

1 *Running head:* TIME SUB-SAMPLES IN DISPARITY-THROUGH-TIME  
2 ANALYSES

3 Time for a rethink: time sub-sampling  
4 methods in disparity-through-time  
5 analyses

6 Thomas Guillerme<sup>1</sup> and Natalie Cooper<sup>2\*</sup>

7 <sup>1</sup>School of Biological Sciences, University of Queensland, St. Lucia, Queensland,  
8 Australia.

9 <sup>2</sup>Department of Life Sciences, Natural History Museum, Cromwell Road, London, SW7  
10 5BD, UK. natalie.cooper@nhm.ac.uk

11 \*Corresponding author

12 *Keywords:* time bin, time-slice, disparity, sub-sampling

13 **ABSTRACT**

14 Disparity-through-time analyses can be used to determine how  
15 morphological diversity changes in response to mass extinctions, or to  
16 investigate the drivers of morphological change. These analyses are  
17 routinely applied to palaeobiological datasets, yet although there is much  
18 discussion about how to best calculate disparity, there has been little  
19 consideration of how taxa should be sub-sampled through time. Standard  
20 practice is to group taxa into discrete time bins, often based on  
21 stratigraphic periods. However, this can introduce biases when bins are of  
22 unequal size, and implicitly assumes a punctuated model of evolution. In  
23 addition, many time bins may have few or no taxa, meaning that disparity  
24 cannot be calculated for the bin and making it harder to complete  
25 downstream analyses. Here we describe a different method to  
26 complement the disparity-through-time tool-kit: time-slicing. This method  
27 uses a time-calibrated phylogenetic tree to sample disparity-through-time  
28 at any fixed point in time rather than binning taxa. It uses all available  
29 data (tips, nodes and branches) to increase the power of the analyses,  
30 specifies the implied model of evolution (punctuated or gradual), and is  
31 implemented in R. We test the time-slicing method on four example  
32 datasets and compare its performance in common disparity-through-time  
33 analyses. We find that the way you time sub-sample your taxa can change  
34 your interpretations of the results of disparity-through-time analyses. We  
35 advise using multiple methods for time sub-sampling taxa, rather than  
36 just time binning, to gain a better understanding disparity-through-time.

37 **INTRODUCTION**

38 Disparity-through-time analyses are common in palaeontology (Gould,  
39 1991; Briggs *et al.*, 1992; Wills *et al.*, 1994; Foote, 1994). They reveal how the  
40 morphological diversity of clades has changed through time, and allow us  
41 to make inferences about the breadth of ecological niches extinct taxa  
42 occupied (Foote, 1997). Results from disparity-through-time studies also  
43 provide insights into the ecological impacts of mass extinctions,  
44 competitive replacements, and the drivers of morphological evolution  
45 (Brusatte *et al.*, 2008b; Foote, 1996; Friedman, 2010). Unfortunately, the way  
46 we perform these analyses may have profound effects on our conclusions.

47 Disparity-through-time analyses have two main analysis  
48 components: calculating disparity, and creating time sub-subsets of the  
49 data. Here we focus on the latter. The nature of disparity (i.e. it is a  
50 diversity metric), means it cannot be calculated using a single individual,  
51 so some way of sub-sampling taxa is required. Changes in  
52 disparity-through-time are generally investigated by calculating the  
53 disparity of taxa present during specific time intervals or time bins (e.g.  
54 Cisneros & Ruta, 2010; Prentice *et al.*, 2011; Hughes *et al.*, 2013; Hopkins,  
55 2013; Benton *et al.*, 2014; Benson & Druckenmiller, 2014). These time bins  
56 are usually defined based on stratigraphy (e.g. Cisneros & Ruta, 2010;  
57 Prentice *et al.*, 2011; Hughes *et al.*, 2013; Benton *et al.*, 2014) but can also be  
58 arbitrarily chosen time bins of equal (or approximately equal) duration  
59 (Butler *et al.*, 2012; Hopkins, 2013; Benson & Druckenmiller, 2014).  
60 However, this approach has several limitations.

61 First, time bins defined by stratigraphy are not of equal size, biasing  
62 higher disparity towards longer stratigraphic periods. This can be dealt

63 with using rarefaction methods, i.e. repeating the analysis while  
64 resampling the taxa to have the same number of taxa in each bin (e.g.  
65 using bootstrapping with limited resampling). This can, however, lead to  
66 large confidence intervals when there are stratigraphic periods with few  
67 species. Other studies split large time bins so they are of roughly equal  
68 size, but this is often an *ad hoc* procedure that can introduce more bias  
69 depending on where bins are split. Second, the time binning approaches  
70 (whether bins are equally sized or not) favour punctuated equilibrium  
71 modes of evolution. Whether the disparity represents an average across  
72 the interval (with no interpretation of if or how it varied within the time  
73 bin), or it is *effectively* postulated to be constant, when analysing the  
74 changes in disparity-through-time, this method will only allow changes in  
75 disparity to occur *between* intervals rather than also allowing for gradual  
76 changes within intervals (a pattern that is fairly common in the fossil  
77 record; Hunt *et al.*, 2015). Third, when investigating changes in disparity  
78 due to events at a specific time point (e.g. a mass extinction), time bins  
79 may have not have high enough resolution to resolve changes at the event,  
80 for example if time bins are every 20 million years it may be hard to  
81 capture the effects of an event five million years into the bin. Finally, time  
82 bin analyses are often limited by the number of taxa in each bin. If there  
83 are insufficient taxa in a time bin, disparity cannot be calculated, so  
84 further analyses, e.g. correlations of disparity with hypothesised drivers  
85 of morphological evolution, are not possible.

86 To address these issues, we propose a “time-slicing” approach that  
87 takes advantage of the wealth of palaeontological datasets which now  
88 have associated phylogenies. Time-slicing uses a phylogenetic tree and  
89 considers subsets of taxa at specific equidistant points in time, as opposed

90 to considering subsets of taxa between two points in time (a similar  
91 approach is outlined in Halliday & Goswami, 2016). This results in  
92 even-sampling across time and permits us to define the underlying model  
93 of character evolution (punctuated or gradual). Time-slicing also includes  
94 any element present in the phylogeny (branches, nodes and tips) at the  
95 time-slice in question as part of the disparity calculation. This allows us to  
96 measure disparity at time points where there are no sampled terminal  
97 taxa, and increases the sample size at each time point, making  
98 downstream analyses of the drivers of disparity much more feasible.

99 Here we present our time-slicing methods using four datasets taken  
100 from the literature. We calculate disparity-through-time for each dataset  
101 using a range of time binning and time-slicing methods, and then  
102 compare these approaches with respect to the relative disparities  
103 calculated, but also investigate how the different approaches influence  
104 biological conclusions. We find that the choice of time sub-sampling  
105 method can have profound effects on the conclusions of  
106 disparity-through-time analyses.

## 107 **MATERIALS AND METHODS**

### 108 *Overview*

109 To test the different time sub-sampling methods, we followed the protocol  
110 below (Fig. 1). All the code needed to reproduce these analyses (along  
111 with detailed instructions) is provided on GitHub  
112 (<https://github.com/nhcooper123/time-slice>).

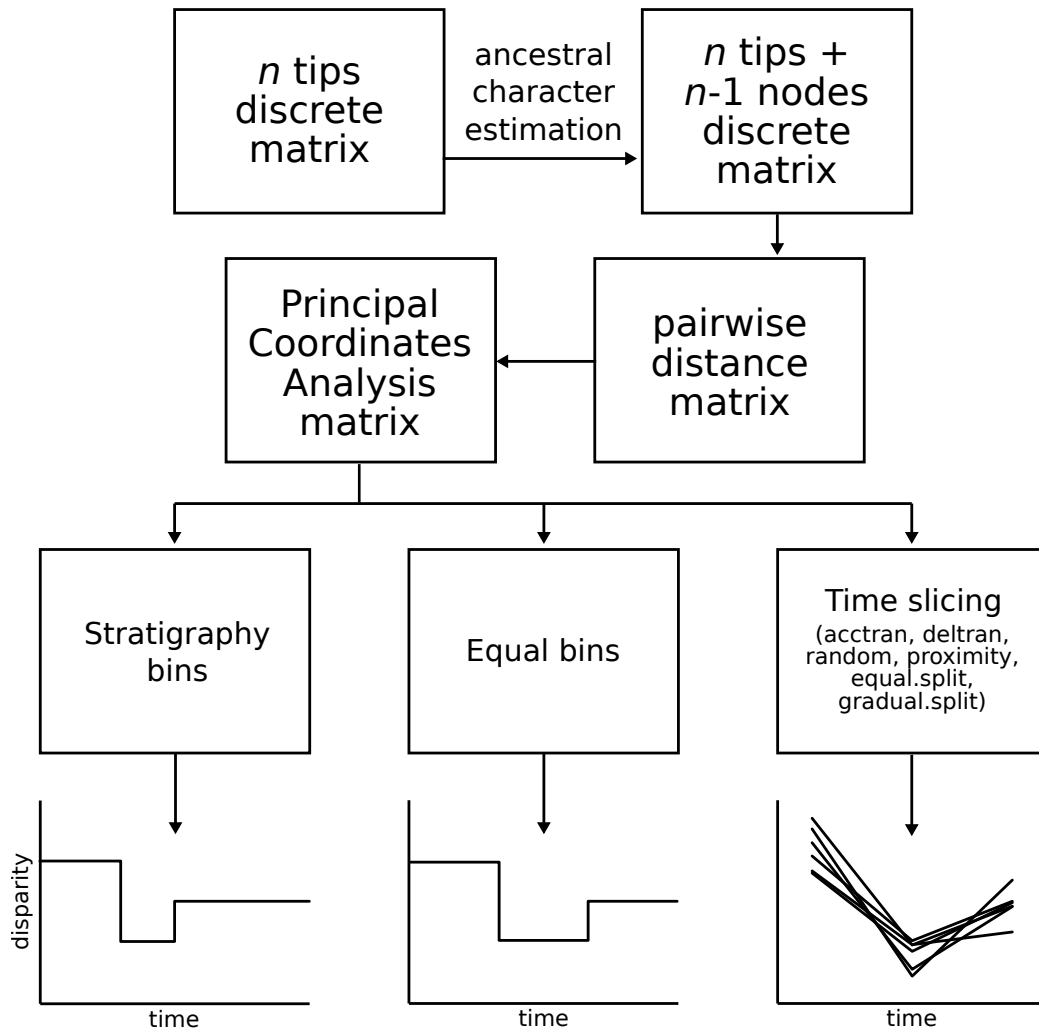


Figure 1: Outline of the disparity-through-time pipeline. 1) We use ancestral character estimation to infer nodal character states; 2) we measure the pairwise Gower distance between the tip character states and nodal character states; 3) we ordinate the distance matrix using principal coordinates analysis (PCoA/PCO); 4) we time sub-sample the PCoA matrix using time bins defined by stratigraphic periods, equally sized time bins and time-slices (using six methods to estimate ordination scores for branches); and finally 5) we measure disparity-through-time for each of these methods.

113 *Example datasets*

114 To test the different time binning/slicing methods we selected four  
115 datasets: a mammal dataset from Beck & Lee (2014), two theropod  
116 datasets from Brusatte *et al.* (2014b) and Bapst *et al.* (2016b), and a crinoid  
117 dataset from Wright (2017b). Table 1 and the Supporting Information  
118 Appendix S1 provide more details. Each dataset consists of first and last  
119 occurrence dates for all taxa, a matrix of morphological characters in  
120 NEXUS format, and a time-scaled phylogeny. These datasets are freely  
121 available with their accompanying papers (Table 1), but for reproducibility  
122 purposes we also provide the data we used on GitHub  
123 (<https://github.com/nhcooper123/time-slice/data>).

Table 1: Details of the datasets used in this study. Age ranges are root time to most recent tip taxon.

	<b>Beck2014</b>	<b>Brusatte2014</b>	<b>Bapst2016</b>	<b>Wright2017</b>
Group	mammals	theropods	theropods	crinoids
# taxa	106	152	89	42
# characters	421	853	374	87
Age range (MYA)	171.8 - 0	168.5 - 66	207.2 - 66	485.4 - 372.2
Mass extinction (MYA)	66 (K-Pg)	NA	NA	443 (O-S)
Reference	Beck & Lee (2014)	Brusatte <i>et al.</i> (2014 <i>b</i> )	Bapst <i>et al.</i> (2016 <i>b</i> )	Wright (2017 <i>b</i> )
Data reference	Beck & Lee (2014)	Brusatte <i>et al.</i> (2014 <i>a</i> )	Bapst <i>et al.</i> (2016 <i>a</i> )	Wright (2017 <i>a</i> )



124 *Preparing the data for disparity-through-time analysis*

125 *Estimating ancestral character states.* For each dataset we estimated the  
126 ancestral character states at each node using the `AncStatesEstMatrix`  
127 function from the `Claddis` R package (Lloyd, 2015; R Core Team, 2015).  
128 This function uses the re-rooting method (Yang *et al.*, 1996; Garland &  
129 Ives, 2000) to get Maximum Likelihood estimates of the ancestral states for  
130 each character at every node in the phylogeny (based on the  
131 `rerootingMethod` function in `phytools`; Revell, 2012). Inapplicable and  
132 missing characters for any taxon were treated as ambiguous characters  
133 (i.e. any possible observed state for the character). To prevent poor  
134 ancestral state estimations from biasing our results, especially when a lot  
135 of error is associated with the estimations, we only included ancestral  
136 state estimations with a scaled Likelihood  $\geq 0.95$ . Ancestral state  
137 estimations with scaled Likelihoods below this threshold were recoded as  
138 missing data (“?”). This allowed our results to be less dependent on the  
139 quality (or the absence thereof) of the ancestral state estimations methods,  
140 especially in parts of the datasets where data were sparse. This approach  
141 is similar to Brusatte *et al.* (2011) but uses model based estimations (rather  
142 than parsimony) allowing us to control for ambiguous (i.e. poorly  
143 estimated) nodes.

144 *Building morphospaces.* To explore disparity-through-time in our datasets,  
145 we used a morphospace approach (e.g. Foote, 1994, 1996; Wesley-Hunt,  
146 2005; Brusatte *et al.*, 2008b; Friedman, 2010; Toljagic & Butler, 2013;  
147 Hughes *et al.*, 2013). Morphospaces can be obtained from any  
148 multidimensional morphological data set but can differ in the data used

149 (e.g. discrete or continuous), and whether they include phylogenetic data  
150 or not. Although empirical morphospaces from discrete or continuous  
151 data have been shown to have similar properties (Foth *et al.*, 2012;  
152 Hetherington *et al.*, 2015), our morphospaces are based on discrete  
153 morphological data (originally collected for phylogenetic analysis; *c.f.*  
154 geometric morphometric data) and include some phylogenetic  
155 information (see above). Mathematically, our morphospaces are  $n$   
156 dimensional objects that summarise the distances between discrete  
157 morphological characters of the taxa present and their ancestors.

158 *Constructing distance matrices.* To estimate the morphospaces for each of our  
159 datasets we first constructed pairwise distance matrices of length  $n$ , where  
160  $n$  is the total number of tips and nodes in the dataset. We calculated the  
161  $n \times n$  distances using the Gower distance (Gower, 1971), i.e. the number of  
162 mismatched characters over the number of shared characters. This allows  
163 us to correct for distances between two taxa that share many characters  
164 and could be closer to each other than to taxa with fewer characters in  
165 common (i.e. because some pairs of taxa share more characters in  
166 common than others, they are more likely to be similar). For discrete  
167 morphological matrices, using this corrected distance is preferable to the  
168 raw Euclidean distance because of its ability to deal with discrete or/and  
169 ordinated characters as well as with missing data (Anderson & Friedman,  
170 2012). However, the Gower distance cannot calculate distances when taxa  
171 have no overlapping data. Therefore, we used the `TrimMorphDistMatrix`  
172 function from the `Claddis` R package to remove pairs of taxa with no  
173 cladistic characters in common. This led to us removing nine taxa from  
174 the Bapst *et al.* (2016b) dataset, and 19 from the Brusatte *et al.* (2014b)

175 dataset, but none from the other two datasets (see Supporting Information  
176 Appendix S1 for details of which species).

177 *Ordination.* After constructing our distance matrices we transformed them  
178 using classical multidimensional scaling (MDS; Torgerson, 1965; Gower,  
179 1966; Cailliez, 1983). This method (also referred to as PCO; e.g. Brusatte  
180 *et al.* 2015; or PCoA; e.g. Paradis *et al.* 2004; but distinguished in Legendre  
181 & Legendre 2012) is an eigen decomposition of the distance matrix.  
182 Because we used Gower distances instead of raw Euclidean distances,  
183 negative eigenvalues can be calculated. To avoid this problem, we first  
184 transformed the distance matrices by applying the Cailliez correction  
185 (Cailliez, 1983) which adds a constant  $c^*$  to the values in a distance matrix  
186 (apart from the diagonal) so that all the Gower distances become  
187 Euclidean ( $d_{Gower} + c^* = d_{Euclidean}$ ; Cailliez 1983). We were then able to  
188 extract  $k$  eigenvectors for each matrix (representing the  $k$  dimensions of  
189 the morphospace) where  $k$  is equal to  $n - 2$ , i.e. the number of taxa in the  
190 matrix ( $n$ ) minus the last two eigenvectors that are always null after  
191 applying the Cailliez correction. Contrary to previous studies (e.g.  
192 Brusatte *et al.*, 2008a; Cisneros & Ruta, 2010; Prentice *et al.*, 2011; Anderson  
193 & Friedman, 2012; Hughes *et al.*, 2013; Benton *et al.*, 2014), we use all  $k$   
194 dimensions of our morphospaces and not a sub-sample representing the  
195 majority of the variance in the distance matrix (e.g. selecting only  $x$   
196 dimensions that represent up to 90% of the variance in the distance matrix;  
197 Brusatte *et al.* 2008b; Toljagic & Butler 2013). Note that our morphospaces  
198 represent an ordination of all possible morphologies coded in each study  
199 through time. It is unlikely that all morphologies will co-occur at each  
200 time point, therefore, the disparity of the whole morphospace is expected

201 to be greater than the disparity at any specific point in time.

202 *Disparity-through-time analyses*

203 Disparity-through-time analyses were performed using the `disPRity` R  
204 package (Guillerme, 2016).

205 *Calculating disparity.* Disparity can be calculated in many different ways  
206 (e.g. Wills *et al.*, 1994; Ciampaglio, 2004; Thorne *et al.*, 2011; Hopkins, 2013;  
207 Huang *et al.*, 2015), however a majority of studies in palaeobiology  
208 estimate disparity using four metrics: the sum and products of ranges and  
209 variances, each of which gives a slightly different estimate of how the data  
210 fits within the morphospace (Foote, 1994; Wills *et al.*, 1994; Brusatte *et al.*,  
211 2008*a,b*; Cisneros & Ruta, 2010; Thorne *et al.*, 2011; Prentice *et al.*, 2011;  
212 Brusatte *et al.*, 2012; Toljagic & Butler, 2013; Ruta *et al.*, 2013; Benton *et al.*,  
213 2014; Benson & Druckenmiller, 2014). However, these metrics have  
214 limitations. First, the range metrics are affected by the uneven sampling of  
215 the fossil record (Butler *et al.*, 2012). Second, because we include all  $k$   
216 dimensions in the analysis (see above), the products of ranges and  
217 variances will tend towards zero since the scores of the last dimension are  
218 usually really close to zero themselves. We therefore use the sum of  
219 variances metric to estimate disparity here:

$$\text{disparity} = \sum \sigma^2 k_i \quad (1)$$

220 where  $\sigma^2 k_i$  is the variance for the  $k_i^{\text{th}}$  dimension ranging from  $n$  to  $n - 2$   
221 with  $n$  being the number of taxa in the dataset. Note that there are still  
222 statistical issues with this metric (such as the co-variance between

223 dimensions not being measured), but for the purposes of comparison with  
224 previous work we decided to use a standard metric for these analyses.

225 *Time sub-sampling* To estimate disparity-through-time we first need to split  
226 the data into time sub-samples. Here we use three time sub-sampling  
227 methods.

- 228 1. Stratigraphic time bins. This is the traditional method, where all the  
229 taxa within each stratigraphic period are included in the disparity  
230 calculation. This often leads to bins of unequal duration. Here we  
231 use stratigraphic stages and epochs.
- 232 2. Equally sized time bins. This is another commonly used method,  
233 where the time frame of interest is split into equally sized time bins,  
234 then all the taxa within each time bin are included in the disparity  
235 calculation.
- 236 3. Time-slicing. We describe this in more detail below, but in brief,  
237 time-slicing uses a phylogeny, and rather than binning the data, it  
238 takes slices through a phylogeny and includes all the taxa and nodes  
239 in that slice within the disparity calculation.

240 *Time-slicing*. The “time-slicing” approach considers subsets of taxa in the  
241 morphospace at specific equidistant points in time, as opposed to  
242 considering subsets of taxa between two points in time. This results in  
243 even-sampling of the morphospace across time and allows us to use  
244 different underlying models of character evolution (punctuated or  
245 gradual).

246 In practice, time-slicing considers the disparity of any element  
247 present in the phylogeny (branches, nodes and tips) at any point in time.  
248 When the phylogenetic elements are nodes or tips, the ordination scores  
249 for the nodes (estimated using ancestral state reconstruction as described  
250 above) or tips are directly used for calculating disparity. When the  
251 phylogenetic elements are branches we choose the ordination score for the  
252 branch using one of two evolutionary models:

253 1. **Punctuated evolution.** This model selects the ordination score from  
254 either the ancestral node or the descendant node/tip of the branch  
255 regardless of the position of the slice along the branch. Similarly to  
256 the time bin approach, this reflects a model of punctuated evolution  
257 where changes in disparity occur either at the start or at the end of a  
258 branch over a relatively short time period, and clades undergo long  
259 periods of stasis during their evolution (Gould & Eldredge, 1977;  
260 Hunt, 2007). We apply this model in four ways:

261 (i) The “acctrans” model, always selecting the ordination score of  
262 the descendant node/tip of the branch.

263 (ii) The “deltrans” model, always selecting the ordination score of  
264 the ancestral node of the branch.

265 (iii) The “random” model, randomly selecting the ordination score  
266 of either the ancestor or the descendant of the branch.

267 (iv) The “proximity” model, selecting the ordination score of the  
268 ancestor if the slice occurs in the first half of the branch, and the  
269 descendant if the slice occurs in the second half of the branch.

270 The two first models assume that changes always occur early

271 (*accelerated transition*) or late along the branches (*delayed transition*).  
272 The third model makes neither assumption and simply selects data  
273 from the ancestor or the descendant at random, and the fourth bases  
274 the selection of either the ancestor or the descendant on where the  
275 slice occurs along the branch. These punctuated models only select  
276 either the ordination score from the ancestor and the descendant  
277 once in the whole disparity analysis. For example, if using the  
278 “random” model, if the data of the ancestor has been randomly  
279 chosen, only this data will be used during the bootstrapping (see  
280 below) and for the disparity calculation.

281 2. **Gradual evolution.** Unlike the punctuated models, the following  
282 models do not select the ordination score of either the ancestor or  
283 the descendant but associate a probability to both. This reflects a  
284 model of gradual evolution where changes in disparity are gradual  
285 and cumulative along the branch.

286 (v) The “equal splits” model (probabilistic), selects the ordination  
287 score from both the ancestor and the descendant with an equal  
288 probability:

$$p(\text{ancestor}) = p(\text{descendant}) = 0.5 \quad (2)$$

289 (vi) The “gradual splits” model (probabilistic), selects the  
290 ordination score from both the ancestor and the descendant  
291 with a probability function of the distance between the

292

nodes/tip at the ends of the branch and the slice:

$$p(\text{ancestor}) = \frac{d(\text{ancestor}, \text{slice})}{d(\text{ancestor}, \text{descendant})} \quad (3)$$

293

$$p(\text{descendant}) = 1 - p(\text{ancestor}) \quad (4)$$

294

where  $d(x, y)$  is the distance between the two elements  $x, y$

295

(ancestor, slice or descendant) measured in units of branch

296

length.

297

In these models, the ordination scores of both the ancestor and

298

descendant contribute to the disparity calculation. For example,

299

using the “gradual splits” model, if the slice occurs in the third

300

quarter of a branch joining node A to node/tip B (75% of the total

301

branch length), after bootstrapping, the disparity results will be

302

based on 25% of the data from A and 75% of the data from B.

303

Because of the probabilistic nature of these models, they are only

304

meaningful when calculating disparity from bootstrapped datasets.

305

It is important to note that the time-slicing method is not an

306

ancestral states estimation method *per se*. This method does not estimate

307

values along a branch applying a model (*c.f.* methods described for

308

ancestral character estimation in the “Preparing the data for

309

disparity-through-time analysis” section above) but rather chooses

310

between the two available pieces of information (the ordination score of

311

the descendant or the ancestor) using the methods described above. This

312

allows the method to be used in post-ordination analysis where the data

313

used in each time-slice is data already present in the morphospace. In



314 other words, this method does not require a re-ordination of the  
315 morphospace every time a slice goes through a branch, thus allowing the  
316 properties of the morphospace (e.g. distance between species, variance of  
317 each axis, etc.) to remain constant. For example, using the “equal.splits”  
318 model on an ancestor and a descendant with PCO<sub>1</sub> values of respectively  
319 0.04 and 0.03, after a sufficient number of bootstrap replicates (e.g. 100)  
320 the value along the branch will be close to  $0.5 \times 0.04 + 0.5 \times 0.03 = 0.035$ .  
321 By estimating this value rather than generating it (i.e. creating a new  
322 element mid-way along the branch that would be the average of the  
323 descendant and ancestor - 0.035) we obtain the same results without  
324 modifying the morphospace properties.

#### 325 *Comparing time sub-sampling methods*

326 To compare the time binning and time-slicing approaches we applied the  
327 methods as follows (see Fig. ??).

328 1. Stratigraphy: time sub-samples defined by stratigraphic periods (Fig.  
329 2).

330 (i) Time bins (unequal). We calculated disparity for the taxa in  
331 each stratigraphic period (stage or epoch). To reduce the  
332 influence of outliers on our disparity estimates, we  
333 bootstrapped each disparity measurement for each time bin by  
334 randomly resampling with replacement a new sub-sample of  
335 taxa from the observed taxa in the bin 100 times. We then  
336 calculated the median disparity value for each time bin along  
337 with the 50% and 95% confidence intervals.

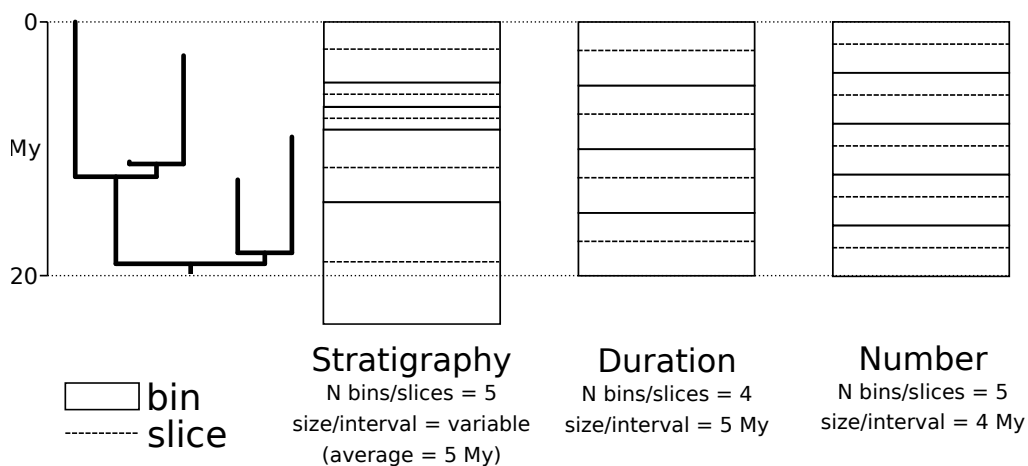


Figure 2: Outline of the three time sub-sampling methods. Stratigraphy: time sub-samples are defined by stratigraphic periods. Here there are five stratigraphic periods in the 20 My (million years) time frame of interest, i.e. five bins/slices with variable sizes/intervals. Duration: time sub-samples are defined based on the mean duration of stratigraphic periods in the time frame of interest. Here, the mean duration of stratigraphic periods is 5 My, so there are four bins/slices of 5 My duration (or four slices with 5 My intervals between them) in the 20 My time frame of interest. Number: time sub-samples are defined based on the number of stratigraphic periods in the time frame of interest. Here, there are five stratigraphic periods, so there are five bins/slices of 4 My duration (or five slices with 4 My intervals between them) in the 20 My time frame of interest.

338 (ii) Time-slices (non-equidistant). We calculated disparity using  
339 our time-slicing approach with slices occurring at the midpoint  
340 of each stratigraphic period (stage or epoch), and using all six  
341 time-slicing methods (acctrans, deltrans, random, proximity,  
342 equal splits and gradual splits). To reduce the influence of  
343 outliers on our disparity estimates, we bootstrapped each  
344 disparity measurement as described above for the stratigraphic  
345 time bins.

346 2. Duration: time sub-samples defined by the duration of stratigraphic  
347 periods (Fig. 2).

348 (i) Time bins (equal). We calculated disparity for the taxa in each  
349 time bin where time bin size was defined by the mean *duration*  
350 of the stratigraphic period (stage or epoch), and bootstrapped  
351 the disparity values as described above.

352 (ii) Time-slices (equidistant). We calculated disparity using our  
353 time-slicing approach where the interval between slices, was  
354 defined by the mean *duration* of the stratigraphic period (stage  
355 or epoch). We used the six time-slicing methods and  
356 bootstrapped as described above.

357 3. Number: time sub-samples defined by the number of stratigraphic  
358 periods (Fig. 2).

359 (i) Time bins (equal). We calculated disparity for the taxa in each  
360 time bin where the number of time bins was defined by the  
361 *number* of stratigraphic periods (ages or epochs) in the time  
362 frame of interest, and bootstrapped the disparity values as

363 described above.

364 (ii) Time-slices (equidistant). We calculated disparity using our  
365 time-slicing approach where the number of slices, was defined  
366 by the *number* of stratigraphic periods (ages or epochs) in the  
367 time frame of interest. We used the six time-slicing methods  
368 and bootstrapped as described above.

369 We also recorded the number of taxa (or taxa and nodes for  
370 time-slicing methods) in each sub-sample as a proxy for taxonomic  
371 diversity.

372 *Testing for differences in the time sub-sampling methods*

373 Testing for statistical differences among the time sub-sampling methods  
374 described above is difficult, as we need to compare similar units, and also  
375 to tackle questions important to the interpretation of  
376 disparity-through-time analyses. We therefore present three different,  
377 simple ways of comparing the time sub-sampling methods as follows.

378 *Systematic differences in disparity-through-time.* To test whether using time bins  
379 or time-slices resulted in significantly different disparity values at  
380 common time points, we used paired Wilcoxon tests to compare the  
381 median bootstrapped disparities obtained in the stratigraphy (time  
382 sub-samples defined by stratigraphic periods), duration (time sub-samples  
383 defined by the duration of stratigraphic periods), and number (time  
384 sub-samples defined by the number of stratigraphic periods) analyses  
385 described above.

386 Due to the uneven spread of taxa across phylogenies, some time bins

387 will contain one or no species, meaning that we cannot estimate disparity  
388 for that time bin. We first, therefore, removed the time bins, and  
389 corresponding time-slices, without disparity estimates. We then  
390 performed paired Wilcoxon tests with Bonferroni corrected p-values, so  
391 that bins and slices for the same time period are being compared.  
392 Significant results suggest that there is a systematic difference in disparity  
393 values at each time point, depending on whether bins or slices are used.

394 *Disparity peaks.* We are perhaps more interested in how the conclusions of  
395 disparity-through-time analyses are influenced by the choice of time  
396 sub-sampling method, rather than the disparities estimated by each  
397 method *per se*, especially as these will be influenced by the number of taxa  
398 (and/or nodes) included in each sub-sample. Therefore, we also  
399 investigated where peaks of disparity occurred in each of our datasets for  
400 the different time sub-sampling methods. We calculated the maximum  
401 bootstrapped disparities for each dataset and for each time sub-sampling  
402 method, along with their associated confidence intervals. Significant shifts  
403 in disparity peaks suggest that the choice of time sub-sampling method  
404 will influence our conclusions about relative changes in the disparity of  
405 our groups through time.

406 *Effects of mass extinction events.* Many analyses of disparity-through-time  
407 aim to demonstrate differences in disparity before and after mass  
408 extinction events. Two of our four datasets contain taxa before and  
409 immediately after a mass extinction (Cretaceous-Paleogene 66 MYA; Beck  
410 & Lee 2014; Ordovician-Silurian 455-430 MYA; Wright 2017b), so we used  
411 Wilcoxon tests with Bonferroni corrected p-values to compare disparity in

412 the time bin/slice prior to the appropriate mass extinction, to that of the  
413 time bin/slice following the extinction event. Significant results suggest  
414 an effect of the mass extinction on disparity in the group. We then  
415 compare these results across the time sub-sampling methods to determine  
416 if our conclusions change depending on the method used. We repeated  
417 these analyses using the two time bins/slices after the one immediately  
418 following the mass extinction event to account for any lag effects of the  
419 mass extinction on disparity.

## 420 **RESULTS**

### 421 *Disparity-through-time analyses*

422 Disparity changes through time for each of our four datasets (Fig. 3,  
423 Supporting Information Appendix S2: Figs A1-A2). Relative disparities  
424 tend to be lower with time binning methods, likely because these contain  
425 fewer taxa than time-slicing methods. The six different time-slicing  
426 methods (acctran, deltran, random, proximity, equal splits and gradual  
427 splits) show similar patterns, so we focus only on the results for one  
428 method with a punctuated model of evolution (specifically the 'proximity'  
429 method), and one method with a gradual model of evolution (specifically  
430 the 'gradual splits' method). Results for all six methods can be found in  
431 Supporting Information Appendix S2: Figures A1-A2.

### 432 *Testing for differences in the time sub-sampling methods*

433 *Systematic differences in disparity-through-time.* There is no overall significant  
434 systematic difference among the disparities calculated using time bins and  
435 those calculated using the time-slicing methods (Table 2, Supporting

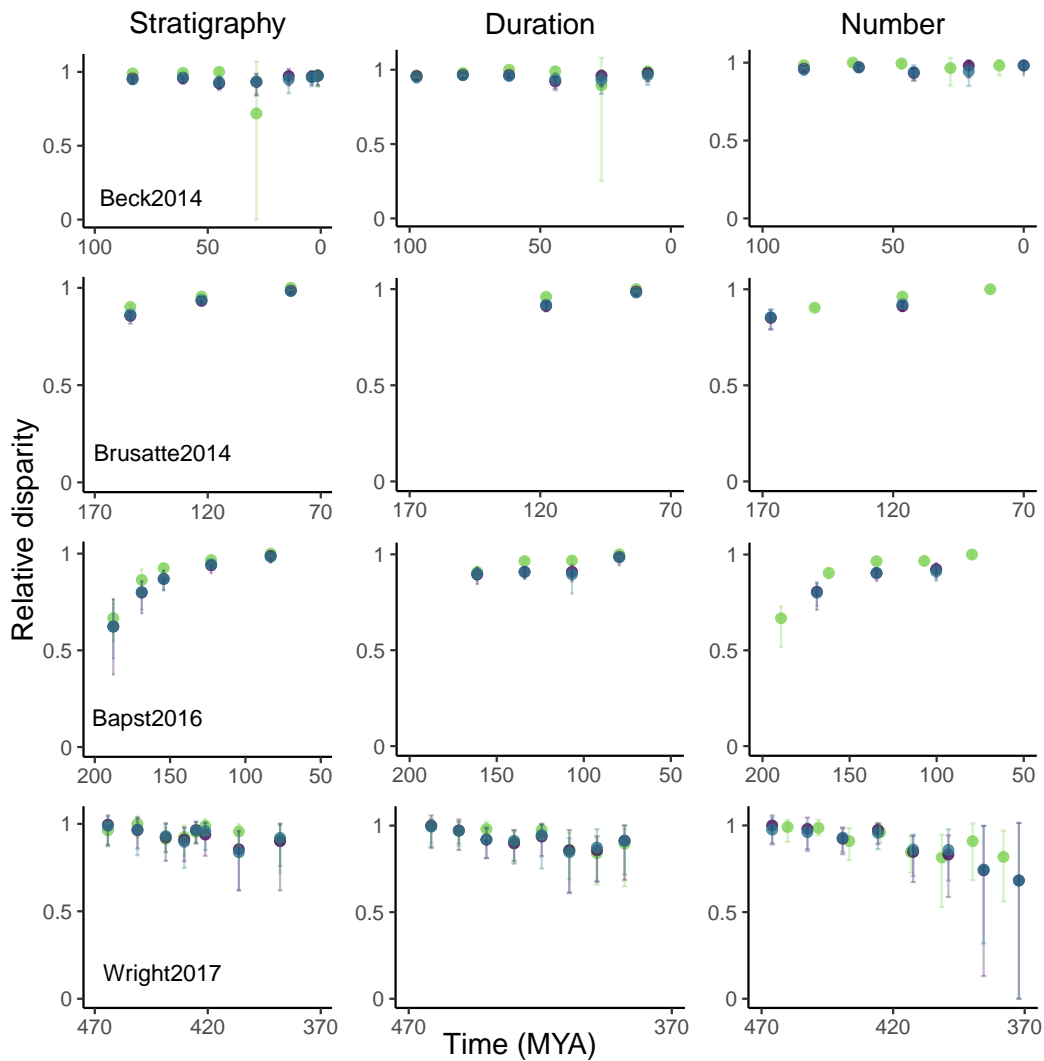


Figure 3: Relative disparity-through-time. Median bootstrapped disparities were calculated using time binning and time-slicing approaches. Green points represent time binning methods, purple points are time-slices with a punctuated model of evolution (‘proximity’ method), and blue points are time-slices with a gradual model of evolution (‘gradual splits’ method). Relative disparities (median bootstrapped disparity divided by the maximum median bootstrapped disparity for a dataset and analysis method) are presented so they can be compared across datasets/methods. Stratigraphy uses unequal time bins or non-equidistant time-slices, where the width of the bin, or the interval between slices, is equivalent to stratigraphic epochs. Duration uses equal time bins or equidistant time-slices, where the width of the bin, or the interval between slices, is the mean duration of stratigraphic epochs in the time frame of the dataset. Number uses equal time bins or equidistant time-slices, where the number of bins, or the number of slices, is the mean number of stratigraphic epochs in the time frame of the dataset. In all cases, time bin disparities are plotted at the midpoint of the bin, and error bars represent the 95% confidence intervals around the bootstrapped median disparity. The four dataset names are on the first plot for each dataset (see Table 1 for details). Results for stratigraphic stages, and for other time-slicing methods, are in the Supporting Information Appendix S2: Figures A1-A2.

436 Information Appendix S2: Table A1). Instead, the differences depend on  
437 the dataset and method in question. For example, the Brusatte *et al.*  
438 (2014b), Bapst *et al.* (2016b) and Wright (2017b) datasets, show significant  
439 differences when using bins versus time-slices defined by stratigraphy, but  
440 the Beck & Lee (2014) dataset appears robust to these different  
441 approaches. Likewise, the Beck & Lee (2014), Brusatte *et al.* (2014b) and  
442 Bapst *et al.* (2016b) datasets have different disparities when the number of  
443 bins or slices is the mean number of stratigraphic periods, but this is not  
444 seen in the Wright (2017b) dataset. Note that for epochs, we find fewer  
445 significant differences simply because the smaller numbers of bins and  
446 slices being compared means we have low power to detect a significant  
447 difference.

448 *Disparity peaks.* In the Beck & Lee (2014) and Bapst *et al.* (2016b) datasets,  
449 disparity peaks occur much at much older ages when time-slicing rather  
450 than time binning approaches are used (Fig. 4; Supporting Information  
451 Appendix S2: Figs A3-A4). This is also true for stratigraphic time bins in  
452 the Wright (2017b) dataset, although when using equal time bins the  
453 peaks are later than the time-slicing methods, or very similar (Fig. 4;  
454 Supporting Information Appendix S2: Figs A3-A4). Across the three time  
455 binning methods, the Brusatte *et al.* (2014b) dataset has similar disparity  
456 peaks whichever method is used, the Wright (2017b) dataset only had  
457 variation in peaks when using unequal time bins (stratigraphy), whereas  
458 in the Bapst *et al.* (2016b) and Beck & Lee (2014) datasets, stratigraphic  
459 (unequal) versus equally sized time bins make a large difference to where  
460 the disparity peak occurs (Fig. 4; Supporting Information Appendix S2:  
461 Figs A3-A4). Additionally, there seem to be small discrepancies within the



Table 2: Results of paired Wilcoxon tests investigating whether disparities calculated using time bins are significantly different to those calculated using time-slices. Time-slices used either a punctuated (proximity method) or gradual (gradual.split method) model of evolution. Stratigraphy uses unequal time bins or non-equidistant time-slices, where the width of the bin, or the interval between slices, is equivalent to stratigraphic ages or epochs. Duration uses equal time bins or equidistant time-slices, where the width of the bin, or the interval between slices, is the average duration of stratigraphic ages or epochs in the time frame of the dataset. Number uses equal time bins or equidistant time-slices, where the number of bins, or the number of slices, is the average number of stratigraphic ages or epochs in the time frame of the dataset. P-values were Bonferroni corrected.  $***p < 0.001$ . Results for other time-slicing methods are in the Supporting Information Appendix S2: Table A1.

Dataset	Period	Method	Stratigraphy	Duration	Number
Beck2014	Age	gradual.split	111	115***	65***
Beck2014	Age	proximity	105	83	68***
Beck2014	Epoch	gradual.split	21	39	43***
Beck2014	Epoch	proximity	21	36	32
Brusatte2014	Age	gradual.split	28***	61***	52***
Brusatte2014	Age	proximity	27***	31	28***
Brusatte2014	Epoch	gradual.split	3	6	6
Brusatte2014	Epoch	proximity	0	5***	5
Bapst2016	Age	gradual.split	93	153	165
Bapst2016	Age	proximity	57***	47	75***
Bapst2016	Epoch	gradual.split	4	6	12
Bapst2016	Epoch	proximity	2	0***	8
Wright2017	Age	gradual.split	152***	155	116
Wright2017	Age	proximity	160***	175***	101
Wright2017	Epoch	gradual.split	28	29	21
Wright2017	Epoch	proximity	23	28	18

462 time-slicing methods (gradual.split *vs.* proximity) except in the Beck &  
463 Lee (2014) dataset where the gradual split model recovered disparity  
464 peaks at younger ages than the proximity model (Fig. 4; Supporting  
465 Information Appendix S2: Figs A3-A4)

466 *Effects of mass extinction events.* Mass extinction events influence disparity in  
467 both the Beck & Lee (2014) and Wright (2017*b*) datasets (Fig. 5). However,  
468 whether this change in disparity is significant or not depends on the  
469 method used to create time sub-samples (Fig. 5), and whether stages or  
470 epochs are used. In general, for the Beck & Lee (2014) dataset, time  
471 binning tended to give more significant results than time-slicing methods,  
472 but this was not the case for the Wright (2017*b*) dataset.

## 473 **DISCUSSION**

474 Disparity-through-time analyses are influenced by the choice of time  
475 sub-sampling method used to divide the taxa. While differences in the  
476 relative disparities calculated among time sub-sampling methods may not  
477 be of much biological importance, these changes can have important  
478 implications for the conclusions of downstream analyses. For example,  
479 using stratigraphic epochs as our reference time period, there are 21  
480 potential methods for time sub-sampling our data (splitting by  
481 stratigraphy, number and duration, see methods, and using time bins or  
482 one of six time-slicing methods). Of these 21 methods, in 16 (76%) we  
483 show that placental mammals (Beck & Lee, 2014) significantly increased in  
484 disparity in the time bin/slice immediately after the K-Pg mass extinction  
485 event, and in 20 (95%) we show that crinoids (Wright, 2017*b*) significantly  
486 decreased in disparity in the time bin/slice immediately after the O-S

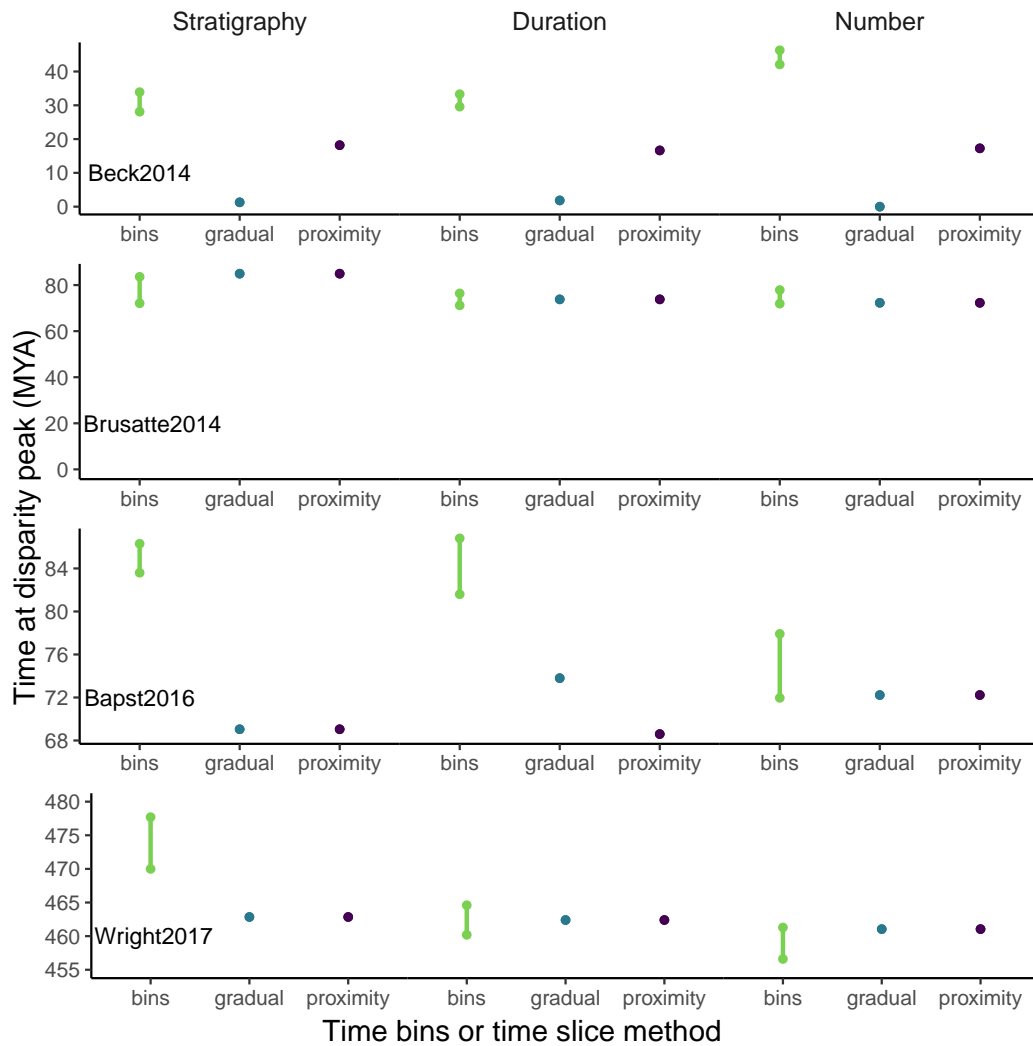


Figure 4: Timing of peak disparity. Median bootstrapped disparities were calculated using time binning and time-slicing approaches. Green points represent time binning methods, blue points are time-slices with a punctuated model of evolution ('proximity' method), and purple points are time-slices with a gradual model of evolution ('gradual splits' method). Stratigraphy uses unequal time bins or non-equidistant time-slices, where the width of the bin, or the interval between slices, is equivalent to stratigraphic epochs. Duration uses equal time bins or equidistant time-slices, where the width of the bin, or the interval between slices, is the mean duration of stratigraphic epochs in the time frame of the dataset. Number uses equal time bins or equidistant time-slices, where the number of bins, or the number of slices, is the mean number of stratigraphic epochs in the time frame of the dataset. For time bins the points indicate the maximum and minimum ages of the time bin within which peak disparities appeared. The four dataset names are on the first plot for each dataset (see Table 1 for details). Results for stratigraphic stages, and for other time-slicing methods, are in the Supporting Information Appendix S2: Figures A3-A4.

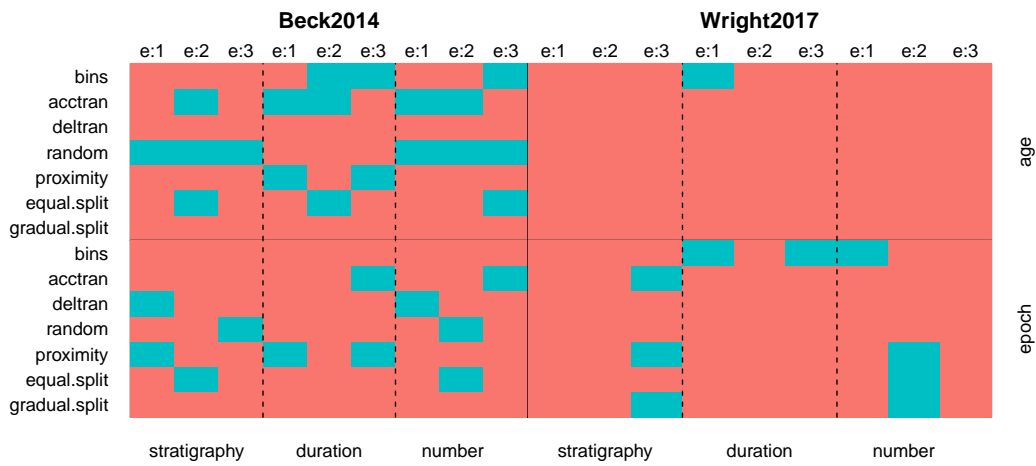


Figure 5: Effects of mass extinction events on disparity. Pink cells and blue cells indicate respectively a significant or non significant change in disparity before and after the mass extinction event (Cretaceous-Paleogene 66 MYA; Beck & Lee 2014; Ordovician-Silurian 455-430 MYA; Wright 2017b).  $e:1$ ,  $e:2$ , and  $e:3$  denote whether the comparison was between the time bin or time-slice immediately after the mass extinction ( $e:1$ ), or the second ( $e:2$ ) or third ( $e:3$ ) bin/slice after the mass extinction to account for any lag effect. The top seven rows use stratigraphic stages and the bottom seven rows use stratigraphic epochs. Labels on the left hand side indicate whether time bins (“bins”) were used, or which of the six time-slicing methods was used.

487 mass extinction event. Given the high congruence (76% and 95%) of these  
488 results, one could argue that time-sub-sampling methods are not  
489 important. However, if we had chosen to investigate crinoid disparity only  
490 using time bins and splitting these so the number of time bins was equal  
491 to the number of epochs (number), we would have concluded that the O-S  
492 extinction had no effect on crinoid disparity. Likewise, the timing of peak  
493 disparity differs among methods. This is particularly evident when  
494 comparing stratigraphic time bins to time slicing methods, where for most  
495 of our datasets we see a much later time to peak disparity. This could  
496 have major implications for our understanding of how morphological  
497 diversity changes through time, for example in response to climate. These  
498 results highlight the sensitivity of disparity-through-time analyses to the  
499 choice of time sub-sampling method. Fortunately this issue is easy to  
500 solve; either disparity-through-time analyses should use, and report  
501 results from, multiple time sub-sampling methods (as demonstrated here),  
502 or great care should be taken in determining the appropriate time  
503 sub-samples to answer the question of interest.

504 Time-slicing has several advantages over time binning (using either  
505 equally or unequally sized bins) approaches. First, it allows us to use as  
506 much of the information available to us, in the form of phylogenetic  
507 relationships and ancestral taxa, as possible. This increases our ability to  
508 investigate key biological questions, such as how do various drivers  
509 influence morphological diversity through time, and how do mass  
510 extinctions influence disparity (Brusatte *et al.*, 2008b; Foote, 1996;  
511 Friedman, 2010), both by increasing the statistical power of analyses and  
512 through the availability of data at key time points in the history of our  
513 groups. Second, we are able to be more explicit about the mode of

514 evolution in our clades; in time-slicing we can apply punctuated or  
515 gradual models of trait change rather than making an assumption of  
516 punctuated evolution. This may be important, as gradual change is a  
517 common pattern of trait evolution in the fossil record (Hunt, 2007).

518       Of course the method also has limitations. The main one of these is a  
519 practical one; it requires a time-calibrated phylogeny and these are not  
520 available for all palaeontological datasets. Furthermore, like most  
521 phylogeny based methods, time-slicing depends on ancestral state  
522 estimations. Care should be taken in interpreting these, as they are highly  
523 dependent on the data and models used for the estimations (Slater *et al.*,  
524 2012; Ekman *et al.*, 2008). The difference between the time-binning and  
525 time-slicing results could also simply be due to the nature of the fossil  
526 record. Rates of sedimentation vary in time and space influencing the  
527 groups found within the rock record and their temporal distribution. In  
528 this case, different beds could represent different “packages” of fauna  
529 through time separated by gaps, resulting in natural “bins” rather than  
530 slices of the data. Slicing through such strata will yield similar results no  
531 matter where in time the slice occurs. It is important to note however, that  
532 the time slicing method also includes ancestral estimations (either through  
533 the nodes or the branches) that are by definition not available in the fossil  
534 record and thus are not influenced by its nature. Additionally, this effect is  
535 likely to be most obvious in groups where the fossil record is “patchy”,  
536 e.g. vertebrates, but less problematic for groups with a more continuous  
537 record like Foraminifera. Finally, Hunt *et al.* (2015) found that time series  
538 are best characterized by gradual directional changes (biased random  
539 walks). In fact, homogeneous directional changes are more likely to be  
540 supported than heterogeneous ones (e.g. punctuated changes) in longer

541 duration series with few samples in each series. In our implementation of  
542 time-slicing, the models are not selected based on any model fit criterion  
543 (e.g. AIC) but merely on researcher assumptions. We thus suggest that  
544 both types of models (punctuated and gradual) are tested during analysis,  
545 unless there is strong independent support for one or the other.

## 546 **CONCLUSIONS**

547 The choice of time sub-sampling method can alter the conclusions we  
548 obtain from disparity-through-time studies. Time-slicing methods, with  
549 explicit models of evolution, provide an alternative to traditional time  
550 binning approaches. Note that while we introduce the time-slicing  
551 methods here, and describe their advantages, we do not suggest  
552 time-slicing is necessarily the “best” method for time sub-sampling in all  
553 cases. As with all methods, the choice of methodology should be  
554 appropriate for the question and data at hand. However, we do strongly  
555 recommend performing disparity-through-time analyses using a series of  
556 appropriate time sub-sampling methods, and comparing these, to ensure  
557 that results are not merely a consequence of the time sub-sampling  
558 method employed.

## 559 **DATA ARCHIVING STATEMENT**

560 Data for this study are available from Bapst *et al.* (2016a); Wright (2017a);  
561 Brusatte *et al.* (2014a); Beck & Lee (2014) but for ease of reproducibility can  
562 also be found on GitHub  
563 (<https://github.com/nhcooper123/time-slice/data>). Supporting  
564 information can be found on Dryad at

565 <https://doi.org/10.5061/dryad.vp4q518> (Guillerme & Cooper, 2018).

## 566 SUPPORTING INFORMATION

567 Additional Supporting information can be found on Dryad at:

568 <https://doi.org/10.5061/dryad.vp4q518> (Guillerme & Cooper, 2018).

569 Appendix S1: Extra details of datasets.

570 Appendix S2: Additional figures.

571 Appendix S3: Additional tables.

572 *Acknowledgments.* NC thanks Mark Sutton and Philip Mannion for the  
573 invitation to contribute to the ‘Evolutionary Modelling’ symposium at The  
574 Palaeontological Association Annual Meeting 2017. TG acknowledges  
575 support from the Australian Discovery Project Grant number DP170103227  
576 awarded to Vera Weisbecker. We thank Dave Bapst, Graeme Lloyd, April  
577 Wright and David Wright for assistance in gathering data for the analyses  
578 and/or discussions about the approach; and Steve Brusatte, Sally Thomas  
579 and one anonymous reviewer for helpful comments on the manuscript.

## 580 REFERENCES

### 581 References

582 ANDERSON, P. S. & FRIEDMAN, M. 2012. Using cladistic characters to  
583 predict functional variety: experiments using early Gnathostomes.  
584 *Journal of Vertebrate Paleontology*, **32**, 1254–1270.

585 BAPST, D., WRIGHT, A., MATZKE, N. & LLOYD, G. 2016a. Data from:  
586 Topology, divergence dates, and macroevolutionary inferences vary



587 between different tip-dating approaches applied to fossil theropods  
588 (dinosauria).

589 ———. 2016b. Topology, divergence dates, and macroevolutionary  
590 inferences vary between different tip-dating approaches applied to  
591 fossil theropods (dinosauria). *Biology letters*, **12**, 20160237.

592 BECK, R. M. & LEE, M. S. 2014. Ancient dates or accelerated rates?  
593 Morphological clocks and the antiquity of placental mammals.  
594 *Proceedings of the Royal Society B: Biological Sciences*, **281**, 1–10.

595 BENSON, R. B. J. & DRUCKENMILLER, P. S. 2014. Faunal turnover of marine  
596 tetrapods during the Jurassic-Cretaceous transition. *Biological Reviews*,  
597 **89**, 1–23.

598 BENTON, M. J., FORTH, J. & LANGER, M. C. 2014. Models for the rise of the  
599 dinosaurs. *Current biology*, **24**, R87–R95.

600 BRIGGS, D. E., FORTEY, R. A. & WILLS, M. A. 1992. Morphological disparity  
601 in the cambrian. *Science*, **256**, 1670–1673.

602 BRUSATTE, S., BUTLER, R. J., PRIETO-MÁRQUEZ, A. & NORELL, M. A. 2012.  
603 Dinosaur morphological diversity and the end-Cretaceous extinction.  
604 *Nature Communications*, **3**, 804–804.

605 BRUSATTE, S., LLOYD, G., WANG, S. & NORELL, M. 2014a. Data from:  
606 Gradual assembly of avian body plan culminated in rapid rates of  
607 evolution across dinosaur-bird transition.

608 BRUSATTE, S. L., BENTON, M. J., RUTA, M. & LLOYD, G. T. 2008a. The first 50  
609 Myr of dinosaur evolution: macroevolutionary pattern and  
610 morphological disparity. *Biology Letters*, **4**, 733–736.

- 611 ———. 2008*b*. Superiority, competition, and opportunism in the  
612 evolutionary radiation of dinosaurs. *Science*, **321**, 1485–1488.
- 613 BRUSATTE, S. L., BUTLER, R. J., BARRETT, P. M., CARRANO, M. T., EVANS,  
614 D. C., LLOYD, G. T., MANNION, P. D., NORELL, M. A., PEPPE, D. J.,  
615 UPCHURCH, P. & WILLIAMSON, T. E. 2015. The extinction of the  
616 dinosaurs. *Biological Reviews*, **90**, 628–642.
- 617 BRUSATTE, S. L., LLOYD, G. T., WANG, S. C. & NORELL, M. A. 2014*b*. Gradual  
618 assembly of avian body plan culminated in rapid rates of evolution  
619 across the dinosaur-bird transition. *Current Biology*, **24**, 2386–2392.
- 620 BRUSATTE, S. L., MONTANARI, S., YI, H.-Y. & NORELL, M. A. 2011.  
621 Phylogenetic corrections for morphological disparity analysis: new  
622 methodology and case studies. *Paleobiology*, **37**, 1–22.
- 623 BUTLER, R. J., BRUSATTE, S. L., ANDRES, B. & BENSON, R. B. J. 2012. How do  
624 geological sampling biases affect studies of morphological evolution in  
625 deep time? a case study of Pterosaur (Reptilia: Archosauria) disparity.  
626 *Evolution*, **66**, 147–162.
- 627 CAILLIEZ, F. 1983. The analytical solution of the additive constant  
628 problem. *Psychometrika*, **48**, 305–308.
- 629 CIAMPAGLIO, C. N. 2004. Measuring changes in articulate brachiopod  
630 morphology before and after the permian mass extinction event: do  
631 developmental constraints limit morphological innovation? *Evolution*  
632 *and Development*, **6**, 260–274.
- 633 CISNEROS, J. C. & RUTA, M. 2010. Morphological diversity and  
634 biogeography of Procolophonids (Amniota: Parareptilia). *Journal of*  
635 *Systematic Palaeontology*, **8**, 607–625.

- 636 EKMAN, S., ANDERSEN, H. L. & WEDIN, M. 2008. The limitations of  
637 ancestral state reconstruction and the evolution of the ascus in the  
638 lecanorales (lichenized ascomycota). *Systematic Biology*, **57**, 141–156.
- 639 FOOTE, M. 1994. Morphological disparity in Ordovician-Devonian crinoids  
640 and the early saturation of morphological space. *Paleobiology*, **20**,  
641 320–344.
- 642 ———. 1996. Ecological controls on the evolutionary recovery of  
643 post-Paleozoic Crinoids. *Science*, **274**, 1492–1495.
- 644 ———. 1997. The evolution of morphological diversity. *Annual Review of*  
645 *Ecology and Systematics*, 129–152.
- 646 FOTH, C., BRUSATTE, S. & BUTLER, R. 2012. Do different disparity proxies  
647 converge on a common signal? Insights from the cranial morphometrics  
648 and evolutionary history of Pterosauria (Diapsida: Archosauria). *Journal*  
649 *of evolutionary biology*, **25**, 904–915.
- 650 FRIEDMAN, M. 2010. Explosive morphological diversification of  
651 spiny-finned teleost fishes in the aftermath of the end-Cretaceous  
652 extinction. *Proceedings of the Royal Society B: Biological Sciences*, **277**,  
653 1675–1683.
- 654 GARLAND, J., THEODORE & IVES, A. R. 2000. Using the past to predict the  
655 present: Confidence intervals for regression equations in phylogenetic  
656 comparative methods. *The American Naturalist*, **155**, 346–364.
- 657 GOULD, S. J. 1991. The disparity of the burgess shale arthropod fauna and  
658 the limits of cladistic analysis: why we must strive to quantify  
659 morphospace. *Paleobiology*, **17**, 411–423.

- 660 GOULD, S. J. & ELDREDGE, N. 1977. Punctuated equilibria: The tempo and  
661 mode of evolution reconsidered. *Paleobiology*, **3**, 115–151.
- 662 GOWER, J. C. 1966. Some distance properties of latent root and vector  
663 methods used in multivariate analysis. *Biometrika*, **53**, 325–338.
- 664 ———. 1971. A general coefficient of similarity and some of its properties.  
665 *Biometrics*, **27**, 857–871.
- 666 GUILLERME, T. 2016. *disprity*: a package for measuring disparity in r.
- 667 GUILLERME, T. & COOPER, N. 2018. *Appendices from: Time for a rethink: time*  
668 *sub-sampling methods in disparity-through-time analyses*. Dryad Digital  
669 Repository. <https://doi.org/10.5061/dryad.vp4q518>.
- 670 HALLIDAY, T. J. D. & GOSWAMI, A. 2016. Eutherian morphological  
671 disparity across the end-cretaceous mass extinction. *Biological Journal of*  
672 *the Linnean Society*, **118**, 152–168.
- 673 HETHERINGTON, A. J., SHERRATT, E., RUTA, M., WILKINSON, M., DELINE, B.  
674 & DONOGHUE, P. C. 2015. Do cladistic and morphometric data capture  
675 common patterns of morphological disparity? *Palaeontology*, **58**,  
676 393–399.
- 677 HOPKINS, M. 2013. Decoupling of taxonomic diversity and morphological  
678 disparity during decline of the Cambrian trilobite family  
679 Pterocephaliidae. *Journal of Evolutionary Biology*, **26**, 1665–1676.
- 680 HUANG, S., ROY, K. & JABLONSKI, D. 2015. Origins, bottlenecks, and  
681 present-day diversity: Patterns of morphospace occupation in marine  
682 bivalves. *Evolution*.

- 683 HUGHES, M., GERBER, S. & WILLS, M. A. 2013. Clades reach highest  
684 morphological disparity early in their evolution. *Proceedings of the*  
685 *National Academy of Sciences of the United States of America*, **110**,  
686 13875–13879.
- 687 HUNT, G. 2007. The relative importance of directional change, random  
688 walks, and stasis in the evolution of fossil lineages. *Proceedings of the*  
689 *National Academy of Sciences of the United States of America*, **104**,  
690 18404–18408.
- 691 HUNT, G., HOPKINS, M. J. & LIDGARD, S. 2015. Simple versus complex  
692 models of trait evolution and stasis as a response to environmental  
693 change. *Proceedings of the National Academy of Sciences of the United States*  
694 *of America*, **112**, 4885–4890.
- 695 LEGENDRE, P. & LEGENDRE, L. F. 2012. *Numerical ecology*, volume 24.  
696 Elsevier.
- 697 LLOYD, G. T. 2015. *Claddis: Measuring Morphological Diversity and*  
698 *Evolutionary Tempo*. R package version 0.1.
- 699 PARADIS, E., CLAUDE, J. & STRIMMER, K. 2004. APE: analyses of  
700 phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.
- 701 PRENTICE, K. C., RUTA, P. & BENTON, M. J. 2011. Evolution of  
702 morphological disparity in Pterosaurs. *Journal of Systematic*  
703 *Palaeontology*, **9**, 337–353.
- 704 R CORE TEAM. 2015. *R: a language and environment for statistical computing*.  
705 R Foundation for Statistical Computing, Vienna, Austria.

- 706 REVELL, L. J. 2012. phytools: An R package for phylogenetic comparative  
707 biology (and other things). *Methods in Ecology and Evolution*, **3**, 217–223.
- 708 RUTA, M., ANGIELCZYK, K. D., FRÖBISCH, J. & BENTON, M. J. 2013.  
709 Decoupling of morphological disparity and taxic diversity during the  
710 adaptive radiation of Anomodont Therapsids. *Proceedings of the Royal  
711 Society of London B: Biological Sciences*, **280**.
- 712 SLATER, G. J., HARMON, L. J. & ALFARO, M. E. 2012. Integrating fossils with  
713 molecular phylogenies improves inference of trait evolution. *Evolution*,  
714 **66**, 3931–3944.
- 715 THORNE, P. M., RUTA, M. & BENTON, M. J. 2011. Resetting the evolution of  
716 marine reptiles at the Triassic-Jurassic boundary. *Proceedings of the  
717 National Academy of Sciences of the United States of America*, **108**,  
718 8339–8344.
- 719 TOLJAGIC, O. & BUTLER, R. J. 2013. Triassic-Jurassic mass extinction as  
720 trigger for the Mesozoic radiation of Crocodylomorphs. *Biology Letters*,  
721 **9**.
- 722 TORGERSON, W. S. 1965. Multidimensional scaling of similarity.  
723 *Psychometrika*, **30**, 379–393.
- 724 WESLEY-HUNT, G. D. 2005. The morphological diversification of carnivores  
725 in North America. *Paleobiology*, **31**, 35–55.
- 726 WILLS, M. A., BRIGGS, D. E. G. & FORTEY, R. A. 1994. Disparity as an  
727 evolutionary index: A comparison of Cambrian and recent arthropods.  
728 *Paleobiology*, **20**, 93–130.

- 729 WRIGHT, D. 2017a. Data from: Bayesian estimation of fossil phylogenies  
730 and the evolution of early to middle paleozoic crinoids (echinodermata).
- 731 WRIGHT, D. F. 2017b. Bayesian estimation of fossil phylogenies and the  
732 evolution of early to middle paleozoic crinoids (echinodermata). *Journal*  
733 *of Paleontology*, 1–16.
- 734 YANG, Z., KUMAR, S. & NEI, M. 1996. A new method of inference of  
735 ancestral nucleotide and amino acid sequences. *Genetics*, **141**, 1641–50.