

# A species-level timeline of mammal evolution integrating phylogenomic data

<https://doi.org/10.1038/s41586-021-04341-1>

Received: 6 July 2021

Accepted: 13 December 2021

Published online: 22 December 2021

 Check for updates

Sandra Álvarez-Carretero<sup>1,2,8</sup>, Asif U. Tamuri<sup>3,4,8</sup>, Matteo Battini<sup>5</sup>, Fabícia F. Nascimento<sup>6</sup>, Emily Carlisle<sup>5</sup>, Robert J. Asher<sup>7</sup>, Ziheng Yang<sup>2</sup>, Philip C. J. Donoghue<sup>5</sup>✉ & Mario dos Reis<sup>8</sup>✉

High-throughput sequencing projects generate genome-scale sequence data for species-level phylogenies<sup>1–3</sup>. However, state-of-the-art Bayesian methods for inferring timetrees are computationally limited to small datasets and cannot exploit the growing number of available genomes<sup>4</sup>. In the case of mammals, molecular-clock analyses of limited datasets have produced conflicting estimates of clade ages with large uncertainties<sup>5,6</sup>, and thus the timescale of placental mammal evolution remains contentious<sup>7–10</sup>. Here we develop a Bayesian molecular-clock dating approach to estimate a timetree of 4,705 mammal species integrating information from 72 mammal genomes. We show that increasingly larger phylogenomic datasets produce diversification time estimates with progressively smaller uncertainties, facilitating precise tests of macroevolutionary hypotheses. For example, we confidently reject an explosive model of placental mammal origination in the Palaeogene<sup>8</sup> and show that crown Placentalia originated in the Late Cretaceous with unambiguous ordinal diversification in the Palaeocene/Eocene. Our Bayesian methodology facilitates analysis of complete genomes and thousands of species within an integrated framework, making it possible to address hitherto intractable research questions on species diversifications. This approach can be used to address other contentious cases of animal and plant diversifications that require analysis of species-level phylogenomic datasets.

High-throughput sequencing projects are generating hundreds<sup>1</sup> to thousands<sup>2</sup> of genome sequences, with imminent plans to sequence more than a million species<sup>11</sup>. However, the accumulation of sequenced genomes is now outpacing the analytical capacity of computer software and many of the tools required to extract information from these vast datasets are lacking<sup>12</sup>. This is particularly the case for Bayesian Markov chain Monte Carlo (MCMC) molecular-clock methods that are used routinely to infer evolutionary timescales<sup>4</sup>, for groups including pathogens<sup>13</sup>, plants<sup>14</sup> and animals<sup>15</sup>, but which are computationally expensive. Consequently, these methods have been limited in their application to datasets comprising dozens of genes for many species<sup>5,16</sup> or many genes for dozens of species<sup>7,17</sup>, constraining the scope of evolutionary questions that can be addressed.

Although fast non-Bayesian clock-dating methods have been developed<sup>18</sup>, these typically do not incorporate uncertainties on evolutionary branch lengths<sup>19</sup> or arbitrary fossil calibration densities<sup>20,21</sup>. However, the Bayesian approach—despite its computational expense—is appealing because it facilitates explicit integration of these uncertainties<sup>4</sup>. Furthermore, large genomic datasets enable inference of precise timelines that can be used to obtain correlations between diversification events and the geological and climatic evolution of our planet<sup>4</sup>. Although increased precision of estimates is not a guarantee that the estimates will be more accurate (particularly if errors in fossil calibrations or

in the clock model are present, in which case the estimates may be biased<sup>21,22</sup>), statistical theory shows that when the prior and model are appropriate, Bayesian estimates of clade ages using genomic data will converge to a limiting distribution centred on their true values<sup>20,22,23</sup>.

The limitations of Bayesian molecular-clock analyses on small datasets have become starkly apparent in studies of mammal diversification. Bayesian estimates using a few genes typically have uncertainties so large that credibility intervals on the ages of ordinal crown-groups straddle the Cretaceous to Palaeogene (K–Pg) boundary<sup>5,6,24,25</sup>, despite a decidedly post-K–Pg fossil record of ordinal crown groups<sup>26–29</sup>. Critically, these Bayesian estimates cannot help to discriminate among competing scenarios of mammal diversification with respect to the K–Pg mass extinction<sup>30–32</sup>. Although Bayesian analyses have been carried out on genome-scale datasets<sup>7,33,34</sup>, only a small number of taxa have been used and, therefore, the increased precision of phylogenomic analyses has not been propagated through to a species-level mammal phylogeny. Thus, despite several decades of research, the precise timeline of mammal evolution remains unresolved<sup>5–9,34</sup>.

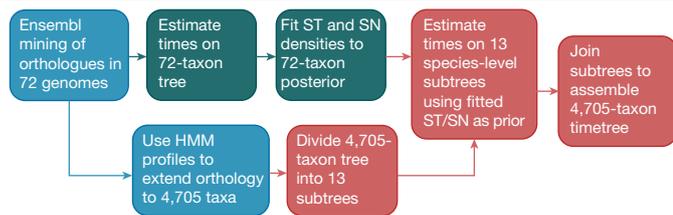
Furthermore, efforts to incorporate species-level alignments into the Bayesian analysis of mammals have been unsatisfactory. For example, in the backbone-and-patch approach, a limited number of genes is used to estimate divergence times on a main tree of few species<sup>6</sup>. Divergence times for key nodes are then used to calibrate the root of densely

<sup>1</sup>School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK. <sup>2</sup>Department of Genetics, Evolution and Environment, University College London, London, UK.

<sup>3</sup>Centre for Advanced Research Computing, University College London, London, UK. <sup>4</sup>EMBL-EBI, Wellcome Genome Campus, Hinxton, UK. <sup>5</sup>School of Earth Sciences, University of Bristol,

Bristol, UK. <sup>6</sup>MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College London, London, UK. <sup>7</sup>Department of Zoology, University of Cambridge, Cambridge,

UK. <sup>8</sup>These authors contributed equally: Sandra Álvarez-Carretero, Asif U. Tamuri. ✉e-mail: phil.donoghue@bristol.ac.uk; m.dosreisbarros@qmul.ac.uk



**Fig. 1 | Summary of the Bayesian sequential subtree dating approach.** The pipeline is divided into molecular data preparation (blue), dating step 1 (green) and dating step 2 (red). The number of taxa ranges from 10 to 72 among genomic loci (50% of loci are present in at least 67 taxa and 90% are present in at least 53 taxa), and from 48 to 3,986 in the 182 gene set. A hidden Markov model<sup>45</sup> (HMM) was used to detect homology and construct the subtree alignments, thus bypassing unreliable homology annotations (Methods). SN, skew-normal distribution; ST, skew- $t$  distribution.

sampled subtrees, resulting in a species-level phylogeny. However, the backbone-and-patch method is not a valid Bayesian approach because the loci used in subtree estimation are the same loci used to estimate the main tree, resulting in duplicate use of the same data and a squaring of the likelihood<sup>35</sup>. In Bayesian clock dating, likelihood squaring leads to convergence to the wrong limiting distribution of node ages<sup>23,36</sup>.

### Sequential Bayesian dating of subtrees

Here we overcome the limitations of previous Bayesian clock-dating studies on small datasets by developing the Bayesian sequential-subtree approach (Fig. 1), which we use to infer a timetree of 4,705 mammal species. First, a genome-scale alignment (15,268 one-to-one orthologues, 33.2 million aligned bases) and a suite of 32 fossil calibrations are used to infer the timetree for 72 species. The resulting posterior distribution of node ages is then used, together with a further set of 60 fossil calibrations, to date 13 subtrees encompassing 4,705 species with new alignments (182 loci, up to  $5.33 \times 10^5$  aligned bases), thus avoiding data duplication in the likelihood (Methods). Our approach is feasible because we use the approximate likelihood calculation<sup>37</sup>, which provides a 1,000 $\times$  speed-up over traditional MCMC timetree inference without loss of accuracy<sup>37,38</sup>. This facilitates analysis of more taxa and much longer alignments than has been possible previously (Table 1). By using the flexible skew- $t$  and skew-normal distributions to model the posterior time estimates from the 72-genome analysis, we accurately transfer information from the genome-scale analysis into the subtree analysis<sup>35</sup>, augmented by the additional subtree-specific fossil calibrations. Our fossil calibrations restrict the minimum ages of clades on the basis of the oldest unequivocal

members of crown groups and, in most cases, also their maximum age through consideration of the presence and absence of stem and sister groups, their palaeoecology, palaeobiogeography and comparative taphonomy<sup>39</sup> (Supplementary Information).

Analyses of small phylogenies<sup>36</sup> and simulated data<sup>23,36</sup> indicate that genome-scale data should lead to asymptotic reduction of uncertainty in divergence time estimates. We demonstrate this for our mammal data by performing random sampling of loci and calculating divergence times on the 72-species phylogeny. By increasing the number of loci analysed from 1 to 15,268, uncertainties in time estimates are progressively reduced irrespective of the relaxed-clock model used (Fig. 2a). Average relative uncertainty on node ages stabilizes at 23.6%–25.0% for the 15,268 loci. This means that, for each 1 million years (Myr) of divergence, approximately 250 thousand years (Kyr) of uncertainty is added to the width of the credibility intervals, which is substantially less than previous Bayesian analyses based on a limited number of genes<sup>5,6</sup> (Table 1). Although reduction in uncertainty is modest beyond 1,000 loci (Fig. 2a), the analysis of the full dataset comes with little extra computational cost because the approximate likelihood calculation depends on the number of taxa, and not on the alignment length<sup>37</sup>.

We next assessed the fit of the relaxed-clock models by using the stepping-stones integrator<sup>40</sup>. This is critical because the competing autocorrelated (geometric Brownian motion<sup>23,41</sup> (GBM)) and independent log-normal<sup>23,42</sup> (ILN) rate models can produce markedly different time estimates when using the same fossil calibrations<sup>35</sup>. Clock-model testing has not previously been conducted at large scale because marginal likelihood inference requires expensive exact likelihood calculation. We overcame this problem by implementing stationary block resampling<sup>43</sup> to obtain reliable estimates of the standard error of the log-marginal likelihood estimates. In this way, we can guarantee the MCMC sample is large enough to obtain an acceptable error for calculating the posterior probabilities of the clock models. We find that, for 71.3% of the 645 loci analysed (Fig. 2b), GBM has a posterior probability above 95%, whereas ILN has a posterior probability above 95% for only 10.7% of loci.

Some topological relationships among major groups of mammals—such as the placement of the placental root, the position of Scandentia with respect to Primates and Glires, and the position of several major groups within Laurasiatheria (Carnivora, Perissodactyla, Chiroptera and Artiodactyla)—have been difficult to resolve<sup>32,33,44</sup>. We selected 7 re-arrangements of these major groups and estimated the divergence times using the 15,268 loci and the 72-species phylogeny. We find these topological re-arrangements have a marginal effect on estimated divergence times (Fig. 2c), apparently because these topological uncertainties are characterized by small internal branches.

**Table 1 | Comparison of molecular-clock dating studies of mammal divergences**

Study	Taxa in molecular alignment <sup>a</sup>	Genes <sup>b</sup>	Alignment length <sup>c</sup>	Estimated age of crown Mammalia <sup>d</sup> (Ma)	Estimated age of Placentalia <sup>d</sup> (Ma)	No. of placental crown orders originating in K (Pg)
Ref. <sup>46</sup>	2,182	66	51,089	166.2 (fixed)	108.7–93.9	9 (7)
<b>Ref. <sup>5</sup></b>	<b>164</b>	<b>26</b>	<b>35,603</b>	<b>238.2–203.3</b>	<b>116.8–92.1</b>	<b>7 (10)</b>
<b>Ref. <sup>7</sup></b>	<b>274 (36)</b>	<b>12 (14,632)</b>	<b>7,370 (20.6 <math>\times 10^6</math>)</b>	<b>191.9–174.1</b>	<b>90.4–87.9</b>	<b>2 (10)</b>
Ref. <sup>8</sup>	46	27	36,860	167.7–164.7	64.85 (fixed)	0 (14)
<b>Ref. <sup>6</sup></b>	<b>4,098</b>	<b>31</b>	<b>39,099</b>	<b>210.9–166.7</b>	<b>105.0–77.4</b>	<b>9 (9)</b>
<b>This study</b>	<b>4,705 (72)</b>	<b>182 (15,268)</b>	<b>&gt; 10<sup>4</sup> (33.2 <math>\times 10^6</math>)</b>	<b>251–165</b>	<b>83.3–77.6</b>	<b>1 (17)</b>

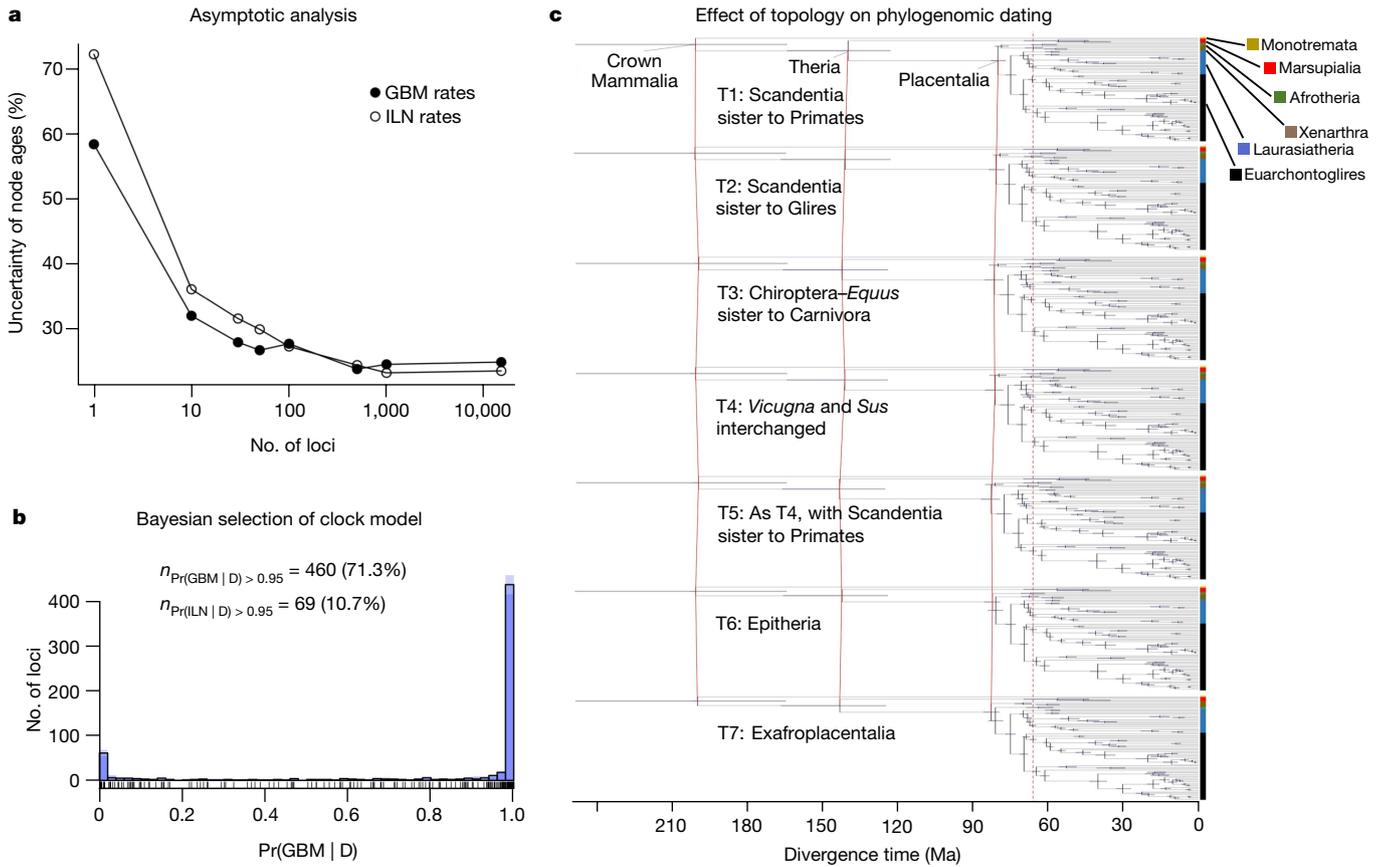
<sup>a</sup>Numbers in parentheses show the number of complete genomes in the alignment.

<sup>b</sup>Numbers in parentheses show the number of genes in the genome-scale part of the alignment.

<sup>c</sup>Numbers in parentheses are the number of nucleotide sites in the genome-scale part of the alignment. In this study, subtree alignment lengths range from  $5.11 \times 10^4$  to  $5.33 \times 10^5$  bases. Missing data range from 46% to 60% in the genomic partitions and from 17% to 99% in the subtree partitions (Methods).

<sup>d</sup>Given as the 95% credibility (for Bayesian studies) or confidence (for non-Bayesian studies) interval.

Studies using Bayesian analysis are shown in bold. K, Cretaceous; Pg, Palaeogene.



**Fig. 2 | Bayesian estimation of mammal divergence times.** **a**, Relative uncertainty in Bayesian estimates of node ages (defined as 95% credibility interval width divided by the posterior mean) is plotted against the number of loci for two relaxed-clock models. Sets of  $n = 1, 10, 30, 50, 100$  and  $1,000$  loci were sampled randomly from the full set of  $15,268$  loci (also included) and grouped into four partitions (except for  $n = 1$ ), with times estimated using approximate likelihood calculation in MCMCtree. **b**, Among the  $15,268$  loci,  $645$  are represented in all  $72$  species. These loci were used to estimate posterior

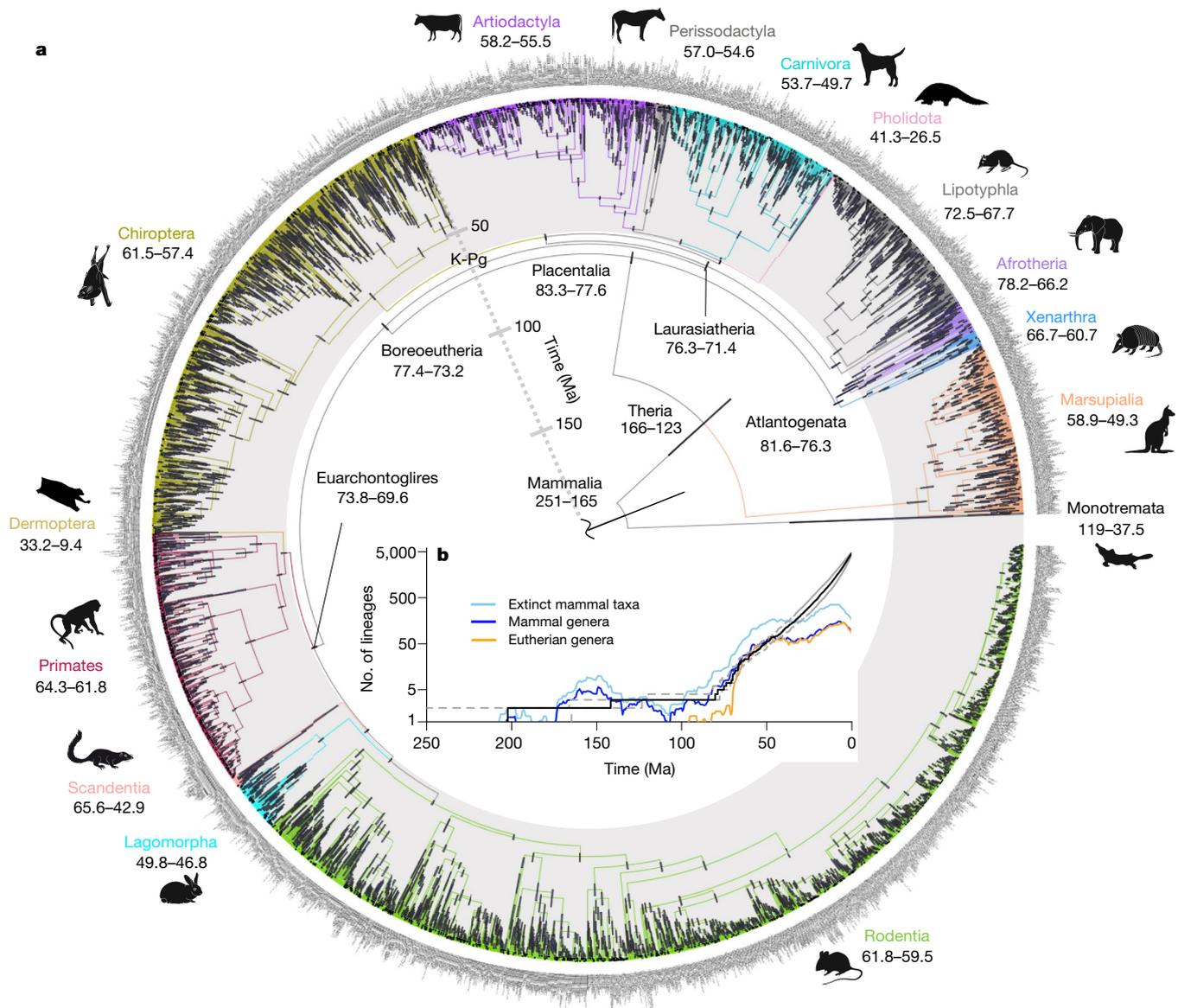
probabilities for the two relaxed-clock models using exact likelihood and the stepping-stones integrator. The histogram of  $\text{Pr}(\text{GBM} | D)$  is shown with a black outline, and the 95% confidence interval of the histogram, obtained using the stationary block bootstrap, are shown in a blue shade. Small vertical black bars are results for individual loci. **c**, Seven dated variations of the mammal topology (T1–T7). T1–T5 have the Atlantogenata rooting of Placentalia<sup>44,49</sup>.  $\text{Pr}(M | D)$ , posterior probability of clock model  $M$  given data  $D$ .  $n_{\text{Pr}(M | D) > 0.95}$ , number of loci for which  $\text{Pr}(M | D) > 0.95$ .

## High-resolution timeline of mammal evolution

Using the results from the 72-genome analysis, we then proceeded with the sequential-subtree approach to date the 4,705-species phylogeny using the GBM rate model. The resulting species-level timetree (Fig. 3a) provides a high-resolution timeline for the diversification of mammals, a timeline with substantially less uncertainty than has been characteristic of previous Bayesian studies, facilitating tests of competing models of ordinal diversification. With this timetree, we have applied the advances set out in ref.<sup>32</sup> to improve the estimated timescale of mammalian evolution, namely: (1) “more contiguous and accurate genome alignments that improve upon detection of orthologous sequences”<sup>45</sup>, (2) “improvements in the calibration of nodes with fossils”<sup>21,35</sup> and (3) “improvements in relaxed-clock methodologies”<sup>35,43</sup>. Furthermore, we have also assessed asymptotics and topological uncertainty to address their impact on the estimated timescale. Generally, we find that, in comparison to previous analyses, clade age estimates are younger and more precise. It appears unlikely, given our asymptotic results (Fig. 2a), that a further increase in precision could be achieved simply by increased sampling of sequence data. Instead, more precise fossil constraints will have a material effect here.

Compared with previous studies, our results show a greater proximity between the origination of ordinal level crown groups and the K–Pg event (Fig. 3a, Extended Data Figs. 1, 2), implying a late Cretaceous

prehistory to ‘modern mammals’ with diversification that continued into the Palaeogene (Fig. 3b). This timescale is incompatible with previous scenarios positing a deep-Cretaceous origin of ordinal level crown placental groups<sup>46</sup>, and is also incompatible with the explosive model, which envisages crown Placentalia originating in the Palaeogene<sup>8</sup>. Among proposed mammal diversification models<sup>32</sup>, our timetree appears compatible with the soft explosive<sup>10</sup> and long fuse<sup>31</sup> models. However, discriminating among these is challenging because, as formulated<sup>32</sup>, these hypotheses lack taxonomic precision. We find crown Placentalia diverged  $83.3$ – $77.6$  million years ago (Ma), whereas the fundamental clades (Boreoeutheria, Laurasiatheria and Euarchontoglires) diverged within the last  $12$  Myr of the Cretaceous. Furthermore,  $17$  out of  $18$  crown placental orders and all crown marsupial orders originated after the K–Pg (Fig. 3a), indicating that the bulk of extant mammal ordinal diversity is a post-K–Pg phenomenon. In terms of inter-ordinal diversity, our results are compatible with both (1) an origin of placental inter-ordinal crown clades following the profound K–Pg extinction in which few lineages survived; and (2) diverse inter-ordinal stem-lineage representatives surviving the K–Pg extinction, but with this diversity later pared back (owing to later extinctions) to the current crown clades that diverged after the K–Pg. Reconciling these competing hypotheses will require integrated co-analysis of living and fossil species<sup>47</sup> to reveal the diversification dynamics that resulted in extant mammal diversity.



**Fig. 3 | Timetree of 4,705 mammal species. a**, Times (in Ma) are estimated with MCMCTree<sup>23</sup> using approximate likelihood<sup>37</sup>. Black bars indicate the 95% highest posterior density credibility interval of node ages, with nodes plotted at the posterior means. Cretaceous–Palaeogene boundary (K–Pg) is indicated.

**b**, Lineages through time plot (black line) with 95% confidence interval (dotted lines), number of extinct mammal species, mammal genera, and eutherian genera through time (mined from Paleodb (<https://paleobiodb.org/>)) are shown.

### Efficient computing in the genomics era

The species-level MCMC sampling required about 80,000 hours of computing time in a high-performance computing cluster and released approximately<sup>48</sup> 16.7 tonnes of CO<sub>2</sub>. Without the technological improvements used here, these analyses would have required hundreds of years of computing time and emitted more than 1.9 thousand tonnes of CO<sub>2</sub> (Supplementary Information). By using existing tools and combining them in a novel way within the sequential-subtree approach, we have demonstrated that hierarchical Bayesian analysis of species-level timetrees integrating genomes is now feasible. Thus, the methodology developed here can be used to address other contentious cases of species diversification that, so far, have been analysed using limited datasets. By integrating our method with the million and more genomes currently planned for sequencing<sup>11</sup>, the prospect of obtaining a reliable evolutionary timescale for the entirety of the tree of life now appears within reach.

### Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-021-04341-1>.

1. Zoonomia Consortium. A comparative genomics multitool for scientific discovery and conservation. *Nature* **587**, 240–245 (2020).
2. Feng, S. et al. Dense sampling of bird diversity increases power of comparative genomics. *Nature* **587**, 252–257 (2020).
3. Harvey, M. G. et al. The evolution of a tropical biodiversity hotspot. *Science* **370**, 1343–1348 (2020).
4. dos Reis, M., Donoghue, P. C. J. & Yang, Z. Bayesian molecular clock dating of species divergences in the genomics era. *Nat. Rev. Genet.* **17**, 71–80 (2016).
5. Meredith, R. W. et al. Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science* **334**, 521–524 (2011).

6. Upham, N. S., Esselstyn, J. A. & Jetz, W. Inferring the mammal tree: species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLoS Biol.* **17**, e3000494 (2019).
7. dos Reis, M. et al. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc. Biol. Sci.* **279**, 3491–3500 (2012).
8. O’Leary, M. A. et al. The placental mammal ancestor and the post-K-Pg radiation of placentals. *Science* **339**, 662–667 (2013).
9. dos Reis, M., Donoghue, P. C. & Yang, Z. Neither phylogenomic nor palaeontological data support a Palaeogene origin of placental mammals. *Biol. Lett.* **10**, 20131003 (2014).
10. Phillips, M. J. Geomolecular dating and the origin of placental mammals. *Syst. Biol.* **65**, 546–557 (2016).
11. Lewin, H. A. et al. Earth BioGenome Project: sequencing life for the future of life. *Proc. Natl Acad. Sci. USA* **115**, 4325–4333 (2018).
12. Siepel, A. Challenges in funding and developing genomic software: roots and remedies. *Genome Biol.* **20**, 147 (2019).
13. Faria, N. R. et al. The early spread and epidemic ignition of HIV-1 in human populations. *Science* **346**, 56–61 (2014).
14. Ramírez-Barahona, S., Sauquet, H. & Magallón, S. The delayed and geographically heterogeneous diversification of flowering plant families. *Nat. Ecol. Evol.* **4**, 1232–1238 (2020).
15. Whelan, N. V. et al. Ctenophore relationships and their placement as the sister group to all other animals. *Nat. Ecol. Evol.* **1**, 1737–1746 (2017).
16. Misof, B. et al. Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763–767 (2014).
17. Jarvis, E. D. et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320–1331 (2014).
18. Tao, Q., Tamura, K. & Kumar, S. in *The Molecular Evolutionary Clock: Theory and Practice* (ed. Ho, S. Y. W.) 197–219 (Springer, 2020).
19. Thorne, J. L. & Kishino, H. in *Statistical Methods in Molecular Evolution* (ed. Nielsen, R.) 235–256 (Springer, 2005).
20. Yang, Z. & Rannala, B. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* **23**, 212–226 (2006).
21. Inoue, J., Donoghue, P. C. J. & Yang, Z. The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. *Syst. Biol.* **59**, 74–89 (2010).
22. dos Reis, M., Zhu, T. & Yang, Z. The impact of the rate prior on Bayesian estimation of divergence times with multiple Loci. *Syst. Biol.* **63**, 555–565 (2014).
23. Rannala, B. & Yang, Z. Inferring speciation times under an episodic molecular clock. *Syst. Biol.* **56**, 453–466 (2007).
24. Springer, M. S., Murphy, W. J., Eizirik, E. & O’Brien, S. J. Placental mammal diversification and the Cretaceous–Tertiary boundary. *Proc. Natl Acad. Sci. USA* **100**, 1056–1061 (2003).
25. Hasegawa, M., Thorne, J. L. & Kishino, H. Time scale of eutherian evolution estimated without assuming a constant rate of molecular evolution. *Genes Genet. Syst.* **78**, 267–283 (2003).
26. Alroy, J. The fossil record of North American mammals: evidence for a Paleocene evolutionary radiation. *Syst. Biol.* **48**, 107–118 (1999).
27. Benton, M. J. Early origins of modern birds and mammals: Molecules vs. morphology. *Bioessays* **21**, 1043–1051 (1999).
28. Hunter, J. P. & Janis, C. M. Spiny Norman in the Garden of Eden? Dispersal and early biogeography of Placentalia. *J. Mamm. Evol.* **13**, 89–123 (2006).
29. Luo, Z. X. Transformation and diversification in early mammal evolution. *Nature* **450**, 1011–1019 (2007).
30. Cooper, A. & Fortey, R. Evolutionary explosions and the phylogenetic fuse. *Trends Ecol. Evol.* **13**, 151–156 (1998).
31. Archibald, J. D. & Deutschman, D. H. Quantitative analysis of the timing of the origin and diversification of extant placental orders. *J. Mamm. Evol.* **8**, 107–124 (2001).
32. Murphy, W. J., Foley, N. M., Bredemeyer, K. R., Gatesy, J. & Springer, M. S. Phylogenomics and the genetic architecture of the placental mammal radiation. *Annu. Rev. Anim. Biosci.* **9**, 29–53 (2021).
33. Tarver, J. E. et al. The interrelationships of placental mammals and the limits of phylogenetic inference. *Genome Biol. Evol.* **8**, 330–344 (2016).
34. Liu, L. et al. Genomic evidence reveals a radiation of placental mammals uninterrupted by the KPg boundary. *Proc. Natl Acad. Sci. USA* **114**, E7282–E7290 (2017).
35. dos Reis, M. et al. Using phylogenomic data to explore the effects of relaxed clocks and calibration strategies on divergence time estimation: primates as a test case. *Syst. Biol.* **67**, 594–615 (2018).
36. dos Reis, M. & Yang, Z. The unbearable uncertainty of Bayesian divergence time estimation. *J. Syst. Evol.* **51**, 30–43 (2013).
37. dos Reis, M. & Yang, Z. Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. *Mol. Biol. Evol.* **28**, 2161–2172 (2011).
38. Battistuzzi, F. U., Billing-Ross, P., Paliwal, A. & Kumar, S. Fast and slow implementations of relaxed-clock methods show similar patterns of accuracy in estimating divergence times. *Mol. Biol. Evol.* **28**, 2439–2442 (2011).
39. Donoghue, P. C. J. & Yang, Z. The evolution of methods for establishing evolutionary timescales. *Philos. Trans. R. Soc. Lond. B* **371**, 20160020 (2016).
40. Xie, W., Lewis, P. O., Fan, Y., Kuo, L. & Chen, M.-H. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Syst. Biol.* **60**, 150–160 (2011).
41. Thorne, J. L., Kishino, H. & Painter, I. S. Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* **15**, 1647–1657 (1998).
42. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, 699–710 (2006).
43. Politis, D. N. & Romano, J. P. The stationary bootstrap. *J. Am. Stat. Assoc.* **89**, 1303–1313 (1994).
44. Nishihara, H., Maruyama, S. & Okada, N. Retroposon analysis and recent geological data suggest near-simultaneous divergence of the three superorders of mammals. *Proc. Natl Acad. Sci. USA* **106**, 5235–5240 (2009).
45. Wheeler, T. J. & Eddy, S. R. nhmmr: DNA homology search with profile HMMs. *Bioinformatics* **29**, 2487–2489 (2013).
46. Bininda-Emonds, O. R. et al. The delayed rise of present-day mammals. *Nature* **446**, 507–512 (2007).
47. Louca, S. & Pennell, M. W. Extant timetrees are consistent with a myriad of diversification histories. *Nature* **580**, 502–505 (2020).
48. Zwart, S. P. The ecological impact of high-performance computing in astrophysics. *Nat. Astron.* **4**, 819–822 (2020).
49. Springer, M. S., Stanhope, M. J., Madsen, O. & de Jong, W. W. Molecules consolidate the placental mammal tree. *Trends Ecol. Evol.* **19**, 430–438 (2004).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2021

## Methods

## Data collection and filtering

**Dataset 1: 72-genome alignment.** We downloaded the set of one-to-one protein-coding orthologues for the 72 mammal genomes available in Ensembl release 98 (<http://www.ensembl.org>, accessed 15 November 2019) using EnsemblBioMarts<sup>50</sup> (<https://m.ensembl.org/biomart/martview>). Sequences that did not meet the following requirements were removed from further analysis: (1) present in both human and mouse, (2) not containing internal stop codons or gene/transcript mismatches, (3) present in at least 10 species, and (4) at least 100 codons in length. This left a total of 15,569 orthologues, which we partitioned into two data blocks: (1) first and second codon positions (12CP) and (2) third codon positions (3CP). For each orthologue, an alignment was built with PRANK v.140603<sup>51</sup> and the best-scoring maximum-likelihood (ML) trees were inferred with RAXML v.8.2.10<sup>52</sup>. We used only the alignments with the 12CP-partition in the subsequent Bayesian molecular-clock analyses.

We further filtered the dataset using the estimated best-scoring ML trees for each gene to identify those having a branch length larger than 60% of the total tree length (the sum of all branch lengths). The relative branch length test is useful to detect misaligned or misidentified orthologues in the alignments<sup>7,53</sup>, which may result in unusually long branch lengths. Let  $b_{ij}$  be the  $i$ -th branch length for gene tree  $j$ , and let  $n$  be the number of branches in the tree, then the relative branch length is

$$r_{ij} = b_{ij} / \sum_{i=1}^n b_{ij}. \quad (1)$$

We identified 133 orthologue alignments associated with at least one relative branch length larger than 60%. These orthologue alignments were removed from further analyses.

Then, we estimated the pairwise distance between each orthologue in *Mus musculus* and *Homo sapiens* using the R (v.3.5 to v.4.0) function `ape::dist.dna`<sup>54</sup> v.5.5. There were 4 genes (ENSG00000132185, ENSG00000204544, ENSG00000120937 and ENSG00000236699) for which the distances were returned as NaN for at least one of the substitution models used (TN93 and JC69) or larger than 0.75 for the raw distance. Furthermore, when we plotted the percentage of the tree length inferred for each orthologue alignment versus the corresponding largest branch length (also as percentage), we found an outlier (ENSG00000176973; Supplementary Fig. 1). We removed these 5 orthologue alignments, resulting in 15,431 orthologous gene alignments. Of those, 163 orthologues were removed as they are used in the construction of dataset 2 (see below). This filtering step resulted in 15,268 orthologue alignments (Supplementary Table 1).

The 15,268 orthologue alignments were sorted from fast- to slow-evolving according to the pairwise-distance estimates and grouped into four partitions with the same number of genes. Each of the four partitions contained the concatenated 12CP of the orthologues for the partition (Supplementary Table 2). The rationale for this partitioning strategy is as follows. In previous work<sup>7,55</sup>, we tested phylogenomic data partitioning according to locus rate, principal component analysis of relative branch lengths and amino acid composition at loci. However, those analyses showed no noticeable differences in posterior time estimates across the partitioning strategies<sup>7,55</sup>. Conversely, we have shown that uncertainty in time estimates is sensitive to the number of partitions used<sup>7,22</sup>, with more partitions producing more precise estimates, but at the cost of additional computation time. It appears that four partitions give a reasonable trade-off between computational speed and precision of estimates. For example, 20 partitions would produce slightly more precise estimates but at 5 times the computational cost.

**Dataset 2: alignments of 4,705 taxa.** We downloaded 832 complete mammal mitochondrial genomes from NCBI RefSeq (accessed

14 January 2016). Twelve extinct and 2 redundant entries were removed, leaving 818 genomes. Twelve protein-coding genes (all but *ND6*) and the two non-coding RNA genes were extracted from each genome. The overlapping region in *ATP8* (position 95 to end) and overlapping codons at the end of *ND4L* were deleted.

To increase the nuclear and mitochondrial datasets, we mined sequences deposited in the European Nucleotide Archive (ENA; <https://www.ebi.ac.uk/ena>). The GenBank taxonomy (<ftp://ftp.ncbi.nih.gov/pub/taxonomy/>) was used to search for non-Ensembl mammalia species (this taxonomy is only used for ENA searches and not for any other analyses). A total of 7,188 taxa were found, 83 of which were extinct. The GenBank identifiers were used to reference the corresponding taxa in the ENA, from which all matching coding and non-coding sequences for non-Ensembl mammal taxa were downloaded (accessed 17 January 2016): we found 6,453 taxa with coding sequences and 3,239 taxa with non-coding sequences (6,606 distinct taxa).

This project started in early 2016. At the time, we downloaded 15,904 nuclear orthologous gene alignments for 43 mammal taxa from Ensembl 83 and used these orthologues to create HMM sequence profiles with HMMER<sup>56</sup>. The HMM profiles were then used to identify orthologues for additional taxa from GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>), bypassing unreliable GenBank homology annotations, and thus allowing reliable construction of large mammal subtrees. In late 2019, we updated the 15,904 orthologues to 72 genomes using Ensembl 98, but the HMM profiles and corresponding homology searches were based on the 2016 mining of Ensembl. HMM profiles were also created for mitochondrial protein-coding and non-coding genes and used for taxa extension of the corresponding alignments. DNA homology searches were performed with nhmmer<sup>45</sup> using the following match criteria: (1) sequences with  $E$ -value  $< 10^{-100}$  for a single gene were collected (that is, sequences with multiple low  $E$ -values for different genes were removed); (2) matched sequences had to be at least 70% as long as the shortest Ensembl sequence in the alignment because many deposited sequences are partial sequences; (3) matches from hybrid and cross species were removed; (4) unspecified species were excluded, unless no other member of the genus was represented (4 taxa included); (5) unconfirmed species were excluded, unless no other member of the genus was represented (1 taxon included); and (6) coding sequences were checked for correct open reading frame and translation.

Nuclear genes resulting in an expanded set of at least 50 taxa were selected, resulting in a set of 168 nuclear genes. These 168 genes correspond to 163 genes in the 2019 Ensembl mining (5 genes did not pass filtering criteria for the 72 taxa, but they did pass the criteria with the 43 taxa). Thus dataset 1, based on the 2019 Ensembl mining, was reduced from 15,431 genes to 15,268 (Supplementary Table 1) to avoid data duplication in the sequential dating approach. For new mined taxa, sequence annotations were extracted, sorted, and visually inspected to help verify homology. Alignments were then extended with homology-matched sequences using PAGAN v.0.61<sup>57</sup>. Sequences were added in order of decreasing length (that is, longest sequences were added to the alignment first). Supplementary Table 3 gives summaries of the numbers of taxa and alignment lengths for datasets 1 and 2.

We then used RAXML to estimate the topology for each one of the 182 loci (168 nuclear and 14 mitochondrial) under the GTR+G model. We then manually inspected the trees and further filtered taxa following these criteria: (1) remove taxa that did not share genes with their order, family, and genus. This is done to avoid unidentifiable positioning of taxa in the subtrees: if a species does not share genes with its close relatives, then several positionings of the species within the subtree will have the same likelihood (also known as 'likelihood terraces'). (2) Keep only one member of each species while maintaining maximum locus coverage, that is, remove redundant subspecies. Many subspecies slow the analysis down and are not informative about deep divergences (for example, *Rangifer tarandus tarandus*). Also, subspecies annotations

are missing for many loci, leading to integrity problems when resolving tips. (3) Outdated taxonomic names according to the literature were removed. (4) Remove taxonomically mismatched or mislabelled taxa. (4) Flag taxa with large topological placement discrepancy with the literature. (5) Outliers with unusually long branches in estimated trees were removed (three sequences in two genes).

Taxa were subdivided according to the following taxonomic groups: Afrotheria, Xenarthra, Marsupialia, Euarchonta, Lagomorpha, Laurasiatheria and Rodentia. Laurasiatheria, Rodentia and Chiroptera, which are species-rich, were further divided into additional subsets to speed up the dating analysis. Monotremata was added as an outgroup to all subtrees. The final dataset has 4,705 taxa and 182 loci divided into 13 subtree alignments. Each alignment was divided into five-partitions: (1) mitochondrial 12CP, (2) mitochondrial 3CP, (3) mitochondrial RNA, (4) nuclear 12CP and (5) nuclear 3CP (Supplementary Tables 4–7). A RAxML analysis (GTR+G) was then run on each subtree with the genome-scale tree used as a backbone constraint. The final phylogeny has two manual adjustments. In the original tree, tenrecs are a sister clade to the rest of Afrotheria but, following recent work<sup>6,33</sup>, we adjusted tenrecs as a sister clade to chrysochlorids. In the original tree, *Dromiciops* is placed as sister to *Notoryctes* but, following recent work<sup>58</sup>, we placed *Dromiciops* basal to the rest of Australidelphia.

### Divergence time estimation

All divergence-time estimation analyses were carried out with the MCMCtree v.4.9h/i dating software from the PAML package<sup>59</sup>. The following analyses were carried out: (1) asymptotic analysis of uncertainties in time estimates as a function of number of loci, (2) Bayesian selection of relaxed-clock model, and (3) analysis of time estimates for seven topological re-arrangements of the mammal tree. Analyses 1–3 were carried out on the 72-taxon dataset. Last, we ran the (4) sequential-subtree analysis, which is divided into two parts: (i) estimation of times in 72-taxon tree, and (ii) estimation of times for the set of subtrees (4,705 taxa) using the time posterior of step 1 as the time prior.

**Asymptotic analysis of uncertainty in time estimates.** To evaluate how our phylogenomic-scale data lead to asymptotic reduction of uncertainty in divergence time estimates, we randomly sampled data subsets with  $n = 1, 10, 30, 50, 100$  and  $1,000$  loci from dataset 1, and grouped them into four partitions with roughly the same number of genes (except for  $n = 1$ ). We then estimated the divergence times using the approximate likelihood calculation in MCMCtree, under both the ILN and the autocorrelated (GBM) rate models for each data subset (see ‘Time estimation on topological rearrangements for 72-taxa’ below for details on approximate likelihood method). Then, for each subset, we calculated the ratio of the 95% credibility interval width over the posterior mean of the node age. The uncertainty ratios across all 71 node ages were then averaged. This provides us with a measure of the average uncertainty in posterior node ages. For example, if the ratio is 20%, it means that, on average, the credibility interval width is equivalent to 20% of the node age, or alternatively, that 20 Myr of uncertainty are added to the credibility interval width for every 100 Myr of divergence.

**Bayesian rate model selection.** We assessed adequacy of the ILN against the GBM rate models by using the stepping-stones approach<sup>40</sup>. Because an MCMC sample is a stationary time series, the stationary bootstrap<sup>43</sup> can be used to estimate the standard error of the log-marginal likelihood estimate while accommodating the autocorrelation of the MCMC. Let  $\log \mathbf{L}_i$  be the vector of log-likelihood values sampled from the  $i$ -th power posterior. We sample, with replacement, blocks of observations from  $\log \mathbf{L}_i$ , and the random blocks are stitched together to form a bootstrap sample  $\log \mathbf{L}_i^*$ . The size of the blocks has a geometric distribution with mean equal to 10% of the length of  $\log \mathbf{L}_i$ . The procedure is repeated for each power-posterior sample and the

log-marginal likelihood is then calculated using the stepping-stones formula<sup>40</sup>. The procedure is repeated 100 times to obtain 100 bootstrap estimates of the marginal-log likelihood, which are then used to estimate the standard error of the estimate. We validated the algorithm by comparing bootstrap standard error estimates against those obtained from brute-force re-calculation of the marginal likelihood, the latter being very computationally expensive as it involves running many independent stepping-stones analyses.

We used the 645 genes that were present in all the 72 taxa. Analyses were carried out using exact likelihood because the approximation is not good in the tails of the likelihood function, and tail values have a large impact on the marginal likelihood estimates. The age of the root was set to have a prior mean of 1 using the gamma density  $\Gamma(100,100)$ . We used diffuse gamma priors on the mean rate,  $\mu \sim \Gamma(2,40)$  and  $\sigma_2 \sim \Gamma(1,10)$ . The birth and death prior was set to  $\lambda = \mu = 1$  and  $\rho = 0.1$ , which generates an approximately uniform density<sup>20</sup>. Analyses were carried out using the main tree topology (Supplementary Fig. 2b, T2). Each gene was analysed separately under the HKY85+G5 nucleotide substitution model<sup>60–62</sup>, and sampling was done over 32 beta points in the power posterior. Choice of beta points, application of the stepping-stones formula and bootstrap block-sampling were done with the mcmc3r v.0.4.3 package<sup>35</sup>. In total, rate model selection required 41,280 MCMC chains (645 genes  $\times$  2 rate models  $\times$  32 beta points), totalling four months of wall time (equivalent to over 2 million hours or 200 years of CPU time) in a high-performance computer cluster.

We also carried out a maximum-likelihood ratio test of the strict molecular clock<sup>63</sup>. The strict clock was rejected in 642 out of 645 loci (after false-discovery rate correction at the 5% level).

**Time estimation on topological rearrangements for 72 taxa.** The seven tree topologies used are shown in Supplementary Fig. 2a–g. Time estimation was carried out using dataset 1 under approximate likelihood calculation. All analysis setups were as in step 1 of the sequential Bayesian approach.

**Sequential Bayesian approach. Hessian calculation to approximate the likelihood on 72 taxa.** We use the approximate likelihood method to speed up computation during MCMC sampling<sup>37</sup>. This involves obtaining the maximum-likelihood estimates (MLEs) of the branch lengths,  $\mathbf{b}$ , on a partition, together with the gradient  $\mathbf{g}$ , and Hessian  $\mathbf{H}$ , of the log-likelihood evaluated at the MLEs. Then  $\mathbf{b}$ ,  $\mathbf{g}$  and  $\mathbf{H}$  are used to approximate the likelihood during MCMC sampling (see ref. <sup>64</sup> for a tutorial). We used BASEML<sup>59</sup> v.4.9h/i to calculate  $\mathbf{b}$ ,  $\mathbf{g}$  and  $\mathbf{H}$  for each of the 4 partitions in the 72-genome alignment using the HKY+G5 substitution model<sup>60,61</sup>. We tested seven different topological relationships among mammals (Supplementary Fig. 2), with each tree topology requiring calculation of its own set of  $\mathbf{b}$ ,  $\mathbf{g}$  and  $\mathbf{H}$ .

**Divergence-time estimation on the 72-taxon tree.** Nodes are calibrated using uniform distributions based on the fossil record. The distributions have soft bounds, that is, there are probabilities,  $p_L$  and  $p_U$ , that the node age falls outside lower and upper calibration bounds. Here we tested two approaches for setting these probabilities: (1) using  $p_L = p_U = 0.025$ , and (2)  $p_L = 0.001$  and  $p_U = 0.1$ . The second approach assumes the probability of violation of the minimum bound is very small (that is, assuming fossil placement and dating are accurate), while allowing for a larger upper-bound probability. We find that choice of  $p_L$  and  $p_U$  have a small impact on time estimates on the 72-genome phylogeny (Supplementary Fig. 3). However, for the second analysis step (subtree time estimation), we find that, when  $p_L = 0.001$  and  $p_U = 0.1$ , the fitted skew- $t$  (ST) calibration densities are too asymmetrical and with heavy tails, leading to convergence problems in the MCMC (for example, see ref. <sup>65</sup> for a discussion of convergence on distribution tails). Thus, we favour the use of ST calibrations based on the posterior using  $p_L = p_U = 0.025$  for the rest of the analyses (see below for details on ST distribution fitting).

Supplementary Table 8 lists the fossil calibrations used for the 72-species phylogeny. Our fossil calibrations include previously published constraints on clade age<sup>33,66</sup>, plus new calibrations that we have formulated following established best practice<sup>67</sup>. In brief, minima are based on the oldest unequivocal member of a crown group and we follow the youngest age interpretations. Maxima are more challenging to establish since absence of fossil evidence cannot simply be interpreted as evidence of the absence of a lineage at a given time interval. Hence, we use evidence of the presence of outgroup lineages with comparable ecology and taphonomy to serve as evidence of the absence of ingroup lineages as, were they present, the preservation of outgroup relatives demonstrates that they should be preserved<sup>39</sup>. However, there remains a non-zero probability that lineages existed before our maxima and hence we implement them as soft constraints, which allows the analysis to explore older ages for the origination of a calibrated clade but at low prior probability. Detailed justifications for all calibrations are provided in the annex of the Supplementary Information. Our fossil calibrations were set at the beginning of the project in 2016. Supplementary Table 8 provides the latest geochronological updates (September 2021). The old calibrations are used for the asymptotic and topology variation analyses, while the updated calibrations are used in the sequential dating approach. The calibration updates are very small, usually below the sampling error of the MCMC, and thus have little effect on time estimates (Supplementary Fig. 4).

The birth and death process<sup>23</sup>, used to specify the time prior for nodes with no calibrations, was set to  $\lambda = \mu = 1$ , and sampling fraction to  $\rho = 0.1$ , which gives an approximately uniform kernel. The GBM rate model is used with a gamma-Dirichlet prior<sup>22</sup> on the mean  $i$ -th partition rate,  $\mu_i \sim \Gamma(2, 40)$ , and on the relaxed-clock parameter,  $\sigma_i^2 \sim \Gamma(1, 10)$ . This setting gives a diffuse prior on the rate that is roughly centered on the average substitution rate of nuclear genes in mammals<sup>22,64</sup>.

We ran MCMCtree without data to sample from the time prior. This is done to verify the prior is sensible and not in conflict with the calibration densities used<sup>21,68</sup>. To ensure MCMC convergence and increase effective sample size (ESS), we ran several MCMC chains of sufficient length. We used Tracer<sup>69</sup> v.1.7 and the R function `effectiveSize`<sup>70</sup> v.0.19.4 to make sure the ESS was larger than 200 for all estimated parameters (Supplementary Table 9). We also used the R function `rstan::monitor`<sup>71</sup> v.2.21.2 to calculate the ESS for bulk and tail quantiles and the potential scale reduction factor on rank normalized split chains (Rhat). Values over 100 for the former are considered good, while Rhat values need to be either smaller than or equal to 1.05 to show chain convergence. We further explored chain convergence by visually plotting the distributions of the different chains ran in MCMCtree with the R package `MCMCtreeR`<sup>72</sup> v.1.1 (Supplementary Fig. 5). Supplementary Figure 6 shows the convergence plots for each tree hypothesis, which show excellent convergence.

Fitting of skew- $t$  distributions to posterior times. We used the posterior time estimates sampled during the MCMC runs under the GBM model to fit ST distributions to the 71 internal nodes of the 72-species tree. This was done with the R function `sn::st.mple`<sup>73</sup> v.2.0.0 under the BFGS method for parameter optimization. To check whether the fitted ST distributions were sensible, we sampled, using MCMCtree, from the new ST-based prior (that is, without the alignment data), and checked whether the sampled prior distributions matched the original ST distributions. This is necessary because of the constraint that nodes are younger than their parents. This means ST calibrations on adjacent nodes could suffer from truncation effects and the resulting prior could be in conflict with the ST densities. We did not observe any such conflict (Supplementary Fig. 7). For the crown-lagomorpha node, however, the corresponding ST calibration caused convergence problems when dating the lagomorpha subtree (dataset 2). This ST calibration, which has a heavy tail, was replaced by an essentially

equivalent skew-normal calibration, which has a light tail, thus solving the convergence problem.

**Time estimation on the 4,705-taxon phylogeny.** Hessian and gradient calculation for each partition on each subtree were done using the HKY+G5 substitution model, as for dataset 1. Subtrees were calibrated using the fitted ST densities and 60 additional soft-bound calibrations (Supplementary Tables 10, 11). These calibrations are also updated according to new geochronology. The same rate and birth–death model priors as in step 1 were used. For each subtree, we ran 32 independent MCMC chains to check convergence and ensure enough samples were collected to approximate the posterior (Supplementary Fig. 8, Supplementary Tables 12, 13), although some of those did not pass quality filters and were not included (Supplementary Fig. 8).

We ran MCMCtree without data to sample from, and verify the integrity of, the prior. We repeated this analysis twice: once including only the ST calibrations and a second time with both ST and soft-bound calibrations in the subtrees. This was necessary to assess whether the soft bounds were in conflict with the ST densities, producing truncation problems. In a few cases, after examining the prior, we observed conflict between the ST densities and the soft bounds. In such cases, calibrations were adjusted so that the resulting prior credibility interval limits were within about 5% of the original ST density quantiles. Adjustments included either nudging the maximum age of a soft-bound calibration or nudging the ST calibration densities themselves.

**Assembly of the 4,705-taxon timetree.** Time calibrated subtrees were attached to the corresponding node in the 72-species mammal phylogeny using a custom Python (v.3.8.5) script. Outgroup (Monotremata) and any marker taxa were removed before merging the subtrees into the main tree. Marker taxa were needed in the Rodentia and Laurasiatheria subtrees to guarantee integrity of ST-calibrated nodes. That is, when splitting these large subtrees for divergence times estimation, some nodes shared with the main 72-species tree would disappear in some subtrees. Consequently, marker taxa (shared with a sister subtree) were added back into the corresponding subtree to retain the calibrated node and guarantee integrity during merging (Supplementary Data). The result of the subtree merging is the fully-dated 4,705-taxon phylogeny. We verified integrity of time estimates by repeating analyses on subsets of shared data cross partitions (Supplementary Information, Supplementary Fig. 9).

### Mining of Paleodb for fossil mammal taxa

The fossil data were downloaded from the Paleobiology Database (<https://paleobiodb.org/>, accessed March 2021), using the API service with resolution set to genus level, excluding uncertain genera, only body fossil taxa (that is, no ichnotaxa), and accepted names only. The data were cleaned by removing individual stratigraphic occurrences that had an age range greater than 20 million years, as this suggests the dating of the occurrence is uncertain or incorrect on the database. The higher clade classifications were added to the genera, and some manual corrections to the fossil classifications were made. The maximum and minimum ages for each genus were extracted, as well as whether the genus is extant or extinct. The genera with age ranges approaching or equal to 0 million years were double checked against the literature to ensure the extinct–extant status was correct. The number of extinct species for each genus was also extracted from the Paleobiology Database and added to the dataset.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

### Data availability

All data required to reproduce the analyses are available at <https://doi.org/10.6084/m9.figshare.14885691>.

## Code availability

A repository containing instructions to reproduce the analyses is available at [http://github.com/sabifo4/mammals\\_dating](http://github.com/sabifo4/mammals_dating) and <https://doi.org/10.5281/zenodo.5736629>. The MCMCtree software and mcmt3r R package are freely available from <http://abacus.gene.ucl.ac.uk/software/paml.html> and <https://github.com/dosreislab>, respectively.

50. Kinsella, R. J. et al. Ensembl BioMart: a hub for data retrieval across taxonomic space. *Database* **2011**, bar030 (2011).
51. Löytynoja, A. Phylogeny-aware alignment with PRANK. *Methods Mol. Biol.* **1079**, 155–170 (2014).
52. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
53. Springer, M. S. & Gatesy, J. On the importance of homology in the age of phylogenomics. *Syst. Biodivers.* **16**, 210–228 (2018).
54. Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
55. dos Reis, M. et al. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. *Curr. Biol.* **25**, 2939–2950 (2015).
56. Eddy, S. R. Profile hidden Markov models. *Bioinformatics* **14**, 755–763 (1998).
57. Löytynoja, A., Vilella, A. J. & Goldman, N. Accurate extension of multiple sequence alignments using a phylogeny-aware graph algorithm. *Bioinformatics* **28**, 1684–1691 (2012).
58. Mitchell, K. J. et al. Molecular phylogeny, biogeography, and habitat preference evolution of marsupials. *Mol. Biol. Evol.* **31**, 2322–2330 (2014).
59. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
60. Hasegawa, M., Kishino, H. & Yano, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160–174 (1985).
61. Hasegawa, M., Yano, T.-A. & Kishino, H. A new molecular clock of mitochondrial DNA and the evolution of hominoids. *Proc. Japan Acad. B* **60**, 95–98 (1984).
62. Yang, Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* **39**, 306–314 (1994).
63. Felsenstein, J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**, 368–376 (1981).
64. dos Reis, M. & Yang, Z. in *Evolutionary Genomics: Statistical and Computational Methods* (ed. Anisimova, M.) 309–330 (Springer, 2019).
65. Rannala, B., Zhu, T. & Yang, Z. Tail paradox, partial identifiability, and influential priors in Bayesian branch length inference. *Mol Biol Evol.* **29**, 325–335 (2012).
66. Benton, M. J. et al. Constraints on the timescale of animal evolutionary history. *Palaeontol. Electron.* **18**, 1–106 (2015).
67. Parham, J. F. et al. Best practices for justifying fossil calibrations. *Syst. Biol.* **61**, 346–359 (2012).
68. Warnock, R. C. M., Parham, J. F., Joyce, W. G., Lyson, T. R. & Donoghue, P. C. J. Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors. *Proc. R. Soc. B* **282**, 20141013 (2015).
69. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
70. Plummer, M., Best, N., Cowles, K. & Vines, K. CODA: convergence diagnosis and output analysis for MCMC. *R News* **6**, 7–11 (2006).
71. RStan: the R interface to Stan. R package. <https://mc-stan.org> (Stan Development Team, 2020).
72. Puttick, M. N. MCMCTreeR: functions to prepare MCMCtree analyses and visualize posterior ages on trees. *Bioinformatics* **35**, 5321–5322 (2019).
73. Azzalini, A. *The R package 'sn': the skew-normal and related distributions such as the skew-t*. <http://azzalini.stat.unipd.it/SN/> (2019).

**Acknowledgements** We thank J. Gilbert and C. G. Faulkes for help with the Rodentia subtrees. This work used computing resources from Queen Mary's Apocrita HPC and University College London Myriad HPC facilities. This work was supported by Biotechnology and Biological Sciences Research Council, UK, awards BB/T01282X/1, BB/T012951/1 and BB/T012773/1.

**Author contributions** M.d.R. conceived the work. M.d.R., Z.Y., P.C.J.D., S.Á.-C. and A.U.T. designed the analysis. S.Á.-C., A.U.T., R.J.A., P.C.J.D., M.B., E.C. and F.F.N. compiled, processed and verified the molecular and fossil data. S.Á.-C., A.U.T. and M.d.R. analysed the data. M.d.R. and P.C.J.D. wrote the paper with input from all authors.

**Competing interests** The authors declare no competing interests.

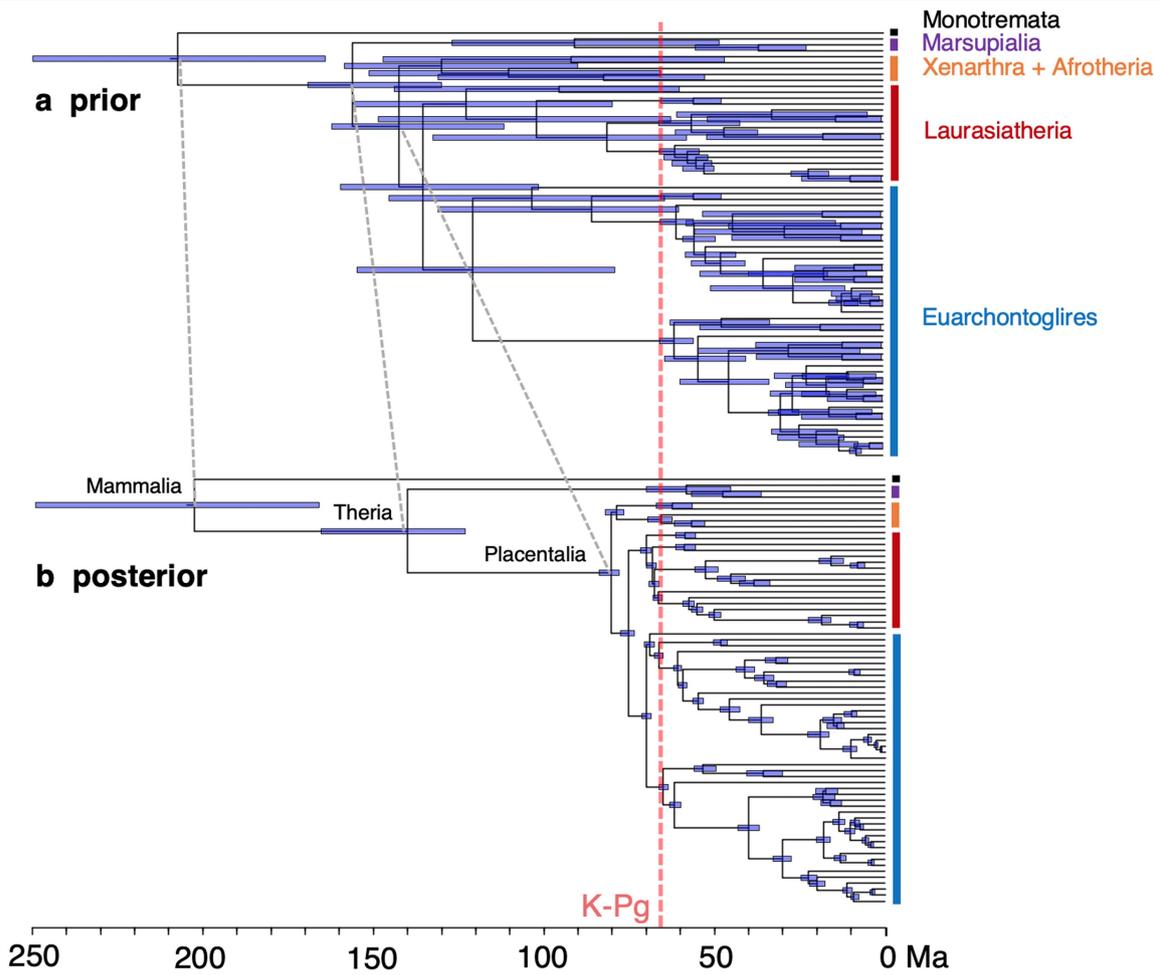
### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41586-021-04341-1>.

**Correspondence and requests for materials** should be addressed to Philip C. J. Donoghue or Mario dos Reis.

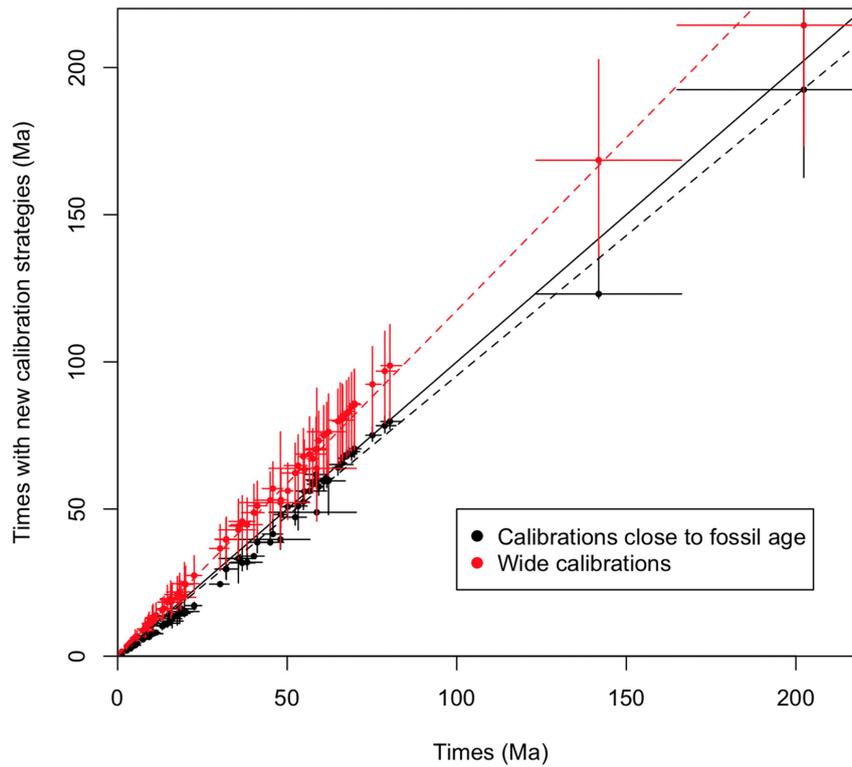
**Peer review information** Nature thanks Olaf Bininda-Emonds, Xing-Xing Shen and the other, anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



**Extended Data Fig.1 | Comparison of prior and posterior times. a,** Prior distribution of node ages generated by MCMC sampling without the molecular alignment. **b,** Posterior distribution of node ages when the 72-genome

alignment is included during MCMC sampling. In both **a** and **b**, nodes are plotted at their posterior mean ages. The blue horizontal bars indicate the 95% credibility intervals of node ages.



**Extended Data Fig. 2 | Impact of fossil calibration strategies on node age estimates.** The posterior of node ages for the 72-taxon phylogeny is estimated using two additional fossil calibration strategies (y-axis) and plotted against the main estimates using best practice in calibration choice<sup>39</sup> (x-axis). In all cases the fossil minima are the same, but the calibration maxima changes. In the first strategy (black dots), calibration densities are narrow and close to the fossil ages. A truncated-Cauchy with a short tail (using  $p = 0$  and  $c = 0.001$ , which extends the tail to about 110% of the fossil age) is used<sup>21</sup>. This strategy assumes the fossil record is a good indicator of the true node ages. In the second

strategy (red dots), a truncated-Cauchy with a heavy tail (using  $p = 0.1$  and  $c = 1$ , which extends the tail to over 900% of the fossil age) is used<sup>21</sup>. This strategy ignores the presence and absence of stem and sister groups, their palaeoecology, palaeobiogeography, and comparative taphonomy<sup>39</sup>; and instead, assumes the node ages can be arbitrarily old. Dots are plotted at the posterior mean ages and vertical and horizontal bars indicate 95% CIs. The solid line is the  $x = y$  line. The dashed lines are the regression lines for the corresponding data points.

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

**Data collection** Molecular sequence from relevant databases was collected using Ensembl BioMart (<https://m.ensembl.org/biomart/martview>) (Note: this is an online tool and could not find version listed in website) and HMMER v3.1b2.

**Data analysis** The following open source software were used: PAML v4.9h-i (MCMCtree and BASEML are part of PAML, both used for divergence time estimation) and mcmc3r v0.4.3 (clock model selection), Prank v140603 (molecular sequence alignment), RaxML v8.2.10 (tree reconstruction), Tracer v1.7 (MCMC diagnostics), FigTree v1.3 (tree visualization), Python v3.8.5 and R v3.5 to 4.0 (scripting and data analysis). The following R packages were used: Ape v5.5, Rstan v2.21.2, MCMCtreeR v1.1, and SN v2.0.0. Instructions and code to reproduce the study are available from [http://github.com/sabifo4/mammals\\_dating](http://github.com/sabifo4/mammals_dating) (DOI: 10.5281/zenodo.5736629).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Molecular sequence data were downloaded from Ensembl ([www.ensembl.org](http://www.ensembl.org)), GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) and ENA (<https://>

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We analysed: Data set 1: 15,268 orthologous genes from 72 mammal species (this was the maximum available at the time from Ensembl). Data set 2: 182 orthologous genes from 4,705 mammal species (this was the maximum we could reliably identify from GenBank and ENA after filtering and data curation). From data set 1, 645 genes were used for clock model selection analyses (this was chosen as these genes were present in all 72 mammal species. This is the maximum sample size that was feasible as this analysis is computationally expensive).
Data exclusions	Molecular sequence data were analysed for miss-identified orthologs in Ensembl (data set 1) and GenBank/ENA (data set 2). 636 genes did not pass the filtering criteria in data set 1 and were removed. We did not keep track of the number of genes removed from data set 2. The criteria used for gene filtering is based on previous work (dos Reis et al. 2012, Proc. Roy. Soc. B, 279: 3491–3500) and new criteria devised by us to minimize gene miss-identification. The long and detailed list of criteria used is provided in the Methods section of the manuscript.
Replication	Bayesian MCMC chains were replicated 32 times. This number was chosen as it gives an average Effective Sample Size (ESS) of ~1,000 for all node ages in all subtree analysis. An ESS of 1,000 is the recommended minimum in the stochastic simulation literature. Note this replication is carried out simply to ensure convergence to the posterior distribution in the MCMC sample, and MCMC samples converged successfully.
Randomization	N/A. No experimental set-ups with controls and cases were used, thus randomization is not necessary.
Blinding	N/A. No experimental set-ups with controls and cases were used, thus randomization is not necessary.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging