FORMATION BINNING: A NEW METHOD FOR INCREASED TEMPORAL RESOLUTION IN REGIONAL STUDIES, APPLIED TO THE LATE CRETACEOUS DINOSAUR FOSSIL RECORD OF NORTH AMERICA

by CHRISTOPHER D. DEAN[1] (iD), A. ALESSANDRO CHIARENZA[2] (iD) *and* SUSANNAH C. R. MAIDMENT[3] (iD)

[1]School of Geography, Earth & Environmental Sciences, University of Birmingham, Birmingham, UK; christopherdaviddean@gmail.com
[2]Perot Museum of Nature & Science, Dallas, TX USA
[3]Department of Earth Sciences, Natural History Museum, London, UK

Abstract: The advent of palaeontological occurrence databases has allowed for detailed reconstruction and analyses of species richness through deep time. While a substantial literature has evolved ensuring that taxa are fairly counted within and between different time periods, how time itself is divided has received less attention. Stage-level or equal-interval age bins have frequently been used for regional and global studies in vertebrate palaeontology. However, when assessing diversity at a regional scale, these resolutions can prove inappropriate with the available data. Herein, we propose a new method of binning geological time for regional studies that intrinsically incorporates the chronostratigraphic heterogeneity of different rock formations to generate unique stratigraphic bins. We use this method to investigate the diversity dynamics of dinosaurs from the Late Cretaceous of the Western Interior of North America prior to the Cretaceous–Palaeogene mass extinction.

Increased resolution through formation binning pinpoints the Maastrichtian diversity decline to between 68 and 66 Ma, coinciding with the retreat of the Western Interior Seaway. Diversity curves are shown to exhibit volatile patterns using different binning methods, supporting claims that heterogeneous biases in this time-frame affect the pre-extinction palaeobiological record. We also show that the apparent high endemicity of dinosaurs in the Campanian is a result of non-contemporaneous geological units within large time bins. This study helps to illustrate the utility of high-resolution, regional studies to supplement our understanding of factors governing global diversity in deep time and ultimately how geology is inherently tied to our understanding of past changes in species richness.

Key words: dinosaur, time bin, temporal resolution, North America, formation, Late Cretaceous.

SINCE the first attempts at qualitative estimates by Phillips (1860), a central goal of palaeontology has been to quantitatively establish patterns of diversity change over time and to understand the rules that govern this change. By counting the number of unique fossil taxa that appear in 'bins' of geological time, it is possible to produce a diversity curve to estimate the variation in diversity throughout Earth's history. However, the inherent heterogeneity in the spatial and temporal preservation of rock outcrop and the sampling of fossil taxa by palaeontologists introduces a number of biases that can non-randomly skew the available data, confounding estimates of true diversity (Raup 1972; Holland 1995; Alroy *et al.* 2001). Whilst a substantial literature using a number of quantitative methods has evolved to ensure that taxa are

fairly counted and represented within and between bins, and thus accurately reflect species richness (Miller & Foote 1996; Smith & McGowan 2007; Alroy *et al.* 2008; Alroy 2010; Vavrek & Larsson 2010; Lloyd 2011; Starrfelt & Liow 2016; Sakamoto *et al.* 2017; Close *et al.* 2018), how time is divided into bins has received relatively less attention despite its equal potential to impact our conclusions (Gibert & Escarguel 2017; Guillerme & Cooper 2018; Rasmussen *et al.* 2019; Fan *et al.* 2020). Palaeobiological studies examining diversity necessarily divide time into a series of arbitrary bins, the resolution of which usually depends to a degree on the length of time, type of data being investigated, or simply what is arbitrarily considered as 'adequate sample size' to provide any statistically significant result. However, different resolutions of

time bins will impact the perceived tempo of evolutionary processes, which can subsequently affect our perception of the mode of diversity dynamics (Simpson 1944). Stage-level bins have frequently been used for regional and global-scale studies in vertebrate palaeontology (Upchurch *et al.* 2011; Dean *et al.* 2016; Cleary *et al.* 2018; Dunne *et al.* 2018; Close *et al.* 2019); these are based on the International Commission on Stratigraphy's agreed time-scale (https://stratigraphy.org/chart) which is in turn based on globally recognized geological events. However, the coarse temporal resolution of these designations can prove inappropriate with the data that are available. It may not, for example, be possible to adequately resolve biological patterns that take place on shorter timescales than those of the bin; if a mass extinction and subsequent biotic recovery occur in a single time bin, high rates of turnover may cause the extinction event to be obscured (Alroy 2008; Benson *et al.* 2016). Furthermore, diversity estimates can be exaggerated when taxa that did not overlap temporally are aggregated into one bin, or when occurrences are assigned to multiple bins based on poor age constraints (Dean *et al.* 2016). The nature of stage-level bins also means that they are unequal in length, which inherently biases longer bins in favour of higher diversity; large changes in bin length between studies can therefore result in discrepancies between diversity results (Rasmussen *et al.* 2019). While researchers have attempted to correct for this issue using equal length time bins (Alroy 2000; Alroy *et al.* 2008; Brusatte *et al.* 2012; Butler *et al.* 2012; Lloyd & Friedman 2013; Newham *et al.* 2014; Silvestro *et al.* 2016; Hopkins 2017; Cleary *et al.* 2018; De Celis *et al.* 2019) this is limited by the resolution at which individual occurrences are dated (Alroy 2008) and the ad hoc nature of this procedure can introduce further bias (Guillerme & Cooper 2018). Combined, these factors result in distortion of diversity signals, and confound our understanding of macroevolutionary processes through time.

These issues are especially apparent within regional contexts, where localized stratigraphic sequences may not bear a strong correspondence to globally-recognized time bins. This problem is exemplified in the Late Cretaceous of North America: the Campanian (which lasts ~11.5 myr) and the Maastrichtian (lasting ~6 myr; Ogg *et al.* 2012) contain the majority of the Late Cretaceous North American dinosaur fossil record (which encompasses ~44 myr of geological history), as well as approximately 51% of global dinosaur occurrences for the Late Cretaceous worldwide (http://fossilworks.org). As such, only two regionally localized time-bins realistically contribute towards estimates of dinosaur diversity in the Late Cretaceous (Brusatte *et al.* 2015). The 'bonanza' effect (Raup 1972) also influences the palaeodiversity record of this pivotal time period (Seilacher *et al.* 1985). Some

chronostratigraphic units are so rich in fossil occurrences that they disproportionately impact observed diversity. A large proportion of all of the dinosaurian records come from the highly fossiliferous Dinosaur Park Formation in Dinosaur Provincial Park (Alberta) and the southern Kaiparotwits, Kirtland and Fruitland assemblages (Colorado, Utah and New Mexico) (Chiarenza *et al.* 2019; Mallon 2019); for example, the Dinosaur Park Formation alone accounts for ~20% of total North American dinosaur diversity (Eberth & Currie 2005; Tennant *et al.* 2018; Chiarenza *et al.* 2019). This extraordinarily abundant record, although geographically localized and temporally restricted (~76.9–75.8 Ma; Fowler 2017), causes a mid-Campanian diversity spike which, when read literally, drives the perceived decline in diversity toward the Maastrichtian from a relatively more abundant Campanian record (Fastovsky *et al.* 2004; Barrett *et al.* 2009; Chiarenza *et al.* 2019).

These factors prove especially problematic when attempting to discern diversity patterns leading up to the major extinction event at the Cretaceous–Palaeogene (K/Pg) boundary (Fastovsky *et al.* 2004; Fastovsky & Sheehan 2005). The drivers of the K/Pg extinction and their effect on fauna are of significant interest (Sloan *et al.* 1986; Sheehan *et al.* 1991; Lillegraven & Eberle 1999; Pearson *et al.* 2002; Wilf *et al.* 2003; Fastovsky & Sheehan 2005; Wilson 2005; Barrett *et al.* 2009; Schulte *et al.* 2010; Brusatte *et al.* 2012, 2015; Mitchell *et al.* 2012; Larson *et al.* 2016; Chiarenza *et al.* 2019) but cannot be adequately constrained without appropriate, high-resolution diversity estimates (Fastovsky & Sheehan 2005 and references therein). Whilst some regions with a marine fossil record have used biostratigraphy to more finely divide time (Pearson *et al.* 2002; Cobban *et al.* 2006), traditional terrestrial biostratigraphy (Russell 1975, 1982) is generally too coarse and spatially restricted to be used reliably at the necessary level of precision. Other researchers have resorted to producing specific high-resolution databases of fossil occurrences to by-pass these issues (Alroy 1998), although this proves time intensive, requires dedicated resources and the resulting databases are often not applicable to other studies.

It is readily apparent that alternative binning methods tailored to specific geographical regions are necessary for resolving regional-scale diversity patterns through time. An objective method to chronologically partition the fossil record is also pivotal for palaeobiologists to fully extend the reproducibility of their results (Close *et al.* 2018). One potential avenue for such a methodology is provided by geological formations: lithostratigraphic units which are mappable in extent (Peters & Foote 2001). Formations: (1) are lithologically, and often environmentally, distinct from overlying and underlying units; (2) contain what is generally accepted to be relatively

contemporaneous fauna (although see the impacts of time averaging; Fowler 2017; Chiarenza *et al.* 2019); (3) have relatively short durations, often less than stage length; (4) lend themselves to easily accessible chronological and geological information, such as start and end dates and depositional environment; (5) are broadly controlled by tectonic and eustatic external factors, and thus when stacked through time will show breaks at similar points (Haubold 1990; Holland & Patzkowsky 2002; Butler *et al.* 2010); and (6) are already often used as the basis for time constraints of vertebrate fauna (Ryan & Evans 2005). The advent of the Paleobiology Database (PBDB; https://paleo biodb.org) and resources such as Macrostrat (http://mac rostrat.org), has made it easier to compare up-to-date formation ages from region to region, and to translate lithostratigraphy into chronostratigraphy. Formations potentially represent useful units of time for comparison in regional studies (e.g. Brocklehurst (2018) used formations from the Permian of Texas as time bins). This is particularly true within the Western Interior, where work by Fowler (2017) has synthesized detailed chronostratigraphic correlation of formations across the basin.

We propose a new method of binning geological time for regional studies using the arrangement of local formations through time to generate unique stratigraphic bins. Formation binning uses start and end dates of formations to assess the most suitable location to draw a time bin boundary, so that the fewest number of formations significantly cross that boundary. Fossil occurrences from the formation as a whole are assigned to bins depending on a user's chosen criteria (see Material and Method, below). This approach allows us to directly explore and incorporate local geology into regionally appropriate units of time, test the effects of bin length on patterns of diversity, and enhance the temporal resolution of palaeobiological studies. We test this method on the dinosaur fauna of the Late Cretaceous (Cenomanian–Maastrichtian) Western Interior of North America due to the relatively high quality of sampling and the well-developed stratigraphic framework of the region (Eberth & Braman 2012; Fowler 2017; Tucker *et al.* 2020). We produce a selection of high-resolution diversity curves that are subsequently compared to traditional stage-level bins. We demonstrate that the temporal resolution at which we view palaeontological data can impact estimates of species richness through time.

MATERIAL AND METHOD

Fossil occurrences

A presence-only fossil occurrence dataset of non-avian dinosaur taxa was obtained from the PBDB. Each occurrence includes taxonomic and geographic locality data, as well as an associated collection with lithological and geological information. Data were extensively screened for problematic records and to ensure taxonomic validation. The number of individual dinosaur genera for each Late Cretaceous formation was calculated by tallying all formal and valid taxonomic names, including multi-specific taxa and those occurrences which had only been identified to the genus level (e.g. *Triceratops* sp. or *Edmontosaurus* sp.) We used dinosaur genera rather than species for two reasons: (1) genera are more taxonomically stable than species for many dinosaur groups, based on uncertainty in species-level taxonomy (Robeck *et al.* 2000); and (2) to increase relative sample sizes per taxon for our sub-sampling methods, resulting in more reliable estimates of generic richness (Benton 2008). Note that genus and species-level diversity patterns generally appear to track each other for Mesozoic tetrapod groups (Barrett *et al.* 2009), and many dinosaur genera are monospecific. Dinosaur genera not associated with a formation were excluded, as were ootaxa and ichnofauna. Occurrences used in this study can be found in Dean *et al.* (2020, SI 1).

Age of formations

Whilst the start and end dates of formations were initially obtained from the PBDB, many of these dates were found to be inaccurate compared to the most recent formally published ages. In several instances, formations known to overlie one another stratigraphically were given the same age range in the PBDB. For example, the Laramie Formation is overlain by the Denver Formation in the Denver Basin (Fowler 2017 and references therein), but both formations are given an age of 70.6–66.043 Ma in the PBDB. Thus, we obtained more accurate age ranges for formations of interest from the primary literature. Each formation has an associated maximum and minimum age, potential error of maximum and minimum age, depositional environment, information on method used to constrain its age, location, and additional notes and reference for age constraints; these are available in Dean *et al.* (2020, SI 1). Diachronous formations (formations which appear in different places at different times) were treated the same as non-diachronous ones, that is, assigned one maximum and one minimum age. To test how differences between PBDB and personally assigned formation ages influenced diversity patterns, we generated stage and sub-stage level diversity curves with both the original PBDB downloaded age data and with the updated dataset including reviewed age and unit assignments.

Formation binning

Binning time. A script was written in R (v. 3.5.0; R Core Team 2018) to produce discrete bins based on formations (Fig. 1). The script repeatedly checks the suitability of drawing a potential time bin boundary in 0.01 myr intervals, with the aim of minimizing the number of formations that lie across boundaries (Fig. 1A). This is accomplished by assessing the chronological position of each formation relative to the proposed time bin boundary. If the formation does not cross the boundary, it gets an automatic maximum binning score of 100. If the formation crosses the boundary, the smallest percentage of the formation that sits either side of that boundary is identified, and the binning score is reduced by that percentage. For example, if a formation spanned from 80 to 90 Ma and the proposed boundary was 81 Ma, 10% of the formation would fall over this boundary, and so the binning score for that formation for that boundary would be reduced by 10%, equaling a final score of 90. As such, these intervals do not represent real, identifiable points within formations, but rather potential points in geochronological time where it is possible for the script to draw a bin boundary. Scores for each formation are automatically recorded in a scoring grid (Score Grid 1; SG1; Fig. 1B), a table of 0.01 myr time intervals in columns and formations of interest in rows. Whilst we acknowledge that a precision of 0.01 myr in terrestrial geological sequences is often unrealistic, this resolution matches the highest possible resolution obtained for formation start/ends dates. For each 0.01 myr time interval, a mean score is generated from the scores of all formations, that reflects the suitability of having a bin boundary at that point in time. Mean scores are then compared throughout the time series in user-defined windows of time (Fig. 1C), for example, 2 myr bins, and the maximum mean suitability score within that user-defined window is recorded as a bin (Fig. 1D). In the event that a bin is under 0.5 myr in duration, the script will amalgamate the bin equally into the bins above and below it, whilst generating a warning for the user.

One potential pitfall of this method is the existence of formations with long durations, which by definition will more strongly influence the location of new bin boundaries than those with short durations. To counteract this effect, a second variation of this methodology was additionally created that ignores formations with ranges longer than the third quantile of all formation ranges (Score Grid 2; SG2). Formations with long durations were only removed from the generation of bins; their occurrences were still binned and included in counts of species richness. Bins created by both of these methods at varying temporal resolutions are compared within our results;

diversity counts were generated using the bins of SG2 (see Dean *et al.* (2020, SI 2) for diversity results generated using the bins of SG1).

Binning occurrences. To investigate diversity patterns, dinosaur occurrences present within formations must themselves be assigned to a bin. However, some formations extend over bin boundaries, and thus it must be determined in which bin to place occurrences. To account for this issue, several methodologies were developed: (M1) occurrences are assigned to all bins spanned/crossed by the formation that contains them; (M2) occurrences are assigned to the bin that the majority of the formation occupies, with formations longer than two times the maximum bin length being ignored; (M3) occurrences for a formation are assigned to all bins that they occur in based on the percentage of the formation that appears within that bin; occurrences are selected at random from the formation occurrence list and not replaced, with the test being repeated according to a user specified number of runs (see Dean *et al.* 2020, SI 1), to give an overall average diversity estimate per bin. To ensure that the method is broadly applicable, individual occurrences are currently not given more precise chronological locations within formations, even if such information is available. Comparisons between these methods are explored in further detail in the Results and Discussion sections. We also generated results using stage and substage-level bins to compare between these new binning techniques and traditional methods. The full, annotated script for Formation Binning is provided in Dean *et al.* (2020, SI 1) and on Github.

Diversity estimates

To estimate changes in diversity through time whilst correcting for uneven sampling intensity, we used Shareholder Quorum Subsampling (SQS; Alroy 2010), a commonly used method of subsampling in vertebrate palaeontology. We implemented SQS using the R package DivDyn (Kocsis *et al.* 2019) in R version 3.5.0, which uses the inexact algorithm of SQS (see Close *et al.* 2018). We elected not to use the optional three-collections-per-reference protocol advocated by Alroy (2010) because: (1) unlike marine invertebrate datasets, dinosaurs do not suffer from over-reporting of common taxa (as in Dunne *et al.* 2018 for Palaeozoic tetrapods); and (2) sample coverage in some intervals is so low that limiting the amount of data drawn (to no more than three-collection-per-reference per trial) prohibited us from obtaining diversity estimates at meaningful quorum levels. We computed coverage-standardized diversity estimates at the genus level.
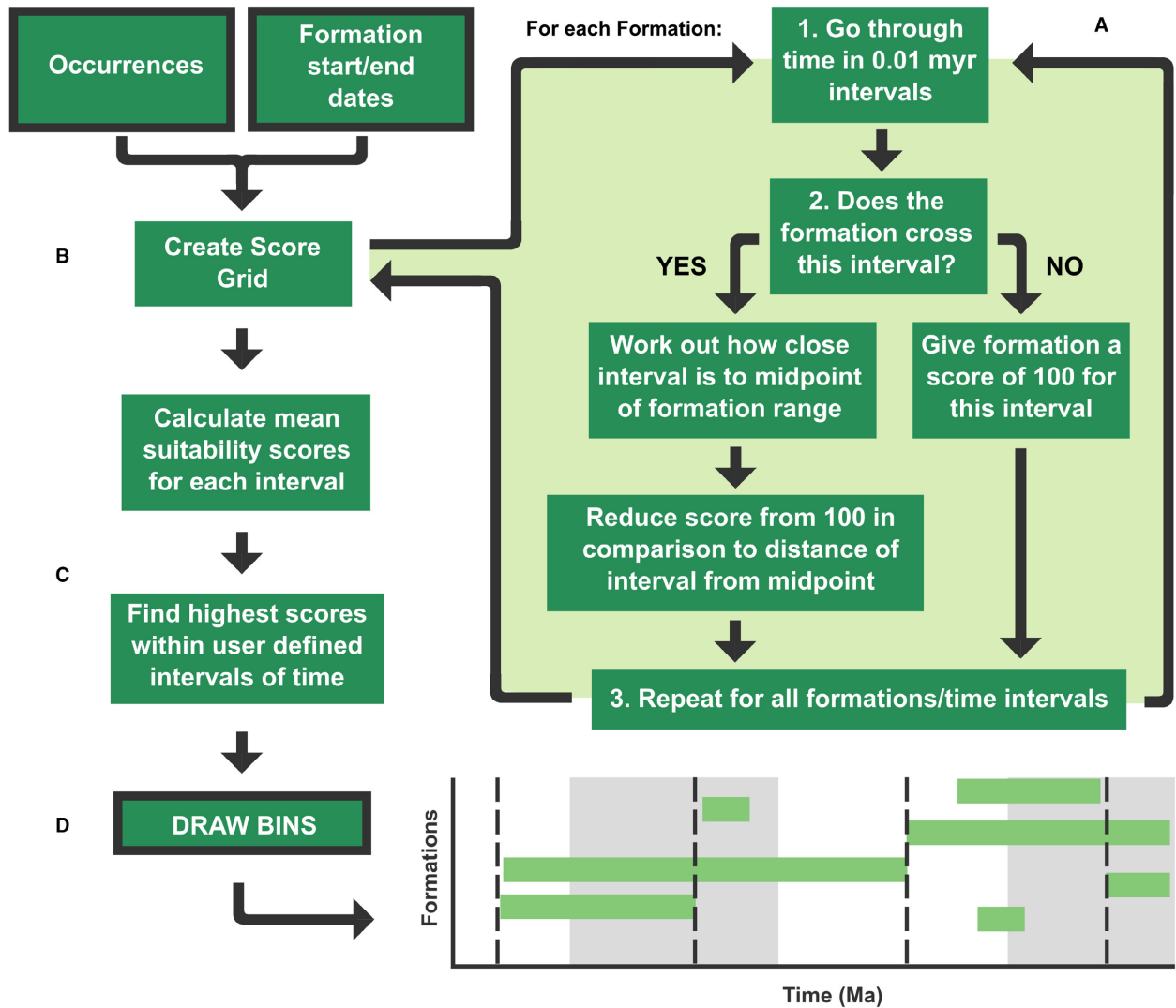
FIG. 1. Diagram showing the process of binning by formation. A .csv file of formations with associated maximum and minimum ages is provided to the formation_binner R script, which for each formation records the suitability of drawing a bin in 0.01 myr increments through time (A), ultimately producing a score grid (B). Mean suitability scores for each for each interval are calculated, and the script uses the highest scores within user specified intervals of time (C) to produce formation bins (D). Colour online.

RESULTS

Formations in time and space

The majority (37/46 = ~80%) of terrestrial dinosaur-bearing formations in the Western Interior appear within the Campanian and Maastrichtian (Fig. 2). Dinosaur-bearing formations appearing before the Campanian are typically lengthy in duration and are restricted to below 45° modern latitude (Fig. 2A). Within the Campanian, formations that contain high dinosaur diversity are latitudinally restricted; one group of formations appear between ~36 and 38° modern latitude, and another between ~43 and 51°. In contrast, formations within the Maastrichtian are

more evenly distributed by latitude, although there is some clustering at 42° (Fig. 2). This is broadly consistent with the findings of Chiarenza *et al.* (2019), who recovered distinct north/south hotspots of dinosaur occurrences within the Campanian and more even distribution within the Maastrichtian.

Formation bin comparisons

Figure 3 shows a comparison between different bins generated when applying formation binning using both SG1 and SG2 methodologies, at user defined intervals of 2, 3 and 4 myr. Unsurprisingly, formation binning at all
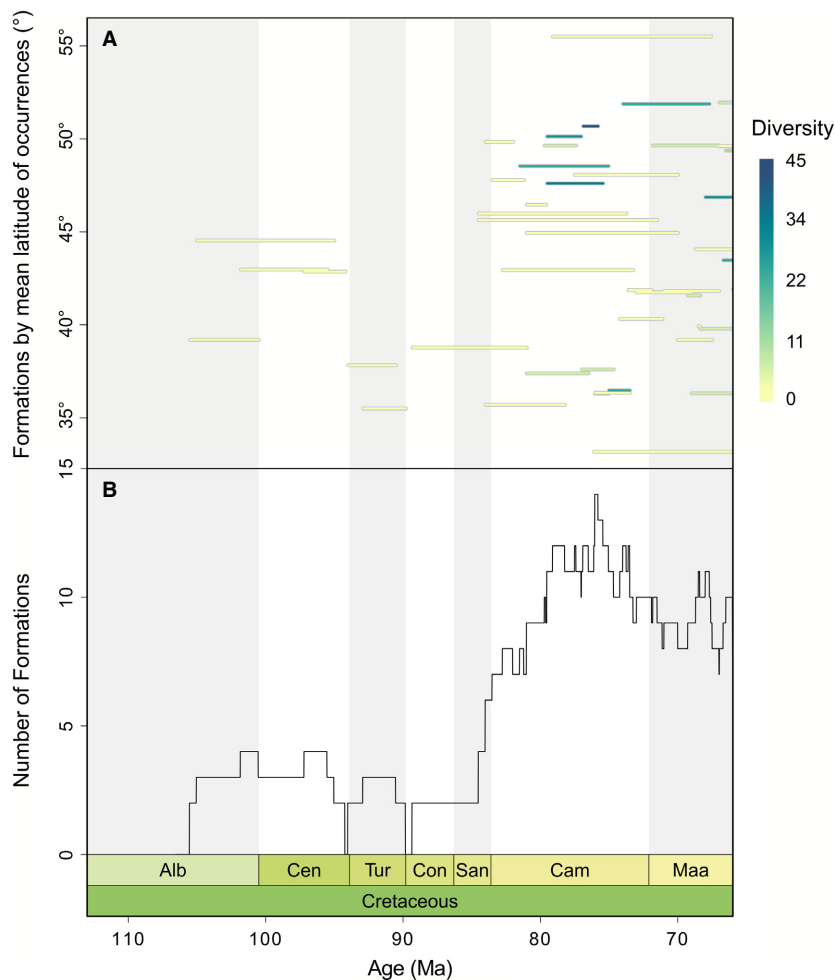
FIG. 2. Plot showing the age range of dinosaur-bearing formations within the Western Interior Basin through time, arranged by mean latitude of occurrences within that formation and coloured by the generic diversity of that formation. Vertical grey bars represent stage-level bins. Colour online.

resolutions produces shorter mean bin lengths and resulting standard deviations than stage level bins (Appendix 1). For each of the binning methods (SG1 and SG2), several bin boundaries remain stable when varying the chosen resolution: For SG1, boundaries at 101.8, 94, 75, and 70 Ma exist at all resolutions, whilst boundaries at 94, 84, 81, 76.9 and 73.5 Ma appear in every resolution for SG2. Comparisons between bin boundaries and traditional stage and sub-stage level bins show that formation bins do not strongly correlate with the positions of stage or sub-stage bins. Only the Cenomanian–Turonian boundary and the early–late Maastrichtian boundary are recovered as boundaries across nearly all formation binning methods, although the early–middle Campanian boundary is recovered using formation binning when using SG2.

Stage and sub-stage level diversity results

To compare the effects of different temporal constraints on diversity curves, we produced stage-level plots using both ages originally assigned to occurrences in the PBDB and age from updated formation information (Fig. 4). To test the effect of binning styles, we also binned occurrences either within all bins that could possibly contain them (Fig. 4), or based on the mid-point of their potential age range (Dean *et al.* 2020, fig. S1).

Stage level raw diversity estimates using ages of occurrences taken from the PBDB (Fig. 4A) show extremely low diversity from the Albian through to the Coniacian. The Santonian shows a sharp increase, from which a gradual increase is seen through the Campanian and Maastrichtian. A similar pattern is observed when applying SQS (Fig. 4B),
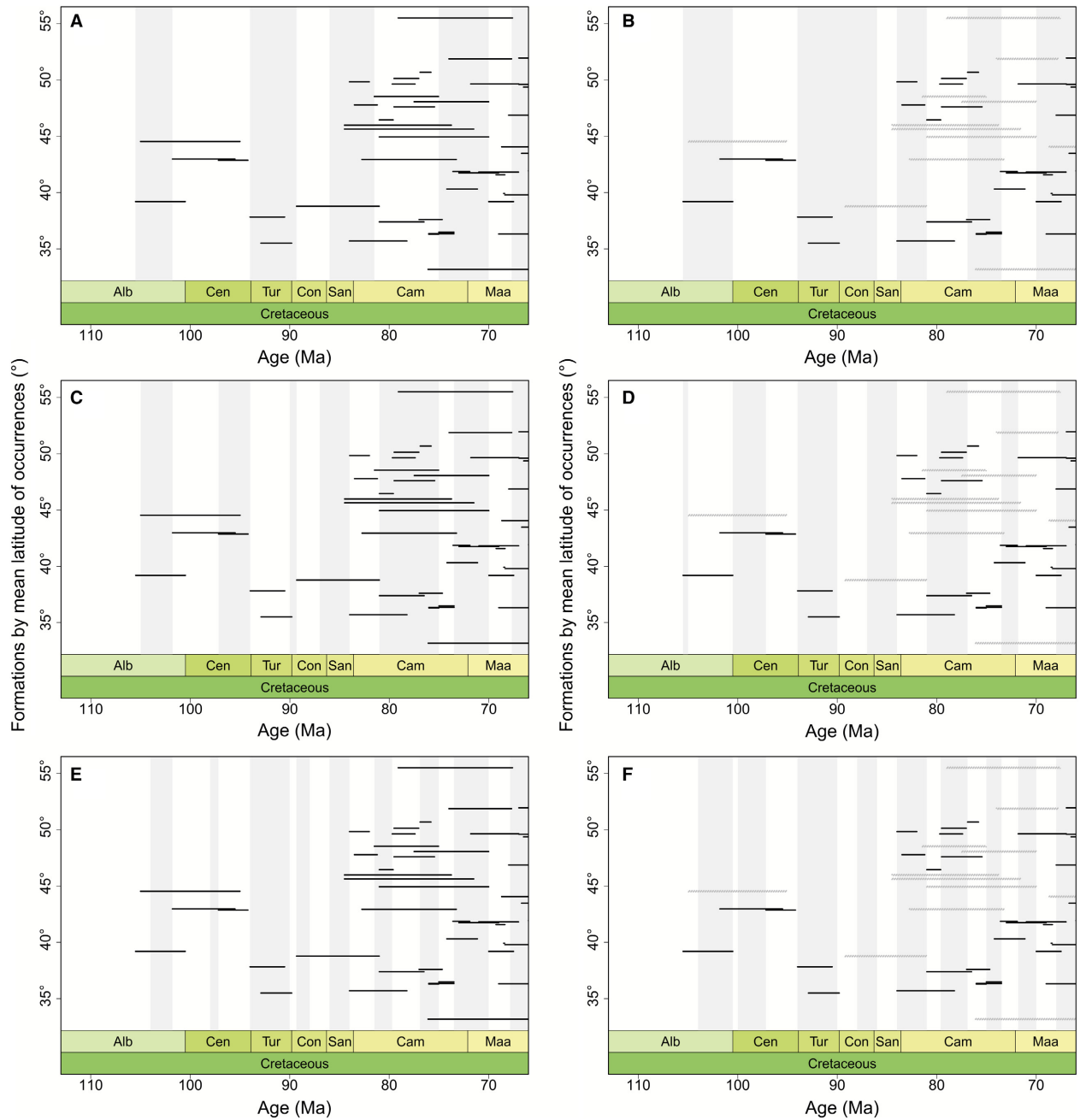
FIG. 3. Plots of dinosaur-bearing formations of the Western Interior Basin through time in association with formation bins (vertical grey bars) at varying user-defined resolutions producing using Score Grid 1 (A, C, E) and Score Grid 2 (B, D, F). For bins produced using Score Grid 2, formations in grey represent long ranging formations that were ignored in the creation of bins. A, formation bins producing using Score Grid 1 at 4 myr resolution. B, formation bins producing using Score Grid 2 at 4 myr resolution. C, formation bins producing using Score Grid 1 at 3 myr resolution. D, formation bins producing using Score Grid 2 at 3 myr resolution. E, formation bins producing using Score Grid 1 at 2 myr resolution. F, formation bins producing using Score Grid 2 at 2 myr resolution. Colour online.

although peak diversity is now recovered in the Santonian. To test the accuracy of diversity curves generated using PBDB age estimates, we also calculated raw and SQS diversity estimates based on our updated formation ages (Fig. 4C–D). Here, a radically different pattern is observed.

This updated information completely removes the initial Santonian diversity increase and lowers the Maastrichtian peak in both raw and subsampled results.

To test the effect of higher resolution time binning, we additionally produced sub-stage level raw and SQS
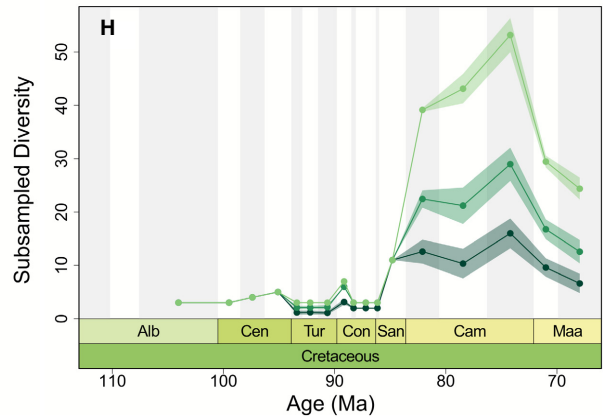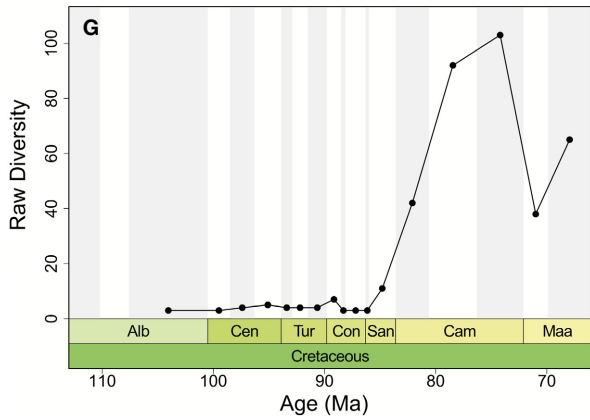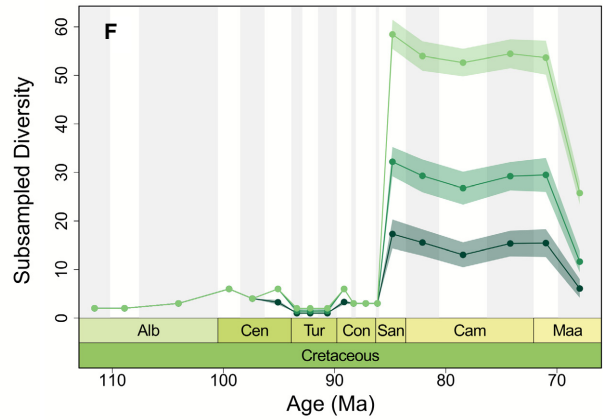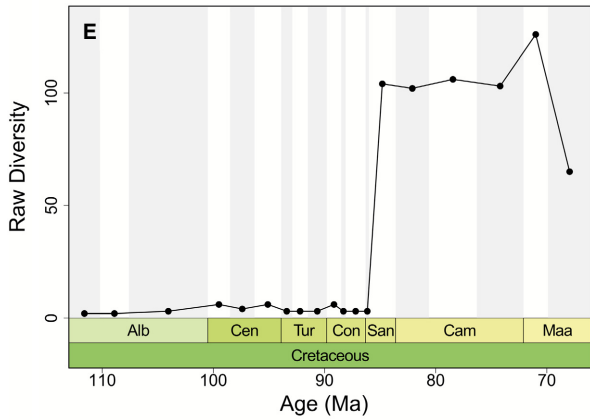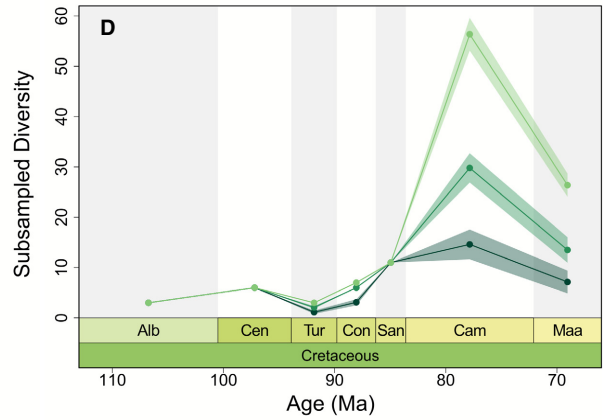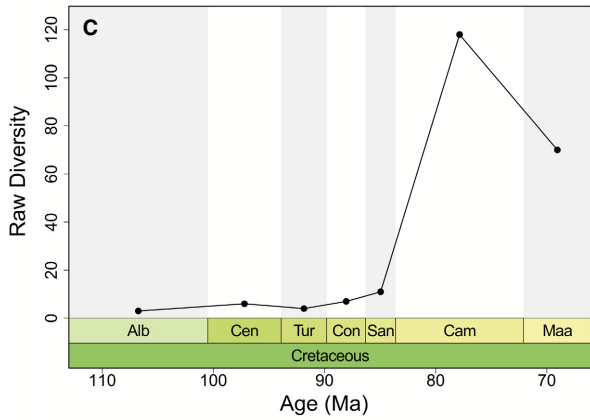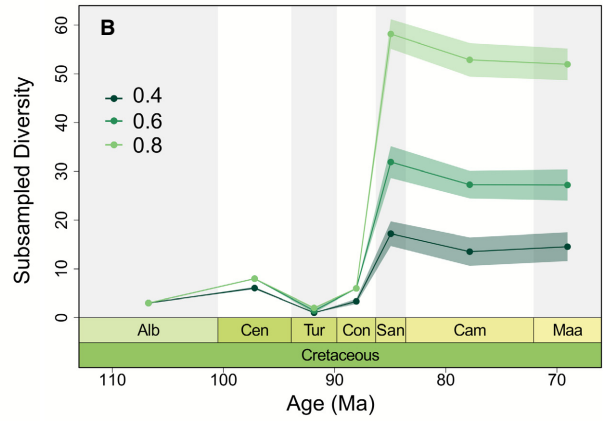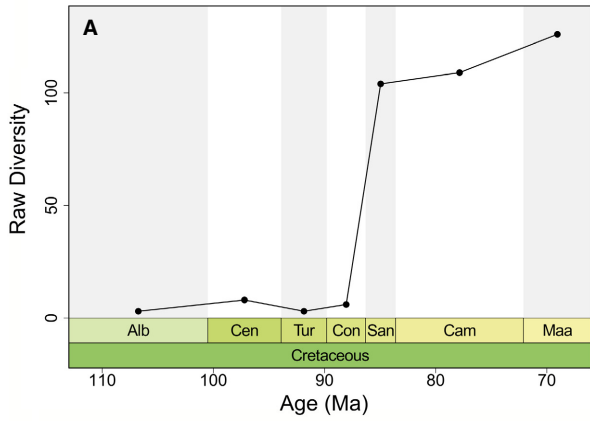
FIG. 4. Plots of stage and sub-stage level raw and SQS diversity curves of Late Cretaceous North American dinosaurs from the Western Interior Basin, produced using either originally assigned PBDB age estimates or updated formation ages (see Material and Method) with occurrences added to all bins they could possibly be found within. A, stage level raw diversity produced using PBDB ages. B, stage level SQS diversity produced using PBDB ages. C, stage level raw diversity produced using updated formation ages. D, stage level SQS diversity produced using updated formation ages. E, sub-stage level raw diversity produced using PBDB ages. F, sub-stage level SQS diversity produced using PBDB ages. G, sub-stage level raw diversity produced using updated formation ages. H, sub-stage level SQS diversity produced using updated formation ages. Colour online.

diversity curves using PBDB age estimates and updated formation age estimates (Fig. 4E–H; see Dean *et al.* (2020, fig. S1E–H) for similar results plotted using mid-points of occurrence age-ranges). Raw diversity using PBDB age estimates (Fig. 4E) is low from the Albian to the Coniacian, before a sharp increase in the Santonian. This high level of diversity continues through the Campanian, before peaking in the early Maastrichtian. The late Maastrichtian sees a sharp decline in species richness. SQS results (Fig. 4F) record a similar pattern to that seen in raw diversity, albeit with a Santonian peak in diversity. When looking at raw diversity using updated formation ages (Fig. 4G), a similar pattern of low and patchy diversity is visible from the Albian to the Coniacian. However, the remaining sub-stages of the Cretaceous show a strongly different curve; instead of a high, flattened profile through the Santonian and Campanian, diversity increases up to the late Campanian, before dropping and then rebounding in the early and late Maastrichtian respectively. SQS results (Fig. 4H) show a more variable pattern than those from PBDB ages, although a mid-Campanian dip in diversity is retained.

Formation binned diversity results

All formation-binned diversity results were carried out using SG2. Results for SG1 can be found in Dean *et al.* (2020, SI 2).

Raw diversity results for M1 at resolutions 4, 3, and 2 myr (Fig. 5A, C, E) reveal increasing time series information compared to stage and sub-stage level results. At a resolution of 4 myr (Fig. 5A) the Campanian raw diversity peak is more tightly constrained, appearing between ~73.5 and 77 Ma, with richness steadily rising in the two prior bins. Increased resolution around the Campanian–Maastrichtian boundary shows an initial decline in diversity, which then rises towards the latest Maastrichtian. This same pattern is observed at 3 myr resolution (Fig. 5C), with a more pronounced diversity plateau in the latest Maastrichtian. At the highest resolution (2 myr; Fig. 5E), the Campanian peak is more tightly constrained to between ~75 and 77 Ma; otherwise diversity appears to gradually increase throughout the latest Cretaceous. SQS results for M1 (Fig. 5B, D, F) show a broadly similar pattern across all resolutions; all report a gradual increase in

diversity from the early Campanian towards a peak between ~75 and 77 Ma, after which diversity declines towards the latest Maastrichtian.

At resolutions 4 and 3 myr for M2 (Fig. 6A, C), diversity shows a mid-Campanian plateau between ~73.5 and 81 Ma instead of a peak. Additionally, the dip and subsequent recovery of diversity from the early to end Maastrichtian is more pronounced; this is particularly observed at resolution 3 myr (Fig. 6C), where diversity flattens and then increases within the final bin before the K/Pg boundary. At a resolution of 2 myr (Fig. 6E), we see a markedly different placement of the Campanian diversity peak, now occurring from ~77 to 80 Ma. This is followed by a longer diversity decline into the early Maastrichtian, which then rebounds until the K/Pg. SQS results at resolutions of 3 and 4 myr for M2 (Fig. 6B, D) are relatively similar to their raw diversity counterparts, showing earlier diversity peaks than the M1 SQS curves and a gradual decline towards the K/Pg boundary. At 2 myr resolution (Fig. 6F), whilst there is a general peak within the mid-Campanian, the profile of the curve at quora 0.4 and 0.6 is much flatter, showing a steep drop off into the latest Maastrichtian.

Raw diversity results for M3 (Fig. 7A, C, E) are similar to those produced using M1, although the early Maastrichtian dip and recovery are both more pronounced. At the highest resolution (2 myr; Fig. 7E) the diversity peak is once again constrained to ~75–77 Ma, and diversity in the early Maastrichtian is equivalently low to that of the early Campanian. SQS results at resolutions 4 myr and 3 myr (Fig. 7B, D) again show a gently peaked profile, with the highest diversity reached during the mid to late Campanian. At 2 myr resolution (Fig. 7F), diversity curves for quora 0.4 and 0.6 are much flatter in overall profile despite small scale fluctuations, and are fairly similar to those produced for M2. Curves produced using M3 all show a sharp decline in the latest stage of the Maastrichtian.

DISCUSSION

Diversity dynamics of North American dinosaurs during the latest Cretaceous

Extinction dynamics in their chronological context are of pivotal importance in our understanding of the nature of these processes (Bowring *et al.* 1999; Brusatte *et al.* 2012).
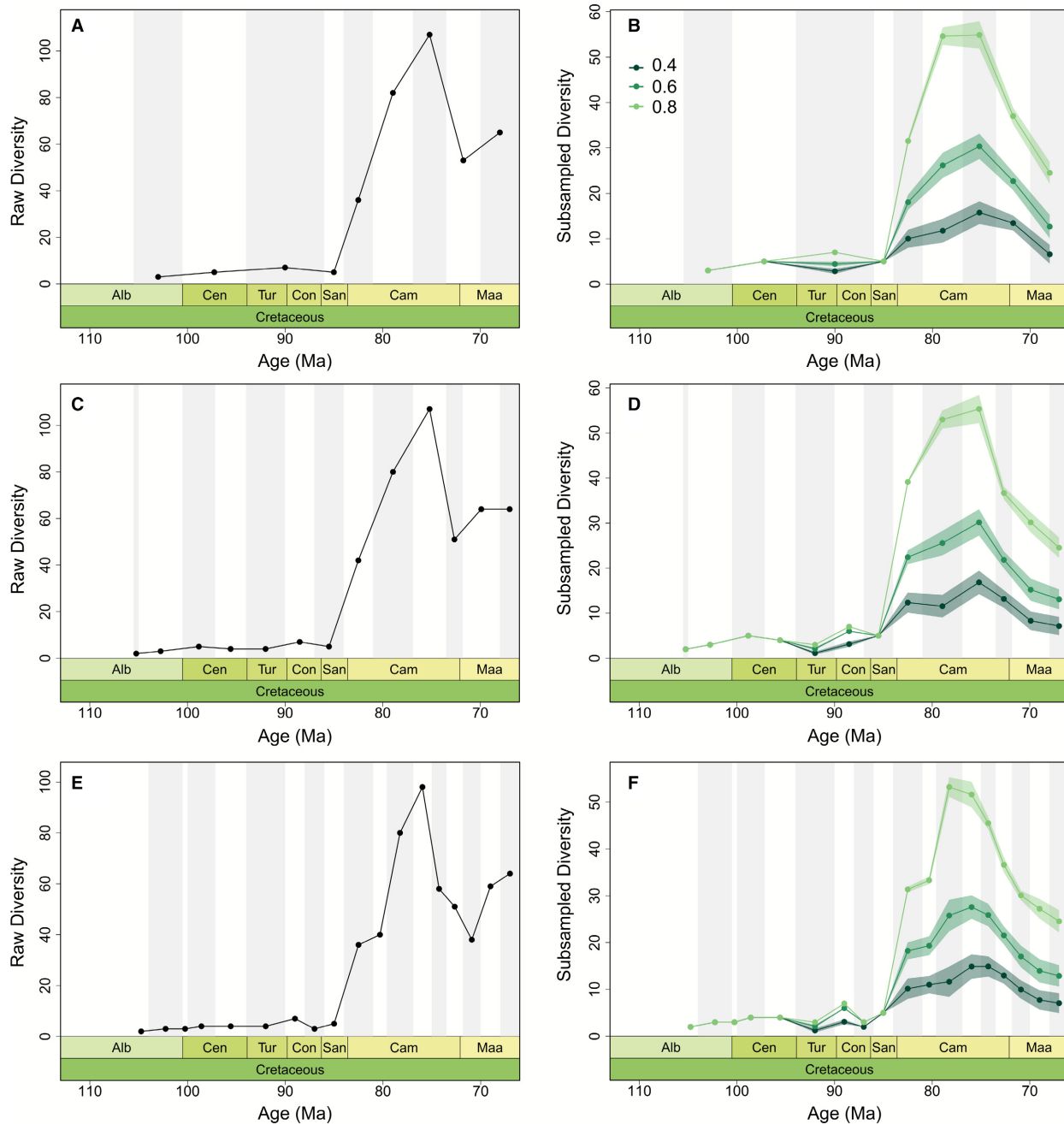
FIG. 5. Plots of raw and SQS diversity curves for Late Cretaceous North American dinosaurs of the Western Interior Basin, produced using formation bins (vertical grey bars) and binned using M1 (see Material and Method for further information). A, raw diversity produced using Score Grid 2 at 4 myr resolution. B, SQS diversity curve produced using Score Grid 2 at 4 myr resolution. C, raw diversity produced using Score Grid 2 at 3 myr resolution. D, SQS diversity curve produced using Score Grid 2 at 3 myr resolution. E, raw diversity produced using Score Grid 2 at 2 myr resolution. F, SQS diversity curve produced using Score Grid 2 at 4 myr resolution. Colour online.

Whilst it is well established that non-avian dinosaurs were wiped out by a bolide impact at the K/Pg boundary 66 million years ago (Schulte *et al.* 2010), their relative diversity trends in the lead-up to this mass extinction have proven to be a more contentious issue. Broadly,

studies have fallen along a spectrum either favouring a sudden demise at the K/Pg boundary (Fastovsky *et al.* 2004; Fastovsky & Sheehan 2005 and studies referenced therein; Brusatte *et al.* 2015; Chiarenza *et al.* 2019) or a gradual decline in diversity throughout the Late
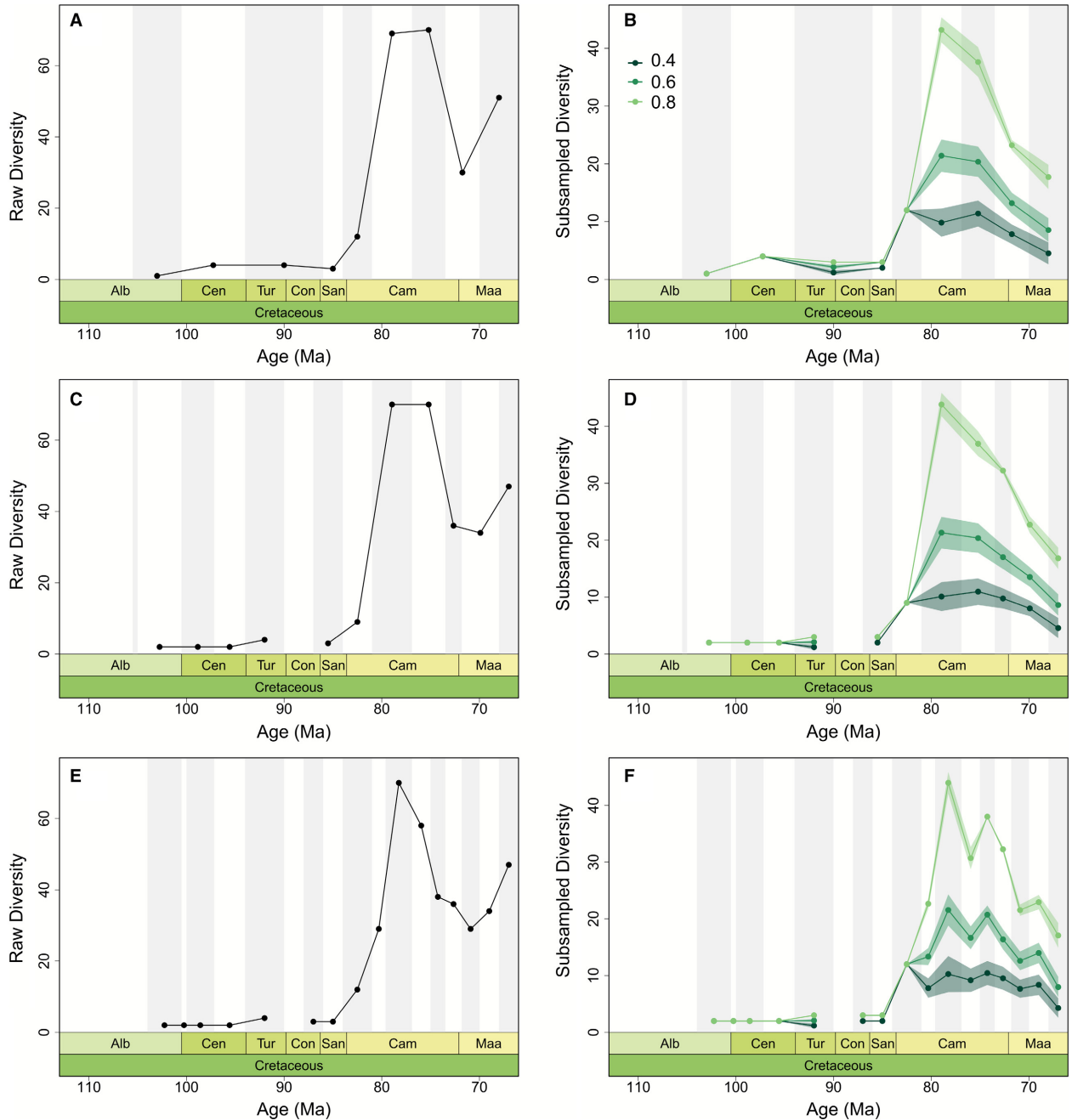
FIG. 6. Plots of raw and SQS diversity curves for Late Cretaceous North American dinosaurs of the Western Interior Basin, produced using formation bins (vertical grey bars) and binned using M2 (see text for further information). A, raw diversity produced using Score Grid 2 at 4 myr resolution. B, SQS diversity curve produced using Score Grid 2 at 4 myr resolution. C, raw diversity produced using Score Grid 2 at 3 myr resolution. D, SQS diversity curve produced using Score Grid 2 at 3 myr resolution. E, raw diversity produced using Score Grid 2 at 2 myr resolution. F, SQS diversity curve produced using Score Grid 2 at 4 myr resolution. Colour online.

Cretaceous (Sarjeant & Currie 2001; Barrett *et al.* 2009; Archibald 2014), caused by climate change, intense volcanism, or changes in regional or global geographic conditions (Sarjeant & Currie 2001; Archibald *et al.* 2010; Archibald 2014). This potential decline in diversity is observed within the North American record from the

Campanian to the Maastrichtian (Brusatte *et al.* 2015) and has been inferred to be genuine due to high sampling intensity in the latter interval (assessed either through relative numbers of dinosaur-bearing formations or number of collections). Several authors have additionally linked this pattern to the overall retreat of the Western Interior
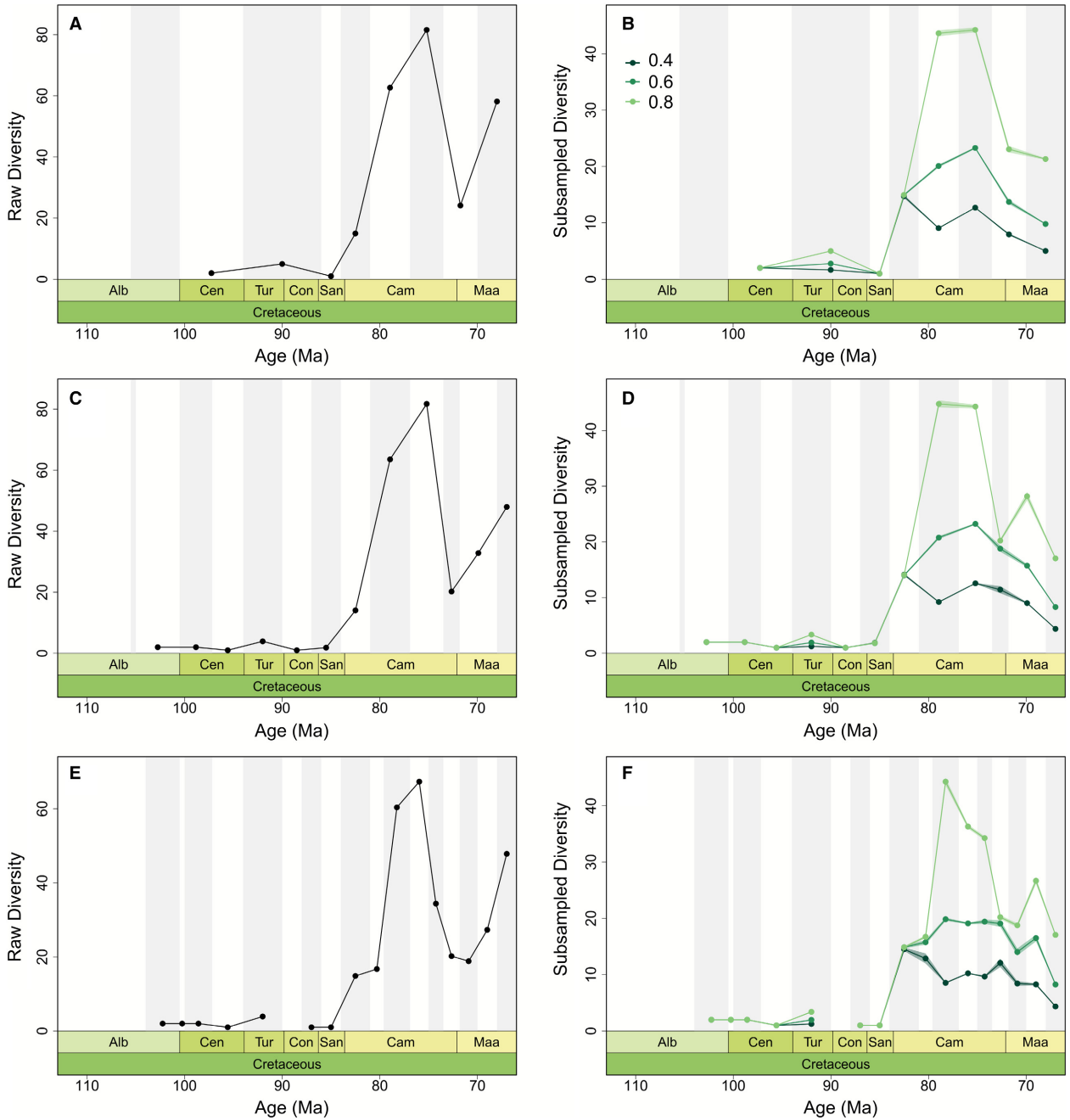
FIG. 7. Plots of raw and SQS diversity curves for Late Cretaceous North American dinosaurs of the Western Interior Basin, produced using formation bins (vertical grey bars) and binned using M3 (see text for further information). A, raw diversity produced using Score Grid 2 at 4 myr resolution. B, SQS diversity curve produced using Score Grid 2 at 4 myr resolution. C, raw diversity produced using Score Grid 2 at 3 myr resolution. D, SQS diversity curve produced using Score Grid 2 at 3 myr resolution. E, raw diversity produced using Score Grid 2 at 2 myr resolution. F, SQS diversity curve produced using Score Grid 2 at 4 myr resolution. Colour online.

Seaway (WIS; Miller *et al.* 2005; Kominz *et al.* 2008; Slattery *et al.* 2015) leading to reduced regional endemism (Sampson *et al.* 2010; Vavrek & Larsson 2010). However, a recent study by Chiarenza *et al.* (2019) suggested that this apparent decline is the result of a strongly

heterogeneous fossil record between the two time intervals. During the Campanian, highstand conditions in the WIS and the onset of the Laramide orogeny resulted in elevated sedimentation rates across a broad latitudinal range (Chiarenza *et al.* 2019; Lyster *et al.* 2020),

increasing the likelihood of fossil preservation. In contrast, whilst several transgressive marine tongues are observed within the Hell Creek formation (Fastovsky & Bercovici 2016), the Maastrichtian overall saw a retreat of the WIS, with possible east–west land connection between the two North American subcontinents (Slattery *et al.* 2015; Berry 2017). This in turn caused a reduction of coastline and a geodynamic reorganization of sedimentary basins, reducing both sediment flux and the available geographical area to preserve fossil material (Chiarenza *et al.* 2019). This pattern can be clearly observed in the relative reduction and geographical spread of formations recording high dinosaur diversity within the latest Maastrichtian (Fig. 2).

The additional resolution provided by formation binning allows for the timing of this apparent diversity decrease to be investigated more closely. SQS results for M1 and M3 at resolutions of 3 myr and 2 myr show generally stable patterns throughout the latest Cretaceous, with the steepest decline in dinosaur diversity occurring within the very latest Maastrichtian, between approximately 68 and 66 Ma, matching closely with constraints of the major retreat of the WIS above the *Jeletzkytes nebrascensis* biozone (~68 Ma; Fowler 2017). This implies that the Maastrichtian decline is intrinsically related to the retreat of the WIS, either through: (1) increased faunal mixing and a reduction in allopatric speciation; (2) a reduction in adequate conditions for fossil preservation (or at least analogue conditions to the mid–late Campanian); or (3) a combination of these effects, known as the 'common cause' hypothesis (Butler *et al.* 2010). Whilst it is not currently possible to distinguish between these potential causes, the results presented herein support the hypothesis that change in regional geological or environmental regime is the dominant control on the perceived diversity of North American dinosaurs in the lead up to the K/Pg extinction. As such, our current attempts to understand the drivers of dinosaur diversity are problematic without explicitly addressing geological issues. It has previously been suggested that spatial partitioning of dinosaur fauna from the Hell Creek and equivalent beds is expressed through lithological association of recovered specimens (Lyson & Longrich 2011). Further work identifying and quantifying relationships of this nature is necessary to establish whether the observed diversity patterns in the Late Cretaceous dinosaur record are a reflection of true biological patterns or a product of change in lithological availability between stages. These results highlight the need for careful evaluation of multiple lines of evidence (geophysical, palaeogeographical, sedimentological and stratigraphic, not just ecological) when considering the tempo and mode of biological events through geological time.

Spatial ecological patterns for North American dinosaurs: the time averaging case

The high species richness of vertebrates recorded in Campanian sediments across a wide range of localities arranged along the western coast of the WIS has generated claims of faunal bioprovinces and enhanced endemism within the Western Interior Basin (WIB). A model of latitudinally-arranged biogeographical provinces in the WIB (Lehman 1987; Gates *et al.* 2010, 2012; Sampson *et al.* 2010; Loewen *et al.* 2013) has recently been questioned by Lucas *et al.* (2016) and Fowler (2017) based on updated age constraints of dinosaur-bearing formations, which reveal the diachroneity of most fossil assemblages in the WIB. As a test of this controversy, results from formation binning allow visualization of the degree of temporal overlap between penecontemporaneous units (Fig. 2), and thus identification of where comparisons between faunas may represent succession rather than true synchronicity.

The diachronous nature of Campanian units in relation to the stage itself is clearly observed when looking at the temporal distribution of formations (Fig. 2); many formations cross the upper and lower boundaries of the stage, and/or span different parts of the same bin. As such, apparently coeval units (defined as falling in the same time bins) are not realistically contemporaneous, but probably contain a succession of faunas within the same bin, hence providing non-comparable faunal associations and probably heightening apparent endemicity between locations. Whilst formation binning on finer temporal resolutions results in truly contemporaneous units being compared within the same time bin (Fig. 3), a large number of formations show some degree of overlap with the time bins on either side of them. For example, for the 2 myr resolution plots (Fig. 3E, F) the two time bins in the middle of the stage have 13 and 14 units overlapping respectively. The large degree of overlap in these time bins is due to the long duration of many units; at least eight of them span for at least half the length of the entire Campanian stage and some also cross the Campanian–Maastrichtian boundary. Inclusion of these long ranging units in faunal comparisons between bins may create an issue of time averaging when trying to analyse faunal distinctiveness between contemporaneous units. This issue of time averaging, where perceived recovery of a hyper-diverse record and distinct bioprovinces proves unjustified based on careful stratigraphic and biostratigraphic analyses, has also been proposed as the cause of 'Stromer's Riddle', which describes the apparent coexistence of numerous large, predatory dinosaurs in the 'mid-Cretaceous' of the African 'Continental Intercalaire' Kem Kem beds (Cavin *et al.* 2010). A number of recent studies have suggested that this recovered pattern is actually a taphonomic artefact among lineages that were ecologically, environmentally and
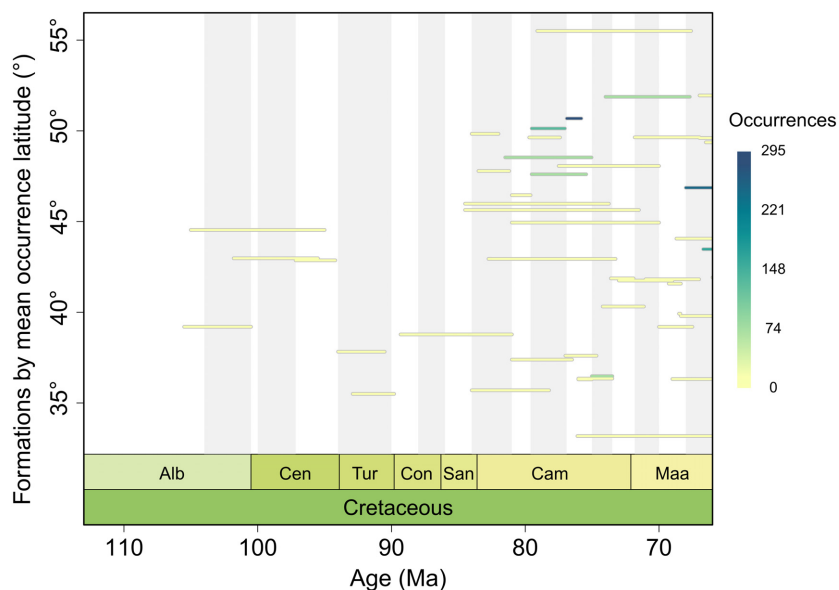
FIG. 8. Plot showing the age range of dinosaur-bearing formations within the Western Interior Basin through time, arranged by mean latitude of occurrences within that formation and coloured by the number of occurrences found within that formation. Grey bars represent formation bins produced using Score Grid 2 at a resolution of 2 myr. Colour online.

chronostratigraphically non-overlapping (Fanti *et al.* 2014; Chiarenza & Cau 2016).

Whilst long-ranging units could potentially be removed from faunal comparison analyses to reduce time averaging, this unfortunately results in an insufficient amount of comparative data within time bins to allow for accurate assessment of the possibility of endemism. In an analysis of Maastrichtian dinosaurs, Vavrek & Larsson (2010) removed all formations with fewer than 100 specimens to ensure that differential sampling intensity did not artificially inflate beta diversity (endemicity). Whilst not a direct comparison, when assessing formations by their total number of occurrences (Fig. 8), it is clear that only a few formations contain a significant record of dinosaur remains. Additionally, there are no Campanian bins at a resolution of 2 myr using SG2 where comparison between fauna would be viable. As such, testing faunal provinces in the Late Cretaceous dinosaur record does not seem feasible currently, and consequently our results support the criticism of the 'faunal endemism' hypothesis as proposed by Lucas *et al.* (2016) and Fowler (2017) who advocated time-averaging as the main factor generating the supposed faunal bioprovincialism in the WIB (see also Fowler & Fowler (2019) for a macroevolutionary perspective on this topic).

Importance of accurate age constraints in palaeodiversity studies

Ideally, diversity studies carried out in deep time allow us to investigate the response of organisms to various environmental drivers to better understand the construction of the modern biosphere. However, without accurate age constraints it is impossible to distinguish cause and effect within the fossil record and, at worst, poor dating of occurrences can lead to erroneous and misleading assumptions about the evolutionary trajectory of the group of interest. There are several steps during the creation of a palaeodiversity curve using data downloaded from the PBDB where temporal constraints are considered. Firstly, when occurrence data are originally added to the PBDB, the enterer will add the geological/chronostratigraphic interval(s) associated with the collection at the sub-stage level or coarser, the start and end dates of which are stored in a mastersheet within the database that is static (Peters & McClennen 2016). These age data are assigned to the collection, and to subsequent occurrences found within that collection. Enterers also have the choice to select specific age constraints, although this is uncommon due to the rarity of precise dating for the majority of published fossil taxa. A second step occurs within the handling of the chosen data downloaded from the PBDB. Researchers using downloaded data necessarily have to decide how to bin that data, as discussed throughout this paper. It is common practice in vertebrate palaeontological studies to bin occurrences within each bin that they could potentially be found in (Dean *et al.* 2016), although recently several studies have used binning approaches similar to method M3 in this paper (Dunhill & Wills 2015; Benevento *et al.* 2019).

Unfortunately, these steps can result in several levels of imprecision being introduced to the analysis, especially

when attempting to bin at finer resolutions. Diversity curves produced using PBDB ages and more specific formation ages exhibit diverging patterns for the majority of the Late Cretaceous (Fig. 4). This is largely due to the effects of the first step described above. The current ICS chronostratigraphic chart (Cohen *et al.* 2020) lists the Santonian–Campanian and Campanian–Maastrichtian boundaries as occurring at 83.6 ± 0.2 Ma and 72.1 ± 0.2 Ma respectively, with latter boundary having been updated from its previous estimate of 70.6 ± 0.6 in 2012. However, these updates have not been reflected in the PBDB; at the time of writing (22 January 2020), the Campanian–Maastrichtian boundary was still recorded as 70.6 Ma on fossilworks.org, whereas the Santonian–Campanian boundary was listed as 84.9 Ma, based on a paper regarding jumping plant lice (Klimaszewski 1998). Consequently, occurrences within the Campanian have associated numerical ages that range from the Santonian to the Maastrichtian when using the current ICS timescale, artificially inflating diversity within these bins through the addition of non-contemporaneous fauna. When the temporal resolution of occurrences is more finely resolved (Fig. 4C, D, G, H) the Santonian drops back down to low diversity in both raw and SQS plots.

Further issues can also arise related to the user chosen method of binning. Whilst diversity curves produced using PBDB mid-point ages (Dean *et al.* 2020, fig. S1) avoid the issue of incorrectly dated stage boundaries, formation dates (or more broadly, associated age dates of occurrences) might not be precisely entered within the PBDB. Age discrepancies of poorly constrained units can impact diversity studies (Dean *et al.* 2016) and thus lack of age precision has the capacity to produce erroneous diversity curves. Although stage level raw and subsampled diversity curves binned using PBDB ages are broadly comparable to those produced using updated formation ages, at the sub-stage level the two diverge; the curves show different patterns of decline and recovery within the Maastrichtian, subsequently leading to differing interpretations of dinosaur diversity in the lead up to the K/Pg boundary.

Whilst the PBDB is an exceptionally useful resource for palaeontological studies (with 360 official publications using its data as of 15 January 2020), we strongly recommend that researchers should carefully consider the affect that poor age constraints and style of binning may have on their analyses, especially if unfamiliar with the chosen time interval or when working at higher temporal resolutions. Incorrect dating or binning methods can strongly affect not only the shape of palaeodiversity curves, but also estimated divergence time, disparity trends, speciation and extinction rates in comparative phylogenetic studies.

Binning by formation: advantages and caveats

Palaeodiversity studies have historically been affected in their impartiality not only by the inherently biased nature of the fossil record but also by the lack of fully reproducible and objective metrics that can reduce the redundancy errors in their final outputs. In this study, we present a binning method that partially bypasses subjective choices (e.g. the use of stages vs 10 myr time bins) in favour of producing a range of chronostratigraphically-based diversity curves that begin to bridge the gap between palaeobiological and stratigraphic tools.

A common critique of palaeodiversity studies is their inability to capture fluctuations in diversity change and ecosystem dynamics at high temporal resolution (Smith & McGowan 2007). Formation binning is able to increase temporal resolution in the diversity curve of Late Cretaceous dinosaurs but, perhaps more importantly, it also provides the opportunity to easily compare multiple temporal frameworks and observe how they affect the shape of the diversity curve. Presenting a range of temporal frameworks can reveal how diversity curves are sensitive to different, objectively-chosen binning methods (Gibert & Escarguel 2017; Fan *et al.* 2020), and identify patterns that are volatile and should not be considered genuine palaeobiological signals. Results from this study (e.g. Fig. 7) show that the higher the resolution of time bins, the less pronounced the mid–late Campanian diversity spike is. This 'smoothening out' of the diversity curve (Fig. 7B–F) is indicative of how the tempo affects the perception of the mode of a diversity dynamic (Simpson 1944). This can be crucial in understanding the lead-up to a mass extinction, highlighting the structuring agents (geological) generating that record. By presenting a consensus of differently resolved diversity curves, we are better positioned to discern whether a sudden drop in diversity is the effect of a geologically instantaneous phenomenon (as is the case for the end-Cretaceous mass extinction) or an arbitrarily introduced bias from binning. We therefore strongly recommend that future studies discussing diversity curves present a range of binning options, even if not using a formation binning approach. Similarly, the method of binning occurrences (e.g. adding occurrences to all bins, the majority bin, or randomly selecting a bin) is shown to impact the shape of diversity curves (Figs 4–7). Consequently, we recommend reporting multiple curves using different occurrence binning methods, to establish which observed patterns are likely to be genuine geological or biological signals rather than artefacts of the chosen methodology.

The quantification of overlapping chronostratigraphical units and their contemporary arrangement in spatial bins (e.g. latitudinal bins; Figs 2–3) is also helpful in

investigating the spatial architecture behind the distribution of these formations, and the fossil record they subsequently provide (Holland 2017). For example, the rhythmic fluctuations in the extent of epeiric seas and the tectonic rearrangements in latitudinal belts of depositional environments can be contemporarily quantified and correlated with spatiotemporal variations in diversity, providing a new, refined tool to investigate the common cause hypothesis (Butler *et al.* 2010). Viewing formation and occurrence data in this format also forces the practitioner to think of the agents shaping the available geological record both spatially and temporally. It is being increasingly recognized that the spatial distribution of fossil occurrences can strongly impact our understanding of diversity and biogeography in deep time (Benson *et al.* 2016; Close *et al.* 2017; Brocklehurst *et al.* 2018; Brocklehurst & Fröbisch 2018; Dean *et al.* 2019). The tests that we applied to faunal endemicity in the WIB could be easily applied to other taxonomic groups and geographical regions through time to ensure that reported palaeobiogeographical signals are based on appropriate and significant fossil and geological evidence.

However, there are some caveats and potential issues associated with binning by formation. Despite resources such as Macrostrat, gathering up-to-date and accurate dating, distributional and correlative data for a large number of geological formations within an area is, whilst quicker than individually dating each fossil occurrence, time consuming and requires potentially sourcing information from multiple sources of data (e.g. museum collections, fossil databases and scientific literature). A similar enterprise may also not be applicable everywhere: while the WIB is famous for its mostly continuous, large-scale stratigraphic and palaeontological record that has been subject of intensive study for the best part of two centuries, other study areas may lack such data, and may in particular be deprived of good, well-dated age constraints. Diversity studies also commonly use multivariate models to establish the primary environmental drivers and controls on macroevolutionary processes (Benson & Mannion 2011), the proxies for which are typically binning at stage level; consequently, the creation of high-resolution formation-based bins may not permit these kinds of data to be used. However, binning by formation has the advantage that depositional environments are easily constrained based on the formations in the time bin. Thus, it is possible to explore correlations between diversity patterns and depositional environments, allowing a clearer understanding of the broad-scale environmental drivers that might be shaping diversity. Issue could also be taken with the use of formations as potential chronostratigraphic units, due to a lack of strict naming conventions and their differences in geographic and temporal scale between countries (Benton *et al.* 2011; Dunhill *et al.*

2018); however, we contend that their use in this approach is primarily for compiling approximately coeval rock units at a regional scale, which is appropriate for this purpose due to their frequent associations with either relational or specific age estimates.

Despite these issues, we argue that binning by formation provides a good first step for approaching palaeontological questions in a way that naturally integrates regional geological and anthropogenic information in order to understand the underlying structure of the fossil record. We recommend the use of this method not only to build palaeodiversity curves, but also as binning method for phylogenetic comparative work, when signals such as evolutionary (both speciation/origination and extinction) rates are impacted strongly by a temporal dimension. The creation of objective time bins at varying resolutions provides the resources to examine data more thoroughly, and hopefully provide researchers a new avenue through which to disentangle unresolved deep-time macroecological questions.

CONCLUSIONS

1. Formation binning represents a new way to divide time within regional studies through providing an objective temporal framework that inherently relies on local geological conditions, producing high resolution diversity curves and providing greater understanding of where fossil occurrence data is sourced in both time and space.

2. Through application of formation binning, the Maastrichtian decline in North American dinosaur diversity is observed to occur between 68 and 66 Ma, concurrent with broad geodynamic reorganization of sedimentary basins and an overall retreat of the Western Interior Seaway. This provides good evidence that these events were inherently linked, with apparent diversity probably reduced through change in regional preservational mode and/or decreased endemism. Further work investigating how lithology or palaeoenvironments are recorded across this interval will help to clarify controls on diversity trends.

3. The quantification of overlap and large time span in chronostratigraphic units within the Campanian supports the criticism of Lucas *et al.* (2016) and Fowler (2017) that hypotheses of faunal endemism in the Late Cretaceous of the WIB are probably the result of compilation of non-contemporaneous geological units.

4. Results produced using different binning methodologies and temporal resolutions confirm that perceived diversity trends can radically change based on the binning methods and temporal constraints used. We

strongly suggest presenting multiple binning methods when carrying out future diversity studies, to better understand the geological and anthropogenic controls on species richness through deep time.

DATA ARCHIVING STATEMENT

All data and R script necessary to run the analyses described in this paper are available on Github: https://github.com/ChristopherDavidDean/Formation_Binner
Data for this study are also available in the Dryad Digital Repository: https://doi.org/10.5061/dryad.tmpg4f4vg

Editor. Philip Mannion

REFERENCES

ALROY, J. 1998. Diachrony of mammalian appearance events: implications for biochronology. *Geology*, **26**, 23–26.

—— 2000. New methods for quantifying macroevolutionary patterns and processes. *Paleobiology*, **26**, 707–733.

—— 2008. Dynamics of origination and extinction in the marine fossil record. *Proceedings of the National Academy of Sciences*, **105**, 11536–11542.

—— 2010. Fair sampling of taxonomic richness and unbiased estimation of origination and extinction rates. *The Paleontological Society Papers*, **16**, 55–80.

—— MARSHALL, C. R., BAMBACH, R. K., BEZUSKO, K., FOOTE, M., FÜRSICH, F. T., HANSEN, T. A., HOLLAND, S. M., IVANY, L. C. and JABLONSKI, D. 2001. Effects of sampling standardization on estimates of Phanerozoic marine diversification. *Proceedings of the National Academy of Sciences*, **98**, 6261–6266.

—— ABERHAN, M., BOTTJER, D. J., FOOTE, M., FÜRSICH, F. T., HARRIES, P. J., HENDY, A. J., HOLLAND, S. M., IVANY, L. C. and KIESSLING, W. 2008. Phanerozoic trends in the global diversity of marine invertebrates. *Science*, **321**, 97–100.

ARCHIBALD, J. D. 2014. What the dinosaur record says about extinction scenarios. *Geological Society of America Special Papers*, **505**, 213–224.

—— CLEMENS, W. A., PADIAN, K., ROWE, T., MACLEOD, N., BARRETT, P. M., GALE, A., HOLROYD, P., SUES, H.-D. and ARENS, N. C. 2010. Cretaceous extinctions: multiple causes. *Science*, **328**, 973.

BARRETT, P. M., McGOWAN, A. J. and PAGE, V. 2009. Dinosaur diversity and the rock record. *Proceedings of the Royal Society B*, **276**, 2667–2674.

BENEVENTO, G. L., BENSON, R. B. and FRIEDMAN, M. 2019. Patterns of mammalian jaw ecomorphological disparity during the Mesozoic/Cenozoic transition. *Proceedings of the Royal Society B*, **286**, 20190347.

BENSON, R. B. and MANNION, P. D. 2011. Multi-variate models are essential for understanding vertebrate diversification in deep time. *Biology Letters*, **8**, 127–130.

—— BUTLER, R. J., ALROY, J., MANNION, P. D., CARRANO, M. T. and LLOYD, G. T. 2016. Near-stasis in the long-term diversification of Mesozoic tetrapods. *PLoS Biology*, **14**, e1002359.

BENTON, M. J. 2008. How to find a dinosaur, and the role of synonymy in biodiversity studies. *Paleobiology*, **34**, 516–533.

—— DUNHILL, A. M., LLOYD, G. T. and MARX, F. G. 2011. Assessing the quality of the fossil record: insights from vertebrates. *Geological Society, London, Special Publications*, **358**, 63–94.

BERRY, K. 2017. New paleontological constraints on the paleogeography of the Western Interior Seaway near the end of the Cretaceous (late Campanian–Maastrichtian) with a special emphasis on the paleogeography of southern Colorado, U.S.A. *Rocky Mountain Geology*, **52**, 1–16.

BOWRING, S. A., ERWIN, D. H. and ISOZAKI, Y. 1999. The tempo of mass extinction and recovery: the end-Permian example. *Proceedings of the National Academy of Sciences*, **96**, 8827–8828.

BROCKLEHURST, N. 2018. An examination of the impact of Olson's extinction on tetrapods from Texas. *PeerJ*, **6**, e4767.

—— and FRÖBISCH, J. 2018. The definition of bioregions in palaeontological studies of diversity and biogeography affects interpretations: Palaeozoic tetrapods as a case study. *Frontiers in Earth Science*, **6**, 200.

—— DUNNE, E. M., CASHMORE, D. D. and FRÖBISCH, J. 2018. Physical and environmental drivers of Paleozoic tetrapod dispersal across Pangaea. *Nature Communications*, **9**, 1–12.

BRUSATTE, S. L., BUTLER, R. J., PRIETO-MÁRQUEZ, A. and NORELL, M. A. 2012. Dinosaur morphological diversity and the end-Cretaceous extinction. *Nature Communications*, **3**, 1–8.

—— —— BARRETT, P. M., CARRANO, M. T., EVANS, D. C., LLOYD, G. T., MANNION, P. D., NORELL, M. A., PEPPE, D. J., UPCHURCH, P. and WILLIAMSON, T. E. 2015. The extinction of the dinosaurs. *Biological Reviews*, **90**, 628–642.

BUTLER, R. J., BENSON, R. B., CARRANO, M. T., MANNION, P. D. and UPCHURCH, P. 2010. Sea level, dinosaur diversity and sampling biases: investigating the 'common cause' hypothesis in the terrestrial realm. *Proceedings of the Royal Society B*, **278**, 1165–1170.

—— BRUSATTE, S. L., ANDRES, B. and BENSON, R. B. 2012. How do geological sampling biases affect studies of morphological evolution in deep time? A case study of pterosaur (Reptilia: Archosauria) disparity. *Evolution*, **66**, 147–162.

CAVIN, L., TONG, H., BOUDAD, L., MEISTER, C., PIUZ, A., TABOUELLE, J., AARAB, M., AMIOT, R., BUFFETAUT, E., DYKE, G., HUA, S. and LE LOEUFF, J. 2010. Vertebrate assemblages from the early Late Cretaceous of southeastern Morocco: an overview. *Journal of African Earth Sciences*, **57**, 391–412.

CHIARENZA, A. A. and CAU, A. 2016. A large abelisaurid (Dinosauria, Theropoda) from Morocco and comments on the Cenomanian theropods from North Africa. *PeerJ*, **4**, e1754.

—— MANNION, P. D., LUNT, D. J., FARNSWORTH, A., JONES, L. A., KELLAND, S.-J. and ALLISON, P. A. 2019. Ecological niche modelling does not support climatically-driven dinosaur diversity decline before the Cretaceous/Paleogene mass extinction. *Nature Communications*, **10**, 1091.

CLEARY, T. J., BENSON, R. B., EVANS, S. E. and BARRETT, P. M. 2018. Lepidosaurian diversity in the Mesozoic–Palaeogene: the potential roles of sampling biases and environmental drivers. *Royal Society Open Science*, **5**, 171830.

CLOSE, R. A., BENSON, R. B., UPCHURCH, P. and BUTLER, R. J. 2017. Controlling for the species-area effect supports constrained long-term Mesozoic terrestrial vertebrate diversification. *Nature Communications*, **8**, 15381.

—— EVERS, S. W., ALROY, J. and BUTLER, R. J. 2018. How should we estimate diversity in the fossil record? Testing richness estimators using sampling-standardised discovery curves. *Methods in Ecology & Evolution*, **9**, 1386–1400.

—— BENSON, R. B., ALROY, J., BEHRENSMEYER, A. K., BENITO, J., CARRANO, M. T., CLEARY, T. J., DUNNE, E. M., MANNION, P. D. and UHEN, M. D. 2019. Diversity dynamics of phanerozoic terrestrial tetrapods at the local-community scale. *Nature Ecology & Evolution*, **3**, 590.

COBBAN, W. A., WALASZCZYK, I., OBRADOVICH, J. D. and MCKINNEY, K. C. 2006. *A USGS zonal table for the Upper Cretaceous middle Cenomanian–Maastrichtian of the Western Interior of the United States Based on ammonites, inoceramids, and radiometric ages*. US Geological Survey, Open-File Report, **1250**, 46 pp.

COHEN, K. M., HARPER, D. A. T. and GIBBARD, P. L. 2020. ICS International Chronostratigraphic Chart 2020/01. International Commission on Stratigraphy, IUGS. https://stratigraphy.org

DE CELIS, A., NARVÁEZ, I. and ORTEGA, F. 2019. Spatiotemporal palaeodiversity patterns of modern crocodiles (Crocodyliformes: Eusuchia). *Zoological Journal of the Linnean Society*, zlz038.

DEAN, C. D., MANNION, P. D. and BUTLER, R. J. 2016. Preservational bias controls the fossil record of pterosaurs. *Palaeontology*, **59**, 225–247.

—— ALLISON, P. A., HAMPSON, G. J. and HILL, J. 2019. Aragonite bias exhibits systematic spatial variation in the Late Cretaceous Western Interior Seaway, North America. *Paleobiology*, **45**, 571–597.

—— CHIARENZA, A. A. and MAIDMENT, S. C. R. 2020. Data from: Formation binning: a new method for increased temporal resolution in regional studies, applied to the Late Cretaceous dinosaur fossil record of North America. *Dryad Digital Repository*. https://doi.org/10.5061/dryad.tmpg4f4vg

DUNHILL, A. M. and WILLS, M. A. 2015. Geographic range did not confer resilience to extinction in terrestrial vertebrates at the end-Triassic crisis. *Nature Communications*, **6**, 1–8.

—— HANNISDAL, B., BROCKLEHURST, N. and BENTON, M. J. 2018. On formation-based sampling proxies and why they should not be used to correct the fossil record. *Palaeontology*, **61**, 119–132.

DUNNE, E. M., CLOSE, R. A., BUTTON, D. J., BROCKLEHURST, N., CASHMORE, D. D., LLOYD, G. T. and BUTLER, R. J. 2018. Diversity change during the rise of tetrapods and the impact of the 'Carboniferous rainforest collapse'. *Proceedings of the Royal Society B*, **285**, 20172730.

EBERTH, D. A. and BRAMAN, D. R. 2012. A revised stratigraphy and depositional history for the Horseshoe Canyon Formation (Upper Cretaceous), southern Alberta plains. *Canadian Journal of Earth Sciences*, **49**, 1053–1086.

—— and CURRIE, P. J. 2005. Vertebrate taphonomy and taphonomic modes. 453–477. *In* CURRIE, P. J. and KOPPELHUS, E. B. (eds). *Dinosaur provincial park: A spectacular ancient ecosystem revealed*. Indiana University Press.

FAN, J., SHEN, S., ERWIN, D. H., SADLER, P. M., MACLEOD, N., CHENG, Q., HOU, X., YANG, J., WANG, X., WANG, Y., ZHANG, H., CHEN, X., LI, G., ZHANG, Y., SHI, Y., YUAN, D., CHEN, Q., ZHANG, L., LI, C. and ZHAO, Y. 2020. A high-resolution summary of Cambrian to Early Triassic marine invertebrate biodiversity. *Science*, **367**, 272–277.

FANTI, F., CAU, A., MARTINELLI, A. and CONTESSI, M. 2014. Integrating palaeoecology and morphology in theropod diversity estimation: a case from the Aptian-Albian of Tunisia. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **410**, 39–57.

FASTOVSKY, D. E. and BERCOVICI, A. 2016. The Hell Creek Formation and its contribution to the Cretaceous–Paleogene extinction: a short primer. *Cretaceous Research*, **57**, 368–390.

—— and SHEEHAN, P. M. 2005. The extinction of the dinosaurs in North America. *GSA Today*, **15**, 4–10.

—— HUANG, Y., HSU, J., MARTIN-McNAUGHTON, J., SHEEHAN, P. M. and WEISHAMPEL, D. B. 2004. Shape of Mesozoic dinosaur richness. *Geology*, **32**, 877–880.

FOWLER, D. W. 2017. Revised geochronology, correlation, and dinosaur stratigraphic ranges of the Santonian-

Maastrichtian (Late Cretaceous) formations of the Western Interior of North America. *PLoS One*, 12, e0188426.

—— and FOWLER, E. A. F. 2019. Transitional evolutionary forms and stratigraphic trends in chasmosaurine ceratopsid dinosaurs: evidence from the Campanian of New Mexico. *BioRχiv*, 854794.

GATES, T. A., SAMPSON, S. D., ZANNO, L. E., ROBERTS, E. M., EATON, J. G., NYDAM, R. L., HUTCHISON, J. H., SMITH, J. A., LOEWEN, M. A. and GETTY, M. A. 2010. Biogeography of terrestrial and freshwater vertebrates from the Late Cretaceous (Campanian) Western Interior of North America. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 291, 371–387.

—— PRIETO-MÁRQUEZ, A. and ZANNO, L. E. 2012. Mountain building triggered Late Cretaceous North American megaherbivore dinosaur radiation. *PLoS One*, 7, e42135.

GIBERT, C. and ESCARGUEL, G. 2017. Evaluating the accuracy of biodiversity changes through geologic times: from simulation to solution. *Paleobiology*, 43, 667–692.

GUILLERME, T. and COOPER, N. 2018. Time for a rethink: time sub-sampling methods in disparity-through-time analyses. *Palaeontology*, 61, 481–493.

HAUBOLD, H. 1990. Dinosaurs and fluctuating sea levels during the Mesozoic. *Historical Biology*, 4, 75–106.

HOLLAND, S. M. 1995. The stratigraphic distribution of fossils. *Paleobiology*, 21, 92–109.

—— 2017. Structure, not Bias. *Journal of Paleontology*, 91, 1315–1317.

—— and PATZKOWSKY, M. E. 2002. Stratigraphic variation in the timing of first and last occurrences. *Palaios*, 17, 134–146.

HOPKINS, M. J. 2017. How well does a part represent the whole? A comparison of cranidial shape evolution with exoskeletal character evolution in the trilobite family Pterocephaliidae *Palaeontology*, 60, 309–318.

KLIMASZEWSKI, S. M. 1998. Further data concerning Dominican amber jumping plant lice [Homoptera, Psylloidea]. *Acta Biologica Silesiana*, 32, 19–27.

KOCSIS, Á. T., REDDIN, C. J., ALROY, J. and KIESSLING, W. 2019. The r package divDyn for quantifying diversity dynamics using fossil sampling data. *Methods in Ecology & Evolution*, 10, 735–743.

KOMINZ, M. A., BROWNING, J. V., MILLER, K. G., SUGARMAN, P. J., MIZINTSEVA, S. and SCOTESE, C. R. 2008. Late Cretaceous to Miocene sea-level estimates from the New Jersey and Delaware coastal plain coreholes: an error analysis. *Basin Research*, 20, 211–226.

LARSON, D. W., BROWN, C. M. and EVANS, D. C. 2016. Dental disparity and ecological stability in bird-like dinosaurs prior to the end-cretaceous mass extinction. *Current Biology*, 26, 1325–1333.

LEHMAN, T. M. 1987. Late Maastrichtian paleoenvironments and dinosaur biogeography in the western interior of North America. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 60, 189–217.

LILLEGRAVEN, J. A. and EBERLE, J. J. 1999. Vertebrate faunal changes through Lancian and Puercan time in southern Wyoming. *Journal of Paleontology*, 73, 691–710.

LLOYD, G. T. 2011. A refined modelling approach to assess the influence of sampling on palaeobiodiversity curves: new support for declining Cretaceous dinosaur richness. *Biology Letters*, 8, 123–126.

—— and FRIEDMAN, M. 2013. A survey of palaeontological sampling biases in fishes based on the Phanerozoic record of Great Britain. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 372, 5–17.

LOEWEN, M. A., IRMIS, R. B., SERTICH, J. J., CURRIE, P. J. and SAMPSON, S. D. 2013. Tyrant dinosaur evolution tracks the rise and fall of Late Cretaceous oceans. *PLoS One*, 8, e79420.

LUCAS, S. G., SULLIVAN, R. M., LICHTIG, A. J., DALMAN, S. G. and JASINSKI, S. E. 2016. Late Cretaceous dinosaur biogeography and endemism in the Western Interior Basin, North America: a critical re-evaluation. *New Mexico Museum of Natural History & Science Bulletin*, 71, 195–213.

LYSON, T. R. and LONGRICH, N. R. 2011. Spatial niche partitioning in dinosaurs from the latest Cretaceous (Maastrichtian) of North America. *Proceedings of the Royal Society B*, 278, 1158–1164.

LYSTER, S. J., WHITTAKER, A. C., ALLISON, P. A., LUNT, D. J. and FARNSWORTH, A. 2020. Predicting sediment discharges and erosion rates in deep time — examples from the late Cretaceous North American continent. *Basin Research*, published online 27 February.

MALLON, J. C. 2019. Competition structured a Late Cretaceous megaherbivorous dinosaur assemblage. *Scientific Reports*, 9, 1–18.

MILLER, A. I. and FOOTE, M. 1996. Calibrating the Ordovician radiation of marine life: implications for Phanerozoic diversity trends. *Paleobiology*, 22, 304–309.

MILLER, K. G., KOMINZ, M. A., BROWNING, J. V., WRIGHT, J. D., MOUNTAIN, G. S., KATZ, M. E., SUGARMAN, P. J., CRAMER, B. S., CHRISTIE-BLICK, N. and PEKAR, S. F. 2005. The Phanerozoic record of global sea-level change. *Science*, 310, 1293–1298.

MITCHELL, J. S., ROOPNARINE, P. D. and ANGIELCZYK, K. D. 2012. Late Cretaceous restructuring of terrestrial communities facilitated the end-Cretaceous mass extinction in North America. *Proceedings of the National Academy of Sciences*, 109, 18857–18861.

NEWHAM, E., BENSON, R., UPCHURCH, P. and GOSWAMI, A. 2014. Mesozoic mammaliaform diversity: the effect of sampling corrections on reconstructions of evolutionary dynamics. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 412, 32–44.

OGG, J. G., HINNOV, L. A. and HUANG, C. 2012. Cretaceous. 793–853. *In* GRADSTEIN, F., OGG, J. G., SCHMITZ, M. D. and OGG, G. M. (eds). *The geologic time scale 2012*. Elsevier.

PEARSON, D. A., SCHAEFER, T., JOHNSON, K. R., NICHOLS, D. J. and HUNTER, J. P. 2002. Vertebrate biostratigraphy of the Hell Creek formation in southwestern North Dakota and northwestern South Dakota. *Geological Society of America Special Paper*, 361, 145–167.

PETERS, S. E. and FOOTE, M. 2001. Biodiversity in the Phanerozoic: a reinterpretation. *Paleobiology*, 27, 583–601.

—— and McCLENNEN, M. 2016. The Paleobiology Database application programming interface. *Paleobiology*, **42**, 1–7.

PHILLIPS, J. 1860. *Life on the Earth: Its origin and succession.* Macmillan & Co., Cambridge.

R CORE TEAM. 2018. *R: a language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org.

RASMUSSEN, C. M. Ø., KRÖGER, B., NIELSEN, M. L. and COLMENAR, J. 2019. Cascading trend of Early Paleozoic marine radiations paused by Late Ordovician extinctions. *Proceedings of the National Academy of Sciences*, **116**, 7207–7213.

RAUP, D. M. 1972. Taxonomic diversity during the Phanerozoic. *Science*, **177**, 1065–1071.

ROBECK, H. E., MALEY, C. C. and DONOGHUE, M. J. 2000. Taxonomy and temporal diversity patterns. *Paleobiology*, **26**, 171–187.

RUSSELL, L. S. 1975. Mammalian faunal succession in the Cretaceous System of western North America. *Geological Association of Canada Special Paper*, **13**, 137–160.

RUSSELL, D. A. 1982. A paleontological consensus on the extinction of the dinosaurs? 401–406. *In* SILVER, L. T. and SCHULTZ, P. H. (eds). *Geological implications of impacts of large asteroids and comets on the Earth.* Geological Society of America, Special Paper, **190**.

RYAN, M. J. and EVANS, D. C. 2005. Ornithischian dinosaurs. 312–348. *In* CURRIE, P. J. and KOPPELHUS, E. B. (eds). *Dinosaur provincial park: A spectacular ancient ecosystem revealed.* Indiana University Press.

SAKAMOTO, M., VENDITTI, C. and BENTON, M. J. 2017. 'Residual diversity estimates' do not correct for sampling bias in palaeodiversity data. *Methods in Ecology & Evolution*, **8**, 453–459.

SAMPSON, S. D., LOEWEN, M. A., FARKE, A. A., ROBERTS, E. M., FORSTER, C. A., SMITH, J. A. and TITUS, A. L. 2010. New horned dinosaurs from Utah provide evidence for intracontinental dinosaur endemism. *PLoS One*, **5**, e12292.

SARJEANT, W. A. S. and CURRIE, P. J. 2001. The "Great Extinction" that never happened: the demise of the dinosaurs considered. *Canadian Journal of Earth Sciences*, **38**, 239–247.

SCHULTE, P., ALEGRET, L., ARENILLAS, I., ARZ, J. A., BARTON, P. J., BOWN, P. R., BRALOWER, T. J., CHRISTESON, G. L., CLAEYS, P., COCKELL, C. S., COLLINS, G. S., DEUTSCH, A., GOLDIN, T. J., GOTO, K., GRAJALES-NISHIMURA, J. M., GRIEVE, R. A. F., GULICK, S. P. S., JOHNSON, K. R., KIESSLING, W., KOEBERL, C., KRING, D. A., MACLEOD, K. G., MATSUI, T., MELOSH, J., MONTANARI, A., MORGAN, J. V., NEAL, C. R., NICHOLS, D. J., NORRIS, R. D., PIERAZZO, E., RAVIZZA, G., REBOLLEDO-VIEYRA, M., REIMOLD, W. U., ROBIN, E., SALGE, T., SPEIJER, R. P., SWEET, A. R., URRUTIA-FUCUGAUCHI, J., VAJDA, V., WHALEN, M. T. and WILLUMSEN, P. S. 2010. The Chicxulub asteroid impact and mass extinction at the Cretaceous-Paleogene boundary. *Science*, **327**, 1214–1218.

SEILACHER, A., REIF, W.-E. and WESTPHAL, F. 1985. Sedimentological, ecological and temporal patterns of fossil Lagerstätten. *Philosophical Transactions of the Royal Society B*, **311**, 5–24.

SHEEHAN, P. M., FASTOVSKY, D. E., HOFFMANN, R. G., BERGHAUS, C. B. and GABRIEL, D. L. 1991. Sudden extinction of the dinosaurs: latest Cretaceous, upper Great Plains, USA. *Science*, **254**, 835–839.

SILVESTRO, D., ZIZKA, A., BACON, C. D., CASCALES-MIÑANA, B., SALAMIN, N. and ANTONELLI, A. 2016. Fossil biogeography: a new model to infer dispersal, extinction and sampling from palaeontological data. *Philosophical Transactions of the Royal Society B*, **371**, 20150225.

SIMPSON, G. G. 1944. *Tempo and mode in evolution.* Columbia University Press.

SLATTERY, J. S., COBBAN, W. A., McKINNEY, K. C., HARRIES, P. J. and SANDNESS, A. L. 2015. Early Cretaceous to Paleocene paleogeography of the Western Interior Seaway: the interaction of eustasy and tectonism. *Wyoming Geological Association Guidebook*, **2015**, 22–60.

SLOAN, R. E., RIGBY, J. K., VAN VALEN, L. M. and GABRIEL, D. 1986. Gradual dinosaur extinction and simultaneous ungulate radiation in the Hell Creek Formation. *Science*, **232**, 629–633.

SMITH, A. B. and McGOWAN, A. J. 2007. The shape of the Phanerozoic marine palaeodiversity curve: how much can be predicted from the sedimentary rock record of Western Europe? *Palaeontology*, **50**, 765–774.

STARRFELT, J. and LIOW, L. H. 2016. How many dinosaur species were there? Fossil bias and true richness estimated using a Poisson sampling model. *Philosophical Transactions of the Royal Society B*, **371**, 20150219.

TENNANT, J. P., CHIARENZA, A. A. and BARON, M. 2018. How has our knowledge of dinosaur diversity through geologic time changed through research history? *PeerJ*, **6**, e4417.

TUCKER, R. T., ZANNO, L. E., HUANG, H.-Q. and MAKOVICKY, P. J. 2020. A refined temporal framework for newly discovered fossil assemblages of the upper Cedar Mountain Formation (Mussentuchit Member), Mussentuchit Wash, Central Utah. *Cretaceous Research*, **110**, 104384.

UPCHURCH, P., MANNION, P. D., BENSON, R. B., BUTLER, R. J. and CARRANO, M. T. 2011. Geological and anthropogenic controls on the sampling of the terrestrial fossil record: a case study from the Dinosauria. *Geological Society, London, Special Publications*, **358**, 209–240.

VAVREK, M. J. and LARSSON, H. C. 2010. Low beta diversity of Maastrichtian dinosaurs of North America. *Proceedings of the National Academy of Sciences*, **107**, 8265–8268.

WILF, P., JOHNSON, K. R. and HUBER, B. T. 2003. Correlated terrestrial and marine evidence for global climate changes before mass extinction at the Cretaceous-Paleogene boundary. *Proceedings of the National Academy of Sciences*, **100**, 599–604.

WILSON, G. P. 2005. Mammalian faunal dynamics during the last 1.8 million years of the Cretaceous in Garfield County, Montana. *Journal of Mammalian Evolution*, **12**, 53–76.

APPENDIX 1

Number, mean length and standard deviation of stage, substage and formation produced bins

Method	Resolution	Number of bins	Mean bin length (myr)	Standard deviation
Stages	NA	7	6.71	3.87
Sub-stages	NA	19	3.48	1.59
SG1	2	18	2.19	0.94
SG1	3	14	2.82	1.49
SG1	4	9	4.39	1.92
SG2	2	18	2.19	0.85
SG2	3	13	3.04	1.09
SG2	4	9	4.39	1.85