- Hedtke SM, Townsend TM, Hillis DM. 2006. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. *Syst Biol.* 55:522–529.
- Kishino H, Thorne JL, Bruno WJ. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Mol Biol Evol.* 18:352–361.
- Kumar S, Hedges SB. 1998. A molecular timescale for vertebrate evolution. *Nature.* 392:917–920.
- Near TJ, Meylan PA, Shaffer HB. 2005. Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. *Am Nat.* 165:137–146.
- Near TJ, Sanderson MJ. 2004. Assessing the quality of molecular divergence time estimates by fossil calibrations and fossil-based model selection. *Philos Trans R Soc Lond B Biol Sci.* 359:1477–1483.
- Peterson KJ, Butterfield NJ. 2005. Origin of the Eumetazoa: testing ecological predictions of molecular clocks against the Proterozoic fossil record. *Proc Natl Acad Sci USA.* 102: 9547–9552.
- Peterson KJ, Lyons JB, Nowak KS, Takacs CM, Wargo MJ, McPeck MA. 2004. Estimating metazoan divergence times with a molecular clock. *Proc Natl Acad Sci USA.* 101: 6536–6541.
- Reisz RR, Muller J. 2004. The comparative method for evaluating fossil calibration dates: a reply to Hedges, Kumar. *Trends Genet.* 20:596–597.
- Roger AJ, Hug LA. 2006. The origin and diversification of eukaryotes: problems with molecular phylogenetics and molecular clock estimation. *Philos Trans R Soc Lond B Biol Sci.* 361:1039–1054.
- Sanderson MJ. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol Biol Evol.* 14:1218–1231.
- Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol Biol Evol.* 19:101–109.
- Sanderson MJ. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics.* 19:301–302.
- Schmidt HA, Strimmer K, Vingron M, von Haeseler A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics.* 18: 502–504.
- Thorne JL, Kishino H. 2002. Divergence time and evolutionary rate estimation with multilocus data. *Syst Biol.* 51:689–702.
- van Tuinen M, Hadly EA. 2004. Error in estimation of rate and time inferred from the early amniote fossil record and avian molecular clocks. *J Mol Evol.* 59:267–276.
- Vinh le S, Von Haeseler A. 2004. IQPNNI: moving fast through tree space and stopping in time. *Mol Biol Evol.* 21: 1565–1571.

- Welch JJ, Bromham L. 2005. Molecular dating when rates vary. *Trends Ecol. Evol.* 20:320–327.
- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *CABIOS.* 13: 555–556.
- Yang Z. 2006. *Computational Molecular Evolution.* Oxford, UK: Oxford University Press.
- Yang Z, Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol Biol Evol.* 23:212–226.
- Yang Z, Yoder AD. 2003. Comparison of likelihood and Bayesian methods for estimating divergence times using multiple gene Loci and calibration points, with application to a radiation of cute-looking mouse lemur species. *Syst Biol.* 52:705–716.
- Zuckerandl E, Pauling L. 1965. Molecules as documents of evolutionary history. *J Theor Biol.* 8:357–366.

Dan Graur, Associate Editor

Accepted May 30, 2007