



THE MOSASAUR FOSSIL RECORD THROUGH THE LENS OF FOSSIL COMPLETENESS

by DANIEL A. DRISCOLL¹ , ALEXANDER M. DUNHILL² , THOMAS L. STUBBS¹  and MICHAEL J. BENTON^{1,*} 

¹School of Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Road, Bristol, BS8 1RJ, UK; dan.driscoll@bristol.ac.uk, tom.stubbs@bristol.ac.uk, mike.benton@bristol.ac.uk

²School of Earth & Environment, University of Leeds, Leeds, LS2 9JT, UK

*Corresponding author

Typescript received 9 January 2018; accepted in revised form 22 April 2018

Abstract: The quality of the fossil record affects our understanding of macroevolutionary patterns. Palaeodiversity is filtered through geological and human processes; efforts to correct for these biases are part of a debate concerning the role of sampling proxies and standardization in biodiversity models. We analyse the fossil record of mosasaurs in terms of fossil completeness as a measure of fossil quality, using three novel, correlating metrics of fossil completeness and 4083 specimens. A new qualitative measure of character completeness (QCM) correlates with the phylogenetic character completeness metric. Mean completeness by species decreases with specimen count; average completeness by substage varies significantly. Mean specimen completeness is higher for species-named fossils than those identified to genus and family. We consider the effect of tooth-only specimens. Importantly, we find that completeness of species does not correlate with completeness of specimens. Completeness varies by palaeogeography: North

American specimens show higher completeness than those from Eurasia and Gondwana. These metrics can be used to identify exceptional preservation; specimen completeness varies significantly by both formation and lithology. The Belgian Ciply Formation displays the highest completeness; clay lithologies show higher completeness values. Neither species diversity nor sea level correlates significantly with fossil completeness. A generalized least squares (GLS) analysis using multiple variables agrees with this result, but reveals two variables with significant predictive value for modelling averaged diversity: sea level, and mosasaur and plesiosaur-bearing formations (the latter is redundant with diversity). Mosasaur completeness is not driven by sea level, nor does completeness limit the mosasaur diversity signal.

Key words: marine reptiles, mosasaur, fossil record quality, fossil completeness, sea level, palaeodiversity.

MOSASAURIDAE was a relatively short-lived, but diverse and abundant clade of marine squamates that radiated in Late Cretaceous epicontinental seas and died out at the Cretaceous–Palaeogene (K/Pg) boundary (Debraga & Carroll 1993). The rise of mosasaurids (called ‘mosasaurs’ throughout the paper) followed dramatic changes in the marine reptile fauna (Stubbs & Benton 2016), including decreases in disparity of plesiosaurs in the Late Jurassic (Benson & Druckenmiller 2014) as well as the extinction of cryptoclidid plesiosaurs, ichthyosaurs (Bardet 1994; Fischer *et al.* 2012) and thalattosuchian crocodiles (Young *et al.* 2010) in the Early to mid-Cretaceous. Mosasaurids (aigialosaurids and dolichosaurids) arose in the Cenomanian as relatively small swimming reptiles, followed by true mosasaurs in the Turonian (Bardet *et al.* 2008). As in other groups of marine reptiles (Massare 1994), the Mosasauridae showed increasing adaptations to the marine environment through time (Motani 2009). The

average body size of mosasaurs increased through the Late Cretaceous, from 1–2 m in early semi-terrestrial forms, to a gigantic 14–17 m in later forms (Polcyn *et al.* 2014; Stubbs & Benton 2016). They became increasingly efficient swimmers and filled niches vacated by some of the aforementioned pelagic marine predators after their extinction (Motani 2005; Lindgren *et al.* 2007, 2009, 2011, 2013; Houssaye *et al.* 2013). Mosasaurs thrived in many marine environments (Kiernan 2002), from rocky shores to pelagic shelves, including fresh water environments (Holmes *et al.* 1999), and by the latest Cretaceous, they were the apex predators in many complex ocean ecosystems (Sørensen *et al.* 2013). Accordingly, mosasaur fossils have a widespread stratigraphical and global geographical distribution in a variety of lithologically distinct Upper Cretaceous marine formations (Russell 1967).

Marine reptiles have figured in several studies that have contributed to the debate about how to address biases in

the fossil record (e.g. Benson *et al.* 2010; Benson & Butler 2011; Cleary *et al.* 2015; Tutin & Butler 2017). Does the fossil record provide a reasonable picture of mosasaur evolution (Polcyn *et al.* 2014), or is the record substantially biased by the idiosyncrasies of preservation and collection (Benson *et al.* 2010; Benson & Butler 2011)? Benson *et al.* (2010) identified serious megabiases affecting all Cretaceous marine reptiles, including mosasaurs, and argued that their palaeodiversity signal was dependent on geological sampling biases, meaning that the raw data said little about their true diversity. Part of this result depended on residual diversity estimates using a method that has since been severely criticized (Dunhill *et al.* 2014, 2018; Brocklehurst 2015; Sakamoto *et al.* 2017). Re-analysis led Benson & Butler (2011) to suggest that shallow marine tetrapods at least, including most mosasaurs, showed close correlation between diversity and sea level and continental area. Benson & Butler (2011) interpreted this as a ‘common cause’ effect (Peters 2005), analogous to a species-area effect; the fossil record and palaeodiversity of marine reptiles fluctuated simultaneously as sea level rose and fell. These alternate viewpoints leave an open question: is the mosasaur fossil record a fair representation of their true biological signal or not?

One approach to understanding inadequacies of the fossil record is to consider the specimens themselves: are they equally complete through all times and places, or do they show variation (Benton *et al.* 2004; Smith 2007)? For example, Mannion & Upchurch (2010) suggested that measures of fossil completeness could be used alongside other sampling proxies to investigate the quality of the fossil record. Fossil completeness studies attempt to quantify the quality of fossil specimens by assigning numerical metrics that reflect the percentage of skeletal or phylogenetic character elements present in individual fossils or whole groups of fossils.

Many recent analyses have used measures of fossil completeness. In taphonomic studies, completeness can reflect post-mortem conditions and transport (Beardmore *et al.* 2012a, b). Aquatic versus terrestrial deposits may preserve differently (Verrière *et al.* 2016) and, more broadly, completeness may be related to lithology (Cleary *et al.* 2015). Completeness may be related to body size; large fossils may be collected more often (Brown *et al.* 2013), or small associated fossils may be preserved better at times (Brocklehurst *et al.* 2012). Completeness may be affected by sea level (Mannion & Upchurch 2010; Cleary *et al.* 2015; Tutin & Butler 2017). Completeness can be used to measure collecting and naming biases through historical time (Benton 2008a, b; Mannion & Upchurch 2010; Walther & Fröbisch 2013; Tutin & Butler 2017) or as a direct metric to assess confidence in fossil record data in a single basin through a key event (Benton *et al.* 2004). Finally, the

fossil record of diversity may be unbiased, or biased by completeness, either inversely (Smith 2007; Brocklehurst & Fröbisch 2014) or directly (Dean *et al.* 2016).

In this study, we explore a database of over 4000 mosasaur specimens and apply novel methods of coding fossil completeness to test whether fossil completeness is biasing the measured richness of these organisms. We find that specimen completeness varies enormously geographically but is not correlated with species diversity or sea level. We find that completeness does not limit the diversity signal in the mosasaur record.

MATERIAL AND METHOD

Data

Specimens. A mosasaur specimen database (Driscoll *et al.* 2018, data A, B) includes all scored specimens of Mosasauridae from collection visits and literature descriptions, comprising 4083 mosasaur specimens. Mosasaur material is housed in at least 112 institutions (Driscoll *et al.* 2018, table S1); 448 specimens were seen first-hand in these collections (Driscoll *et al.* 2018, table S2), including many referred, cited and holotype specimens. Examination confirmed their description in the literature, even if some of the elements showed abrasion or minor disintegration. In a few cases, elements originally described with the specimen were not found on visiting the museum, and this was noted in assessing skeletal completeness. Most specimens were identifiable, and scorings of the holotype in the literature and observed first hand were identical, providing confidence that measurements taken from the literature can be accurate. Catalogue descriptions as well as photos from museum online collections databases (AMNH, GPIT, MCZ, SDMNH, TMP, UAVPL, UCMP, USNM, UVER and YPM) were also used, and files containing museum databases were obtained from LACM, FMNH, ALNHM and TMP.

Additional specimen data were obtained from publications and monographs, including original descriptions of holotypes (Driscoll *et al.* 2018, table S3), as well as secondary descriptions of non-type materials (e.g. Lydekker 1888; Camp 1942; Russell 1967; Schultze *et al.* 1985; Kuypers *et al.* 1998; Bardet 2012; and others listed in Driscoll *et al.* 2018, table S7). No publicly inaccessible or undocumented material was used in the study. In total, over 4300 specimens were identified for study, but some were excluded because of poor morphological data or lack of illustration.

In this study, we used different subsets of the specimen lists. In many cases, we considered all 4083 specimens. In other cases, we considered just those specimens that could be assigned to named species, and excluded those that

were assigned to genus alone (e.g. *Mosasaurus* sp.) or to an even more general taxon (e.g. Mosasauridae indet.) 1044 specimens were attributed to Mosasauridae indeterminate (i.e. family level), 731 specimens were identified to generic level, and 2308 to species level. In the specimen list (Driscoll *et al.* 2018, data B), the specimens 1–843 and 1878–4073 are assigned to a named genus or named genus and species, and specimens 844–1887 are termed simply ‘Mosasauridae indet.’ Specimens 2426–2554, for example are ‘*Mosasaurus* sp.’ More than 1100 of the 4083 specimens consist only of isolated teeth, and these were included and excluded in different analyses.

Stratigraphic ranges. The stratigraphical position of many historical mosasaur specimens is unknown. In some formations in which mosasaurs commonly occur, the stratigraphy and age have been revised (e.g. Everhart 2001; Jagt 2005), and the revised date was used for allocation to time bins. Mosasaurs generally occur in marine rocks, and often in close association with zone fossils such as belemnites, ammonites or foraminifera, so enabling correlation with short-term time zones that can be tied to radioisotopic ages in the standard marine time scale. We compiled a list of 135 mosasaur-bearing formations from the specimen search and cross-checked the age and stratigraphy of formations with the stratigraphic literature (Driscoll *et al.* 2018, data H). Many formation ages were already accurately represented in the primary literature.

The specimens were datable to different degrees of precision. 1726 mosasaur specimens were datable to substage, and 2357 specimens were dated at best to two or more substages. Because of the large amount of data, there were no substage time bins that did not contain precisely assignable specimens.

Species list. A list of valid species was assembled based on the primary scientific literature (see Appendix), paying special attention to apomorphy-based descriptions. Only species with clear taxonomic assignment and little disagreement on taxonomy were used in this study. Our species list includes 74 valid species, and it agrees broadly with a recent, independent compilation (Polcyn *et al.* 2014).

Specimen completeness metrics

Background. One of the most exact methods for scoring skeletal completeness in vertebrates is to count the number of elements present compared to the total number of bones in the skeleton. This has been done in some taphonomic studies, including *Archaeopteryx* (Kemp & Unwin 1997), a Triassic prolacertiform (Casey *et al.* 2007), and a

Miocene salamander (McNamara *et al.* 2012). However, this method is time-consuming and impractical when many specimens are compared.

Other quantitative methods have been developed for dealing with larger sample sizes. Mannion & Upchurch (2010) presented two approaches to measure fossil completeness in sauropods: a Skeletal Completeness Metric (SCM) that records the proportional completeness of skeletons against a roster of elements that ought to be present, and a Character Completeness Metric (CCM) that reports the number of phylogenetically informative characters that are reported for each taxon. They suggested that SCM might be a more useful metric in taphonomic studies comparing preservation in different geographical zones or facies, etc., and CCM would be a better tool for comparing diversity patterns through time.

Both SCM and CCM were subdivided into three individual measures: the best specimen of a taxon, termed SCM1 or CCM1, the type specimen SCMs or CCMs and a composite specimen that includes all preserved elements of the taxon from any number of specimens, termed SCM2 and CCM2. These scores can be averaged over all taxa in a time bin, or all taxa in a geological formation or geographical region, or for all representatives of a species or genus, whether they occur in a single time bin or not.

Another method, designed by Beardmore *et al.* (2012a, b) for scoring fossil preservation in marine crocodylomorphs, compares disarticulation and completeness, which are related to environmental and preservational factors that were present at the time of death and burial. The unmodified Beardmore index divides the skeleton into anatomical regions, giving each region equal weight. This method can be quantitative, scoring every element present, but also allows estimation of proportions of regions present; this might therefore be called a semi-quantitative scoring system.

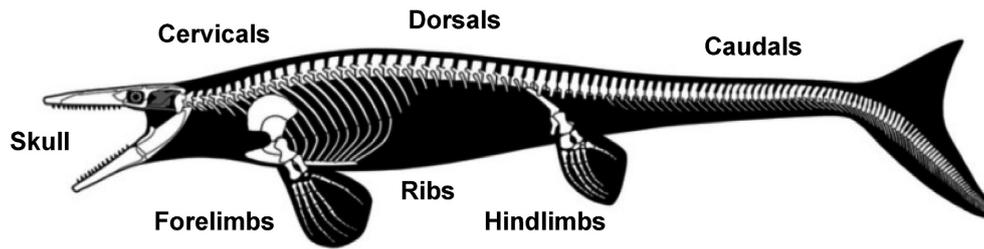
Cleary *et al.* (2015) used SCM1 and SCM2 (modified for the laterally crushed nature of ichthyosaur fossils), but they also implemented a modified Beardmore Skeletal Completeness Metric (BSCM) in an investigation of fossil completeness in ichthyosaurs. These authors also divided BSCM into best (BSCM1) and composite (BSCM2) specimen per species, and averaged these values over all species assigned to stage level time bins.

Qualitative approaches can also be used to score fossil quality. For example, Benton *et al.* (2004) and Benton (2008b) measured dinosaur specimen completeness using the ratio of incomplete material (isolated elements or collections of bones) to complete material, such as skulls or complete skeletons. This approach has been used successfully in several studies (Fountain *et al.* 2005; Smith 2007). Metrics such as SCM and CCM are more accurate than qualitative scores (Brocklehurst *et al.* 2012), but

qualitative metrics can be useful for comparisons of diverse taxa or large sample sizes.

Completeness metrics. We used three completeness metrics. The Taphonomic Completeness Metric (TCM) is based on Beardmore (2012a, b) and is a non-weighted method (Fig. 1). The mosasaur skeleton is divided into nine anatomical regions, namely the skull, limbs (two forelimbs and two hindlimbs), vertebral column (cervical, dorsal, caudal) and ribs, and each region is given an arbitrary maximum score of 4, giving a total possible TCM of 36.

The Qualitative Completeness Metric (QCM) is based on Benton’s (Benton *et al.* 2004; Benton 2008a, b) qualitative description of dinosaur completeness and is weighted so that skulls and jaws are afforded a higher weight than post-cranial elements, which is in proportion to the distributions of characters used in phylogenetic analysis (e.g. Bell 1993). QCM is presented here (Table 1) as an estimate of character completeness when it is not possible to examine every character present on individual elements. This is in accordance with some previous studies (albeit using CCM) where each anatomical element present was similarly assumed to contain all its characters



Taphonomic Completeness Metric (TCM)

Score:	0	1	2	3	4
Skull	Head is absent	A few elements remaining	Approx. 1/2 elements remaining	A few elements missing	Complete skull
Cervical vertebrae	0 vertebrae	1 or 2	3 or 4	5 or 6	7 vertebrae
Forelimbs	All bones absent	1 of 4 subunits present	2 of 4 subunits absent	3 of 4 subunits present	Humerus, radius, ulna and most phalanges
Ribs	No ribs	Extensive loss of ribs	Approx. 1/2 ribs	Approx. 3/4 ribs	Most ribs present
Dorsal vertebrae	0–2 vertebrae	10–25%	25–50%	50–75%	At least 12 present
Hindlimbs	All bones absent	1 of 4 subunits present	2 of 4 subunits present	3 of 4 subunits present	Femur, tibia, fibula and most phalanges
Caudal vertebrae	0–4 vertebrae	10–25%	25–50%	50–75%	>75%

FIG. 1. Beardmore scoring method for mosasaur taphonomic completeness metric (TCM). Each complete region of the skeleton (skull, ribs, two forelimbs, two hindlimbs and vertebrae including cervicals, dorsals, or caudals) is each worth 4 points, for a maximum possible score of 36. Beardmore scoring can assess taphonomy. Scoring is as follows: (1) Count or approximate number of elements for each region; (2) In incomplete skeletons, score one for any girdle elements; (3) The vertebral column can achieve a maximum score of 12, if all vertebrae are present and complete, and a minimum score of 2, for multiple undifferentiated vertebrae; (4) Any portion of a skull + any portion of a jaw or tooth = 2. 5. Sum scores for each region. Skeletal image © Scott Hartman.

TABLE 1. QCM method for scoring completeness.

	Skull	Skeleton
Fragments	2	1
Incomplete	3	2
Almost complete	5	
Complete	6	3

Presence or absence of a skull weights the score greatly. The skull contains 61% of characters used in the phylogenetic analysis of mosasaurs. A jaw fragment or tooth scores two points. Add total of parts present. There are nine points possible.

(e.g. Brocklehurst *et al.* 2012). In contrast, other analyses (e.g. Dean *et al.* 2016) used only the number of characters that could be observed.

In general terms, QCM is like CCM. Regions with higher numbers of phylogenetic characters are given greater weight in both. The phylogenetic character list was derived from the character matrix of Bell (Bell 1997; Bell & Polcyn 2005). This cladistic data matrix was selected because it has more mosasaur characters than other matrices (e.g. Leblanc *et al.* 2012). A table of anatomical elements and the number of their associated characters was compiled by anatomical region for a test subset of 26 specimens of representative species (Driscoll *et al.* 2018, data D). For each of these specimens, the total score over all anatomical regions was compared to the QCM fossil completeness metric (Driscoll *et al.* 2018, data E). This comparison tests the pre-weighted character total per specimen against an estimate of character completeness provided by the QCM. Although not necessary for the analysis, a weighted value of the character scores (assuming a maximum value of 9) for each specimen is listed also, for comparison to QCM.

The final scoring method, Informal Completeness Metric (ICM), allows the inclusion of specimens that are associated only with general descriptions such as ‘skull’, or ‘axial elements’, or ‘appendicular skeleton’. The total possible ICM score is set arbitrarily at 5, with any mention of a skull scoring three points and any mention of axial and appendicular parts scoring one point each (Driscoll *et al.* 2018, table S4).

Of the 4083 mosasaur specimens, 375 could be scored for only one or two of the three metrics (TCM, QCM and ICM). We compared all three methods as measures of fossil completeness. The equivalence of TCM values using all specimens versus those exactly datable to single substages was also tested. Completeness scores were assigned to all holotype specimens (TCMh, QCMh and ICMh) and the best specimens (TCMb, QCMb and ICMb) of each species. In addition, a composite score (TCMc, QCMc and ICMc) for each species was calculated (Driscoll *et al.* 2018, data C–E).

Mosasaur fossil completeness

Time series. Mean completeness scores were compiled and divided into time bins equivalent to Upper Cretaceous stratigraphic substages (Gradstein *et al.* 2012). For both species and genus-level specimens, sampled in-bin taxa were compiled at the substage level and diversity was calculated. The terms ‘richness’ and ‘diversity’ are considered to be equivalent in this paper. Mean sea level for each Upper Cretaceous substage was calculated from Miller *et al.* (2005).

We assessed mean completeness for all specimens (TCMall, QCMall, ICMall) and for those specimens identified to species level (TCMsp, QCMsp, ICMsp), so excluding material only identifiable to higher taxonomic levels. Since there were so many specimens consisting of teeth alone, we compared completeness values across the above two time series, both with and without specimens consisting of teeth alone.

The mean completeness of all specimens for each species (TCMtot) was averaged over the time bins where those species occur (TCMav, QCMav, ICMav). This time series was compared to those derived from all specimens (TCMall, QCMall, ICMall), and to specimens named to species level (TCMsp, QCMsp, ICMsp). This compared the utility of using average completeness values assigned to whole species (in many cases from various time bins) to those derived from sampled-in-bin specimens. For clarity, a description of all the completeness metrics described in this study is listed (Table 2). No best, holotype or composite specimen scores were used for any time series analysis.

Mosasaur species and generic diversities are calculated based on specimen occurrences only, in substage-level time bins, so we do not use first-to-last ranges or include any Lazarus taxa, in this part of the analysis. In addition, we included five mosasaur species that were only assignable imprecisely to a range of two or three time bins, so we present also an ‘averaged’ species diversity curve that includes these taxa counted as fractions. For example, *Plotosaurus bennisoni* is dated to the latest early Maastrichtian and/or earliest late Maastrichtian, so its diversity is counted as 0.5 in both substages. We present the exact species and genus diversity curves as well as the ‘averaged’ species diversity curve, together with comparisons among all curves and with sea level and completeness. TCMsp, QCMsp and ICMsp were compared with species and generic diversity, but not to ‘averaged’ diversity so as not to make spurious comparisons between time bins that do not contain equivalent specimens.

Predicting diversity and completeness. To more fully understand the relationship between diversity and completeness, we used a multiple regression technique to compare the relationships between explanatory variables. A substage-level sampling proxy for explaining diversity and completeness

Metric	Sub-metric	Comment
TCM		Taphonomic Completeness Metric; total of scores from 8 regions
QCM		Qualitative Completeness Metric; regions are weighted by phylogenetic character density
ICM		Informal Completeness Metric; scored using only skull, axial and appendicular portions as regions
	(TCM, QCM, ICM)tot	Total mean completeness of a species, disregarding time bins
	(TCM, QCM, ICM)sp	Mean metric from specimens named to species assignable to single time bins
	(TCM, QCM, ICM)all	Mean metric from all specimens assignable to single bins, regardless of taxonomy
	(TCM, QCM, ICM)av	The 'tot' metric is calculated for each species, and this is averaged over every species in a time bin
	(TCM, QCM, ICM)h	The metric of the holotype specimen
	(TCM, QCM, ICM)b	The metric of the best specimen
	(TCM, QCM, ICM)c	The metric of the composite; calculated using the best specimen plus any extra elements found in other specimens.

TABLE 2. Description of completeness metrics used in this paper.

was created and tested using mosasaur and plesiosaur-bearing formations (MPBF). These formations (Driscoll *et al.* 2018, data 1) were drawn from our mosasaur database and Upper Cretaceous plesiosaur data from unpublished research. We used generalized least squares (GLS) to check the relationship between mean TCM, diversity, sea level, formations and age by modelling TCM and diversity as a function of the other variables (i.e. $TCM_{sp} \sim \text{species diversity} + \text{sea level} + \text{MPBF} + \text{age}$; and 'averaged' diversity $\sim TCM_{av} + \text{sea level} + \text{MPBF} + \text{age}$). We provide the raw data used for this analysis (Driscoll *et al.* 2018, tables S5A, S6B).

Determinants of fossil completeness. Mean completeness scores were compared across several classes of taxonomic, biological, palaeogeographical, and lithological variables. TCM (instead of QCM) was used as a measure of preservation as affected by taphonomy. For most of these categorical variables, the data were much richer for specimens labelled as species, so TCM_{sp} was the completeness metric used. Mean TCM was compared between specimens allocated to a maximum taxonomic resolution of family, genus or species. Tests both included and excluded specimens consisting of only teeth.

Mean TCM_{sp} was assessed across different lithologies by assigning mosasaur-bearing formations to the categories chalk, sandstone, limestone, or clay, based on the predominant rock type of each formation. The specific lithology of individual specimens was not used. Mean TCM_{sp} values of a few well-known and prolific mosasaur-bearing formations were calculated and compared. Differences in

mean TCM_{sp} between palaeogeographical regions (i.e. Eurasia, Gondwana, North America) were also analysed.

Finally, estimated body size for sample species was compared to the species mean completeness using mosasaur length estimates taken from Polcyn *et al.* (2014). For this analysis, mosasaurs were divided into three informal size groups: small (1–4 m), medium (4.5–7.5 m) and large (8 m or longer) because we did not have good quality individual measurements for each taxon, and for those with large sample sizes, we would have to consider a range of body sizes.

Sampling. A few representative sampling metrics were compiled to test the relationship between palaeontological sampling effort and fossil completeness. The number of specimens per species was used as one measure of sampling because it could be related to collector effort or availability of samples. The number of Google Scholar 'hits' was tested as a measure of scientific interest (we recorded these on 1 December 2017). The number of years since first discovery (i.e. naming of a species) was used to test scientific effort over historical time. Sampling or study effort could be related to absolute body length (Polcyn *et al.* 2014; Driscoll *et al.* 2018, data A), and this was also compared with other variables.

Sampling and/or completeness might be related to rock outcrop area. For North American formations, this information is available at <https://macrostrat.org/>. The maps for representative North American mosasaur-bearing formations were double-checked against actual specimen locations. The average TCM_{sp} by formation was

compared to the rock area of these formations and their mosasaur species diversity. The number of formations (n) necessary to be confident about our results using the lowest p -value (0.35) and highest r_s (0.6) was calculated using the method of Bonett & Wright (2000). Their work showed that a value of $n = 4$ is the smallest sample size that is adequate at this level of confidence; we have a value of $n = 5$ for our data.

Statistical tests

Differences in specimen completeness among categorical data (i.e. taxonomic rank, body size, lithology, palaeogeography etc.) were assessed using Wilcoxon tests and Kruskal–Wallis tests. Relationships between numerical data and paired time series were assessed using Spearman rank correlation tests. The correlation between completeness values across specimens was double-checked using the Kendall's tau-b test, which corrects for ties in the ranks across the thousands of specimens included. Time series were detrended using generalized differencing prior to correlation tests (using the Graeme Lloyd's `gen.diff` function (<http://www.graemetlloyd.com/methgd.html>), and these were corrected for false discovery rate (FDR) using the method of Benjamini & Hochberg (1995). Time series of completeness metrics were correlated with mosasaur diversity, sea level, and the various sampling proxies, with the aim of determining whether specimen completeness has any bearing on mosasaur diversity, and whether specimen completeness is driven by external factors such as sea level or sampling intensity. All statistical analyses were performed in R (R Core Team 2016).

Generalized least squares (GLS) is a multiple regression method for estimating the unknown parameters in a linear regression model, and it can be used when there is a certain degree of correlation between the residuals in a regression model. GLS has an advantage over pairwise tests of correlation as it allows multiple explanatory variables to be examined simultaneously and allows the addition or removal of additional variables to be assessed quantitatively. Variables tested included diversity, sea level, TCM, age and formations.

GLS models were fitted in R using the package `nlme` (Pinheiro *et al.* 2017). As there was evidence for heterogeneity in the spread of the residuals in some of the explanatory variables, we applied a number of variance structures to the data and tested for the best fitting model using the Akaike information criterion (AIC). The best fitting model for predicting diversity contains a power of the covariate variance structure applied to the age data and the best fitting model for predicting TCM contains a fixed variance structure applied to the age data. Models were also fitted with an auto-regressive model of order 1 (AR-1)

correlation structure, which models the residual at time s as a function of the residual of time $s - 1$ (Zuur *et al.* 2009). The models with the AR-1 structure were worse fits than the models without. This is because of the common increasing trend of diversity, formations and sea level through time. We therefore present both sets of models, with and without the autocorrelation structure applied to the age parameter. Model fitting was achieved by comparing the full models with models that drop each explanatory variable in turn and perform a likelihood ratio test. This informs whether the dropped term has a significant influence on the fit of the model (Zuur *et al.* 2009).

Institutional abbreviations. ALNHM, Alabama Natural History Museum, Tuscaloosa, Alabama, USA; AMNH, American Museum of Natural History, New York, USA; FMNH, Field Museum of Natural History, Chicago, Illinois, USA; GPIT, University of Tübingen, Tübingen, Germany; LACM, Los Angeles County Museum of Natural History, Los Angeles, California, USA; MCZ, Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts, USA; SDMNH, San Diego Museum of Natural History, San Diego, California, USA; TMP, Royal Tyrell Museum of Palaeontology, Drumheller, Alberta, Canada; UAVPL, University of Alberta Vertebrate Paleontology Lab, Edmonton, Alberta, Canada; UCMP, University of California Museum of Paleontology, Berkeley, California, USA; USNM, Smithsonian Institution National Museum of Natural History, Washington DC, USA; UPI, Uppsala University Palaeontological Institute, Uppsala, Sweden; UVER, University of Vermont Zadock Thompson Natural History Collection, Burlington, Vermont, USA; YPM, Yale Peabody Museum, New Haven, Connecticut, USA.

RESULTS

Completeness scores

The mean completeness scores for all specimens per species show a broad range of values for different taxa: averages per taxon range as follows: TCM (1–21 out of 36), QCM (1–7 out of 9) and ICM (1–5 out of 5). Summaries are given of the overall mean TCM, QCM and ICM scores for all mosasaur species (Driscoll *et al.* 2018, data A), for all mosasaur specimens (Driscoll *et al.* 2018, data B) and by holotype, best specimen and species composite scores (Driscoll *et al.* 2018, data F, G and H, respectively). An overview of the completeness of the various species and exemplary specimens is reviewed below and summarized in Table 3.

When the total character scores from the mosasaur phylogenetic matrix are compared to QCM there is a very highly significant positive correlation (Spearman: $r_s = 0.925$, $p \ll 0.001$) derived from our 26 representative specimens, which remains significant after FDR correction for multiple comparisons.

TABLE 3. Summary of representative mosasaur completeness scores.

	Metric	Species	Specimen	Value
Highest mean TCM	TCMtot	<i>H. bernardi</i>		13.67
Average holotype TCM	TCMh	All		8.1
Average composite TCM	TCMc	All		15.7
Highest holotype TCM	TCMh	<i>E. sternbergii</i>	UPI R163	32
Most complete specimen	TCM	<i>P. tympaniticus</i>	YPM 58129	36
Highest mean QCM	QCMtot	<i>T. nopscai</i>		5.6
Average holotype QCM	QCMh	All		8
Highest best specimens	QCMb	<i>T. proriger</i> , <i>P. tympaniticus</i>	YPM 58129, AMNH FR221	9
Mean composite score	QCMc	All		6.2
Highest composite score	QCMc	Many		36

Statistical comparison of the completeness scores (TCM, QCM, ICM) across all specimens shows highly significant positive correlations that were also significant after FDR correction: TCM vs ICM (Spearman: $r_s = 0.74$, $p \ll 0.001$), QCM vs ICM (Spearman: $r_s = 0.72$, $p \ll 0.001$) and TCM vs QCM (Spearman: $r_s = 0.49$, $p \ll 0.001$). The correlations were also very highly significant using Kendall's tau-b test, with all p-values < 0.001 . The metrics are so closely correlated with each other that all can be regarded as equivalent metrics for recording fossil completeness data.

A comparison of the completeness data (TCM) for all 4083 specimens versus the 1726 specimens datable to substage shows significant discrimination, as indicated by the Kruskal–Wallis test ($\chi^2 = 81.174$, $df = 35$, $p \ll 0.001$). These sets are not equivalent. At first, this seems surprising, since the medians for these values are both 1; the mean TCM for single substage specimens is 2.18, and for all specimens is 2.28. But this result is influenced by the fact that the distribution of precisely datable specimens is skewed according to the level of taxonomic assignment.

Only 154/1034 (15%) of taxa assigned to Mosasauridae could be dated precisely to substage, whereas for those identified to genus, this rises to 146/731 (20%), and to 62% for specimens identified to species. However, when one compares the two sets of data using only specimens named to species, there is no significant difference when using the TCM of all specimens and those datable to precise substages (Kruskal–Wallis: $\chi^2 = 41.94$, $df = 32$, $p = 0.124$). This shows that TCMsp, which necessarily leaves out many un-sampled and/or un-datable specimens, can be trusted as a fair representation of the mean TCM for the set of all specimens identified to species level. Note that comparisons like this, where some data sets compared are subsets of each other, might be inadvisable; but in this case, the null hypothesis was that the partial set of datable specimens should be equivalent to the set of all specimens. We confirm this here in the comparison of substage-dated specimens with the sample of all specimens.

Mosasaur fossil completeness

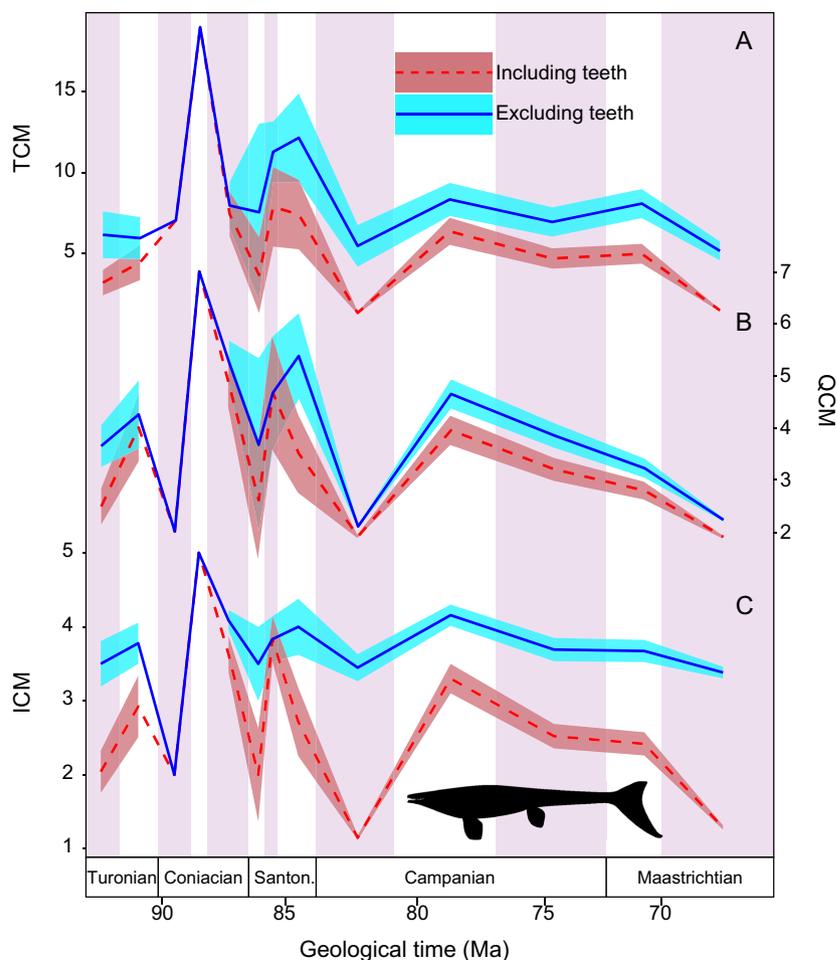
Time series. Mean completeness of all mosasaur specimens varies through time, and there are close similarities in the overall patterns through time for all three completeness metrics (Fig. 2). The same is true for mosasaur specimens named to species through time (Fig. 3).

The time series of completeness metrics by substage all correlate significantly with each other, both for all specimens and for those named as species. These correlations remain significant after FDR correction. There is no bias in the overall pattern of mosasaur completeness according to whether specimens have been assigned to named species or not. Further, many of the rises and falls in the respective time series (Figs 2, 3) show statistically significant differences between substages, both for named species and for all specimens.

All patterns for the different metrics appear broadly similar, whether isolated teeth are included or not, but the values without such teeth are inevitably always higher (over 1400 of the 4083 specimens comprise isolated teeth only). The differences between time bins continue to be significant (Figs 2, 3) regardless of whether the data include or exclude specimens consisting of only a single tooth. The metric that is least changed by the removal of tooth-only specimens is QCM.

A difference between the 'with teeth' and 'without teeth' time series occurs in the late Santonian for QCM, where the value excluding teeth is considerably higher (Fig. 2B). In the early Campanian, a disproportionate number of tooth-only specimens (probably from *Prognathodon lutugini*) shift the ICM curve lower (Fig. 2C). The results are comparable also for taxa named to species (Fig. 3), although the data set is smaller.

FIG. 2. Mosasaur specimen completeness by substage, for specimens whose age is known. Mean completeness by substage was calculated, according to: A, TCMall; B, QCMall; C, ICMall. Confidence intervals are 95%, except for the early–middle Coniacian because there are very few specimens. Completeness was plotted including and excluding specimens consisting of only teeth. Statistics for specimens including teeth: TCMall vs QCMall, Spearman, $r_s = 0.71$, $p < 0.01$; TCMall vs ICMall, Spearman, $r_s = 0.77$, $p < 0.01$; QCMall vs ICMall, Spearman, $r_s = 0.97$, $p \ll 0.001$. When comparing differences between time bins: TCMall, Kruskal–Wallis, $\chi^2 = 597.73$, $df = 12$, $p < 0.001$; QCMall, Kruskal–Wallis, $\chi^2 = 219.08$, $df = 12$, $p < 0.001$; ICMall, Kruskal–Wallis, $\chi^2 = 529.56$, $df = 12$, $p < 0.001$. For specimens excluding teeth alone, when comparing differences between time bins: TCMall, Kruskal–Wallis, $\chi^2 = 30.05$, $df = 12$, $p = 0.003$; QCMall, Kruskal–Wallis, $\chi^2 = 537.22$, $df = 12$, $p < 0.001$; ICMall, Kruskal–Wallis, $\chi^2 = 38.38$, $df = 12$, $p = 0.001$. (Silhouette: Matt Crick; PhyloPic.) Colour online.



There is an overall slightly declining trend in fossil completeness through time, with completeness scores in the middle Cretaceous somewhat higher than those in the Maastrichtian. However, the trend is modest, and perhaps dominated by the downturn from the early to late Maastrichtian. For the whole data set, all three metrics show a high point in the middle Coniacian (Fig. 2), but this is based on a single specimen, and so the value is hardly meaningful. Further, there are no named mosasaur species datable exactly to the early Coniacian (Fig. 3). For all specimens, the lowest average completeness is in the early Santonian. The earliest true mosasaurs, found in the Turonian, have average completeness. Later, completeness peaks in the late Santonian and drops to its lowest points in the early Campanian and late Maastrichtian.

For all three metrics, mean completeness (Table 4) for named species (TCMsp, QCMsp, ICMsp) through time is correlated with overall completeness (TCMall, QCMall, ICMall) and remains significant after FDR correction. Note that TCMsp and ICMsp are significantly correlated with TCMav and ICMav after FDR correction, but other completeness metrics, whether based on named specimens

or all specimens, do not show significant correlations with average species completeness (TCMav, QCMav, ICMav) analysed in specific time bins, after FDR correction. For example, QCMsp does not correlate with QCMav. This indicates that caution is required in interpreting time series that assume completeness values derived from whole species-based values, such as averaged species completeness (TCMav, QCMav, ICMav) are equivalent to specimen-based completeness averaged in single time bins.

Both the sampled in-bin and ‘averaged’ species diversity curves correlate significantly with the generic diversity curves even after FDR correction (Table 4, Fig. 4A). Generic diversity rises and falls, then upturns sharply from the middle Santonian onwards, and gently rises through the Campanian to Maastrichtian, with only a slight increase through that span of nearly 20 myr. The species curves roughly follow the same pattern early on. The two curves show a dramatic drop in diversity after the Turonian, corresponding to a low number of assignable specimens; and in the early Coniacian none can be named to species level (e.g. *Tylosaurus* indet.) Mosasaurid species diversity rises during the middle Coniacian, dropping in the early

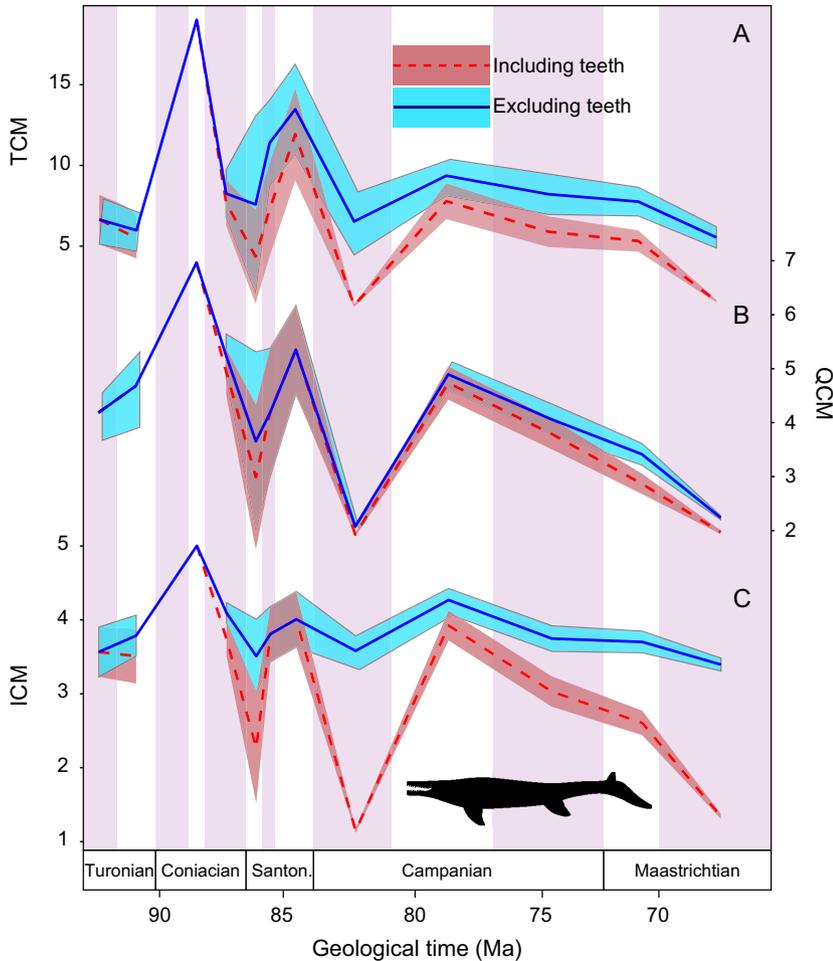


FIG. 3. Comparing mosasaur species completeness by substage. Mean and 95% confidence intervals are plotted, and curves are plotted with and without teeth. Statistics for all specimens including only teeth: TCMsp vs QCMsp, Spearman, $r_s = 0.85$, $p < 0.001$; TCMsp vs ICMsp, Spearman, $r_s = 0.89$, $p \ll 0.001$; QCMsp vs ICMsp, Spearman, $r_s = 0.96$, $p \ll 0.001$. In comparing completeness between time bins: TCMsp, Kruskal–Wallis, $\chi^2 = 598.7904$, $df = 11$, $p < 0.001$; QCMsp, Kruskal–Wallis, $\chi^2 = 332.5136$, $df = 11$, $p < 0.001$; ICMsp, Kruskal–Wallis: $\chi^2 = 596.8376$, $df = 11$, $p < 0.001$. Statistics for specimens excluding teeth alone: TCMsp, Kruskal–Wallis, $\chi^2 = 28.13$, $df = 11$, $p = 0.003$; QCMsp, Kruskal–Wallis, $\chi^2 = 518.28$, $df = 11$, $p < 0.001$; ICMsp, Kruskal–Wallis, $\chi^2 = 30.87$, $df = 11$, $p = 0.001$. (Silhouette: Iain Reid; PhyloPic.) Colour online.

Santonian, but then generally rises to the K/Pg boundary, with a slight drop in the early Maastrichtian. Note that species diversity is at its highest during the late Maastrichtian, with no hint of a pre-mass extinction diversity drop.

For comparison, sea level (Fig. 4B) fluctuates in the Turonian through early Santonian, concurrently with the variability in mosasaur diversity. The lowest sea level occurs in the early Santonian, but it rises until the early Campanian, when it reaches its highest level. A drop in sea level occurs in the early Maastrichtian, at the same time as a small drop in species diversity. Species diversity is high during some times of relatively high sea level, but none of the three diversity time series curves correlates in a statistically significant way with mean sea level (Table 4) in this comparison. In like manner, none of the measures of completeness shows any statistically significant correlation with sea level (Table 4).

Predicting diversity and completeness. GLS model fitting shows that a combination of all variables (i.e. sea level, MPBF, TCM, age) best predict averaged diversity (Driscoll *et al.* 2018, table S6A). While sea level and MPBFs

appear to be positively related to averaged diversity (i.e. higher sea level or more MPBF sampled equals higher diversity), age is negatively related to diversity, that is, mosasaur diversity increases through time (Driscoll *et al.* 2018, table S6B). However, once we account for autocorrelation, we find that the best fitting model contains only sea level and MPBF, both of which are positively related to averaged diversity (Tables 5, 6). The best fitting model for predicting TCMsp consists of all variables (e.g. species diversity, sea level, MPBF and age) (Driscoll *et al.* 2018, table S6A). However, none of these variables appears to be significantly associated with TCMsp (Driscoll *et al.* 2018, table S6B). When we account for autocorrelation, the best fitting model for TCMsp contains all variables, but none is significant (Tables 5, 6).

Determinants of fossil completeness

Taxonomic rank. As might be predicted, mean TCM increases as one narrows taxonomy from family to genus to species (Fig. 5), and these differences are

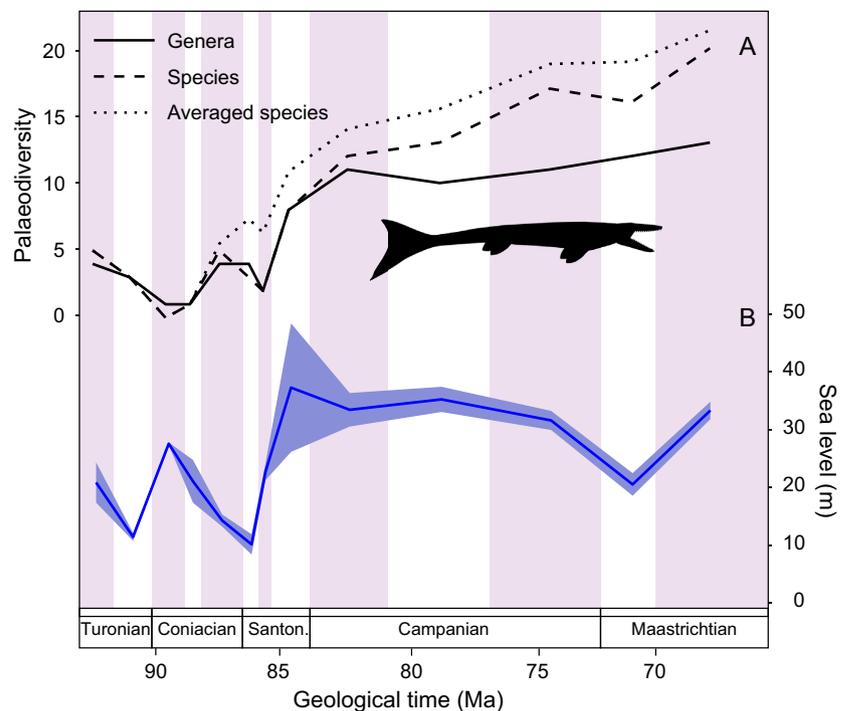
TABLE 4. Mean completeness comparisons by substage.

	Sea level	TCMall	QCMall	ICMall	TCMsp	QCMsp	ICMsp	Genus	Species
TCMall	-0.06								
QCMall	-0.2								
ICMall	-0.24								
TCMsp	0.04	0.75**							
QCMsp	-0.06		0.84**						
ICMsp	-0.01			0.69**					
Genus	0.04	-0.17	0.03	-0.04	0.15	0.24	0.37		
Species	0.22	-0.43	-0.01	-0.13	0.05	0.24	0.26	0.8**	
Averaged	0.13	-0.49	-0.04	-0.15				0.8**	0.98**
TCMav		0.58			0.81**				
QCMav			0.68*			0.53			
ICMav				0.6			0.64*		

Mean substage sea level (Miller *et al.* 2005), species diversity (Species), generic diversity (Genus), and time-averaged species diversity (Averaged) curves were compared. TCMav, QCMav and ICMav represent mean substage completeness using average completeness for all specimens per species (TCMtot, QCMtot, ICMtot) in the substage. TCMall, QCMall and ICMall are averages of all specimens assigned to substage, regardless of species status. TCMsp, QCMsp, and ICMsp are average completeness values per substage using only specimens with a designated species epithet.

*Significant at $p < 0.05$; **significant after false discovery rate (FDR) correction using method of Benjamini & Hochberg (1995).

FIG. 4. Mosasaur diversity and sea level through time. A, generic and species diversity lines include only specimens with an exact substage assignment; the ‘averaged species’ curve includes both in-bin species records, Lazarus taxa plus species based on specimens that could not be assigned to a stratigraphic substage with confidence, and so are averaged over all possible bins (e.g. two possible time bins, each species rated 0.5 per bin; three possible time bins, each species rated 0.33 per bin). B, mean substage sea level from data in Miller *et al.* (2005), and showing 95% confidence intervals. (Silhouette: Craig Dylke; PhyloPic.) Colour online.



statistically highly significant. Interestingly though, when all specimens including teeth were used in the analysis, it was difficult to discriminate a significant difference in completeness between genera and species (Fig. 5B), even though there was discrimination among all three taxonomic categories when tested together. But, when the tests were repeated excluding tooth-only

specimens, there was clear discrimination (Fig. 5A). Among all 4083 specimens, 1044 were attributable to family only (Mosasauridae indet.), 731 to genera, and 2304 to species.

Lithology. There are highly significant differences in TCMsp among specimens preserved in different

TABLE 5. Summary of GLS multiple regression analysis, showing the full and best models for predicting both diversity and TCM with autocorrelation structure for age parameter.

Model	Parameters	AIC	BIC	Log likelihood
Full averaged diversity	TCM Sea level MPBF	66.183	66.739	-26.091
Best averaged diversity	Sea level MPBF	60.251	61.435	-24.126
Full TCMsp	Diversity Sea level MPBF	78.753	79.309	-32.376
Best TCMsp	Diversity Sea level MPBF	78.753	79.309	-32.376

AIC, Akaike information criterion; BIC, Bayesian information criterion.

TABLE 6. Summary of best-fitting GLS multiple regression models for predicting averaged diversity and TCMsp with autocorrelation structure for age parameter.

Response	Parameters	Value	SE	t	p
Averaged diversity	Intercept	-5.79	1.531	-3.782	0.004
	Sea level	0.246	0.05	5.039	0.001
	MPBF	0.521	0.024	21.774	<0.001
TCMsp	Intercept	7.01	4.312	1.625	0.14
	Species diversity	-0.707	0.587	1.131	0.29
	Sea level	0.17	0.15	-1.205	0.26
	MPBF	0.116	0.333	0.347	0.74

SE, standard error; t, coefficient of standard error.

lithologies, with fossils preserved in clays displaying higher completeness than those preserved in carbonate or coarse siliciclastic deposits (Fig. 6). These differences are much smaller when specimens consisting only of teeth are left out of the analysis.

Palaeogeographical region. The mean TCM of specimens identified as species varies significantly by palaeocontinental region, with specimens from North America showing higher completeness than those from Eurasia and Gondwana (Fig. 7). When tooth-only specimens are excluded, there are no statistical differences.

Formation. Fossil completeness as measured by TCMsp varies significantly between different geological formations. The Pierre Shale Formation in the western interior of the USA and the Craie de Ciply in Belgium have the most complete fossils (Fig. 8).

The Maastrichtian formations of the New Jersey Greensand and Maastricht Chalk yield the least complete specimens.

Body size. There are no significant differences in total mean species completeness (TCMtot) between different body size classes derived from the estimated average body length of the individual species concerned (Fig. 9).

Sampling. Average species completeness correlates significantly and inversely with the number of years elapsed since description, and inversely also with the number of specimens per species (Table 7). The average completeness per species compared to the completeness of the best specimen in a species correlates strongly for all three metrics. The best specimen influences the average for a whole species. The total number of specimens per species shows statistically significant positive variation with the number of years since description. The number of Google Scholar 'hits' for a species correlates strongly with the number of specimens, as well as years elapsed since description. There is a trend for Google Scholar 'hits' to increase with estimated mosasaur body length, which is not quite statistically significant using Spearman's rho ($r_s = 0.26$, $p = 0.052$).

Results for North American rock outcrop area (Table 7) show no correlations between regional diversity, formational outcrop area, or mean TCM by formation.

Overview of mosasaur fossil completeness. The mean completeness (TCMtot) by species using individual specimens ranged from 1.0 for those species known only from teeth, jaws or individual bones to a high of 13.67 for *Hainosaurus bernardi* (Driscoll *et al.* 2018, data A). The scores by specimen type (Table 3; Driscoll *et al.* 2018, data F-H) show that the average holotype completeness is 8.1 for TCMh, which is approximately equivalent to 25% of the skeleton in the average mosasaur holotype. The highest scoring holotype is 32 for UPI R163, *Eonatator sternbergii*. The lowest score for TCMh is 1.0 for several holotypes. The most complete specimen is not always a holotype. *Platecarpus tympaniticus* (YPM 58129) from the Kansas Chalk, with a TCMc of 36, is the most complete specimen in the database. However, at least five other specimens scored over 30. Several species have perfect composite scores, including some with soft tissue preservation (e.g. *Platecarpus tympaniticus* and *Tylosaurus proriger*).

The highest mean QCMtot was 5.6 for individual specimens of *Tethysaurus nopscai*. The mean holotype completeness (QCMh) is 4.5; thus 50% of the phylogenetic characters occur in the average type specimen. The highest QCMh is 8 (equivalent to a skull and most

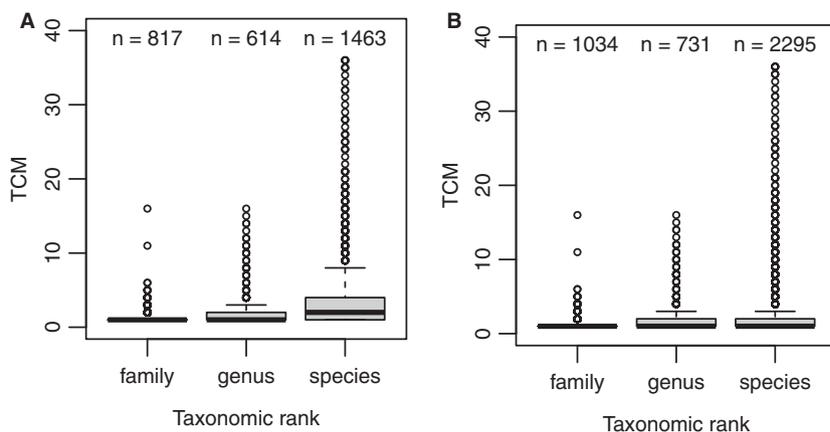
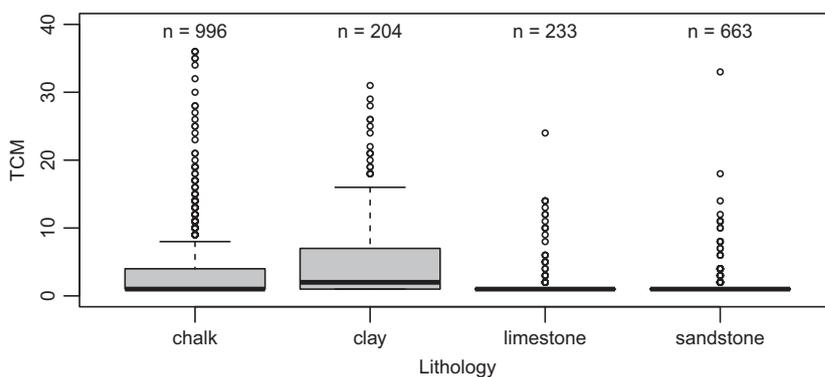


FIG. 5. Completeness by taxonomic rank. The mean completeness (TCM) was calculated for specimens in each category: Mososauridae indeterminate, specimens identified to genus or identified to species. There are highly significant differences in TCM when comparing all three different taxonomic ranks (Kruskal–Wallis, $\chi^2 = 95.62$, $df = 2$, $p < 0.001$). A, plot for specimens excluding those consisting only of teeth (Kruskal–Wallis: $\chi^2 = 248.64$, $df = 38$, $p < 2.2 \times 10^{-16}$); there are highly significant differences between groups, including genus and species. B, results for all specimens including those consisting of a single tooth; in this case, there was no significant difference in completeness between genus and species specimens, because the median for each group = 1 (Wilcoxon: $W = 731\ 307$, $p = 0.09$).

FIG. 6. Completeness by lithology. Where possible, specimens identified to species were assigned to the main lithology of their formation of origin. There are highly significant differences in TCM between different lithologies (Kruskal–Wallis, $\chi^2 = 364.44$, $df = 3$, $p < 0.001$). Differences remain when specimens consisting of only teeth are left out, but are barely significant (Kruskal–Wallis, $\chi^2 = 7.63$, $df = 3$, $p < 0.05$; plot not shown).



of the skeleton) for several holotypes: *Eonatator sternbergii*, *Clidastes propyhton*, *Mosasaurus missouriensis*, *Plotosaurus bennisoni* and *Latoplatecarpus willistoni*. The lowest QCMh for a type specimen is 1.0 (but this a lectotype) for *Goronyosaurus nigeriensis*. Of note, the composite score (QCMc) of *G. nigeriensis* is 8; multiple specimens make up for most of the elements missing in most individual fossils. The best specimens of both *Tylosaurus proriger* and *Platecarpus tympaniticus* both have QCMb scores of 9. The mean QCMc for all species is 6.2. This indicates that the composite character completeness of the average mosasaur species is approximately equivalent to the score for a skull of that species. Multiple species have a perfect QCMc. It should be noted, that at the time of this compilation, there were three species with a QCMc equal to only 2.0, the lowest composite score, equivalent to a jaw element, namely *Carinodens belgicus*, *Carinodens minimalmimar* and *Igdamanosaurus aegyptiacus*.

The average TCMc for composite specimens is 15.7 (about 44% of the skeleton). The mean composite score for ICMc is 4.5/6, equivalent to some skull, axial and limb elements available to describe the average mosasaur species.

Discussion

Describing the mosasaur fossil record. In this study, we have addressed one of the richest vertebrate fossil records. Mosasaurs have attracted study for over two centuries, with the first find, *Mosasaurus hoffmanni*, being described by Cuvier in 1808 (Russell 1967). Later collectors noted their huge abundance: it is said that O. C. Marsh collected over 2000 Kansas mosasaur specimens (Everhart 2000), and Ikejiri *et al.* (2013) counted 1563 Alabama mosasaur specimens. An estimate of ‘literally thousands’ of specimens of *Platecarpus* from Kansas has been

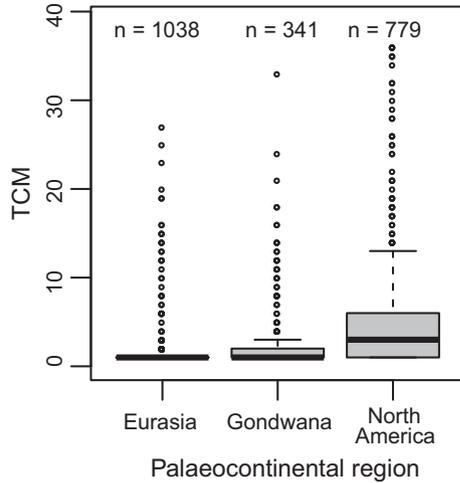


FIG. 7. Completeness by palaeogeographical region. Where possible, fossils named to species were divided by geographical origin. Because of a relative paucity of specimens, those from Africa, South America, Australia, New Zealand and Antarctica were included in a ‘Gondwana’ group. Mean TCM of species-named specimens showed highly significant differences between groups (Kruskal–Wallis, $\chi^2 = 701.46$, $df = 2$, $p < 0.001$). When tooth-only specimens are left out of analysis, the differences are no longer significant (Kruskal–Wallis, $\chi^2 = 0.9303$, $df = 2$, $p = 0.63$; plot not shown).

suggested (Konishi & Caldwell 2007). Our analysis of mosasaur diversity through time complements previous studies (Ross 2009; Polcyn *et al.* 2014).

It has been argued that skeletal completeness metrics can evaluate confidence in palaeontological data: as knowledge of the anatomy of a taxon becomes more complete, with increased numbers of specimens or more complete skeletons, confidence in taxonomic assignments improves (Mannion & Upchurch 2010). In an ideal world, palaeontologists would wait for relatively complete

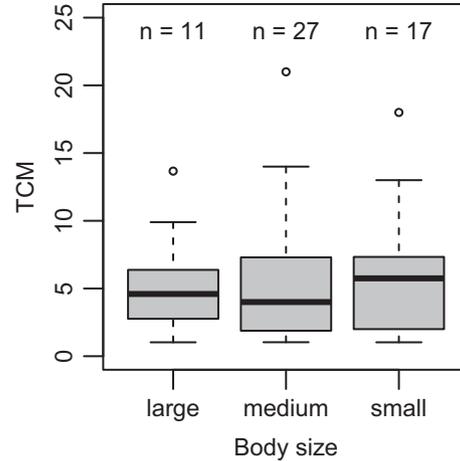


FIG. 9. Completeness by body size groups. There were no statistical differences between mean completeness (TCMtot) values of species among small, medium or large mosasaurs.

specimens before applying new taxonomic names, but in fact new genera and species are often based on poor material. For example, in the case of echinoids (Smith 2007), incomplete fossils were named more frequently than complete specimens, and in the case of dinosaurs (Benton 2008a) the naming of species in the nineteenth century was prodigious but quite inaccurate; holotypes were on average much more incomplete before 1960 than after that date (Benton 2008b). On the other hand, Brocklehurst & Fröbisch (2014) found that pelycosaurs named before 1900 were on average much more complete than those named after that date.

Early in mosasaur palaeontology, many species were named based on inadequate material, as can be seen by a perusal of invalid names listed by Russell (1967). Perhaps in the nineteenth century, names applied to scrappy material might by chance have been correct, as

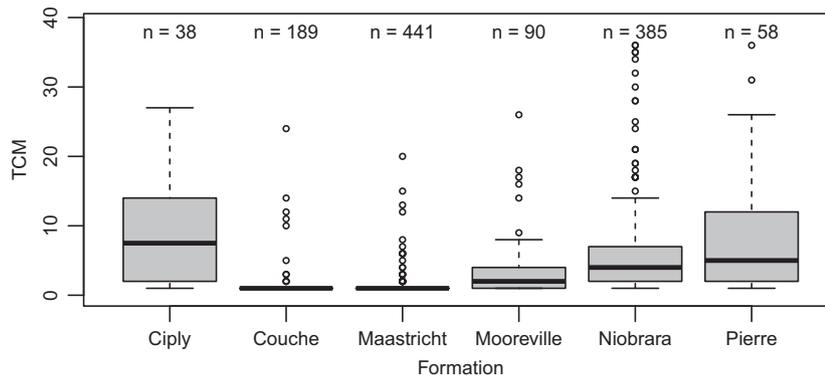


FIG. 8. Completeness by well-known formations. There were over 100 formations to choose from. In this case, some of the best-known formations were compared. The mean TCM values for specimens named to species between formation groups showed highly significant differences (Kruskal–Wallis: $\chi^2 = 595.89$, $df = 5$, $p < 0.001$). Differences remain when tooth-only specimens are left out (Kruskal–Wallis: $\chi^2 = 32.05$, $df = 5$, $p < 0.001$; plot not included).

TABLE 7. Correlation between species properties and completeness measures, showing Spearman correlation coefficients.

	Specimens	Google	Length	TCM	QCM	ICM
Years	0.47**	0.55**	0.25	-0.28*	-0.34*	-0.30
Specimens		0.49**	0.25	-0.41**	-0.49**	-0.50**
Google			0.26	-0.05	-0.14	-0.19
Length				-0.08	-0.12	-0.19
TCMb				0.66**		
TCMc				0.04		
QCMb					0.50**	
QCMc					0.07	
ICMb						0.37**
ICMc						0.05

Correlation tests were run over the following variables by species: average overall completeness by species (TCM, QCM and ICM); body length (Length) from Polcyn *et al.* (2014); Google Scholar 'hits' (Google) for the species name; number of specimens analysed for that species (Specimens); and years since species first described (Years). Best of species (TCMb, QCMb, ICMb) and composite completeness (TCMc, QCMc, ICMc) were also compared to average species completeness values.

*Significant at $p < 0.05$; **significant after false discovery rate correction using method of Benjamini & Hochberg (1995).

palaeontologists were naming the first ever mosasaur finds from newly identified geological formations, but today, it is likely that new names applied to scrappy material risk being synonyms of already named taxa.

Using completeness scores in mosasaurs. This is a specimen rather than taxon-based study. Occurrence-based records that depend on presumed ranges of species were not used to place specimens in time bins in the completeness calculations. This paper shows the utility of this method, which might not have had enough power for statistical testing; the greatest risk was with TCM, since the average completeness values are quite low. However, the analysis was possible because of the large number of specimens in the database.

All three completeness metrics (TCM, QCM, TCM) correlated with each other, both specimen by specimen and through the time series. A similar result was found with ichthyosaur completeness, where SCM correlated with BSCM (Cleary *et al.* 2015), a metric similar to those used in this study. Our results suggest that even qualitative measures, such as QCM and ICM, can be useful for comparing specimens and, because they correlate with TCM (a quantitative metric) and with each other, any one of these metrics could be used to score mosasaur fossils.

TCM is based on true specimen in-bin averages, and thus it is probably driven by taphonomy (Mannion & Upchurch 2010; Beardmore *et al.* 2012a, b). TCM is similar to SCM, but it is not weighted volumetrically, but equally by anatomical region. Weighting by size may introduce an assumption that larger elements or regions are preserved more readily than smaller ones. Disallowing

such weighting then allows TCM to be used to test taphonomic or preservational hypotheses.

QCM was developed as a proxy for phylogenetic completeness and is somewhat equivalent to CCM. QCM estimates phylogenetic completeness without having to score characters on every element of a fossil specimen because QCM is pre-weighted by character density. In terms of the time involved in scoring, QCM can be assessed quickly from a photograph or a fair description, whereas methods such as CCM require careful coding of all skeletal elements. ICM, although less quantitative than the other metrics, was easily scored and could discriminate mosasaur completeness in line with TCM and QCM values, even when specimens could not be examined directly, or photos or more specific descriptions were not available. This confirms its usefulness.

It is important to note that there were some differences in our results when compared to other studies using the SCM and CCM metrics. In most other fossil completeness studies (e.g. Mannion & Upchurch 2010; Brocklehurst *et al.* 2012; Brocklehurst & Fröbisch 2014; Cleary *et al.* 2015; Dean *et al.* 2016; Tutin & Butler 2017) best and composite completeness values for a species are calculated and then these values are generally assigned to time bins of the species temporal range (usually based on first and last appearances). These completeness values are then averaged in the various time bins. If there were only a single specimen representing a species that is only assignable imprecisely to several time bins, there would be no other alternative but to use this method. If the best specimen of a species or the composite specimen cannot be

assigned to an individual time bin, the result is the same as if the mean completeness for a species did not vary over time bins. The large size of our data set allowed for analysis using exactly assignable in-bin specimens only and avoided the need for proxies of specimen completeness such as SCM or CCM.

In our study, we chose not to include composite (TCM_c, QCM_c, ICM_c), best (TCM_b, QCM_b, ICM_b), or holotype (TCM_h, QCM_h, ICM_h) completeness metrics in the time series analyses, to avoid calculating estimates of completeness from un-sampled specimens. In some studies that used stage-level time bins (e.g. Cleary *et al.* 2015; Dean *et al.* 2016), best specimen or composite specimen values were used, but this involves some risk of amalgamating disparate data across time bins. We provide data for holotype completeness (TCM_h, QCM_h, ICM_h), equivalent to SCMTs, as well as best and composite specimen scores average scores, lowest scores, etc. (Table 3), only for comparative purposes.

Completeness and the mosasaur fossil record. Most mosasaur species are very complete (Driscoll *et al.* 2018, data F–H), especially if one considers composite completeness by species. On average, over 65% of the phylogenetic information is available for the average mosasaur species. This is better than for anomodonts (Walther & Fröbisch 2013); otherwise assumed to be rather complete, anomodont skulls yield 82% of phylogenetic characters on average, but postcranial characters account for only 4–9% of possible totals. Our data show that through the history of mosasaur collecting there has not generally been a bias in selecting well-preserved fossils. This is demonstrated by the fact that museums curate thousands of incomplete specimens, indicated by the wide range of TCM and QCM values (Driscoll *et al.* 2018, data B). It should be noted that most of the QCM values are low because there was not an over-representation of, say, skull material that would show bias in collecting. QCM does not correlate with diversity (see below), and this is also an argument that the best specimens do not bias the mosasaur record.

Because we included over 4000 specimens of all completeness values, half of which have TCM and QCM scores of 1 or 2, we were not sure at first whether the inclusion of such low-scoring singleton specimens would distort our conclusions. On the other hand, we reasoned that the inclusion of low-scoring elements should contain valuable information concerning taphonomic drivers of preservation. Especially worrisome was the fact that so many teeth were included as individual specimens.

These concerns were tested in several ways. Kendall's tau-b correlation analysis comparing completeness values over all specimens showed that, even though there were many incomplete specimens, ties in the ranks did not

affect analyses. All metrics are equally useful for scoring. In any case, we analysed time series with and without teeth, and comparisons remained statistically significant. It makes taphonomic sense that the inclusion of teeth in the analysis of lithological variables should increase the discrimination between rock types. The same type of result occurred with the analysis of formational data, which again must vary by rock type. Interestingly, leaving out tooth-only specimens obliterated the statistical differences between palaeogeographical regions. European collections certainly do contain more teeth (Driscoll *et al.* 2018, data B) and perhaps European scientists have always identified more specimens from teeth alone.

It is reassuring that completeness scores are inversely proportional to category-level discrimination, being best for specimens identified to species level, then poorer for those identified to genus level, and worst for those assigned only to family. As noted before, the difference between species and genus completeness is greatly enhanced when specimens consisting only of teeth are excluded from the analysis. A few taxa have a very low completeness score (e.g. *Tylosaurus ivoensis*) but all species with a species epithet have at least some material that is separable by apomorphic characters (Bell 1993), the minimum requirement for naming a new taxon (Parham *et al.* 2012). It is perhaps true that a species can be identified by its teeth (Lindgren & Siverson 2002; Bardet *et al.* 2015) but in marine reptiles the teeth are often homodont and lack variability, being in many cases convergently adapted to diet (Massare 1987), and so may be of limited taxonomic use. When designing future specimen-based studies, analyses with and without teeth would be recommended.

A key discovery was that completeness of species as measured by TCM_{av} does not necessarily correlate with completeness of specimens. Most studies on the completeness of the fossils have assigned various completeness scores to each species but have treated these scores as a measure of preservation quality. The fact that the species-level scores for mosasaurs do not necessarily represent the quality of the preservation of the individual fossil specimens has important implications for how the results of these studies should be interpreted. The use of any whole species proxies for completeness that are derived from data outside of the time bin where the data is averaged will not necessarily be equivalent to analyses compiled from specimens in their home time bins. Results from species and specimen-based studies are likely to be more disparate with larger samples and shorter time bins. In addition, there could be links to correlations between completeness, the number of specimens and year since description: the completeness scores assigned to whole species will be a result of the accumulation of specimens assigned to that species, which we have shown are

influenced by the history of discovery (= number of years since description).

Mosasauro completeness through time. Hundreds of specimens of mosasaurs are datable to substage, and the analysis shows that this subset is a good representation of overall mosasauro species completeness. There are significant differences in completeness over time; but values in mean completeness from substage to substage are not unexpected, as the conditions for fossil preservation must vary in complex ways from fossil to fossil, formation to formation and taphonomic microenvironment to microenvironment. Because there are so many specimens in these time bins, from a wide geographical range, it is difficult to recognize any individual collections or formations that are driving these curves. The differences between time bins represent true in-bin mean values. We show here that using the average completeness of a species group (TCMav, QCMav and ICMav) to calculate overall time bin completeness (TCMtot, QCMtot and ICMtot) is not generally warranted, at least for QCM in this data set. Surprisingly, when TCMav was used to estimate species completeness by substage, it did correlate significantly with the mean TCMsp of the individual specimens in the time bin. This may indicate that multiple specimens of a single species tend to fossilize in similar ways.

It might have been predicted that mosasauro completeness would depend on sea level, as is the case for ichthyosaurs (Cleary *et al.* 2015) and plesiosaurs (Tutin & Butler 2017). However, we found no relationship between mosasauro skeletal completeness and average sea level in any of the time series analyses. There were some negative correlations (Table 4), but the correlation coefficients were extremely low, and not even near significant. Similarly, in GLS analysis, even though the best fitting model for predicting TCMsp included sea level, its predictive value was no better than the null model.

In cases where specimen quality depends on sea level, it might be predicted that the relationship would be positive, in that deep-water settings should provide better conditions for preservation than shallow waters, because the deep oceans are less subject to high-energy deposition, except through the medium of turbidity currents, and there are fewer scavengers than on the marine shelf. However, for ichthyosaurs (Cleary *et al.* 2015) and plesiosaurs (Tutin & Butler 2017), completeness is inversely proportional to sea level, significantly so for the former, but not the latter. This inverse statistically insignificant relationship may also occur in marine crocodiles (DAD, unpub. data) but the reasons for this relationship are not clear. As mentioned by Tutin & Butler (2017), the marine reptile fossil record is not particularly well sampled in the Jurassic and earliest Cretaceous, which might bias results.

It is not clear whether this idea is confirmed by the absence of such a trend in the more intensively sampled and time-limited sample of mosasaurs, or whether different marine reptile groups show different preservation conditions.

We suggest here that the mosasauro fossil record is not much affected by lack of sampling (the exception being the early and middle Coniacian) and there is no correlation to changing sea level. For terrestrial tetrapods, a negative relationship between completeness and sea level was found for sauropod dinosaurs (Mannion & Upchurch 2010), which was explained by differences in sauropodomorph ecology; but there was no correlation for Mesozoic birds (Brocklehurst *et al.* 2012) or pterosaurs (Dean *et al.* 2016).

Explaining mosasauro macroevolution. In our first analysis, correlation results show no direct relationship between species or genus diversity and sea level, but our GLS results do show a significant contribution by sea level in the best fitting model explaining diversity. This compares with Polcyn *et al.* (2014), who argued that sea level at least partially drove mosasauro diversity, as mosasauro richness in their analysis trended in the same direction as sea level. The initial expansion of the clade might well have been triggered by the onset of major continental flooding in the early Late Cretaceous (Caldwell 2002). We suggest that any model of mosasauro macroevolution using environmental drivers will have to take more than sea level into account. The increase of mosasauro species richness combined with the quality of their fossil record makes a strong case for a model of marine reptile evolution in which mosasauro species steadily filled specific niches or expanded steadily into different biogeographical regions, once variability in global marine environmental drivers became stable in the Santonian. The almost level generic diversity curve in the latest Cretaceous shows that mosasaurs had become long-term and stable residents of the Cretaceous seas right up to the late Maastrichtian.

Neither species nor generic diversity through time correlated with skeletal completeness in mosasaurs for any of our metrics. In GLS modelling, the best-fitting auto-correlation model of completeness (TCMsp), species diversity was a predictive variable, but it was not statistically significant. This lack of correlation, confirming what Cleary *et al.* (2015) found for ichthyosaurs, suggests that the quality of fossils does not drive our models of marine reptile diversity and it would be hard to construct a case that apparent changes in diversity are simply artefacts of the quality of fossils or the quality of nomenclature based on those fossils.

Sampling and redundancy. In most previous fossil completeness studies, sampling proxies have been used in multiple regression analysis of fossil completeness to help

understand what is driving measured values of diversity and completeness. Variables such as collections, fossiliferous marine formations, dinosaur-bearing formations, marine tetrapod-bearing formations, pterosaur-bearing formations and other proxies have all been used. Much of this data is relatively accessible from the Paleobiology Database and/or the primary literature, but most of it is tallied at stage level. In our study, using substage-dated specimens, it was not possible to include the above proxies in our multiple regression analyses. Such a comparison will be interesting once a narrower time range analysis is possible. Instead, we developed an Upper Cretaceous substage-level proxy from MPBF. Almost all plesiosaur-bearing formations also contained mosasaurs, so the data overlapped.

In all our best fitting GLS models, with and without auto-correlation, MPBF correlated highly significantly with diversity. In the past, counts of fossiliferous formation were used as a proxy for sampling that combined geological and human biases (Benson *et al.* 2010). If MPBF is considered as a proxy for geological megabiases, then our results could indicate that none of our diversity data is reliable enough to compare with any other time series, including sea level or completeness. However, the shape of the diversity curve, lack of evidence for lithological or regional correlates with specimen completeness, and the thousands of sampled fossils argue against jumping to this conclusion. Further, the data on fossil occurrence (collections, specimen counts, localities, formations) were collected at the same time as the data on diversity, and so there is a risk of tallying rock and fossil data that describe the same history of discovery, so pointing to redundancy (Benton *et al.* 2011; Benton 2015). The redundancy hypothesis for highly correlated rock and fossil data was confirmed in the case of the fossil records of the UK and the world by using statistical methods that identify not only correlation but also directionality of causation (Dunhill *et al.* 2014, 2018). Therefore, we cannot use the MPBF count as a sampling proxy because it is not an independent yardstick that represents either geological or human sampling. Our other variables, including fossil completeness, diversity and sea-level, are independent of one another. We have shown that fossil preservation, as measured by specimen completeness metrics, does not bias the fossil record of mosasaurs.

Explaining mosasaur completeness. We have compared completeness in the best-known mosasaur-bearing formations. Factors that might explain the differences include lithology, rock exposure and collecting biases. Comparing completeness among different outcrops and formations can be used as an aid to understanding Lagerstätten effects.

Our results for mosasaurs show many agreements with the study of the ichthyosaur fossil record by Cleary *et al.*

(2015). In both studies, skeletons were more complete in fine-grained than coarse-grained sediments (Fig. 6), and this was expected because fossil completeness is partially dependent on taphonomy (Beardmore *et al.* 2012a, b) and post-depositional geological factors. This is supported by the fact that when low-completeness specimens consisting only of teeth are left out of calculations of mean completeness, the differences between lithologies are less evident. We expect, for example, that since sandstones are deposited in high-energy environments, which toss and abrade bone, specimens in sandstones would have smaller mean completeness values than those in lower-energy mudstones. Fine-grained sediments should preserve more detail. In fact, the New Jersey Greensands have the lowest completeness of the formations considered, and the Pierre Shale has a high average mean completeness value. A further contributor to the high quality of specimens in fine-grained sediments such as the Pierre Shale are their anoxic environments, with little scavenging (Kauffman & Sageman 1988).

The Pierre Shale covers thousands of square miles of the North American western interior, and produces some almost complete articulated fossils with soft tissue (pers. obs.; Carpenter 2006, 2008). The Pierre has yielded fewer specimens than the Niobrara or Greensand, but considering its greater mean completeness, the sheer size of the Pierre outcrop (311 000 km²) in comparison to that of the Niobrara and Greensand (21 000 km² each) and its relative inaccessibility in remote regions of the North American western interior, suggests that complete specimens may yet be found. The Pierre Shale fossils have a higher mean completeness score than those from the Niobrara Chalk, but the latter formation is often considered to be a Lagerstätte (Bottjer 2002), and indeed some mosasaur soft tissue impressions have been found (Lindgren *et al.* 2010). If average completeness could be considered one measure of a Lagerstätte, the Pierre Shale should also be considered as such.

The Niobrara Chalk has experienced a great deal of collector effort (thousands of specimens; over 150 years of effort by hundreds of people) and is still yielding fresh finds, but no new species barring those re-described (such as *Tylosaurus kansasensis*). Exposure (desert badlands) and accessibility are high. Most of the Niobrara species have probably been collected. Depending on average lithology and depositional environment, there may be a limit to the skeletal quality within any geological formation, and no amount of additional collecting can improve that. The fact that the mean TCM values between different formations are significantly different with and without tooth-only specimens supports the idea that in highly collected formations there may be a limit to average preservation values. Once enough rock is exposed and collected, lithofacies biodiversity reaches a peak (e.g. Smith & Benson 2013) and the known biodiversity then is limited by

TABLE 8. North American outcrop area vs species diversity and mean species completeness per formation.

Group	Area (km ²)	Diversity	TCM
Pierre (NA)	310 728	11	8.51
Niobrara (KS)	21 091	10	5.35
Mooreville (AL)	315 788	12	3.35
Monmouth (NJ)	21 506	6	1.93
Moreno (CA)	23 358	2	4.95

There were no significant correlations (Area vs Diversity, $r_s = 0.60$ $p = 0.35$; Area vs TCM, $r_s = 0$, $p = 1$; Diversity vs TCM, $r_s = 0.2$, $p = 0.78$).

the ecology of the ancient environment and the preservation potential of the rocks, assuming collector effort and accessibility is high. Bones from the same family probably have similar preservation potential (Smith & McGowan 2011), so it is doubtful that there are missing taxa based on preservation alone.

Further to this theme, it might have been predicted that skeletal completeness and diversity would be related in some way to outcrop or exposure area; perhaps, for example, when the overall area of a geological formation is high, more skeletons of all kinds of completeness might be found, and so the mean completeness score might then rise, and thus perhaps biodiversity. In our preliminary analysis, the results show no significant correlation between completeness and diversity or outcrop area for North American formations (Table 8). This supports the idea that each formation is associated with an upper limit on preservation potential if there has been adequate exposure and collector effort.

The Craie de Ciply formation from the Mons Basin in Belgium has the highest average skeletal completeness score. This Ciply chalk has produced many holotypes (Dollo 1904), and the blocks from that formation at the Royal Belgian Institute of Natural Sciences (Brussels) contain highly articulated and well-preserved specimens, and very few single elements or partial fossils. This is striking when compared, for example, to the chalk at Maastricht, which has yielded many hundreds of disarticulated specimens, but the explanation, presumably to do with mode of deposition and rate of burial of the carcasses, is not clear. In this case, with all the almost complete skeletons available, perhaps less spectacular specimens were not deemed worthy of collection, or perhaps they do not exist. Neither lithology, outcrop area, nor the amount of collecting explains the completeness of these Belgian fossils, limited in area to quarries in a relatively small region.

The average completeness of mosasaur specimens has tended to decrease through research time, which was initially unexpected; specimens described and named many

years ago tend to be more complete than those named more recently. The holotypes of species currently regarded as valid are typically rather complete specimens, and subsequently identified materials of many of these species may on average be less complete, and now more easily identifiable. Such specimens are not typically considered to be publishable material, and studies that use only published material to describe historical trends in fossil quality may not show the same result. The average-quality material found in many museums outnumbers more complete material. The inverse completeness trend may reflect the fact that the holotypes of taxa named in former centuries were substantially complete and have been preferentially retained, whereas less complete material was disposed of, or perhaps not collected at all in the early days of palaeontology when collectors were perhaps less assiduous in recording everything. Today, on the other hand, perhaps holotypes are of similar completeness, but museums retain enormous collections of less complete, referred specimens. Again, completeness does not continuously rise for a species as more specimens are collected, but we have not explored historical differences in completeness for specific formations.

Known fossil completeness of mosasaurs is best in North America. Surprisingly, the well-known very complete European specimens do not significantly drive fossil completeness in Eurasia, nor does the relative number of specimens. When the tooth-alone specimens are left out of the analysis, there are no significant differences in completeness between the continents. North American collections in this analysis are relatively devoid of tooth specimens. We were not able to make a significant comparison of completeness in northern versus southern hemispheres, as there are too few of the latter. For ichthyosaurs, Cleary *et al.* (2015) showed that the well-studied northern hemisphere produced fossils of significantly higher quality than the southern hemisphere. The differences above are probably all sampling artefacts.

Larger mosasaurs do not show higher skeletal completeness than either small or medium-sized ones. One might hypothesize that larger specimens would be more complete, as in some dinosaurs (Brown *et al.* 2013). The situation here is different from that seen in other, smaller taxa such as birds or pterosaurs (Brocklehurst *et al.* 2012; Dean *et al.* 2016), where Lagerstätten may selectively preserve smaller specimens better than large specimens found in other deposits. This could be explained by the greater weight of their bones, the higher energy required by sedimentary flows to disarticulate a skeleton, the fact that larger specimens are easier to find, or that they are preferentially collected. It is interesting to note that the number of Google Scholar hits per species showed a trend (although, not quite significant) with estimated body length, perhaps indicating preferential study of larger

mosasaurs. For ichthyosaurs, Cleary *et al.* (2015) rather surprisingly found that medium-sized specimens were significantly more complete than small or large taxa: the incompleteness of small specimens was expected, but it was a surprise that larger specimens were also relatively incomplete.

CONCLUSIONS

Palaeobiology has been built on the idea that, in spite of limitations of the fossil record, biological information including patterns of diversity and macroevolution might be demonstrated with the proper analytical techniques. The mosasaur fossil record has been explored in terms of skeletal completeness, a study enabled and strengthened by the great abundance and quality of specimens. New completeness metrics, introduced here, adequately describe the preservation of the mosasaur fossil record. QCM, a novel and quick method for estimating fossil completeness correlates with true phylogenetic character completeness and can be used as a proxy for it.

Mosasaur fossils are found in all stratigraphic substages throughout their evolution, and neither skeletal nor phylogenetic completeness explains their diversity; fossil completeness does not bias the fossil record of mosasaurs and cannot be used as a proxy for diversity. A huge amount of both incomplete and well-preserved mosasaur material is identifiable, which is not the case for some other Mesozoic tetrapod groups. The mosasaur fossil specimen record contains thousands of teeth, which do not affect the general utility of the methods but improve the resolution of completeness values in taphonomically related comparisons. Outcrop area, where data is available, does not explain mosasaur diversity. However, lithology has a role: skeletal completeness is higher in fine-grained than in coarse-grained sediments. There are no correlations that suggest that sea level is related in any way to mosasaur fossil completeness or is a direct driver of mosasaur diversity. We do not detect any geological megabiases driving the fossil record of mosasaurs. Mosasaur species richness, based on specimens assignable to a single substage, rises steadily and smoothly from the late Coniacian to the late Maastrichtian and correlates with the generic richness curve. Although ambiguous in this study, sea level may play a role in further models of mosasaur diversity. Low sampling in the middle Cretaceous makes the analysis of completeness difficult through this time range. Even considering this, mosasaurs appear unique among marine tetrapods in terms of the reliability of their fossil record.

Acknowledgements. None of this work would have been possible without the input and contributions of many generous and knowledgeable people. Thank you, Susan Beardmore for sending

pictures of marine crocodile specimens, which taught the first author how to use the completeness metric; and this in turn inspired ideas on how to apply it to mosasaurs. We thank Scott Hartman for giving permission to use the mosasaur skeleton drawing and silhouette in Figure 1. Mosasaur silhouettes in Figures 2–4 are used courtesy of PhyloPic (<http://phylopic.org/>). For discussions on mosasaur specimens, we thank M. Polcyn and T. Ikejiri. We are grateful to the following institutions and personnel who allowed access to their collections and provided hospitality and mosasaur knowledge: D. Burnham and D. Miao at University of Kansas; S. Chapman at British Natural History Museum; P. Viegas at University of Bristol, for introduction to The Royal Museum of Natural History, Brussels and to its personnel, A. Folie, including the invitation by P. Godefroit; D. Winkler at the Schuler Museum, Southern Methodist University; K. Morton at the Perot Museum, Dallas; J.C. Sagebiel and staff at the University of Texas at Austin Vertebrate Paleontology Lab; and J. Sertich and L. Ivy at the Denver Museum of Nature and Science. Thanks go out to the museum specimen database information and communications from D. Ehret, L. Ellington at the Alabama Natural History Museum; W. Simpson at the Field Museum of Natural History; T. Konishi, J. Gardner, B. Strilisky, and G. Housebo at the Royal Tyrell Museum; and M. Walsh at the Los Angeles County Natural History Museum. We thank Christopher Dean for useful comments and criticisms of the initial manuscript of this paper, Neil Brocklehurst and Jörg Fröbisch as reviewers, and Roger Benson and Sally Thomas for editorial improvements.

DATA ARCHIVING STATEMENT

Data for this study are available in the Dryad Digital Repository: <https://doi.org/10.5061/dryad.5rr99>

Editor. Roger Benson

REFERENCES

- BARDET, N. 1994. Extinction events among Mesozoic marine reptiles. *Historical Biology*, **7**, 313–324.
- 2012. The mosasaur collections of the Muséum National d'Histoire Naturelle of Paris. *Bulletin de la Société Géologique de France*, **183**, 35–53.
- HOUSSAYE, A., RAGE, J.-C. and SUPERBIOLA, X. P. 2008. The Cenomanian-Turonian (Late Cretaceous) radiation of marine squamates (Reptilia), the role of the Mediterranean Tethys. *Bulletin de la Société Géologique de France*, **79**, 605–622.
- — VINCENT, P., SUPERBIOLA, X. P., AMAGH-ZAZ, M., JOURANI, E. and MESLOUH, S. 2015. Mosasaurs (Squamata) from the Maastrichtian phosphates of Morocco, biodiversity, palaeobiogeography and palaeoecology based on tooth morphoguilds. *Gondwana Research*, **27**, 1068–1078.
- BEARDMORE, S. R., ORR, P. J., MANZOCCHI, T. and FURRER, H. 2012a. Float or sink: modelling the

- taphonomic pathway of marine crocodiles (Mesoeucrocodylia, Thalattsuchia) during the death–burial interval. *Palaeobiodiversity & Palaeoenvironments*, **92**, 83–98.
- — — — — and JOHNSON C. 2012b. Death, decay and disarticulation: modelling the skeletal taphonomy of marine reptiles demonstrated using *Serpianosaurus* (Reptilia; Sauropterygia). *Palaeogeography, Palaeoclimatology, Palaeoecology*, **337–338**, 1–13.
- BELL, G. L. 1993. A phylogenetic revision of Mosasuroidea (Squamata). PhD thesis, The University of Texas at Austin.
- 1997. A phylogenetic revision of North American and Adriatic Mosasuroidea. 293–332. In CALLAWAY, J. M. and NICHOLLS, E. L. (eds). *Ancient marine reptiles*. Academic Press, San Diego, 501 pp.
- and POLCYN, M. J. 2005. *Dallasaurus turneri*, a new primitive mosasuroid from the Middle Turonian of Texas and comments on the phylogeny of Mosasauridae (Squamata). *Netherlands Journal of Geosciences*, **84**, 177–194.
- BENJAMINI, Y. and HOCHBERG, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B (Methodological)*, **1**, 289–300.
- BENSON, R. B. J. and BUTLER, R. J. 2011. Uncovering the diversification history of marine tetrapods, ecology influences the effect of geological sampling biases. *Geological Society of London, Special Publications*, **358**, 191–208.
- and DRUCKENMILLER, P. S. 2014. Faunal turnover of marine tetrapods during the Jurassic-Cretaceous transition. *Biological Reviews*, **89**, 1–23.
- BUTLER, R. J., LINDGREN, J. and SMITH, A. S. 2010. Mesozoic marine tetrapod diversity, mass extinctions and temporal heterogeneity in geological megabiases affects vertebrates. *Proceedings of the Royal Society B*, **277**, 828–834.
- BENTON, M. J. 2008a. How to find a dinosaur, and the role of synonymy in biodiversity studies. *Paleobiology*, **34**, 516–533.
- 2008b. Fossil quality and naming dinosaurs. *Biology Letters*, **4**, 729–732.
- 2015. Palaeodiversity and formation counts, redundancy or bias? *Palaeontology*, **58**, 1003–1029.
- TVERDOKHLEBOV, V. P. and SURKOV, M. V. 2004. Ecosystem remodeling among vertebrates at the Permian-Triassic boundary in Russia. *Nature*, **432**, 97–100.
- DUNHILL, A. M. and LLOYD, G. T. 2011. Assessing the quality of the fossil record, insight from vertebrates. *Geological Society of London, Special Publications*, **358**, 63–94.
- BONETT, D. G. and WRIGHT, T. A. 2000. Sample size requirements for estimating Pearson, Kendall and Spearman correlations. *Psychometrika*, **65**, 23–28.
- BOTTJER, D. J. 2002. Smoky Hill Chalk, Spectacular Cretaceous Marine Fauna. 352–364. In BOTTJER, D. J., ETTER, W., HAGGERDORN, J. W. and TANG, C. M. (eds). *Exceptional fossil preservation, a unique view on the evolution of marine life*. Columbia University Press, New York, 403 pp.
- BROCKLEHURST, N. P. 2015. A simulation-based examination of residual diversity estimates as a method for correcting for sampling bias. *Palaeontologica Electronica*, **18**, 1–15.
- and FRÖBISCH, J. 2014. Current and historical perspectives on the completeness of the fossil record of pelycosaurian-grade synapsids. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **399**, 114–126.
- UPCHURCH, P., MANNION, P. D. and O'CONNOR, J. 2012. The completeness of the fossil record of Mesozoic birds, implications for early avian evolution. *PLoS One*, **7**, e39056.
- BROWN, C. M., EVANS, D. C., CAMPIONE, N. E., O'BRIEN, L. J. and EBERTH, D. A. 2013. Evidence for taphonomic size bias in the Dinosaur Park Formation (Campanian, Alberta), a model Mesozoic terrestrial alluvial-paralic system. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **372**, 108–122.
- CALDWELL, M. 2002. From fins to limbs to fins, limb evolution in fossil marine reptiles. *American Journal of Medical Genetics*, **112**, 236–249.
- CAMP, C. L. 1942. California mosasaurs. *Memoirs of the University of California*, **13**, 1–68.
- CARPENTER, K. 2006. Comparative vertebrate taphonomy of the Pembina and Sharon Springs members (middle Campanian) of the Pierre Shale, Western Interior. *Paludicola*, **5**, 125–149.
- 2008. Vertebrate biostratigraphy of the Smoky Hill Chalk (Niobrara Formation) and the Sharon Springs Member (Pierre Shale). 421–437. In HARRIES, P. J. (ed.) *Approaches in high-resolution stratigraphic paleontology*. Kluwer, 474 pp.
- CASEY, M. M., FRASER, N. C. and KOWALEWSKI, M. 2007. Quantitative taphonomy of a Triassic reptile *Tanytrachelos ahynis* from the Cow Branch Formation, Dan River Basin, Solite Quarry, Virginia. *Palaio*, **22**, 598–611.
- CLEARY, T. J., MOON, B. C., DUNHILL, A. M. and BENTON, M. J. 2015. The fossil record of ichthyosaurs, completeness metrics and sampling biases. *Palaeontology*, **58**, 521–536.
- DEAN, C. D., MANNION, P. D. and BUTLER, R. J. 2016. Preservation bias controls the fossil record of pterosaurs. *Palaeontology*, **59**, 225–247.
- DEBRAGA, M. and CARROLL, R. L. 1993. The origin of mosasaurs as a model of macroevolutionary patterns and processes. *Evolutionary Biology*, **27**, 245–322.
- DOLLO, L. 1904. Les mosasauriens de la Belgique. *Bulletin de la Société Belge de Géologie de Paléontologie et d'Hydrogéologie, Mémoires*, **18**, 207–216.
- DRISCOLL, D. A., DUNHILL, A. M., STUBBS, T. L. and BENTON, M. J. 2018. Data from 'The mosasaur fossil record through the lens of fossil completeness'. *Dryad Digital Repository*. <https://doi.org/10.5061/dryad.5rr99>
- DUNHILL, A. M., HANNISDAL, B. and BENTON, M. J. 2014. Disentangling geological megabiases and common-cause from redundancy in the British fossil record. *Nature Communications*, **5**, 4818.
- — — BROCKLEHURST, N. and BENTON, M. J. 2018. Testing the accuracy of formation-based sampling proxies to correct the fossil record. *Palaeontology*, **61**, 119–132.
- EVERHART, M. J. 2000. Mosasaurs, last of the great marine reptiles. *Prehistoric Times*, **44**, 29–31.
- 2001. Revisions to the biostratigraphy of the Mosasauridae (Squamata) in the Smoky Hill Chalk Member of the Niobrara

- Chalk (Late Cretaceous) of Kansas. *Transactions of the Kansas Academy of Science*, **104**, 59–78.
- FISCHER, V., MAISCH, M. W., NAISH, D., KOSMA, R., LISTON, J., JOGER, U., KRÜGER, F. J., PÉREZ, J., TAINSH, J. and APPLEBY, R. M. 2012. New ophthalmosaurid ichthyosaurs from the European Lower Cretaceous demonstrate extensive ichthyosaur survival across the Jurassic–Cretaceous boundary. *PLoS One*, **7**, e29234.
- FOUNTAINE, T. M. R., BENTON, M. J., DYKE, G. J. and NUDDS, R. L. 2005. The quality of the fossil record of Mesozoic birds. *Proceedings of the Royal Society B*, **272**, 289–294.
- GRADSTEIN, F. M., OGG, J. G. and SCHMITZ, M. (eds). 2012. *The geologic time scale 2012*. Elsevier, 1176 pp.
- HOLMES, R., CALDWELL, M. W. and CUMBAA, S. L. 1999. A new specimen of *Plioplatecarpus* (Mosasauridae) from the lower Maastrichtian of Alberta, comments on allometry, functional morphology, and paleoecology. *Canadian Journal of Earth Sciences*, **36**, 363–369.
- HOUSSAYE, A., LINDGREN, J., PELLEGRINI, R., LEE, A. H., GERMAIN, D. and POLCYN, M. J. 2013. Microanatomical and histological features in the long bones of mosasaurine mosasaurs (Reptilia, Squamata)—implications for aquatic adaptation and growth rates. *PLoS One*, **8**, e76741.
- IKEJIRI, T., EBERSOLE, J. A., BLEWITT, H. L. and EBERSOLE, S. M. 2013. An overview of Late Cretaceous vertebrates from Alabama. *Bulletin of the Alabama Museum of Natural History*, **31**, 46–71.
- JAGT, J. W. M. 2005. Stratigraphic ranges of mosasaurs in Belgium and the Netherlands (Late Cretaceous) and cephalopod-based correlations with North America. *Netherlands Journal of Geosciences*, **84**, 288–301.
- KAUFFMAN, E. G. and SAGEMAN, B. B. 1988. Biological sensing of benthic environments in dark shales and related oxygen-restricted facies. 121–138. In GINSBERG, R. N. and BEAUDOIN, B. (eds). *Cretaceous resources, events and rhythms*. Kluwer, 352 pp.
- KEMP, R. A. and UNWIN, D. M. 1997. The skeletal taphonomy of *Archaeopteryx*, a quantitative approach. *Lethaia*, **30**, 229–238.
- KIERNAN, C. R. 2002. Stratigraphic distribution and habitat segregation of mosasaurs in the Upper Cretaceous of western and central Alabama, with an historical review of Alabama mosasaur discoveries. *Journal of Vertebrate Paleontology*, **22**, 91–103.
- KONISHI, T. and CALDWELL, M. W. 2007. New specimens of *Platecarpus planifrons* (Cope, 1874) (Squamata, Mosasauridae) and a revised taxonomy of the genus. *Journal of Vertebrate Paleontology*, **27**, 59–72.
- KUYPERS, M. M. M., JAGT, J. W. M., PEETERS, H. H. G., DE GRAAF, D. T., DORTANGE, R. W., DECKERS, M. J. M., EYSERMANS, D. and ARPOT, L. 1998. Laat-kretaceische mosasauriërs uit Luik-Limburg, nieuwe vondsten leiden tot nieuwe inzichten. *Publicaties van het Natuurhistorisch Genootschap in Limburg*, **41**, 4–47.
- LEBLANC, A. R., CALDWELL, M. W. and BARDET, N. 2012. A new mosasaurine from the Maastrichtian (Upper Cretaceous) phosphates of Morocco and its implications for mosasaurine systematics. *Journal of Vertebrate Paleontology*, **32**, 82–104.
- LINDGREN, J. and SIVERSON, M. 2002. *Tylosaurus ivoensis*: a giant mosasaur from the early Campanian of Sweden. *Earth & Environmental Transactions of the Royal Society of Edinburgh*, **93**, 73–93.
- JAGT, J. W. M. and CALDWELL, M. W. 2007. A fishy mosasaur, the axial skeleton of *Plotosaurus* (Reptilia, Squamata) reassessed. *Lethaia*, **40**, 153–160.
- ALWMARK, C., CALDWELL, M. W. and FIORILLO, A. R. 2009. Skin of the Cretaceous mosasaur *Plotosaurus*: implications for aquatic adaptations in giant marine reptiles. *Biology Letters*, **5**, 528–531.
- CALDWELL, M. W., KONISHI, T. and CHIAPPE, L. M. 2010. Convergent evolution in aquatic tetrapods: insights from an exceptional fossil mosasaur. *PLoS One*, **5**, e11998.
- POLCYN, M. J. and YUNG, B. A. 2011. Landlubbers to leviathans, evolution of swimming in mosasaurine mosasaurs. *Paleobiology*, **37**, 445–469.
- KADDUMI, H. F. and POLCYN, M. J. 2013. Soft tissue preservation in a fossil marine lizard with a bilobed tail fin. *Nature Communications*, **4**, 2423.
- LYDEKKER, R. 1888. *Catalogue of the fossil Reptilia and Amphibia in the British Museum (Natural History)*. Part 2. British Museum (Natural History), London, 307 pp.
- MANNION, P. D. and UPCHURCH, P. 2010. Completeness metrics and the quality of the sauropodomorph fossil record through geological and historical time. *Paleobiology*, **36**, 283–302.
- MASSARE, J. A. 1987. Tooth morphology and prey preference of Mesozoic marine reptiles. *Journal of Vertebrate Paleontology*, **7**, 121–137.
- 1994. Swimming capabilities of Mesozoic marine reptiles, a review. 133–149. In MADDOCK, L., BONE, Q. and RAYNER, J. M. V. (eds). *Mechanics and physiology of animal swimming*. Cambridge University Press, 251 pp.
- McNAMARA, M. E., ORR, P. J., ALCALÁ, L., ANADÓN, P. and PEÑALVER, E. 2012. What controls the taphonomy of exceptionally preserved taxa-environment or biology? A case study using frogs from the Miocene Libros konservat-lagerstätte (Teruel, Spain). *Palaio*, **27**, 63–77.
- 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, B. and PEKAR, S. F. 2005. The Phanerozoic record of global sea-level change. *Science*, **310**, 1293–1298.
- MOTANI, R. 2005. Evolution of fish-shaped reptiles (Reptilia, Ichthyopterygia) in their physical environments and constraints. *Annual Review of Earth & Planetary Sciences*, **33**, 395–420.
- 2009. The evolution of marine reptiles. *Evolution: Education & Outreach*, **2**, 224–235.
- PARHAM, J. F., DONOGHUE, P. C. J., BELL, C. J., CALWAY, T. D., HEAD, J. J., HOLROYD, P. A., INOUE, J. G., IRMIS, R. B., JOYCE, W. G., KSEPKA, D. T., PATANE, J. S. L., SMITH, N. D., TARVER, J. E.,

- TUINEN, M. VAN, YANG, Z., ANGIELCZYK, K. D., GREENWOOD, J., HIPSLEY, C. A., JACOBS, L., MAKOVICKY, P. J., MÜLLER, J., SMITH, K. T., THEODOR, J. M., WARNOCK, R. C. M. and BENTON, M. J. 2012. Best practices for justifying fossil calibrations. *Systematic Biology*, **61**, 346–359.
- PETERS, S. E. 2005. Geologic constraints on the macroevolutionary history of marine animals. *Proceedings of the National Academy of Science*, **102**, 12326–12331.
- PINHEIRO, J., BATES, D., DEBROY, S., SARKAR, D. and R CORE TEAM 2017. nlme, Linear and Nonlinear Mixed Effects Models. Version 3.1–131. <https://CRAN.R-project.org/package=nlme>
- POLCYN, M. J., JACOBS, L. L., ARAÚJO, R., SCHULP, A. S. and MATEUS, O. 2014. Physical drivers of mosasaur evolution. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **400**, 17–27.
- R CORE TEAM 2016. *R: a language and environment for statistical computing*. Version 3.3.0. R Foundation for Statistical Computing. <https://www.R-project.org>
- ROSS, M. R. 2009. Charting the Late Cretaceous seas, mosasaur richness and morphological diversification. *Journal of Vertebrate Paleontology*, **29**, 409–416.
- RUSSELL, D. A. 1967. Systematics and morphology of American mosasaurs (Reptilia, Sauria). *Bulletin of the Peabody Museum of Natural History*, **23**, 1–237.
- 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.
- SCHULTZE, H.-P., HUNT, L., CHORN, J. and NEUNER, A. M. 1985. Type and figured specimens of fossil vertebrates in the collection of the University of Kansas Museum of Natural History. Part II. Fossil amphibians and reptiles. *The University of Kansas, Museum of Natural History, Miscellaneous Publication*, **77**, 1–66.
- SMITH, A. B. 2007. Intrinsic vs. extrinsic biases in the fossil record, contrasting the fossil record of echinoids in the Triassic and early Jurassic using sampling data, phylogenetic analysis, and molecular clocks. *Paleobiology*, **33**, 310–323.
- and BENSON, R. B. J. 2013. Marine diversity in the geological record and its relationship to surviving bedrock area, lithofacies diversity, and original marine shelf area. *Geology*, **41**, 171–174.
- and MCGOWAN, A. J. 2011. The ties linking rock and fossil records and why they are important for palaeobiodiversity studies. *Geological Society of London, Special Publications*, **358**, 1–7.
- SØRENSEN, A. M., SURLYK, F. and LINDGREN, J. 2013. Food resources and habitat selection of a diverse vertebrate fauna from the upper lower Campanian of the Kristianstad Basin, southern Sweden. *Cretaceous Research*, **42**, 85–92.
- STUBBS, T. L. and BENTON, M. J. 2016. Ecomorphological diversifications of Mesozoic marine reptiles, the roles of ecological opportunity and extinction. *Paleobiology*, **42**, 547–573.
- TUTIN, S. L. and BUTLER, R. J. 2017. The completeness of the fossil record of plesiosaurs, marine reptiles from the Mesozoic. *Acta Palaeontologica Polonica*, **62**, 563–573.
- VERRIÈRE, A., BROCKLEHURST, N. and FRÖBISCH, J. 2016. Assessing the completeness of the fossil record, comparison of different methods applied to parareptilian tetrapods (Vertebrata, Sauropsida). *Paleobiology*, **42**, 680–695.
- WALTHER, M. and FRÖBISCH, J. 2013. The quality of the fossil record of anomodonts (Synapsida, Therapsida). *Comptes Rendus Palevol*, **12**, 495–504.
- YOUNG, M. T., BRUSATTE, S. L., RUTA, M. and DE ANDRADE, M. B. 2010. The evolution of Metriorhynchoidea (Mesoeucrocodylia, Thalattosuchia), an integrated approach using geometric morphometrics, analysis of disparity, and biomechanics. *Zoological Journal of the Linnean Society*, **158**, 801–859.
- ZUUR, A. F., IENO, E. N., WALKER, N. J., SAVELIEV, A. A. and SMITH, G. M. 2009. *Mixed effects models and extensions in ecology with R*. Springer, 574 pp.

APPENDIX: SPECIES LIST

Species list for this study (Squamata, Mosasauridae), including the number of specimens included in the analysis and mean completeness scores (taphonomic (TCM), qualitative (QCM) or informal (ICM) completeness metrics). The range is based on individual specimens that are datable to specific substages, except those with a ‘?’ whose range is near a boundary, or ‘??’ whose range is imprecise. *Abbreviations*: Camp, Campanian; Con, Coniacian; Maas, Maastrichtian; Sant, Santonian; Tur, Turonian. Completeness figures are rounded to two decimal places.

Genus	Species	Number	TCM	QCM	TCM	Range
<i>Eonatator</i>	sp.	1	2	1	1	Late Sant – Early Camp??
<i>Eonatator</i>	<i>sternbergii</i>	87	1.51	1.92	1.22	Early Camp
<i>Eonatator</i>	cf. <i>sternbergii</i>	1	2	3	3	Late Sant – Early Camp??
<i>Halisaurus</i>	<i>arambourgi</i>	33	5.64	2.76	3.25	Early – Late Maas
<i>Halisaurus</i>	<i>platyspondylus</i>	10	2.4	2	2	Early – Late Maas
<i>Phosphorosaurus</i>	<i>ortliebi</i>	1	2	3	3	Early Maas
<i>Phosphorosaurus</i>	<i>ponpetelegans</i>	1	10	7	4	Early Maas
<i>Carinodens</i>	<i>belgicus</i>	38	1.03	2	1.21	Late Maas

(continued)

APPENDIX. (Continued)

Genus	Species	Number	TCM	QCM	TCM	Range
<i>Carinodens</i>	<i>minalmamar</i>	2	1.5	2	2	Late Maas
<i>Clidastes</i>	<i>liodontus</i>	21	8.47	5	3.67	Late Con – Early Camp
<i>Clidastes</i>	<i>moorevillensis</i>	4	6.67	4	3.67	Late Sant – Early Camp
<i>Clidastes</i>	<i>propyhton</i>	196	3.68	2.89	2.35	Early–Late Camp
<i>Dallasaurus</i>	<i>turneri</i>	3	7.33	3.33	3.33	Mid Tur
<i>Globidens</i>	<i>alabamensis</i>	16	1.47	2.44	1.53	Early–Late Camp
<i>Globidens</i>	<i>dakotensis</i>	1	8	6	4	Mid Camp
<i>Globidens</i>	<i>phosphaticus</i>	63	1.14	2.08	1.1	Early–Late Maas
<i>Globidens</i>	<i>schurmanni</i>	1	21	7	5	Late Camp
<i>Igdamanosaurus</i>	<i>aegyptiacus</i>	5	1	2	1	Early–Late Maas
<i>Kourisodon</i>	sp.	1	2	3	3	Late Camp
<i>Kourisodon</i>	<i>puntledgensis</i>	1	18	7	5	Late Sant
<i>Moanasaurus</i>	<i>mangahouangae</i>	5	6.6	5.2	3.6	Mid–Late Camp?
<i>Mosasaurus</i>	<i>beaugei</i>	110	1.03	2.02	1.04	Late Maas
<i>Mosasaurus</i>	<i>conodon</i>	123	3.53	2.49	1.81	Late Camp – Late Maas
<i>Mosasaurus</i>	<i>hobetsuensis</i>	1	13	3	3	Early Maas
<i>Mosasaurus</i>	<i>hoffmanni</i>	291	1.42	1.9	1.46	Early – Late Maas
<i>Mosasaurus</i>	<i>missouriensis</i>	11	9.9	5.4	4	Mid Camp – Late Camp
<i>Mosasaurus</i>	<i>mokoroa</i>	2	3.5	4.5	2.5	Mid Camp
<i>Mosasaurus</i>	<i>prismaticus</i>	1	2	3	3	Late Camp – Late Maas??
<i>Plotosaurus</i>	sp.	1	1	2	1	Late Maas
<i>Plotosaurus</i>	<i>bennisoni</i>	50	4.59	2.19	1.86	Early Camp – Early Maas??
<i>Plesiotylosaurus</i>	<i>crassidens</i>	4	6.25	6	4	Early–Late Maas?
<i>Prognathodon</i>	<i>anceps</i>	6	1.17	2.17	1.67	Early Camp
<i>Prognathodon</i>	<i>currii</i>	8	2	2.63	1.38	Late Camp – Late Maas
<i>Prognathodon</i>	<i>kianda</i>	6	7	3.67	3.17	Late Maas
<i>Prognathodon</i>	<i>giganteus</i>	5	3	2.6	1.8	Early Maas
<i>Prognathodon</i>	<i>lutugini</i>	75	1.17	2.01	1.05	Early–Late Camp
<i>Prognathodon</i>	<i>mosasauroides</i>	2	1.5	2.5	2	Late Maas
<i>Prognathodon</i>	<i>overtoni</i>	12	8.58	5	4.08	Early–Late Camp
<i>Prognathodon</i>	<i>rapax</i>	13	1.91	1.92	2.27	Early–Late Maas
<i>Prognathodon</i>	<i>saturator</i>	4	5.75	3.25	2	Late Maas
<i>Prognathodon</i>	<i>sectorius</i>	30	1.03	2.03	1.07	Late Camp – Late Maas
<i>Prognathodon</i>	<i>solvayi</i>	7	5.71	3.29	3	Late Camp – Early Maas
<i>Prognathodon</i>	<i>waiparaensis</i>	3	4	4	3	Late Maas
<i>Eremiasaurus</i>	<i>heterodontus</i>	2	19	7	4.5	Late Maas
<i>Angolasaurus?</i>	sp.	5	1	1	1	Late Maas
<i>Angolasaurus</i>	<i>bocagei</i>	4	6.5	5.5	3.75	Late Tur
<i>Ectenosaurus</i>	sp.	2	2	5.5	3	Late Con – Late Sant??
<i>Ectenosaurus</i>	<i>clidastoides</i>	7	9.33	3.57	3.57	Late Con – Mid Sant
<i>Goronyosaurus</i>	<i>nigeriensis</i>	26	1.42	1.69	1.35	Late Maas
<i>Latoplatecarpus</i>	sp.	1	N/A	N/A	5	Mid–Late Camp
<i>Latoplatecarpus</i>	<i>nichollsae</i>	15	6.25	5	4.33	Late Sant – Mid Camp
<i>Latoplatecarpus</i>	<i>willistoni</i>	5	8.2	6.4	5	Mid Camp
<i>Platecarpus</i>	<i>ptychodon</i>	8	2	2.63	2	Late Maas
<i>Platecarpus</i>	<i>tympaniticus</i>	224	5.12	3.64	2.87	Early Sant – Mid Camp
<i>Plesioplatecarpus</i>	<i>planifrons</i>	30	6.33	4.14	3.76	Mid Con – Mid Sant
<i>Plioplatecarpus</i>	<i>houzeaui</i>	7	9.57	4.71	4.14	Early Maas
<i>Plioplatecarpus</i>	<i>marshi</i>	114	1.39	1.62	1.23	Late Maas
<i>Plioplatecarpus</i>	<i>depressus</i>	8	1.75	1.63	1.63	Early Maas
<i>Plioplatecarpus</i>	<i>peckensis</i>	1	11	6	4	Late Camp
<i>Plioplatecarpus</i>	<i>primaevus</i>	6	3.67	3.5	2.67	Mid–Late Camp

(continued)

APPENDIX. (Continued)

Genus	Species	Number	TCM	QCM	TCM	Range
<i>Selmasaurus</i>	<i>johnsoni</i>	1	13	7	4	Early Sant
<i>Selmasaurus</i>	<i>russelli</i>	3	3	3.67	3	Late Sant – Early Camp
<i>Tethysaurus</i>	<i>nopcsai</i>	5	9.5	5.6	4.25	Mid Tur
<i>Pannoniasaurus</i>	<i>inexpectatus</i>	118	1	1.19	1.36	Early–Late Sant??
<i>Hainosaurus</i>	<i>bernardi</i>	3	13.67	4.67	4.33	Early Maas
<i>Taniwhasaurus</i>	<i>antarcticus</i>	1	14	6	4	Late Camp
<i>Taniwhasaurus</i>	<i>mikasaensis</i>	4	2	2	2	Late Sant – Early Camp??
<i>Taniwhasaurus</i>	<i>oweni</i>	17	1.76	1.88	1.65	Mid–Late Camp
<i>Tylosaurus</i>	<i>capensis</i>	1	2	3	3	Late Con – Early Sant??
<i>Tylosaurus</i>	<i>gaudryi</i>	1	3	3	3	Late Sant
<i>Tylosaurus</i>	<i>iembeensis</i>	2	2.5	2.5	3	Late Tur
<i>Tylosaurus</i>	<i>ivoensis</i>	206	1	1.94	1.01	Early Camp
<i>Tylosaurus</i>	<i>kansasensis</i>	16	8.86	5.38	3.88	Late Con
<i>Tylosaurus</i>	<i>nepaeolicus</i>	39	3.97	3.8	3.58	Late Con – Early Sant
<i>Tylosaurus</i>	<i>pembinensis</i>	12	4.25	3.5	3.42	Mid Camp
<i>Tylosaurus</i>	<i>proriger</i>	163	4.98	3.40	2.75	Early Sant – Late Camp
<i>Romeosaurus</i>	<i>fumanensis</i>	4	7	3.25	3.5	Mid Tur
<i>Romeosaurus</i>	<i>sorbinii</i>	1	2	3	3	Mid Tur
<i>Russellosaurus</i>	<i>coheni</i>	1	4	5	3	Mid Tur
<i>Yaguarasaurus</i>	<i>columbianus</i>	4	5.75	5	3.5	Mid Tur