# A Species-Level Timeline of Mammal Evolution Integrating Phylogenomic Data

Sandra Álvarez-Carretero[1,2], Asif U. Tamuri[2,3], Matteo Battini[4], Fabrícia F. Nascimento[5], Emily Carlisle[4], Robert J. Asher[6], Ziheng Yang[2], Philip C. J. Donoghue[4] & Mario dos Reis[1].

1. Queen Mary University of London, 2. University College London, 3. EMBL-EBI, 4. University of Bristol, 5. Imperial College London, 6. University of Cambridge.
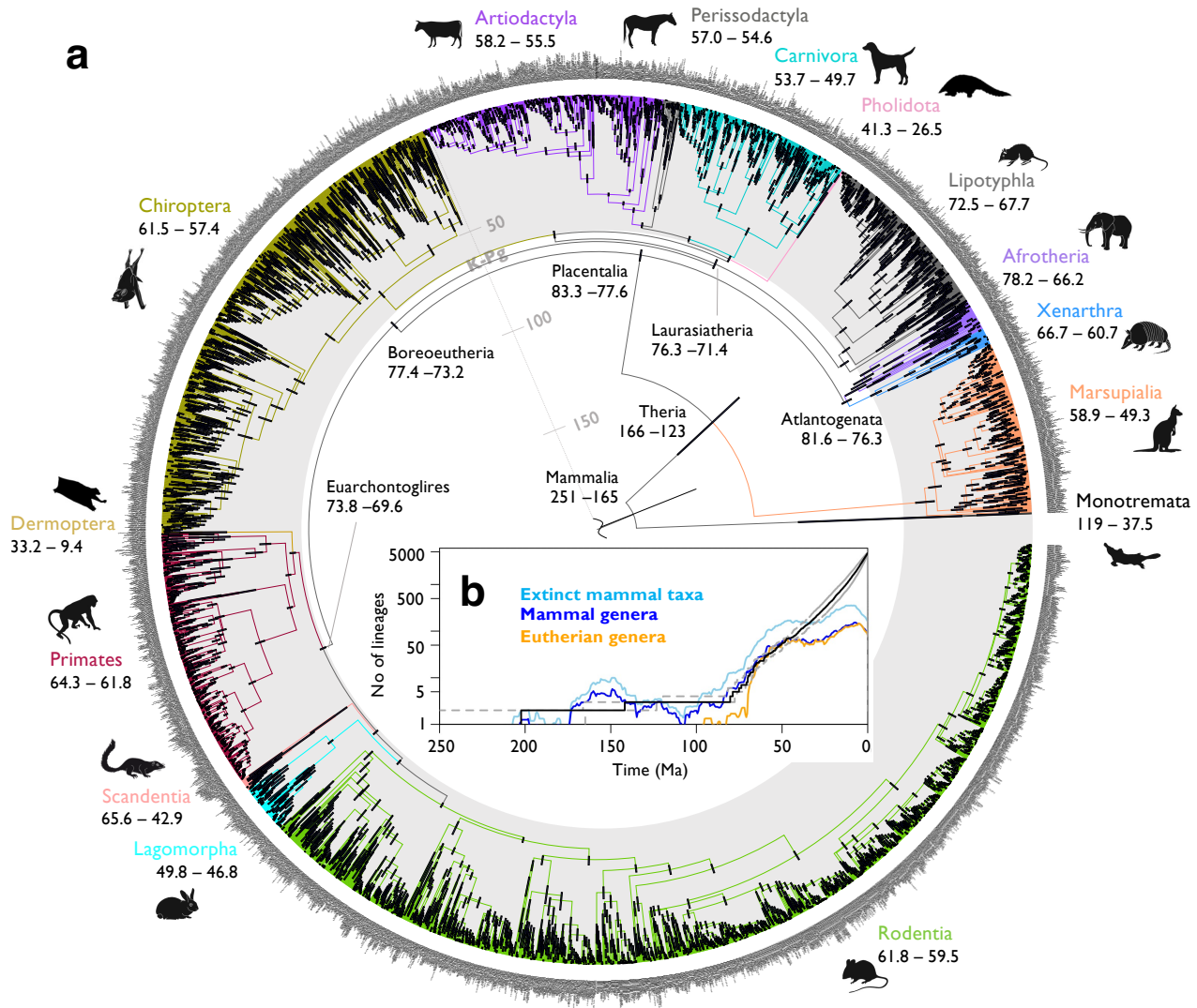
Timetree of 4,705 mammal species. a, *Times (in Ma) are estimated with MCMCtree using approximate likelihood. Black bars indicate the 95% HPD CI of node ages, with nodes plotted at the posterior means.* **b**, *Lineages through time plot (black line) with 95% CI (dotted lines), number of extinct mammal species, mammal genera, and eutherian genera through time (mined from PaleoDB) are shown as smoothed histograms.*

Summary. High-throughput sequencing projects generate genome-scale sequence data for species-level phylogenies. However, state-of-the-art Bayesian methods for inferring timetrees are computationally limited to small datasets and cannot exploit the deluge of new genomes. In the case of mammals, molecular-clock analyses of limited datasets have produced conflicting estimates of clade ages with large uncertainties and, thus, the timescale of placental mammal evolution remains contentious. 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 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 Paleogene and show crown Placentalia originated in the Late Cretaceous with unambiguous ordinal diversification in the Paleocene/Eocene. Our new 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. As such, our approach can be used to tackle other contentious cases of animal and plant diversifications that require analysis of species-level phylogenomic datasets.
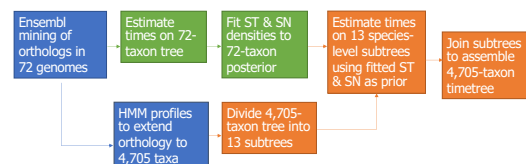
Summary of the Bayesian sequential subtree dating approach. *The pipeline is divided into molecular data preparation (blue), dating step 1 (green) and dating step 2 (orange). Number of taxa ranges from 10 to 72 among genomic loci (50% of loci have at least 67 taxa and 90% at least 53 taxa), and from 48 to 3,986 in the 182 gene set. A hidden-Markov model (HMM) was used to detect homology and construct the subtree alignments, thus bypassing unreliable homology annotations. SN and ST: Skew-normal and Skew-t distributions.*