



Cite this article: Lloyd GT, Bapst DW, Friedman M, Davis KE. 2016 Probabilistic divergence time estimation without branch lengths: dating the origins of dinosaurs, avian flight and crown birds. *Biol. Lett.* **12**: 20160609. <http://dx.doi.org/10.1098/rsbl.2016.0609>

Received: 20 July 2016

Accepted: 14 October 2016

Subject Areas:

evolution, palaeontology, taxonomy and systematics

Keywords:

birds, dinosaurs, divergence time, phylogeny

Author for correspondence:

G. T. Lloyd

e-mail: graemetlloyd@gmail.com

[†]Present address: School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK.

[‡]Present address: Museum of Paleontology and Department of Earth and Environmental Science, University of Michigan, Ann Arbor, MI 48109, USA.

An invited contribution to the special feature 'Putting fossils in trees: combining morphology, time, and molecules to estimate phylogenies and divergence times'.

Electronic supplementary material is available online at <https://dx.doi.org/10.6084/m9.figshare.c.3552543>.

Probabilistic divergence time estimation without branch lengths: dating the origins of dinosaurs, avian flight and crown birds

G. T. Lloyd^{1,†}, D. W. Bapst^{2,3}, M. Friedman^{4,‡} and K. E. Davis⁵

¹Department of Biological Sciences, Macquarie University, New South Wales 2109, Australia

²Department of Geology, University of California Davis, One Shields Avenue, Davis, CA 94568, USA

³Department of Geology and Geological Engineering, South Dakota School of Mines and Technology, Rapid City, SD 57701, USA

⁴Department of Earth Sciences, University of Oxford, South Parks Road, Oxford OX1 3AN, UK

⁵Department of Biology, University of York, Wentworth Way, York YO10 5DD, UK

id GTL, 0000-0001-6887-3981; DWB, 0000-0002-9087-1103; MF, 0000-0002-0114-7384; KED, 0000-0002-9235-7853

Branch lengths—measured in character changes—are an essential requirement of clock-based divergence estimation, regardless of whether the fossil calibrations used represent nodes or tips. However, a separate set of divergence time approaches are typically used to date palaeontological trees, which may lack such branch lengths. Among these methods, sophisticated probabilistic approaches have recently emerged, in contrast with simpler algorithms relying on minimum node ages. Here, using a novel phylogenetic hypothesis for Mesozoic dinosaurs, we apply two such approaches to estimate divergence times for: (i) Dinosauria, (ii) Avialae (the earliest birds) and (iii) Neornithes (crown birds). We find: (i) the plausibility of a Permian origin for dinosaurs to be dependent on whether *Nyasasaurus* is the oldest dinosaur, (ii) a Middle to Late Jurassic origin of avian flight regardless of whether *Archaeopteryx* or *Aurornis* is considered the first bird and (iii) a Late Cretaceous origin for Neornithes that is broadly congruent with other node- and tip-dating estimates. Demonstrating the feasibility of probabilistic time-scaling further opens up divergence estimation to the rich histories of extinct biodiversity in the fossil record, even in the absence of detailed character data.

1. Background

Divergence times are often estimated by combining fossil information with a phylogenetic hypothesis. In a classical clock-based 'node-dating' framework, a tree of extant taxa with branch lengths representing character changes is dated with reference to a set of fossil calibrations that constrain the minimum age for particular nodes. However, multiple important divergences within the tree of life are not bracketed by extant lineages (e.g. origin of birds) and thus cannot be estimated using molecular data. More recently, 'tip-dating' approaches have allowed extinct taxa to be included as terminals, with phylogenetic inference and divergence time estimation occurring simultaneously without reference to node calibrations. This opens up the possibility for using character change from molecular or morphological sources (or both) when estimating divergences between extinct and extant lineages, or even among entirely extinct lineages. Independently of these advancements, palaeontologists have been using stratigraphic ages to directly date divergences on existing phylogenies. Often constructed from both

morphological cladograms and taxonomy, these lack measures of character change [1], with their strength instead relying on the stratigraphic distribution of fossils [2]. We term these ‘*a posteriori*’ time-scaling (APT) approaches.

Most of these APT approaches work independently of inferred amounts of character change (but see [3]), relying solely on occurrence data. Broadly speaking, three types are typically applied: (i) minimum-age dating [4], (ii) extending branch durations by adding a constant [5] and (iii) branch duration sharing [3]. The latter two utilize a minimum age (based on first appearances) dated tree as a preliminary step. These approaches suffer from arbitrary choices of required variables and make strong assumptions of the quality of the fossil record without reference to said fossil record. While Bayesian tip-dating methods have recently become accessible to completely extinct clades ([6]; other papers in Special Feature), their availability has also coincided with improvements in APT approaches [7].

Here, we assemble a novel phylogenetic hypothesis for Mesozoic dinosaurs and time-scale it using two different *probabilistic* APT methods—cal3 [7] and a new method developed from the node-dating approach of Hedman [8]. We use these results to ask three questions involving major evolutionary transitions: (i) When did Dinosauria originate?, (ii) When did birds originate? and (iii) How old is the avian crown? The first two involve dating the divergence of an extinct lineage, but the latter considers a split among extant taxa, thus permitting comparisons between our estimates and those from published clock-based analyses.

2. Material and methods

A novel ‘metatree’ approach (electronic supplementary material, figures S1, S2)—which operates in a similar way to formal supertree approaches, but generates source trees directly from character–taxon matrices rather than published figures (see the electronic supplementary material)—generated 1000 phylogenetic hypotheses containing 962 separate operational taxonomic units (OTUs). Of these, 100 were sampled at random to account for phylogenetic uncertainty (a larger number being computationally prohibitive). Ages came from fossil occurrence data in the Paleobiology Database and primary literature. Divergence times were then estimated for three nodes on the tree: (i) the origin of dinosaurs (inclusive and exclusive of the potential oldest dinosaur, *Nyasaosaurus* [9]), (ii) the origin of Avialae (inclusive or exclusive of the purported earliest bird, *Aurornis* [10]) and (iii) the origin of crown birds (Neornithes).

We first applied the cal3 method of Bapst [7], which requires *a priori* estimates of diversification and sampling rates to draw likely divergence dates under a birth–death–sampling model and operates in a similar manner to the fossilized birth–death (FBD) process [11]. Sampling and extinction estimates were obtained by stochastically sampling sets of congruent taxon ranges from the occurrence data, via the function ‘seqTimeList’ in the R package palaeotree [12]. We calculated maximum-likelihood estimates of sampling and extinction rates using the resulting range frequency distributions [13], and used our extinction rate estimates as a proxy for speciation rate. To account for the uncertainty in these rate estimates, each cal3 tree was time-scaled with a different set of estimated rates. As the cal3 approach is stochastic, we applied it 1000 times and aggregated the results into distributions.

Next, we applied a novel algorithm based on the node-dating approach of Hedman [8]. Not all nodes can be appropriately dated using this algorithm (electronic supplementary material, figure S4), and thus, we utilized a novel approach to obtain missing dates via a randomization process (electronic supplementary

material). Again, 1000 dates were estimated for each node on each tree. Thus for the three nodes distributions were produced from 100 000 point estimates for both the cal3 and Hedman approaches.

Tree searches were performed in TNT [14] and all other analyses in R [15].

3. Results

The main results are summarized in figure 1 and electronic supplementary material, table S2. Overall, the cal3 and Hedman approaches broadly agree (mean difference between median estimates = 5.4 Ma). However, the shapes of the distributions often vary (figure 1), with the Hedman approach giving less precise estimates (mean highest posterior density (HPD) width is 6.3 Ma greater than for cal3). The probability of a Permian origin for dinosaurs depends on whether *Nyasaosaurus* is (2.6%, cal3; 9.6%, Hedman), or is not (0.2%, cal3; 1.0%, Hedman) considered a dinosaur. Conversely, the difference in the median age estimate is minimal regardless of whether *Archaeopteryx* or *Aurornis* is considered the earliest bird (2.8 Ma, cal3; 0.4 Ma, Hedman), despite a 7 Ma difference in their lower bounds. Finally, the 95% HPD width for crown birds is the largest for any node (23.4 Ma, cal3; 39.5 Ma, Hedman) due to both greater phylogenetic uncertainty and a poorer fossil record, creating overlap with multiple published clock-based estimates.

4. Discussion

Application of our metatree approach results in a well resolved strict consensus (electronic supplementary material, figure S3) which we attribute to the inclusion of a taxonomy ‘tree’ and our weighting scheme (see the electronic supplementary material), which create superior overlap and conflict resolution (compared with formal supertrees), respectively. Remaining topological uncertainty is distal to the main nodes under consideration, situated primarily among the long-necked sauropodomorphs and extinct bird clade Enantiornithes.

If *Nyasaosaurus* is the oldest dinosaur, it significantly increases the age of the dinosaurian node. However, even if this taxon falls just outside the clade the upper 95% HPD does not exclude an Early Triassic age—an older value than most macroevolutionary studies apply [16]. Proper resolution of this positional uncertainty would require a larger archosaur phylogeny and illustrates the general difficulty of estimating root values. The absence of phylogenetic uncertainty at this node also coincides with the narrowest HPD width (greatest precision) and closest agreement between cal3 and Hedman estimates. This is to be expected as the major source of difference between cal3 and Hedman is the latter approach’s greater sensitivity to the order of outgroups.

Although *Aurornis* may be a troodontid rather than a bird [17], either position does not seem to affect the estimated bird origin age here. Given the demonstrated volatility of dinosaur phylogeny [18], this is an encouraging result for the robustness of both approaches. The bird node also confirms that there is no consistent, and hence predictable, bias between both methods in terms of median age: here (unlike the root) cal3 returns the older median value.

Published age estimates for Neornithes vary considerably (figure 1). This uncertainty is in part due to a worse fossil record in crown- versus stem-birds, as well as the credibility

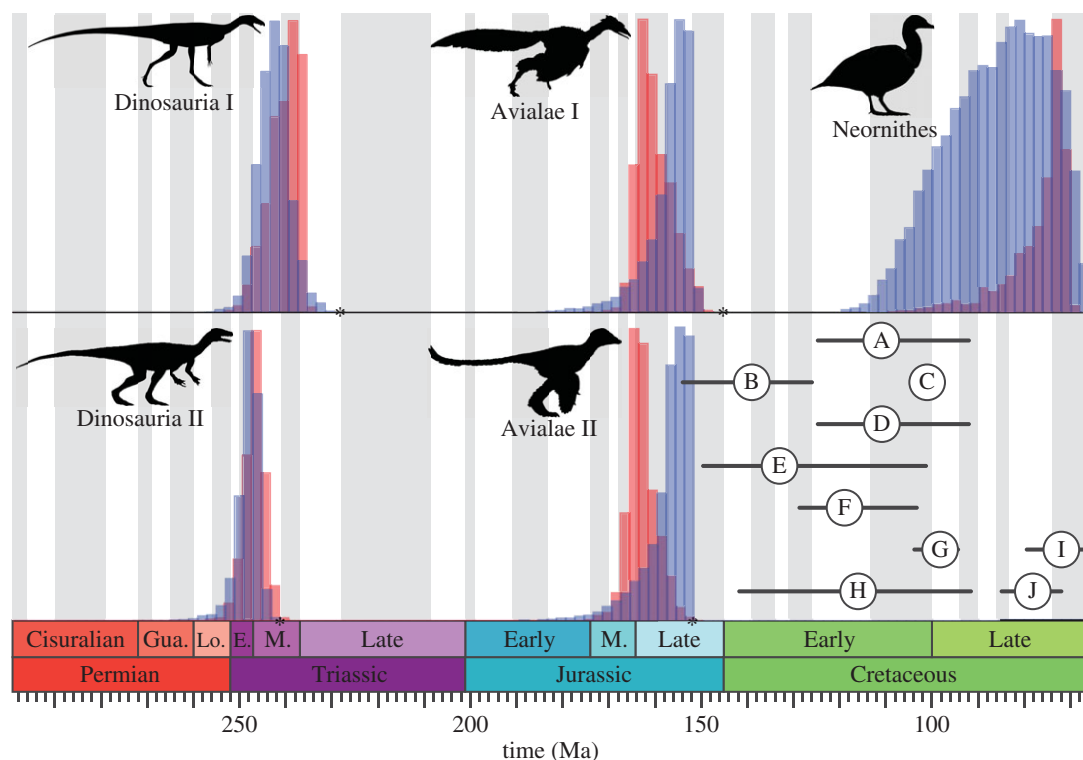


Figure 1. Comparison of probabilistic APT dates (red bars, cal3; blue bars, Hedman; see the text) for key nodes in dinosaur phylogeny: Dinosauria I (*Nyasasaurus* as sister to Dinosauria), Dinosauria II (*Nyasasaurus* nested within Dinosauria), Avialae I (*Archaeopteryx* as first bird), Avialae II (*Aurornis* as first bird) and Neornithes (crown birds). Asterisks mark minimum bound or ‘traditional’ palaeontological estimate. Molecular and morphological clock dates for Neornithes are shown in the lower right corner (A–J; electronic supplementary material, table S3): circles indicate mean and horizontal bars the 95% HPD. Silhouettes were taken from public domain images on phylopic.org (*Aurornis*, Gareth Monger; *Eoraptor*, Scott Hartman; *Vegavis*, Matt Martyniuk), or modified with kind permission from works by Sergio Pérez (*Archaeopteryx*) and Nobu Tamura (*Nyasasaurus*).

of the neornithine affinities of some Late Cretaceous taxa [19] which make calibrating clock-based approaches difficult. Our methods also seem to capture this signal, as suggested by the largest HPD widths for both cal3 and Hedman. Both approaches also share the same level of accuracy as their clock-based comparisons: 95% HPDs overlap with all but one published estimate. Furthermore, both APT median age estimates fall within the range of mean clock-based ages. The neornithine node also illustrates the larger variability in distribution shape for the Hedman approach. This reflects the reliance on the order and age of outgroups that produce the distributions on which the Hedman approach is conditioned.

Overall both of our probabilistic APT approaches are broadly congruent with clock-based estimates and each other, suggestive of robustness. Thus, not only is it feasible to apply cal3 to vertebrate data (see also [20]), cal3 and Hedman should be preferred over APTs that ignore fossil record quality. Between these approaches and the FBD model [11], adequate divergence dating is available independently of whether lineages are still extant or have character information available. It is also conceivable that these distributions could inform priors on specific divergences in clock-based approaches, such as node- or tip-dating constraints. However, there is still room for improvement. Ideally, phylogenies dated with APT methods contain almost all taxa known from the fossil

record, like the metatree constructed here, but this may be impossible for some groups. (For example, many invertebrate macrofossil species have never been included in a phylogenetic hypothesis [2].) Both APT approaches used here also make simplistic assumptions about sampling, which may vary considerably over time and space [21], and future approaches should relax such constraints as some implementations of the FBD model already allow. Regardless, the increasing ability to calculate robust divergence dates for phylogenies of fossil taxa opens more intersections on the tree of life to dating and subsequent macroevolutionary analyses.

Data accessibility. Source trees: graemetlloyd.com/matr.html; other data and code: <http://dx.doi.org/10.5061/dryad.p660m> [22].

Authors' contributions. G.T.L. compiled the data; M.F. and G.T.L. developed the Hedman time-scaling approach; K.E.D. performed the tree searches; D.W.B. and G.T.L. time-scaled the trees. All authors wrote the manuscript, gave final approval for publication and agree to be held accountable for its contents.

Competing interests. We have no competing interests.

Funding. This research was funded by BBSRC grant no. BB/K006754/1 to K.E.D.; NSF grant no. EAR-1147537 to D.W.B.; NERC grant no. NE/I005536/1 to M.F. and ARC grant no. DE140101879 to G.T.L.

Acknowledgements. We thank numerous authors and colleagues who provided matrices and publications, Steve Brusatte for discussion, and several anonymous reviewers for their helpful comments.

References

- Hunt G, Carrano M. 2010 Models and methods for analyzing phenotypic evolution in lineages and clades. In *Short course on quantitative methods in paleobiology* (eds J Alroy, G Hunt), pp. 245–269. New Haven, CT: Paleontological Society.

2. Soul L, Friedman M. 2015 Taxonomy and phylogeny can yield comparable results in comparative paleontological analyses. *Syst. Biol.* **64**, 608–620. (doi:10.1093/sysbio/syv015)
3. Ruta M, Wagner P, Coates M. 2006 Evolutionary patterns in early tetrapods. I. Rapid initial diversification followed by decrease in rates of character change. *Proc. R. Soc. B* **273**, 2107–2111. (doi:10.1098/rspb.2006.3577)
4. Smith A. 1994 *Systematics and the fossil record*. London, UK: Blackwell Scientific.
5. Laurin M. 2004 The evolution of body size, Cope's Rule and the origin of amniotes. *Syst. Biol.* **53**, 594–622. (doi:10.1080/10635150490445706)
6. Lee M, Cau A, Naish D, Dyke G. 2014 Sustained miniaturization and anatomical innovation in the dinosaurian ancestors of birds. *Science* **345**, 562–566. (doi:10.1126/science.1252243)
7. Bapst D. 2013 A stochastic rate-calibrated method for time-scaling phylogenies of fossil taxa. *Methods Ecol. Evol.* **4**, 724–733. (doi:10.1111/2041-210X.12081)
8. Hedman M. 2010 Constraints on clade ages from fossil outgroups. *Paleobiology* **36**, 16–31. (doi:10.1666/0094-8373-36.1.16)
9. Nesbitt S, Barrett P, Werning S, Sidor C, Charig A. 2013 The oldest dinosaur? A Middle Triassic dinosauriform from Tanzania. *Biol. Lett.* **9**, 20120949. (doi:10.1098/rsbl.2012.0949)
10. Godefroit P, Cau A, Hu D, Escuillié F, Wu W, Dyke G. 2013 A Jurassic avialan dinosaur from China resolves the early phylogenetic history of birds. *Nature* **498**, 359–362. (doi:10.1038/nature12168)
11. Heath T, Huelsenbeck J, Stadler T. 2014 The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proc. Natl Acad. Sci. USA* **111**, E2957–E2966. (doi:10.1073/pnas.1319091111)
12. Bapst D. 2012 paleotree: an R package for paleontological and phylogenetic analyses of evolution. *Methods Ecol. Evol.* **3**, 803–807. (doi:10.1111/j.2041-210X.2012.00223.x)
13. Foote M. 1997. Estimating taxonomic durations and preservation probability. *Paleobiology* **23**, 278–300. (doi:10.1017/S0094837300019692)
14. Goloboff P, Farris J, Nixon K. 2008 TNT, a free program for phylogenetic analysis. *Cladistics* **24**, 774–786. (doi:10.1111/j.1096-0031.2008.00217.x)
15. R Core Team. 2015. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. See <http://www.R-project.org/>.
16. Sakamoto M, Benton M, Venditti C. 2016 Dinosaurs in decline tens of millions of years before their final extinction. *Proc. Natl Acad. Sci. USA* **113**, 5036–5040. (doi:10.1073/pnas.1521478113)
17. Brusatte S, Lloyd G, Wang S, Norell M. 2014 Gradual assembly of avian body plan culminated in rapid rates of evolution across dinosaur–bird transition. *Curr. Biol.* **24**, 2386–2392. (doi:10.1016/j.cub.2014.08.034)
18. Tarver J, Donoghue P, Benton M. 2011 Is evolutionary history repeatedly rewritten in light of new fossil discoveries? *Proc. R. Soc. B* **278**, 599–604. (doi:10.1098/rspb.2010.0663)
19. Smith N, Ksepka D. 2015 Five well-supported fossil calibrations within the 'Waterbird' assemblage (Tetrapoda, Aves). *Palaeontol. Electronica* **18.1.7FC**, 1–21.
20. Halliday T, Goswami A. 2016 The impact of phylogenetic dating method on interpreting trait evolution: a case study of Cretaceous–Palaeogene eutherian body-size evolution. *Biol. Lett.* **12**, 20160051. (doi:10.1098/rsbl.2016.0051)
21. Wagner P, Marcot J. 2013 Modelling distributions of fossil sampling rates over time, space and taxa: assessment and implications for macroevolutionary studies. *Methods Ecol. Evol.* **4**, 703–713. (doi:10.1111/2041-210X.12088)
22. Lloyd GT, Bapst DW, Friedman M, Davis KE. 2016 Data from: Probabilistic divergence time estimation without branch lengths: dating the origins of dinosaurs, avian flight and crown birds. Dryad Digital Repository. (<http://dx.doi.org/10.5061/dryad.p660m>)