

1

2

3

4

5

Title:

6

**How many dinosaur species were there?**

7

**Fossil bias and true richness estimated using a Poisson sampling model (TRiPS)**

8

9

Jostein Starrfelt\*<sup>1</sup> and Lee Hsiang Liow<sup>1</sup>

10

<sup>1</sup>Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo,

11

Norway, P.O. Box 1066 Blindern, NO-0316 OSLO, Norway.

12

\*Corresponding author. Email: [jostein.starrfelt@ibv.uio.no](mailto:jostein.starrfelt@ibv.uio.no)

13

14 **Abstract**

15 The fossil record is a rich source of information about biological diversity in the past. However, the fossil  
16 record is not only incomplete but has inherent biases due to geological, physical, chemical and biological  
17 factors. Our knowledge of past life is also biased because of differences in academic and amateur interests  
18 and sampling efforts. As a result, not all individuals or species that lived in the past are equally likely to  
19 be discovered at any point in time or space. To reconstruct temporal dynamics of diversity using the fossil  
20 record, biased sampling must be explicitly taken into account. Here, we introduce an approach that  
21 utilizes the variation in the number of times each species is observed in the fossil record to estimate both  
22 sampling bias and true richness. We term our technique TRiPS (Trueness Richness estimated using a Poisson  
23 Sampling model) and explore its robustness to violation of its assumptions via simulations. We then  
24 venture to estimate sampling bias and absolute species richness of dinosaurs in the geological stages of  
25 the Mesozoic. Using TRiPS, we estimate that 1936 (1543-2468) species of dinosaurs roamed the Earth  
26 during the Mesozoic. We also present improved estimates of species richness trajectories of the three  
27 major dinosaur clades; the sauropodomorphs, ornithischians and theropods, casting doubt on the Jurassic-  
28 Cretaceous extinction event and demonstrating that all dinosaur groups are subject to considerable  
29 sampling bias throughout the Mesozoic.

30 Keywords: Dinosauria, Poisson, taxonomic richness, fossil record, sampling bias, diversity curves.

31

32

### 33 **Introduction**

34 One of the main goals of palaeobiology is to reconstruct diversity using information from the fossil record.  
35 While the patterns of diversity in space and through time are interesting in themselves, understanding the  
36 dynamics of taxon richness is also the first step in elucidating the biotic and abiotic forces that shape the  
37 spatial and temporal variation in taxon diversity. In other words, we need an accurate picture of patterns  
38 of past diversity to understand processes that operate on long time scales. As in all study systems where  
39 data samples in themselves cannot be assumed to represent a complete picture of the underlying  
40 population, richness studies based on the fossil record must consider the incompleteness of the fossil  
41 record.

42 Not all organisms enter the fossil record or have the same potential of doing so. Once created, a fossil  
43 record (a physical record of the existence of organisms that were alive in the past) is subject to eternal  
44 loss through physical processes such as erosion. Whether or not a fossilized organism can be found is  
45 also affected by variability in outcrop accessibility. Last but not least, sampling intensity encompassing  
46 factors including academic/commercial interest, geographic location and sampling design also influence  
47 information from the fossil record we have access to. While some of these factors contribute to noise in  
48 our inference of historical patterns and processes, and thus only cloud biological signals, others may  
49 cause systematic bias so as to yield misleading results if the data are interpreted at face value or with  
50 inappropriate methods.

51 Several classes of approaches for estimating richness using an incomplete fossil record have been  
52 developed. These might be loosely grouped into subsampling approaches, phylogenetic corrections and  
53 residual approaches. It is not our purpose to give a full overview of the approaches available, which have  
54 variously been reviewed elsewhere (see e.g. 1,2), but we briefly describe these in order to clarify why we  
55 have developed a new approach here. Subsampling approaches, including rarefaction (reviewed in 1) and  
56 Shareholder Quorum Subsampling (SQS; 3,4), attempt to standardize temporal (or spatial) samples so as  
57 to achieve comparable relative richness across samples. Phylogenetic approaches use phylogenetic  
58 hypotheses of the clade in question to infer ghost lineages unobserved in the fossil record but that must  
59 have existed as implied by the given phylogenetic hypothesis (5). These ghost lineages are thus assumed  
60 to give a minimum estimate of the number of lineages we have failed to observe in the fossil record. The  
61 residual approach (6–9) assumes that a given proxy for sampling (e.g. outcrop area or number of fossil  
62 bearing collections) captures the biases that might influence our observations and uses the proxy to model  
63 how a signal driven entirely by sampling would appear. Deviations from such a model are thought to  
64 reveal the real troughs and peaks in richness. In all of these approaches, we can only hope to estimate

65 relative richness through time and not true richness. Additionally, none of these approaches attempts to  
66 estimate the bias itself, i.e. the differential sampling across time, space or taxa. Without an estimate of  
67 sampling bias that is separate from richness estimates, it is difficult to shed light on the Common Cause  
68 Hypothesis, which states that a common factor affects both biological dynamics and sampling (10–13).  
69 Here, we introduce an approach that explicitly models the sampling process while estimating richness,  
70 using multiple observations of fossils belonging to an organismal group. We named it TRiPS (TTrue  
71 Richness estimated using a Poisson Sampling model). While we and others have used the simultaneous  
72 estimation of extinction, speciation and sampling processes to study diversification processes (14–17),  
73 there has not been a direct attempt to use multiple observations of fossil species to estimate true richness,  
74 rather than relative richness, while simultaneously and explicitly estimating sampling, as far as we are  
75 aware. Specifically, TRiPS assumes that species observed multiple times in a given time interval have a  
76 relatively high probability of fossilization and modern day discovery. We use this type of information  
77 across species that are likely to have similar fossilization potential and modern day discovery rates to  
78 estimate the number of species we might be missing and hence the true number of species that might have  
79 existed.

80 Dinosaurs are used as an example to illustrate our approach, not least because there is a lot of interest in  
81 estimating both the absolute (18–20) and relative temporal richness (21–25) of dinosaur taxa. As earlier  
82 analyses suggest that the three major dinosaur groups Sauropodomorpha, Ornithischia and Theropoda  
83 exhibit both different diversity dynamics and differential impact of sampling bias (9,25,26), we estimate  
84 sampling rates and true richness for all dinosaurs as well as these groups independently. We present  
85 stage-specific dinosaur sampling rates (i.e. bias) and dinosaur species richness through the Mesozoic as  
86 estimated from TRiPS, compare our estimates with those discussed in the literature and present  
87 simulations that explore the power of our approach and the sensitivity of TRiPS to violations of key  
88 assumptions.

## 89 **Methods and data.**

### 90 **Data**

91 We downloaded records of Dinosauria, Ornithischia, Sauropodomorpha and Theropoda separately from  
92 the Paleobiology Database (PaleoDB, <https://paleobiodb.org/#/>, download August 13<sup>th</sup> 2015) using the R  
93 package paleobioDB (27). Each row of data downloaded from the PaleoDB is associated with an  
94 observation of a taxon, its location and age range of the outcrop it was found in, and their metadata. In  
95 each download, we first discarded all records that were not identified to the species level by requiring that

96 the ‘matched rank’ of the entry was “species.” Second, we removed all *nomina dubia* and *nomina nuda*  
97 by comparing observations with a list of invalid taxa from PaleoDB. Third, we removed ichno- and  
98 ootaxa by matching our downloaded data with recently compiled lists of such taxa (28). We further  
99 identified four ichno-genera (*Harpedactylus*, *Yunnanpus*, *Saurichnium* and *Tetrapodium*), when  
100 comparing the download of all dinosaur observations with the subclades. Lastly we required that each  
101 observation was temporally assigned to one or more geological stages, i.e. observations only dated to  
102 epochs or periods were discarded. The final four datasets had 3122 observations of Dinosauria (974  
103 genera and 1124 species), 1150 observations of Ornithischia (306 genera, 360 species), 602 observations  
104 of Sauropodomorpha (215 genera with 246 species) and 1366 observations of Theropoda (451 genera  
105 with 516 species).

106 For each species, we tallied the number of observations per species in each stage in the Mesozoic,  
107 generating an observation count matrix. As already noted, the reported age range of a given record can  
108 span more than one geological stage. In such cases, we assigned a stage within the reported age range  
109 with a probability that is proportional to the duration of those stages. Because of this probabilistic  
110 assignment of records to stages within given age ranges, we performed TRiPS analyses (described below)  
111 on 100 replicated observation count matrices and used the median estimated sampling rate for species  
112 richness estimation. We also analysed genus level data but because both richness and sampling dynamics  
113 are similar to species level dynamics, we refer readers to the Electronic Supplementary Material (ESM)  
114 for genus level estimates.

### 115 **Modelling fossil sampling as a Poisson process**

116 Here, we treat the process of fossil sampling, which we will estimate from records from the PaleoDB (see  
117 previous section) as the combined processes of fossilization and detection. In other words, a sampling  
118 event is one fossil observation of a particular species, subsuming everything that has to occur from the  
119 point an individual died in the past, through fossilization and detection and finally ending up as an  
120 observation in a scientific database. Bias in the fossil record is then interpreted as systematic differences  
121 in the rate of sampling across time, space or organismal groups.

122 We assume that sampling can be viewed as a homogenous Poisson process inside a particular time  
123 interval. While others have also treated fossil sampling as a Poisson process (e.g. 29–33), ours, to the best  
124 of our knowledge, is the first attempt to use such an approach to estimate true richness. Formally, let the  
125 Poisson intensity (or rate)  $\lambda_t$  be the parameter controlling the sampling process in a given time interval  $t$ .  
126 The number of observations  $O_{i,t}$  for a species  $i$  in that time interval  $t$  with duration  $d_t$  has a Poisson

127 distribution with mean  $\lambda_t d_t$ . The likelihood of the sampling rate  $\lambda_t$  given  $O_{i,t}$  observations in that interval is  
 128 then

$$L(\lambda_t | O_{i,t}, d_t) = \frac{(\lambda_t d_t)^{O_{i,t}}}{O_{i,t}!} e^{-\lambda_t d_t} \quad (1)$$

129 Here we explicitly assume that a species detected in a time interval is extant during that whole time  
 130 interval. Because any species that is represented in the database must have left at least one detected fossil  
 131 we must condition the likelihood of  $\lambda_t$  on  $O_{i,t} > 0$ . The likelihood of  $\lambda_t$  is then

$$L(\lambda_t | O_{i,t}, d_t) = \frac{(\lambda_t d_t)^{O_{i,t}} e^{-\lambda_t d_t}}{1 - e^{-\lambda_t d_t}} \quad (2)$$

132 The maximum likelihood estimate for the sampling rate of a group of species in a given interval is found  
 133 by maximizing the product of eq 2 over all the observed species ( $N_t$ ) belonging to that group;

$$\hat{\lambda}_t = \max_{\lambda_t} \prod_{i=1}^{N_t} \frac{(\lambda_t d_t)^{O_{i,t}} e^{-\lambda_t d_t}}{1 - e^{-\lambda_t d_t}} \quad (3)$$

134 If our data consist of only single records (i.e.  $O_{i,t} = 1$ , for all  $i$ ), estimating  $\lambda_t$  using maximum likelihood  
 135 will yield an estimate of 0. Hence, the minimum data requirement for estimating the fossilization rate is a  
 136 dataset where at least one of the species has more than one observation.

137 We assume that sampling rates estimated are constant for all species within a clade in the same time  
 138 interval (i.e. the sampling rates estimated are time-specific but not species-specific). We can then  
 139 estimate the probability of detecting a species from this group as  $1 - Poiss(0, \hat{\lambda}_t d_t)$ , i.e. one minus the  
 140 probability of not detecting a species if it was actually extant, according to the Poisson process. We  
 141 further use this binomial probability in deriving the most likely true richness. The binomial probability of  
 142 a species sampled during an interval  $d_t$  is

$$p_{binom,t} = 1 - Poiss(0, \hat{\lambda}_t d_t) = 1 - e^{-\hat{\lambda}_t d_t} \quad (4)$$

143  
 144 where  $Poiss(0, \hat{\lambda}_t d_t)$  is the probability of zero sampling events in one lineage with a rate  $\hat{\lambda}_t$  in a bin of  
 145 duration  $d_t$ . Note that sampling rate ( $\lambda_t$ ) and sampling probability ( $p_{binom,t}$ ) are different; while sampling  
 146 rate has units observations per lineage per Ma, the sampling probability is a stage-specific probability of  
 147 detecting a species that was extant during that interval. The last step in estimating the true richness in a  
 148 given time interval is to find the true species richness ( $N_{true}$ ) that maximizes the binomial likelihood

$$L(N_{true} | p_{binom,t}, N_t) = \binom{N_{true}}{N_t} p_{binom,t}^{N_t} (1 - p_{binom,t})^{(N_{true}-N_t)} \quad (5)$$

149

150 where  $p_{binom,t}$  is the binomial probability calculated from the estimated sampling rate (eq 4) using  
151 maximum likelihood estimates (eq 3) and  $N_t$  is the observed number of species in the time interval. Thus  
152 the value  $N_{true}$  that maximizes eq 5 is the maximum likelihood estimate of the true richness where  $N_t$   
153 species were observed.

154 To quantify the uncertainty surrounding the estimate of the sampling rate and the true species richness we  
155 utilize the relationship between the  $\chi^2$  distribution and log likelihood profiles (see e.g. 29). For the  
156 confidence bounds on the maximum likelihood estimate  $\hat{\lambda}_t$  we find the range of values for  $\lambda$  that satisfy  
157 the inequality

$$2 \left( \log \left( L(\hat{\lambda}_t | O_{i,t}, d_t) \right) - \log \left( L(\lambda | O_{i,t}, d_t) \right) \right) < \chi_1^2(\alpha) \quad (6)$$

158

159 where  $\chi_1^2(\alpha)$  is the upper quantile function of the  $\chi^2$  distribution with one degree of freedom. Similarly  
160 the upper and lower confidence bounds for the estimated true richness  $N_{true}$  is found using the lower and  
161 upper confidence bounds on the sampling probability ( $p_{binom,t}$ ).

162 TRiPS thus yields maximum likelihood estimates and confidence intervals of true species richness for a  
163 given time interval by estimating a sampling rate (observations per species per Ma). This sampling rate  
164 can be transformed into a time interval specific sampling probability (probability of fossil detection per  
165 species) and thereby appropriately take the duration of the time interval into account. In other words, we  
166 do not need to conform the data to equal durations as commonly done (22,34). The sampling rates  
167 estimated from TRiPS are thus directly comparable across geological intervals of unequal durations. Note  
168 that while we have described TRiPS using species observations it can also be directly applied to genera or  
169 groups of taxa defined in other ways. In fact, any grouping of taxa thought to exhibit similar sampling  
170 rates might be combined, whether or not they actually are taxonomic clades.

### 171 **Simulations using a birth-death-fossilization process**

172 To evaluate our method's applicability and power we performed a large number of continuous time birth-  
173 death (BD) simulations, coupled with a fossilization scheme, which we interpret as sampling. In a classic  
174 BD process a lineage either gives rise to a new species or goes extinct at a certain rate; our fossilization  
175 scheme adds a third potential event: that of a lineage leaving a fossil. We are thus simulating a 'fossil

176 record' given a set of parameters controlling the dynamics and sampling of the simulated clade, and then  
177 using TRiPS to estimate the true number of species in these simulations

178 In case of no changes in species richness within a given time interval and identical sampling rates for all  
179 lineages, TRiPS will consistently recover true richness. Using simulations, we explicitly investigate the  
180 robustness of our approach to violations of TRiPS's two main assumptions, 1) equal sampling rates (per  
181 lineage per Ma) for all species in the clade in question and 2) and negligible species turnover within a  
182 time interval (i.e. all lineages span the entire interval in which they are observed). We then use the results  
183 of our simulations to aid interpretation of our estimates based on dinosaur records (see below).

184 Our birth-death-fossilize model has six parameters; speciation and extinction rates (in per species per time  
185 unit), the number of species at the start of the simulation, the duration of time units (in continuous time),  
186 mean sampling rate (fossils per species per time unit), and a parameter that scales the variability of  
187 sampling rates among individual species.

188 For our simulations, speciation and extinction rates were drawn from log-uniform distributions spanning  
189 0.001 to 0.15 per species per time unit. For comparison, the maximum mean speciation rate for birds in  
190 the Cenozoic has been estimated to be approximately 0.15 (35). Note that having non-zero speciation and  
191 extinction rates allow for changes in species richness, which explicitly violates the assumption of all  
192 lineages spanning the whole interval in which they are found. The initial number of species in the  
193 simulations ranged from 10 to 250, drawn from a uniform distribution. Durations of time units simulated  
194 ranged from 2 to 20, drawn from a uniform distribution mimicking the durations of the geological stages  
195 in the Mesozoic where the Hettangian (2 Ma) and Norian (19.5 Ma) are the shortest and longest stages,  
196 respectively. Mean sampling rates ranged from 0.001 to 1.5 fossils per species per time unit as drawn  
197 from a uniform distribution. Variation in sampling rate among species ranged from 0 to 0.3, also drawn  
198 from a uniform distribution; setting this parameter to 0.3, for example, gives each lineage a sampling rate  
199 ranging from 0.7 to 1.3 times the clade mean (this is equivalent to twice the coefficient of variation). Each  
200 species thus has its own unique sampling rate drawn from a uniform distribution with means and variance  
201 differing across simulations, explicitly violating the TRiPS assumption that lineages have the same  
202 sampling rate. We ran 100 000 simulations.

203 We evaluated TRiPS's ability to infer sampling rates and true species richness by 1) tabulating the  
204 number of simulations in which the true species richness was inside the predicted confidence interval  
205 (which we call success rate), 2) estimating the bias in the maximum likelihood prediction of species  
206 richness, the mean scaled error  $MSE = \frac{1}{n} \sum_{i=1}^n \frac{N_{true} - N_{TRiPS}}{N_{true}}$ , and 3) Pearson's product-moment correlation



207 ( $\rho$ ) between true and estimated richness in the simulations. We compiled the same statistics for estimated  
208 sampling rates and sampling probabilities. Together these three metrics inform us of the precision, bias  
209 and relative accuracy of our methodology.

## 210 **Estimating dinosaur sampling bias and species richness**

211 For each stage in the Mesozoic we estimate the sampling rate (i.e. the bias) and the species richness for  
212 each of the four clades using TRiPS. Combined with confidence intervals, these estimates allow for sound  
213 statistical comparisons of both bias and richness between geological stages. Additionally, to give a crude  
214 estimate of the total richness across the Mesozoic, we calculate a Mesozoic sampling probability as a  
215 mean of binomial probabilities, weighted by estimated richness in each stage. Mesozoic mean sampling  
216 rates are calculated similarly. Using mean Mesozoic sampling probability and the number of unique  
217 observed species we could also estimate the total richness of the dinosaur clades throughout the Mesozoic.

## 218 **Implementation**

219 All data analysis was performed in *R* (36). Code necessary for the analysis, combined with scripts to  
220 directly download relevant (and thus updated) data from Paleobiology Database in addition to simpler *R*  
221 scripts showing how other datasets can be analysed using TRiPS is available on the authors' website and  
222 Dryad. Output from our simulations and functions and scripts to rerun simulations are also  
223 deposited. [AUTHOR COMMENT: Dryad will not accept data before acceptance of manuscript].

## 224 **Results**

### 225 **Dinosaur sampling estimates among clades and through time**

226 Ornithischian species have the highest Mesozoic mean sampling rate (0.241 sampling events per lineage  
227 per Ma (0.172 – 0.355), see table 1). Sauropodomorph and theropod species have similar mean sampling  
228 rates with 0.156 (0.084 – 0.304) and 0.153 (0.104 – 0.243) respectively, and all dinosaurs have on average  
229 been sampled 0.192 (0.146 – 0.262) times per species per Ma. Mesozoic mean species sampling  
230 probabilities are also different among clades; theropods (0.463; 0.335 – 0.628) and sauropodomorphs  
231 (0.479; 0.278 – 0.753) have lower sampling probabilities than ornithischians (0.708; 0.575 – 0.847).

232 TRiPS based stage-specific sampling rates and probabilities for dinosaurs do not monotonically increase  
233 through the Mesozoic, but exhibit a combination of high and low sampling regimes (figure 1). This  
234 observation runs counter to the commonly held belief that younger geological strata exhibit a higher level  
235 of fossil sampling (e.g. 21,23). Sampling probabilities are particularly high during the Hettangian (201.3 –  
236 199.3 Ma) and Sinemurian (199.3-190.8 Ma), the Tithonian (152.1-145 Ma), the Albian (113-100.5 Ma)

237 and the Maastrichtian (72.1-66 Ma) but these high sampling intervals are interspersed with lower ones.  
238 Note that sampling rates ( $\lambda$ , top panel figure 1) and sampling probabilities ( $p_{binom,t}$ , lower panel figure 1)  
239 while showing some commonalities, are not the same. For instance, the Norian (228-202.5 Ma) has  
240 relatively few sampling events per lineage per million years, but the probability of a species being  
241 sampled, given that it was extant, is quite high ( $>0.8$  for all groups, see figure 1). In general, the relative  
242 changes in sampling dynamics are similar for our genus level analyses although sampling rates and  
243 probabilities are naturally higher for genera (ESM figures S5 and S6).

244 The three clades have notably different sampling estimates from stage to stage, with binomial sampling  
245 probabilities spanning from about 0.1 to almost 1. Theropods show higher sampling rates relative to  
246 ornithischians and sauropodomorphs in the Triassic (Carnian, 237 – 227 Ma, and Norian) but comparably  
247 lower rates in parts of the Early Cretaceous (Valanginian, 139.8 – 132.9, through Aptian, 125 – 112 Ma).  
248 This runs counter to earlier conclusions that richness trajectories of Theropoda and Ornithischia seems to  
249 largely be driven by sampling bias, whereas Sauropodomorpha are less affected by bias in the fossil  
250 record (25,26).

251 The different sampling rates across stages give a very different picture of bias than what the residual  
252 approach does. To reiterate, the residual approach assumes that a chosen time series fully captures the  
253 sampling bias, and uses a model of fixed diversity to predict how richness would look *if only* biased  
254 sampling drove the detected signal (8,9). The number of fossil collections from different intervals that  
255 contains at least one dinosaur (DBC) are often used as a sampling proxy for dinosaurs (23,26,37). We  
256 compare our proxy-free sampling estimates to DBCs to check how much sampling bias DBCs capture.  
257 We performed Pearson correlation tests of the binomial sampling probabilities estimated and the linearly  
258 detrended log10 number of collections for all downloaded dinosaur observations (see ESM table S2).  
259 These estimates indicate that there is a common signal in our estimated sampling probabilities and DBCs  
260 with correlations of 0.50 ( $p=0.0090$ ), 0.64 ( $p=0.003$ ), 0.71 ( $p=9e-5$ ) and 0.60 ( $p=0.002$ ), for Dinosauria,  
261 Ornithischia, Sauropodomorpha and Theropoda, respectively. However, there is substantial remaining  
262 sampling bias not captured by DBCs (since  $R^2 < 0.51$  for all). Using DBCs to remove bias in richness  
263 estimation of dinosaurs (e.g. 38) will add a substantial amount of error in correcting diversity curves.

## 264 **Dinosaur richness in the Mesozoic**

265 The whole Mesozoic is estimated to have seen 1936 (1543 – 2468, table 1) dinosaur species. A Mesozoic  
266 mean sampling probability at the genus level (again, weighted by estimated genus richness from each  
267 stage) yields estimates of total dinosaur genera richness at 1536 (1255 – 1929). Total species richness for

268 the subclades is estimated to be 508 (409 – 668) for Ornithischia, 513 (307 – 983) for Sauropodomorpha  
269 and 1115 (780 – 1653) for Theropoda (table 1).

270 The species richness estimates from TRiPS shares dynamics with those painted by both the raw counts of  
271 species and range-through species richness using the same dataset (figure 2). However, only in about half  
272 the stages are the range-through estimates within the confidence interval of TRiPS estimates. Genus  
273 richness dynamics are similar to species dynamics (ESM figure S6) and indicate that for at least this  
274 dataset using these analyses, genus level estimates can be a proxy for species estimates, corroborating  
275 Jablonski and Finarelli's findings (39). While genus richness estimates are lower, they are similar to  
276 species estimates, unsurprisingly, given there are few dinosaur species per genera (approximately 1.15  
277 identified species per genera in our data).

278 Estimated species richness shows an increase in the late Triassic, with a peak in the Rhaetian, though with  
279 a considerably wide confidence interval (figure 2, ESM table S2). This indicates that the initial  
280 diversification of dinosaur continued for up to 40 Ma, in contrast to the raw counts which peak in the  
281 Norian. Comparing stages within the Early Jurassic shows that the elevated raw richness in the  
282 Sinemurian for all clades is not corroborated by our estimates; TRiPS estimate richness for this stage well  
283 within the confidence intervals for richness in the Hettangian, Pliensbachian (190.8 – 182.7 Ma) and the  
284 Toarcian (182.7 – 174.1 Ma). In the Mid and Late Jurassic, also contrary to the raw species counts  
285 indicating a richness peak in the Tithonian, TRiPS estimates that all clades exhibit elevated richness in the  
286 Oxfordian due to the relatively low sampling rate in this interval for all clades. Finally, the trough in raw  
287 richness in the Coniacian (89.8 – 86.3 Ma) - Santonian with a peak in the two final stages of the  
288 Cretaceous for all clades seems to be a sampling artefact; while there is still a signal of reduced richness  
289 in the Coniacian for all dinosaurs, this is most likely driven by the ornithischians; the Coniacian-  
290 Santonian trough is not particularly strong for sauropodomorphs and theropods when inspecting the  
291 confidence intervals for these periods.

## 292 **Simulation results**

293 Most simulated species do not span the whole interval in which they were sampled, unless speciation and  
294 extinction rates are both zero. Using TRiPS to estimate the true richness is in such cases (the majority of  
295 our simulations) is thus a test of the degree to which our approach is robust to deviations from the  
296 assumption that an observed lineage spans the whole interval in question. Across the whole parameter  
297 space simulated (see above) TRiPS yielded confidence intervals including the true richness in 36% of  
298 simulations and true sampling rates were inside the confidence interval in 39% of simulations (see ESM  
299 for details). Though these are low absolute success rates, the correlations between true and estimated

300 richness is still high ( $\rho = 0.99$ ) and the mean scaled error is also low ( $MSE < 0.07$ ). This indicates that the  
301 maximum likelihood estimate of richness correlates very strongly with the true richness (indicating  
302 robustness in terms of *relative* richness) and that the richness estimate is, on average, only 7% smaller  
303 than the true richness. Unsurprisingly, the success rate and MSE of TRiPS are also highly dependent on  
304 all parameter settings, except the degree of variation in sampling rate between lineages in a single  
305 simulation (see ESM). In cases where individual lineages differ in sampling rates, TRiPS effectively  
306 estimates the group mean, and there is no first order effect on bias, precision or correlation of estimates  
307 when increasing the variability of sampling rates between lineages. The duration of the interval directly  
308 scales the potential for changes in species richness; the longer the duration, the more time for speciation  
309 and extinction within the time interval, leading to larger deviations from our assumption of lineages  
310 spanning the entire interval, and a higher failure rate of capturing the true richness within the confidence  
311 interval predicted by TRiPS. This underscores that TRiPS will perform better when applied to empirical  
312 data with narrower time intervals. Secondly, the sampling rate itself affect the success rate of TRiPS;  
313 TRiPS works best where sampling rates are relatively low, and even so when speciation and extinction  
314 rates are  $> 0$  (figure 3). We note, however, that the majority of our empirical dinosaur estimates are in  
315 regions of parameter space in which TRiPS has a relatively high success rate (figure 3).

## 316 **Discussion**

317 Fossil observation data are readily available in public databases, such as the PaleoDB. Yet, estimating  
318 taxon richness using such databases is not trivial as fossilization, outcrop exposure and modern day  
319 sampling and data compilation are heterogeneous processes. Unlike approaches such as subsampling  
320 (1,34,40) and bias-corrected residual analysis (8,9) our approach, TRiPS, estimates true rather than  
321 relative richness by utilizing information on sampling which is inherent in PaleoDB. In addition, unlike  
322 the residual approach, we do not make presuppositions that an external time series can be used to correct  
323 for sampling. This is important because such external time series (e.g. amount of outcrop, sea level) may  
324 constitute a factor driving both richness and sampling as postulated by the Common Cause Hypothesis  
325 (10,11) or be an effect of a third factor. The residual approach can also suffer from redundancy in the data  
326 used (41,42), which would lead to erroneously corrected richness estimates.

327 In TRiPS, we tackle bias in the fossil record directly by estimating rates of sampling. This also allows us  
328 to disentangle sampling and richness dynamics such that tests of links between potential drivers can be  
329 done on sampling and richness independently (see also 43). An advantage of TRiPS is that our treatment  
330 of sampling allows sampling probabilities to be directly comparable between intervals of unequal  
331 duration (but see below). One assumption we do explicitly make which cannot be true most of the time, is

332 that a species detected in a given time interval is extant during that whole time interval. This is because  
333 most species are unlikely to become extinct exactly at the late boundary of a time interval or originate  
334 exactly at the early boundary of a time interval. While other methods for estimating richness also assume  
335 that turnover is clumped at interval boundaries (see e.g. 40, p. 74), we explicitly examined the robustness  
336 of our estimates to the violation of this crucial assumption.

337 Comparing our empirical sampling rate estimates with simulations that violated key assumptions of  
338 TRiPS (figure 3), we find most of our empirical estimates fall within parameter ranges in which we are  
339 able to retrieve true richness estimates reliably. This is with the caveat that the simulated speciation and  
340 extinction rates are realistic for dinosaurs.

341 However, we note other caveats to the estimates from TRiPS. First, although the ability to estimate true  
342 richness is relatively robust to deviations from our assumptions under our simulations, TRiPS does give  
343 biased estimates of both sampling and richness when there are within-bin dynamics. When applying  
344 TRiPS to cases with non-zero speciation and extinction rates and increasingly wide time intervals, our  
345 ‘data’ will consist more and more of observations of lineages that did not span the entire interval. This  
346 generates a bias in our estimated sampling rates such that these estimates are lower than true sampling  
347 rates (see ESM figure S2), and more so for longer durations and higher levels of speciation and extinction  
348 rates. On the other hand, transforming this rate into a binomial probability (eq 4), assuming all lineages  
349 lasts longer than they actually do, leads to an overestimation of the binomial sampling probability when  
350 compared to the fraction of species actually observed in our simulations (see ESM figure S3). These two  
351 opposing biases in our methodology in sum lead to a slight negative bias in the reconstructed richness (on  
352 average 7% below true richness in our simulations). So, sampling rates and richness are slightly  
353 underestimated, but sampling probabilities are overestimated.

354 The estimated richness is thus best treated as minimum richness estimates, particularly for intervals in  
355 which there is reason to believe that within interval changes in true richness have been substantial, such as  
356 in long geological stages. Second, with longer intervals (which gives more time for in-bin dynamics) and  
357 higher sampling rates TRiPS fails more often in terms of capturing true richness inside confidence  
358 intervals (figure 3) due to high precision combined with bias. Note, however, that the maximum  
359 likelihood estimate of true richness itself is still not too far off target if speciation and extinction rates are  
360 not too high (see ESM figure S1). On the other hand, one benefit of our explicit approach is that it is  
361 straightforward to simulate a birth-death-fossilization process to check if the empirical estimates of  
362 sampling rates and richness can be considered robust to violation of the assumption of negligible turnover  
363 within an interval (see below). It is also worth highlighting that other approaches for reconstructing past

364 richness also fall victim to deviations from constant richness (see e.g. 40), even though such violations  
365 have not been explicitly examined in published simulations, as far as we know.

366 The obvious solution to these caveats is to apply TRiPS only for temporally well-resolved data, and to  
367 strive for better and more accurate dating of fossils. Other potential avenues for improvement lie in  
368 development of the methodology itself. For instance, extending TRiPS to estimate sampling rates in a  
369 hierarchical manner, where instead of estimating a single sampling rate, one could estimate a distribution  
370 of sampling rates within a time interval, or link sampling rates for subclades in a hierarchical fashion,  
371 possibly including more phylogenetic information. Other directions could be to include estimates of  
372 lineage durations from independent sources instead of assuming that all lineages span the interval in  
373 which they are observed.

374 In summary our simulations shows that TRiPS is prone to errors if its assumptions are substantially  
375 violated. Even in applications where we have reason to believe there are high levels of species turnover,  
376 estimates of richness will still be a reasonable approximation of the true richness, although the associated  
377 confidence intervals should be treated with caution; they are in general too narrow for such cases. In  
378 applying TRiPS on shorter intervals, i.e. when there is more reason to believe that lineages sampled  
379 actually do span most of the interval, TRiPS confidence intervals will in most cases encompass the true  
380 richness. With the pros and cons of TRiPS in mind, we can now compare our dinosaur estimates to  
381 previous studies.

382 Dodson (19) estimated the total number of dinosaur genera to be 900-1200 for the whole Mesozoic, with  
383 about 100 genera at any one geological stage. Our estimated genus richness of 1536 (1255 – 1929) are  
384 more in line with Wang and Dodson's (18) estimates, which inferred that the entire Mesozoic saw 1844  
385 dinosaur genera. For the final stages of the Cretaceous Wang and Dodson (18) estimated 200-300 genera  
386 of dinosaurs roaming our planet, corroborated by our estimates of 221 and 227 genera for Campanian and  
387 Maastrichtian respectively (see ESM figure S5, ESM table S2).

388 It is currently accepted that dinosaurs did not rapidly diversify when they appeared around the start of the  
389 Late Triassic. Rather, sauropodomorphs diversified during the final part of the Triassic, while  
390 ornithischians and theropods increased in richness in the early Jurassic (21,38). While this pattern is in  
391 part corroborated by our analysis for sauropodomorphs, a species richness of 80 for the whole dinosaur  
392 clade is estimated for the last stage in the Triassic, which is relatively high compared with the rest of the  
393 Mesozoic. Sampling rates for ornithischians cannot be estimated with confidence for any interval in the  
394 Triassic. Sauropodomorphs exhibit rather high levels of both observed and estimated species richness  
395 already in the Norian (228 – 208.5 Ma), and our estimate of sauropodomorph species richness during the



396 Rhaetian (208.5 – 201.3 Ma) is so high that it was not even surpassed by the diversity in the final stages  
397 of the Late Cretaceous, and only barely so for the diversity peaks in the Late Jurassic (44) and Early  
398 Cretaceous. In other words, our results indicate that both sauropodomorph and dinosaur richness peaks in  
399 the Rhaetian, and not in the Norian as earlier studies (26), due to the estimated low sampling this final  
400 stage of the Triassic. This is in contrast to previous residual based analysis which indicate the Rhaetian  
401 shows a marked decrease in diversity (38, but also see 9).

402 It is largely accepted that the Oxfordian (163.5-157.3 Ma) exhibits remarkably low diversity; perhaps it  
403 could even be considered the most depauperate stage throughout the Age of the Dinosaurs (21,26,38,44),  
404 and particularly so for sauropodomorphs (26). Our approach, in contrast, estimates the sampling  
405 probability in this particular stage to be the culprit of this trough and especially so for sauropodomorphs  
406 (see figure 1). Instead of inferring low species richness in this stage, our estimates indicate the  
407 sauropodomorph richness has doubled from the previous stage (Callovian), and a great richness increase  
408 is also evident in ornithischians and the dinosaurs as a whole. More intense sampling efforts, both in the  
409 field and in museum collections, and detailed analysis on the observations from the Oxfordian are clearly  
410 needed.

411 Some have argued that the Jurassic-Cretaceous (J/K, Tithonian - Berriasian) boundary (~145 Ma)  
412 demonstrated a clear diversity loss particularly pronounced for sauropodomorphs (25,26,38,45), while  
413 others have claimed otherwise (46,47). Though the raw species counts partially corroborate a decline in  
414 richness across the J/K boundary, the sampling rates for the early part of the Cretaceous are estimated to  
415 be much lower than late Jurassic (figure 1), and the difference in sampling rates are particularly  
416 pronounced for the sauropodomorph clade. While the richness for all dinosaurs is estimated to decrease  
417 (figure 2 top panel), this signal is equivocal for any of the subclades whose confidence intervals for  
418 richness in the Berriasian are wide (and span the richness in the Tithonian) due to low and uncertain  
419 sampling probabilities. Genus level analyses (see ESM Table S2) estimate that the number of  
420 sauropodomorph genera in Berriasian is in fact the same before and after the J/K boundary. It is also  
421 worth emphasizing that the confidence intervals for the estimated species diversity are much wider in  
422 Berriasian compared to Tithonian, implying that an increase in the true richness across this boundary  
423 cannot be excluded. Lower sampling rates during the Berriasian across all dinosaur clades and regardless  
424 of whether species or genus level data are used, suggest that this “clear diversity loss”, might be an  
425 artefact of sampling bias and that the ‘major extinction’ of dinosaurs across the Jurassic-Cretaceous  
426 boundary (25,26,44,48,49) might be less severe than previously thought (see also 45,46).

427 There current view is that not only is there no longer-term (i.e. epoch-level) decline in richness prior to  
428 the end-Cretaceous extinction, there is in fact a sharp increase in richness for the final two stages in the  
429 Cretaceous (50). The latter is not corroborated by our estimates; while the two final stages of the  
430 Cretaceous are high in raw species counts for all clades, estimated richness in Campanian and  
431 Maastrichtian are well within confidence intervals for earlier stages in the Cretaceous for  
432 sauropodomorphs, and also partially for ornithischians and theropods. Rather than the two final stages  
433 exhibiting elevated richness levels, TRiPS estimates that similar level of richness was reached in the  
434 previous stage (Santonian), at least for Theropoda and Ornithischia. Ornithischia do seem to have a real  
435 trough in the Coniacian (89.8 – 86.3 Ma), but sauropodomorphs and theropods seem instead to have  
436 steady, but slow, decreases and increases in richness, respectively, across the Late Cretaceous.

437 Late Cretaceous increase in dinosaur diversity has also been framed as a debate on whether or not  
438 dinosaurs showed a decline in species richness prior to the Cretaceous-Paleogene extinction event  
439 (9,18,21,24,25,50). Brusatte et al. (50) argued that, while there was no global long-term decline prior to  
440 the end-Cretaceous extinction, there is evidence for ceratopsids and hadrosaurids (members of  
441 Ornithischia) exhibiting declines in both species richness and morphological disparity in the final 15 Ma  
442 of the Cretaceous (50). On the other hand, Lloyd (9) claimed that sauropodomorphs and ornithischians  
443 both show long-term declines throughout most of the Cretaceous (using the residual approach) whereas  
444 Barrett et al. (25) highlighted a negative trend in taxic diversity for theropods and ornithischians in the  
445 last two stages of the Cretaceous, but suggested a ‘radiation’ of sauropodomorphs in the Late Cretaceous  
446 (see also 44). TRiPS estimates the final stage of the Cretaceous to be about 13 % less speciose for  
447 ornithischians (figure 2, ESM table S2) and a reduction in species richness by 17% is estimated to have  
448 occurred for sauropodomorphs, though with less confidence, while we estimate a significantly higher  
449 richness in the Maastrichtian than the Campanian for theropods (see ESM table S2).

## 450 **Conclusions and future directions**

451 To paint an accurate picture of past species richness and to identify of periods of high diversification and  
452 major extinction events, the bias inherent the fossil record that may mislead and confound our inferences  
453 needs to be taken into account. Here we have detailed TRiPS, a new approach for estimating both  
454 temporally varying sampling and species richness. The application of TRiPS to a global dataset of  
455 dinosaur records indicates that several of the commonly held ideas about the species richness trajectory of  
456 dinosaurs might be effects of either sampling bias or the use of methods that might have introduced new  
457 biases to richness estimates through their assumptions.



458 As a tool that estimates both sampling rates *and* true richness directly, TRiPS is pregnant with  
459 possibilities and have applicability to a range of other palaeontological questions. Richness and sampling  
460 estimates from TRiPS allow us to test the Common Cause Hypothesis in a straightforward manner if  
461 potential common drivers can be measured in the geological record. Estimates of sampling can be used in  
462 predicting true ranges of a given species, if we can make the assumption that species have the same  
463 temporally varying sampling rates. The two forms of sampling estimates may help palaeontologists focus  
464 their sampling and taxonomic efforts in time intervals (or geographic regions) which require most effort  
465 given the specific questions we wish to answer. Additionally, sampling rates are needed for better  
466 calibration of phylogenetic trees using fossil observations, and for inference of rates of speciation and  
467 extinction using phylogenies on extinct taxa. While the application of TRiPS we demonstrated here  
468 attempts to estimate global richness of dinosaurs and its major subclades, TRiPS can be applied to any  
469 collection of lineages that are assumed or shown to have similar sampling rates, and could also be used to  
470 estimate taxonomic richness on local to continental scales.

#### 471 **Data accessibility statement.**

472 All data were downloaded from Paleobiology Database (<https://paleobiodb.org/#/>). A copy of the download is  
473 available on Dryad [ref to Dryad here].

#### 474 **Competing interests statement.**

475 The authors declare no competing interests.

#### 476 **Authors' contributions**

477 JS conceived the study, performed the analysis and wrote the manuscript with substantial contributions to writing  
478 and interpretation by LHL. Both authors gave final approval for publication.

#### 479 **Acknowledgements:**

480 Kjetil Lysne Voje is acknowledged for fruitful discussions surrounding the initial idea, and Fakutsi is recognized for  
481 inspiration. We are also thankful for comments and criticisms made by Philip Mannion and Manabu Sakamoto  
482 acting as reviewers. Comments on the manuscript and ideas inside it by Roger Benson, Matthew Clapham, Gene  
483 Hunt, Trond Reitan and Steve Wang are also acknowledged. We also thank Benson for sharing his lists of ichno-  
484 and ootaxa. Credits to Andrew A. Farke, Scott Hartman and Conty, via PhyloPic, for the silhouettes in figure 2. This  
485 work would not be possible without vertebrate palaeontologists who contributed primary data and the authorizers  
486 and enterers of the PaleoDB who compiled these primary data. Major contributors to the downloaded PaleoDB data  
487 are Matthew Carrano (59,8%), John Alroy (13,9%), Mark Uhen (8,9%), Anna Behrensmeyer (4,6%) and Philip  
488 Mannion (3,1%). This is PaleoDB Publication XXX.

#### 489 **Funding Statement**

490 The research reported herein was funded by Norwegian Research Council Grant 235073/F20 to L.H. Liow.

#### 491 **Electronic supplementary material (ESM)**

492 Electronic supplementary material contains details of the simulations, results regarding these simulations, , results  
493 from estimating sampling from all 100 replicated datasets detailed in the main text. An Excel file  
494 (ESM\_Table\_S2.xlsx) contains all estimated sampling rates, sampling probabilities and true richness for each clade  
495 in each stage at both species and genus level. Data and R code to replicate analysis and simulations is available on  
496 Dryad [ref to dryad here]

497

## 498 **References**

- 499 1. Bush AM, Markey MJ, Marshall CR. Removing bias from diversity curves: the effects of spatially  
500 organized biodiversity on sampling-standardization. *Paleobiology*. 2004;30(4):666–86.  
501
- 502 2. Smith AB, McGowan AJ. The ties linking rock and fossil records and why they are important for  
503 palaeobiodiversity studies. *Geol Soc London, Spec Publ*. 2011 Dec 13;358(1):1–7.  
504
- 505 3. Alroy J. The shifting balance of diversity among major marine animal groups. *Science*.  
506 2010;329(5996):1191–4.  
507
- 508 4. Alroy J. Geographical, environmental and intrinsic biotic controls on Phanerozoic marine  
509 diversification. *Palaeontology*. 2010 Nov 16;53(6):1211–35.  
510
- 511 5. Norell MA, Novacek MJ. Congruence Between Superpositional and Phylogenetic Patterns:  
512 Comparing Cladistic Patterns With Fossil Records. *Cladistics*. 1992;8(4):319–37.  
513
- 514 6. Smith AB, Lloyd GT, McGowan AJ. Phanerozoic marine biodiversity: rock record modelling  
515 provides an independent test of large-scale trends. 2012;(September):4489–95.  
516
- 517 7. McGowan AJ, Smith AB. Comparing the Geological and Fossil Records: Implications for  
518 Biodiversity Studies. *Geological Society of London*; 2011. 247 p.  
519
- 520 8. Smith AB, McGowan AJ. The shape of the Phanerozoic marine paleodiversity curve: How much  
521 can be predicted from the sedimentary rock record of western Europe? *Palaeontology*.  
522 2007;50:765–74.  
523
- 524 9. Lloyd GT. A refined modelling approach to assess the influence of sampling on palaeobiodiversity  
525 curves: new support for declining Cretaceous dinosaur richness. *Biol Lett*. 2012;8(1):123–6.  
526
- 527 10. Peters SE. Geologic constraints on the macroevolutionary history of marine animals. *Proc Natl*  
528 *Acad Sci U S A*. 2005;102(35):12326–31.  
529
- 530 11. Peters SE, Heim NA. The geological completeness of paleontological sampling in North America.  
531 *Paleobiology*. *The Paleontological Society*; 2010 Jan;36(1):61–79.  
532

- 533 12. Hannisdal B, Peters SE. Phanerozoic Earth System Evolution and Marine Biodiversity. *Science*.  
534 2011 Dec 25;334(6059):1121–4.  
535
- 536 13. Hannisdal B. Non-parametric inference of causal interactions from geological records. *Am J Sci*.  
537 2011 Aug 15;311(4):315–34.  
538
- 539 14. Conroy MJ, Nichols JD. Testing for Variation in Taxonomic Extinction Probabilities: A Suggested  
540 Methodology and Some Results. *Paleobiology*. 1984;10(3):228–337.  
541
- 542 15. Nichols JD, Morris RW, Brownie C, Pollock KH. Sources of variation in extinction rates, turnover,  
543 and diversity of marine invertebrate families during the Paleozoic. *Paleobiology*. 1986;12(4):421–  
544 32.  
545
- 546 16. Foote M. Origination and Extinction through the Phanerozoic: A New Approach. *J Geol*.  
547 2003;111(2):125–48.  
548
- 549 17. Liow LH, Finarelli JA. A dynamic global equilibrium in carnivoran diversification over 20 million  
550 years. *Proc R Soc London, B*. 2014;281:20132312.  
551
- 552 18. Wang SC, Dodson P. Estimating the diversity of dinosaurs. *Proc Natl Acad Sci U S A*. 2006 Sep  
553 12;103(37):13601–5.  
554
- 555 19. Dodson P. Counting dinosaurs: how many kinds were there? *Proc Natl Acad Sci U S A*.  
556 1990;87(19):7608–12.  
557
- 558 20. Russell DA. China and the lost worlds of the dinosaurian era. *Hist Biol*. 1995;10(1):3–12.  
559
- 560 21. Brusatte SL. Macroevolution and Extinction. *Dinosaur Paleobiology*. John Wiley & Sons, Ltd;  
561 2012. p. 242–61.  
562
- 563 22. Lloyd GT, Davis KE, Pisani D, Tarver JE, Ruta M, Sakamoto M, et al. Dinosaurs and the  
564 Cretaceous Terrestrial Revolution. *Proc R Soc London, B*. 2008 Nov 7;275(1650):2483–90.  
565
- 566 23. Butler RJ, Benson RBJ, Carrano MT, Mannion PD, Upchurch P. Sea level, dinosaur diversity and  
567 sampling biases: investigating the “common cause” hypothesis in the terrestrial realm. *Proc Biol*  
568 *Sci*. 2011;278(September 2010):1165–70.  
569
- 570 24. Fastovsky DE, Huang Y, Hsu J, Martin-McNaughton J, Sheehan PM, Weishampel DB. Shape of  
571 Mesozoic dinosaur richness. *Geology*. 2004;32(10):877–80.  
572
- 573 25. Barrett PM, McGowan AJ, Page V. Dinosaur diversity and the rock record. *Proc Biol Sci*. 2009 Jul

- 574 22;276(1667):2667–74.  
575
- 576 26. Mannion PD, Upchurch P, Carrano MT, Barrett PM. Testing the effect of the rock record on  
577 diversity: a multidisciplinary approach to elucidating the generic richness of sauropodomorph  
578 dinosaurs through time. *Biol Rev Camb Philos Soc*. 2011 Feb;86(1):157–81.  
579
- 580 27. Varela S, González-Hernández J, Sgarbi LF, Marshall C, Uhen MD, Peters SE, et al. paleobioDB:  
581 an R package for downloading, visualizing and processing data from the Paleobiology Database.  
582 *Ecography (Cop)*. 2015 Apr 2;38(4):419–25.  
583
- 584 28. Benson RBJ, Butler RJ, Alroy J, Mannion PD, Carrano MT, Lloyd GT. Near-stasis in the long-  
585 term diversification of Mesozoic tetrapods. *PLoS Biol*.  
586
- 587 29. Solow AR, Smith WK. On Fossil Preservation and the Stratigraphic Ranges of Taxa On fossil  
588 preservation and the stratigraphic ranges of taxa. *Paleobiology*. 1997;23(3):271–7.  
589
- 590 30. Foote M. Estimating Taxonomic Durations and Preservation Probability. *Paleobiology*.  
591 1997;23(3):278–300.  
592
- 593 31. Stadler T. Sampling-through-time in birth-death trees. *J Theor Biol. Elsevier*; 2010 Dec  
594 7;267(3):396–404.  
595
- 596 32. Silvestro D, Salamin N, Schnitzler J. PyRate: A new program to estimate speciation and extinction  
597 rates from incomplete fossil data. *Methods Ecol Evol*. 2014 Sep;5(10):1126–31.  
598
- 599 33. Heath TA, Huelsenbeck JP, Stadler T. The fossilized birth-death process for coherent calibration  
600 of divergence-time estimates. *Proc Natl Acad Sci*. 2014;111(29):E2957–66.  
601
- 602 34. Alroy J, Marshall CR, Bambach RK, Bezusko K, Foote M, Hansen TA, et al. Effects of Sampling  
603 Standardization on Estimates of Phanerozoic Marine Diversification. *Proc Natl Acad Sci U S A*.  
604 2001;11:6261–6.  
605
- 606 35. Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO. The global diversity of birds in space and  
607 time. *Nature*. 2012 Nov 15;491(7424):444–8.  
608
- 609 36. Team RC. R: A Language and Environment for Statistical Computing. Vienna, Austria: R  
610 Foundation for Statistical Computing; 2015.  
611
- 612 37. Brusatte SL, Benton MJ, Ruta M, Lloyd GT. Superiority, competition, and opportunism in the  
613 evolutionary radiation of dinosaurs. *Science*. 2008;321(5895):1485–8.  
614

- 615 38. Upchurch P, Mannion PD, Benson RBJ, Butler RJ, Carrano MT. Geological and anthropogenic  
616 controls on the sampling of the terrestrial fossil record : a case study from the Dinosauria. In:  
617 McGowan AJ, Smith AB, editors. Comparing the Geological and Fossil Records. 2011. p. 209–40.  
618
- 619 39. Jablonski D, Finarelli JA. Congruence of morphologically-defined genera with molecular  
620 phylogenies. *Proc Natl Acad Sci U S A*. 2009 May 19;106(20):8262–6.  
621
- 622 40. Alroy J. Fair sampling of taxonomic richness and unbiased estimation of origination and  
623 extinction rates. *Quant methods Paleobiol*. 2010;16:55–80.  
624
- 625 41. Benton MJ. Palaeodiversity and formation counts: redundancy or bias? *Palaeontology*.  
626 2015;58(6):1003–29.  
627
- 628 42. Benton MJ, Dunhill AM, Lloyd GT, Marx FG. Assessing the quality of the fossil record: insights  
629 from vertebrates. *Geol Soc London, Spec Publ*. 2011 Dec 13;358(1):63–94.  
630
- 631 43. Liow LH, Reitan T, Harnik PG. Ecological interactions on macroevolutionary time scales: clams  
632 and brachiopods are more than ships that pass in the night. *Ecol Lett*. 2015;18(10):1030–9.  
633
- 634 44. Upchurch P, Barrett PM. Phylogenetic and taxic perspectives on sauropod diversity. *The*  
635 *Sauropods: evolution and paleobiology*. Berkeley: University of California Press; 2005. p. 104–24.  
636
- 637 45. Raup DM, Sepkoski JJ. Periodicity of extinctions in the geologic past. *Proc Natl Acad Sci U S A*.  
638 1984;  
639
- 640 46. Benson RBJ, Druckenmiller PS. Faunal turnover of marine tetrapods during the Jurassic-  
641 Cretaceous transition. *Biol Rev*. 2014;89(1):1–23.  
642
- 643 47. Benson RBJ, Mannion PD, Butler RJ, Upchurch P, Goswami A, Evans SE. Cretaceous tetrapod  
644 fossil record sampling and faunal turnover: Implications for biogeography and the rise of modern  
645 clades. *Palaeogeogr Palaeoclimatol Palaeoecol*. Elsevier B.V.; 2013 Feb;372:88–107.  
646
- 647 48. Benson RBJ, Mannion PD. Multi-variate models are essential for understanding vertebrate  
648 diversification in deep time. *Biol Lett*. 2012;8:127–30.  
649
- 650 49. Mannion PD, Benson RBJ, Butler RJ. Vertebrate palaeobiodiversity patterns and the impact of  
651 sampling bias. *Palaeogeogr Palaeoclimatol Palaeoecol*. 2013;372:1–4.  
652
- 653 50. Brusatte SL, Butler RJ, Barrett PM, Carrano MT, Evans DC, Lloyd GT, et al. The extinction of the  
654 dinosaurs. *Biol Rev*. 2015 Jul 28;90(2):628–42.  
655

657 **Figures legends.**

658

659 Figure 1. Species sampling estimates from TRiPS for all dinosaurs, ornithischian, sauropodomorph and  
660 theropod. Top: Estimated sampling rates ( $\lambda_i$ ) in sampling events per species per million years. Bottom:  
661 Estimated binomial sampling probabilities ( $p_{binom,t}$ ) of species within the plotted time interval. In both  
662 panels, estimates are in red (squares and full line) for all dinosaurs, blue for Ornithischia (circle and  
663 dashed line), green for Sauropodomorpha (triangles with dotted line) and purple for Theropoda (diamonds  
664 with dash-dotted line). Confidence intervals on all rates and probabilities and estimates from 100  
665 replicated observation count matrices (see main text) are reported in the Electronic Supplementary  
666 Information. Abbreviations used: Lad – Ladinian, Car – Carnian, Nor – Norian, Rha – Rhaetian, Het –  
667 Hettangian, Sin – Sinemurian, Pli – Pliensbachian, Toa – Toarcian, Aal – Aalenian, Baj – Bajacian, Bat –  
668 Bathonian, Cal – Callovian, Oxf – Oxfordian, Kim – Kimmeridgian, Tit – Tithonian, Ber – Berriasian,  
669 Val – Valanginian, Hau – Hauterivian, Bar – Barremian, Apt – Aptian, Alb – Albian, Cen – Cenomanian,  
670 Tur – Turonian, Con – Coniacian, San – Santonian, Cam – Campanian, Maa – Maastrichtian.

671

672 Figure 2. Species richness estimates from TRiPS. Black circles connected by full line indicate observed  
673 species counts, triangles connected by dotted line indicate range-through species counts while coloured  
674 line and shading indicate maximum likelihood estimate and 95% confidence intervals for true species  
675 richness estimated using TRiPS. Corresponding sampling estimates can be seen in figure 1. All estimates  
676 with confidence intervals are also in the Electronic Supplementary Information.

677

678 Figure 3: Simulation and estimation results. The grayscale indicates TRiPS's success rate (the fraction of  
679 simulations that had confidence intervals that span the true value) in the simulations, lighter areas indicate  
680 higher success rate. For each combination of duration and sampling rate in the figure above, the full range  
681 of all other parameters are represented (i.e. the success rate is the mean value over sampled speciation and  
682 extinction rates, as well degree of individual lineage variability in sampling rate and number of initial  
683 lineages, see main text). Estimated sampling rates for Dinosauria (red squares), Ornithischia (blue circles),  
684 Sauropodomorpha (green triangles), Theropoda (purple diamonds) are plotted against their corresponding  
685 stage durations.

686

687 Table:

688

<b>Clade</b>	<b>Mesozoic sampling rate</b>	<b>Mesozoic sampling probability</b>	<b>Total Mesozoic species richness</b>
<b>Dinosauria</b>	0.192 (0.146 – 0.262)	0.580 (0.474 – 0.706)	1936 (1543 – 2468)
<b>Ornithischia</b>	0.241 (0.172 – 0.355)	0.708 (0.575 – 0.847)	508 (409 – 668)
<b>Sauropodomorpha</b>	0.156 (0.084 – 0.304)	0.479 (0.278 – 0.753)	513 (307 – 983)
<b>Theropoda</b>	0.153 (0.103 – 0.245)	0.463 (0.335 – 0.628)	1115 (780 – 1653)

689

690 Table 1. Mesozoic mean sampling rates (observations per lineage per Ma), sampling probability and total  
691 richness for each of the four clades, with 95% confidence intervals in parentheses. Note that the mean  
692 rates and probabilities were calculated weighted by the estimated richness (see ESM, table S2). Total  
693 Mesozoic richness was calculated using the Mesozoic mean sampling probability and the total number of  
694 species in each clade. For stage specific sampling estimates see figure 1 and ESM.

695









