

## Electronic Supplementary Material

### Pires *et al.* - Diversification dynamics of mammalian clades during the K-Pg mass

#### extinction

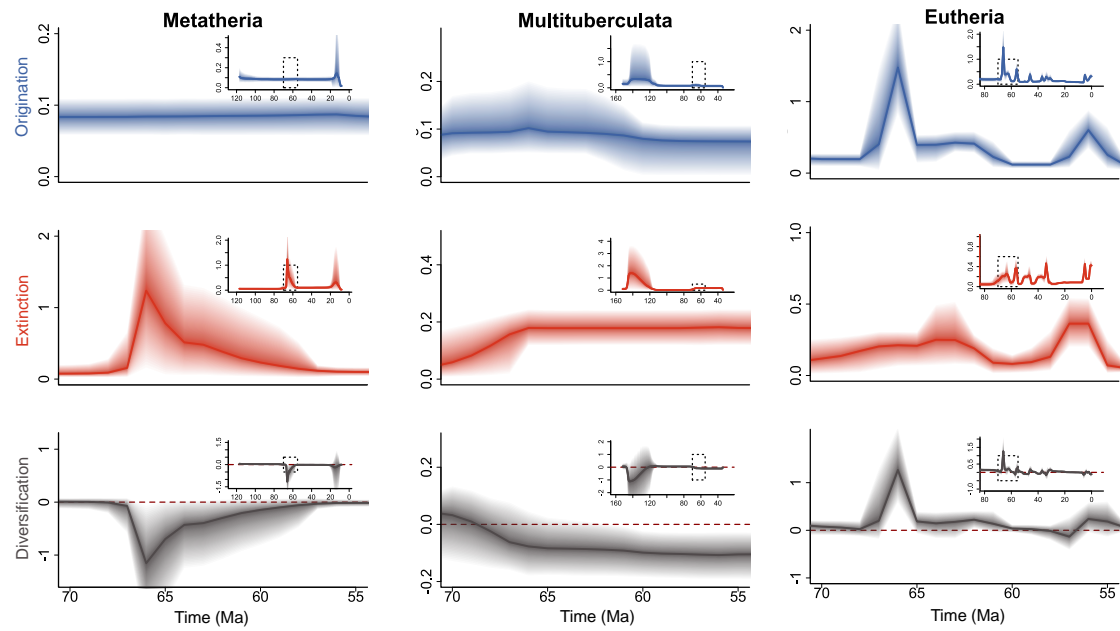


figure S1. Origination, extinction, and diversification rates for the three mammalian clades in North America estimated when considering all data available in the Paleobiology Database. The main panels depict the rates only for the interval of interest. The inset panels show the rate estimates from the root age of the clade until present and the dotted polygon delimits the time interval considered in the main panels. Solid lines and shaded areas depict the mean posterior rates and 95% credible interval.

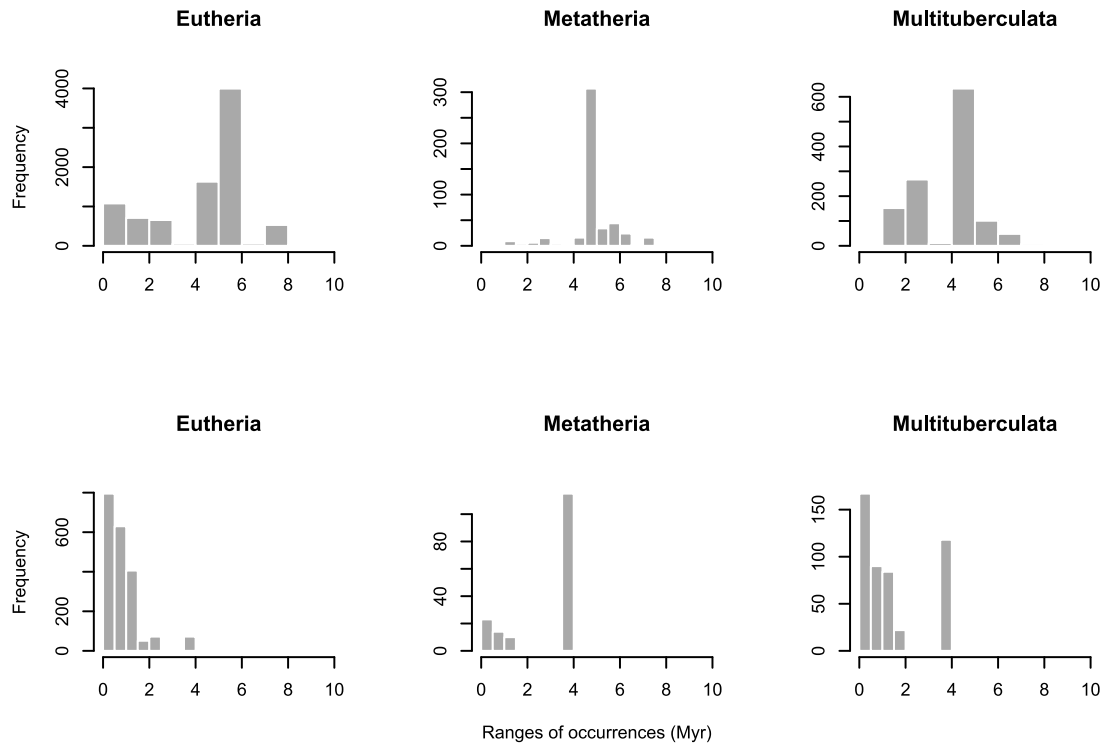


figure S2. Time resolution of fossil occurrences in the full PBDB dataset (top panel) and main dataset compiled for this study (bottom panel). Only occurrences from taxa that appear in both datasets are considered for comparison.

## Supplementary Methods

### *PyRate analyses*

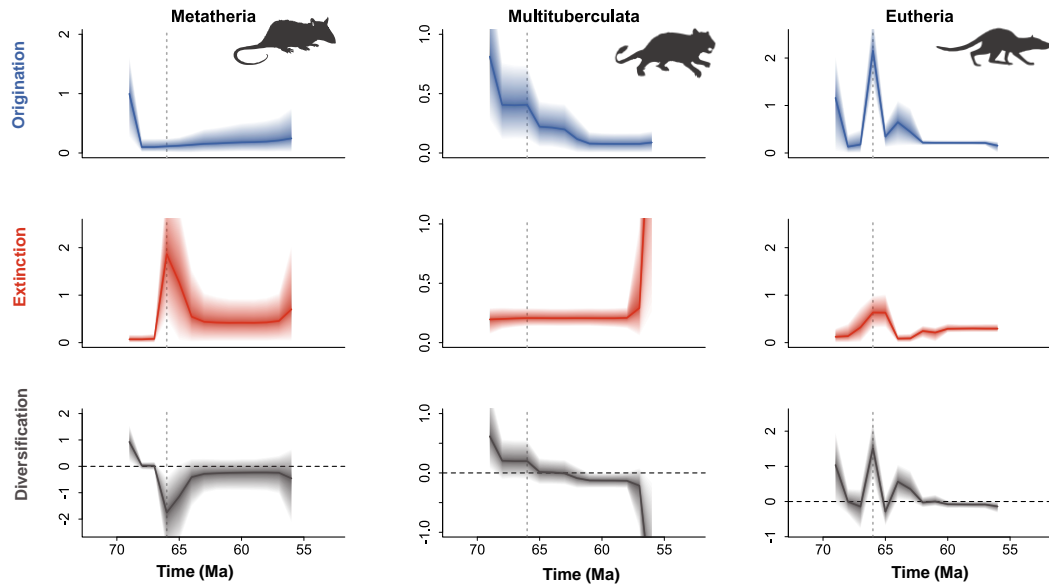
We estimated origination and extinction rates from fossil occurrences using PyRate, a hierarchical Bayesian approach that allows testing multiple birth-death models for the occurrence data while considering the preservation process [1,2]. For each taxon in the dataset PyRate estimates the times of origination and extinction using all its fossil occurrences. Using these estimates, it obtains posterior estimates for the origination ( $\lambda$ ) and extinction rates ( $\mu$ ), and the number and temporal placement of rate shifts. PyRate is an open source program and all code is freely available in GitHub (<http://github.com/dsilvestro/PyRate>). A thorough derivation of the equations and an explanation about the MCMC algorithms implemented in PyRate can be found in Silvestro et al [1,2].

Because we restricted our analysis to a particular time window we implemented novel features within PyRate to deal with potential edge effects. One of these features allows establishing a priori times of rate shift at the edges, while estimating rates and rate shifts within the window. This method consequently ignores what happens beyond the boundaries. Without this restriction, edge effects can cause apparent rate shifts at the edges of the time range reflecting a sampling bias (see figure S3). We also informed the models, which taxa are known to have originated before the analyzed time frame and those that survived beyond the boundary. This reduces the chances that edge effects will affect rate estimates within the time frame.

Previous versions of PyRate modeled preservation as either a homogeneous Poisson process (HPP), in which the preservation rate is constant through time, or a non-homogeneous Poisson process of preservation (NHPP), which assumes preservation rates may change during the lifespan of each lineage. In the most recent versions an alternative model of preservation assumes that preservation rates are constant within a time frame defined a priori (e.g., geological epochs), but can vary across time frames. Because we know mass extinctions left strong signatures on preservation we used this feature to allow preservation rates to vary over time. To do that we combined some time intervals to establish four time bins with enough occurrences (see table S2) within which we estimated independent preservation rates. This can be done using the command `-qShift` and providing a file containing the times that delimit the different time windows. An example of rate estimates without defining time-varying preservation rates can be found in figure S3 for comparison. Note that rate estimates at the boundaries of the interval in figure S3 show the signatures of edge effects when compared to figure 1 in the main document.

Another novel feature of PyRate used here is the Reversible Jump Markov Chain Monte Carlo (RJMCMC) which was shown to outperform alternative algorithms such as the birth-death Markov Chain Monte Carlo (BDMCMC) implemented in PyRate, being more likely to converge and to detect true rate shifts [3]. The RJMCMC is the default method in the PyRate since version 2.0 was released. In appendix S1 we provide an example of the commands used to

compute the rates using PyRate and explain each of the commands. More thorough tutorials with other examples can be found in <http://github.com/dsilvestro/PyRate/tree/master/tutorials>



**figure S3.** Origination, extinction and diversification rates estimated for the three mammalian clades in North America near the K-PG boundary. In these analyses preservation rate is assumed to be independent of time and boundaries are not properly controlled for edge effects. Solid lines and shaded areas depict the mean posterior rates and 95% credible interval. The vertical dotted lines denote the K-Pg boundary (66Ma).

**table S2.** Time windows within which we estimated independent preservation rates and the number of occurrences within each time window.

Time window	Eutheria occurrences	Metatheria occurrences	Multituberculata occurrences
69.9 to 66.04	72	115	118
66.04 to 62.11	752	18	153
62.11 to 58.9	740	19	159
58.9 to 55.5	462	11	51

**Appendix S1.** Example of command to execute PyRate to estimate origination and extinction rates

*Command*

```
python PyRate.py 'inputfile.py' -j 1 -N -1 -A 4 -out outputname -mG -trait_file  
'taxon_survival.txt' -bound 69.9 55.85 -qShift 'preservationWindows.txt' -s 10000 -n 30000000.
```

*Command list*

Inputfile.py is a '.py' file created by pyrateutilities.R (included in the PyRate package). The file is built from a table that includes occurrences, status and minimum and maximum ages (see table S1)

- j determines the data replicate to be used (pyrateutilities.R resamples the ages of each fossil occurrence from the respective temporal ranges generating replicates of the dataset)
- N number of extant taxa used to constrain the hyperprior of origination and extinction. -1 is advised if the clade is extinct or if the number of extant taxa is not representative of the clade at the investigated interval.
- A determines the MCMC algorithm to be used. -A 4 (the default option in the current version) runs a RJMCMC analysis.
- n sets the number of MCMC iterations
- s determines MCMC sampling frequency
- mG allows heterogeneity of the mean preservation rate across lineages
- qShift is used to pass a list of time windows ('preservationWindows.txt') at which preservation rates are estimated, allowing time varying preservation rates
- bound specifies that the data are truncated at 69.9 and 55.85 Ma
- trait\_file is used to pass a list of species traits to be considered for estimation of times of origination and extinctions. Here we defined a text file indicating species that originated or survived beyond the boundaries of the analyzed interval. (see worksheet 4, table S1).

### *Diversification rates estimators*

Different methods have been proposed to estimate origination and extinction rates from fossil occurrence data. These estimators differ in precision and how they deal with biases such as Signor-Lipps and edge effects [4]. We used the R-package *divDyn* (<http://github.com/adamkocsis/divDyn>) to estimate rates using different estimators [5–8] and examined how estimated rates compare with those obtained using *PyRate*. Figures S3-S5 show rate estimates for each clade using four different methods. We used 1Myr time bins to compute rates so that the rate-through-time plots are comparable with those generated by *PyRate*, but using NALMA bins, result in very similar patterns.

The main differences between these estimators and *PyRate* are that *PyRate* 1) corrects times of extinction and speciation taking into account the estimated preservation rate, thus incorporating our uncertainty about the times of origination and speciation and 2) estimates the temporal placement of rate shift in continuous time without relying on predefined bins. Additionally, because we randomly sample the ages of occurrences within the limits of each occurrence range we analyze multiple possible scenarios, incorporating the uncertainty related to occurrence dating. Finally, because rates and rate shifts are estimated within a Bayesian framework in *PyRate*, different models are evaluated and rate shifts are accepted depending on model likelihood. Thus, only rate shifts significantly supported by the data are accepted, making the detection of shifts more conservative than in those other methods, which tend to display more variability in rates over time.

Rate estimates for Eutheria and Multituberculata using the estimators (figures S4 and S6) tend to converge to the mean rates estimated using *PyRate*, although they cannot obtain estimates for all time bins, especially at the edges of the analyzed interval (figure 1). The estimators fail in detecting reasonable rates for metatherians (figure S5), suggesting a delayed peak in extinction around 62 Ma. These differences in rate estimates are caused by differences in how the methods deal with occurrence data, especially single-interval taxa. Because those estimators typically need to discard the first and last bins and ignore taxa that occur only within

bins, they tend to be more sensitive to fewer occurrences and may not obtain rate estimates for some intervals.

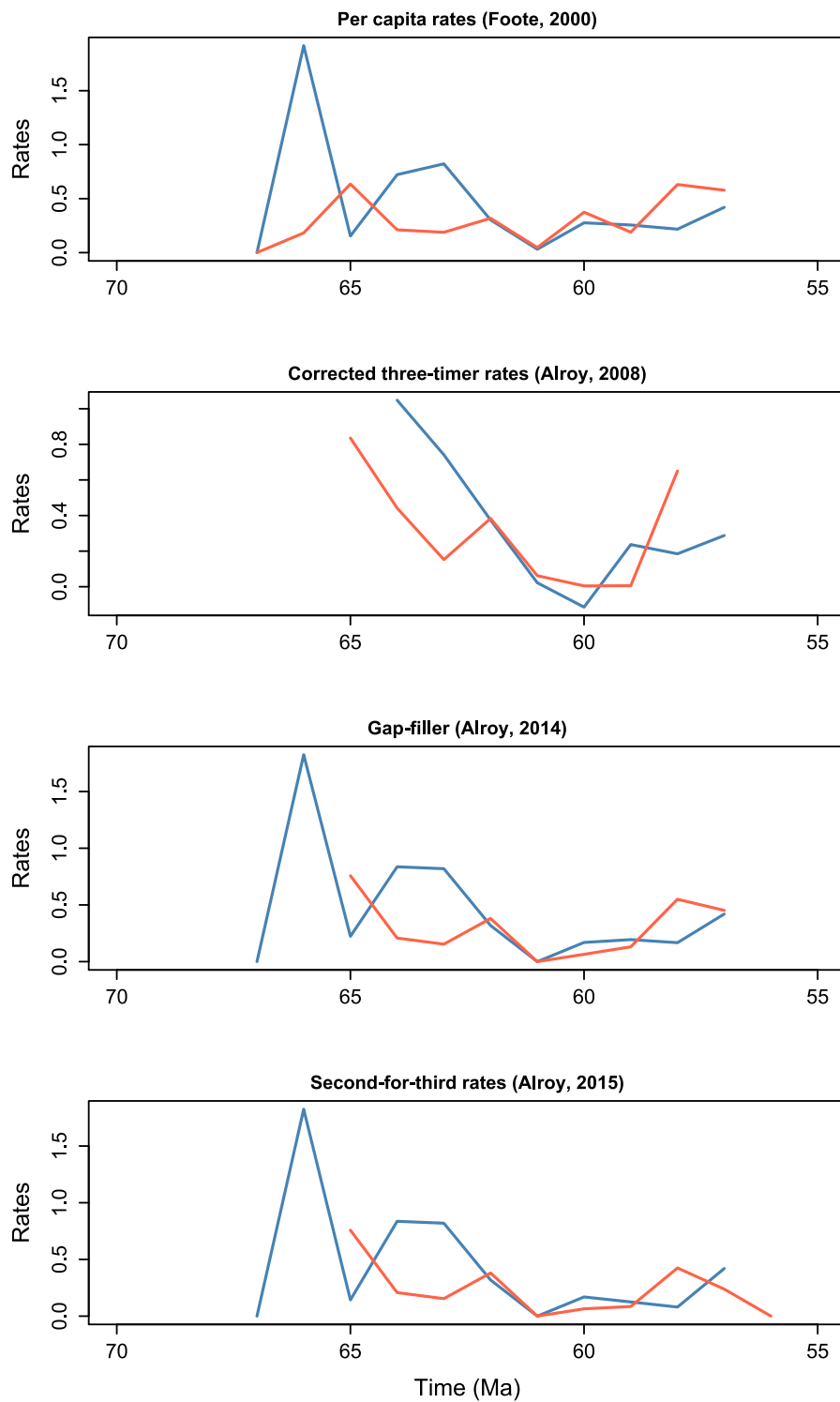


figure S4. Rates of origination (blue) and extinction (red) of Eutheria in North America around the K-Pg boundary according to four different rate estimators.

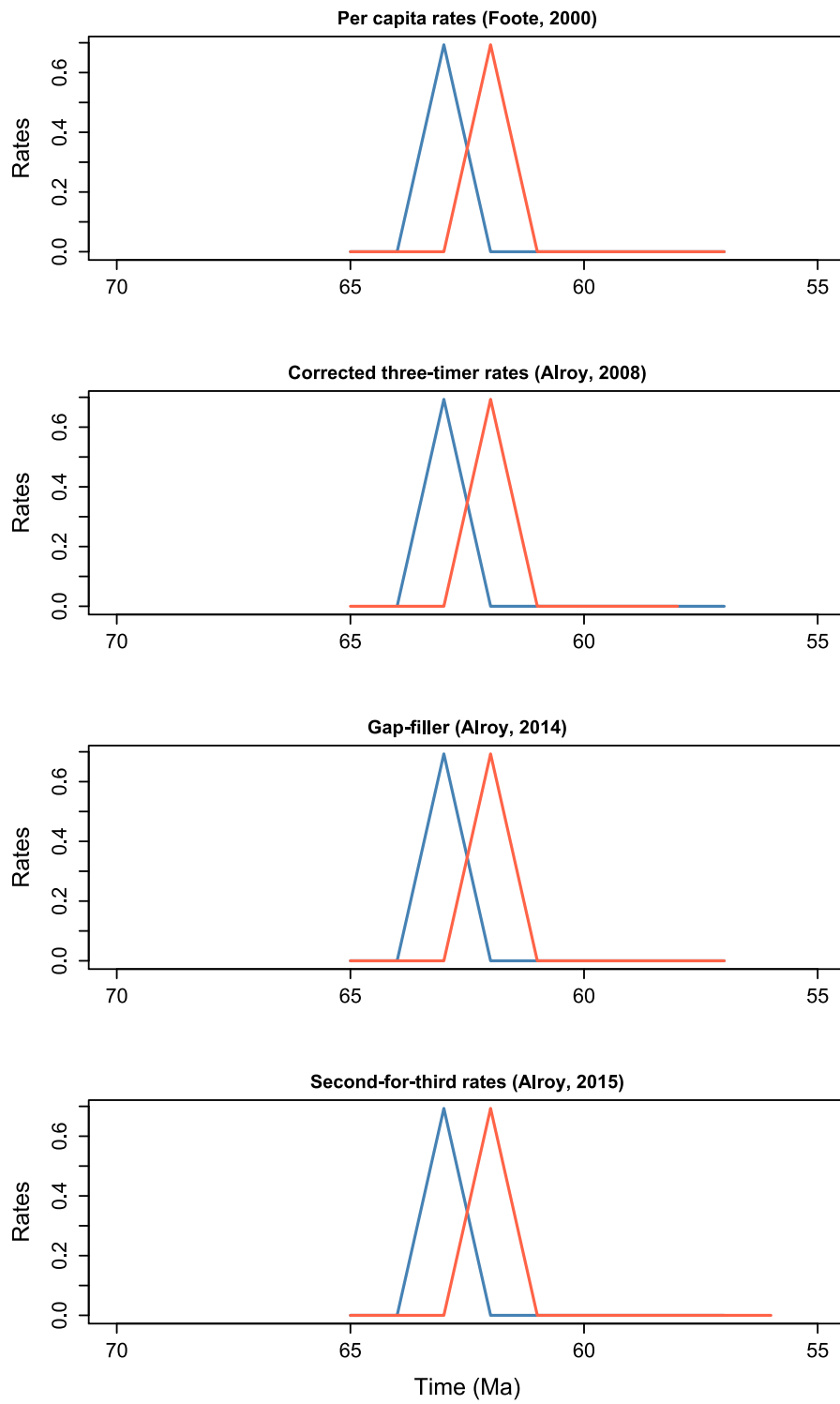


figure S5. Rates of origination (blue) and extinction (red) of Metatheria in North America around the K-Pg boundary according to four different rate estimators.

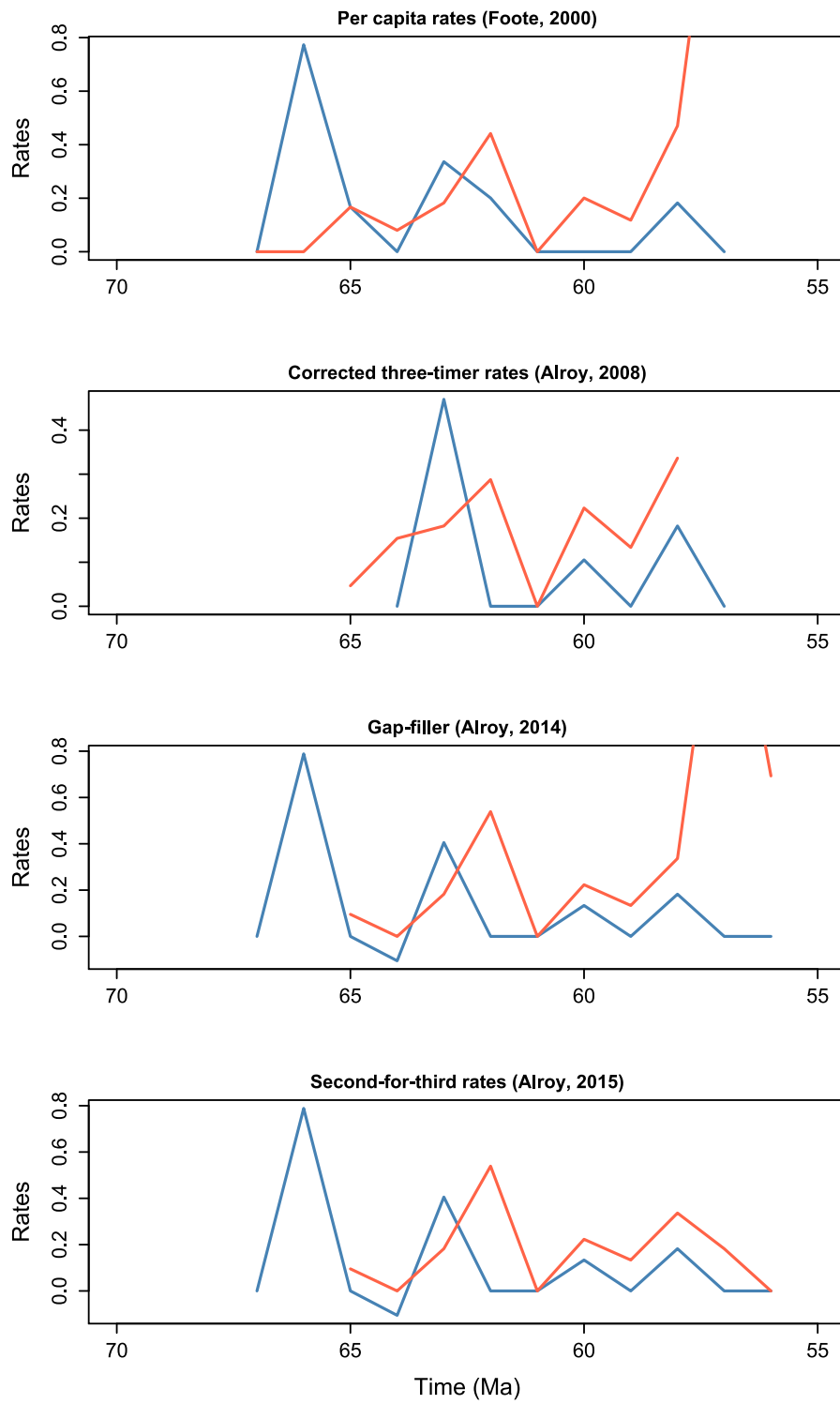


figure S6. Rates of origination (blue) and extinction (red) of Multituberculata in North America around the K-Pg boundary according to four different rate estimators.

### *Diversification of Mammalia*

When all three clades are considered together we recover a rise and fall pattern for both rates around the K-Pg (figure S7) that is similar to that found by Alroy [9] when analyzing species-level data using other estimators. These diversification dynamics are similar to those we found for eutherians alone. Since eutherians have many more occurrences than the other clades, the resulting rate estimates for the combined data are mainly driven by the changes in taxonomic diversity of the Eutheria.

When using a finer resolution where we analyzed the three clades individually we uncovered contrasting dynamics for the individual clades. This underscores that the analyses of diversification dynamics at a broad resolution may mask the dynamics occurring within each clade [10]. The analyses of origination and extinction rates subdividing the clades further into subclades would not be informative in this case because the taxonomy of subclades and their composition in terms of the constituent taxa is not well resolved for most subclades; also, the number of known taxa would be too small to obtain rate estimates for most subclades. However, using the available knowledge on ecomorphological features of taxa within each group [11,12] we can devise hypotheses to explain differences in diversification in terms of ecological aspects such as diet and dietary specialization, thus allowing inferences about the processes underpinning differences in diversification patterns.

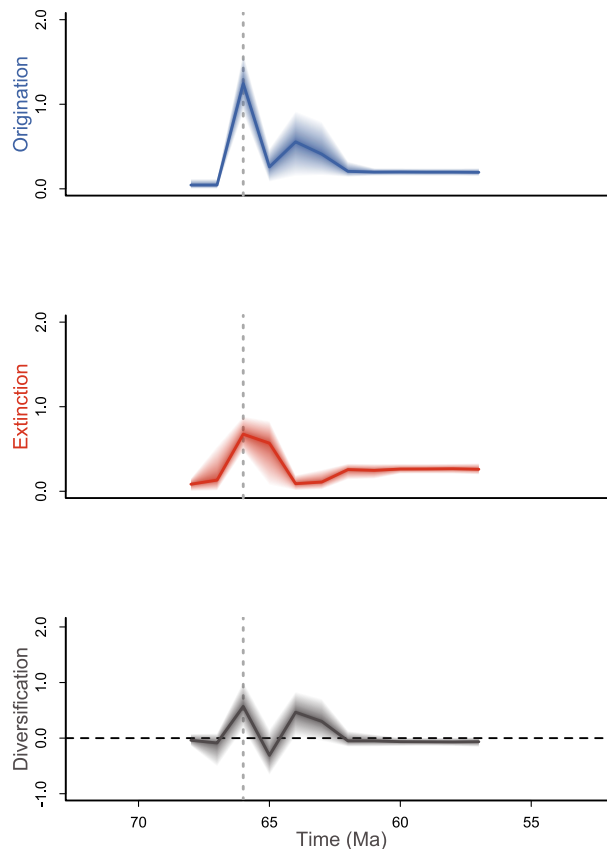


figure S7. Origination, extinction and diversification rates for the three mammalian clades analyzed as a single clade. Solid lines and shaded areas depict the mean posterior rates and 95% credible interval. The vertical dotted lines denote the K-Pg boundary (66 Ma).

### Supplementary references

1. Silvestro D, Schnitzler J, Liow LH, Antonelli A, Salamin N. 2014 Bayesian estimation of speciation and extinction from incomplete fossil occurrence data. *Syst. Biol.* **63**, 349–367.
2. Silvestro D, Salamin N, Schnitzler J. 2014 PyRate: a new program to estimate speciation and extinction rates from incomplete fossil data. *Methods Ecol. Evol.* **5**, 1126–1131.
3. Silvestro D, Antonelli A, Salamin N, Meyer X. 2018 Improved estimation of macroevolutionary rates from fossil data using a Bayesian framework. *bioRxiv* **316992**.
4. Alroy J. 2015 A more precise speciation and extinction rate estimator. *Paleobiology* **41**, 633–639.

5. Foote M. 2000 Origination and extinction components of taxonomic diversity: Paleozoic and post-Paleozoic dynamics. *Paleobiology* **26**, 578–605.
6. Alroy J. 2014 Accurate and precise estimates of origination and extinction rates. *Paleobiology* **40**, 374–397.
7. Alroy J. 2015 A more precise speciation and extinction rate estimator. *Paleobiology* **41**, 633–639.
8. Alroy J. 2008 Dynamics of origination and extinction in the marine fossil record. *Proc. Natl. Acad. Sci.* **105**, 11536–11542.
9. Alroy J. 1999 The fossil record of North American mammals: evidence for a Paleocene evolutionary radiation. *Syst. Biol.* **48**, 107–118.
10. Pires MM, Silvestro D, Quental TB. 2017 Interactions within and between clades shaped the diversification of terrestrial carnivores. *Evolution* **71**, 1855–1864.
11. Wilson GP. 2014 Mammalian extinction, survival, and recovery dynamics across the Cretaceous-Paleogene boundary in northeastern Montana, USA. *Spec. Pap. Geol. Soc. Am.* **503**, 365–392.
12. Grossnickle DM, Newham E. 2016 Therian mammals experience an ecomorphological radiation during the Late Cretaceous and selective extinction at the K–Pg boundary. *Proc. R. Soc. B Biol. Sci.* **283**, 20160256.