



THE FOSSIL RECORD OF ICHTHYOSAURS, COMPLETENESS METRICS AND SAMPLING BIASES

by TERRI J. CLEARY^{1,2}, BENJAMIN C. MOON¹, ALEXANDER M. DUNHILL³ and
MICHAEL J. BENTON^{1*}

¹School of Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Road, Bristol, BS8 1RJ, UK; e-mails: tc12618.2012@my.bristol.ac.uk, Benjamin.Moon@bristol.ac.uk, mike.benton@bristol.ac.uk

²Current address: Department of Earth Sciences, Natural History Museum, Cromwell Road, London, SW7 5BD

³School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, UK; e-mail: a.dunhill@leeds.ac.uk

*Corresponding author

Typescript received 4 November 2014; accepted in revised form 9 February 2015

Abstract: Ichthyosaurs were highly successful marine reptiles with an abundant and well-studied fossil record. However, their occurrences through geological time and space are sporadic, and it is important to understand whether times of apparent species richness and rarity are real or the result of sampling bias. Here, we explore the skeletal completeness of 351 dated and identified ichthyosaur specimens, belonging to all 102 species, the first time that such a study has been carried out on vertebrates from the marine realm. No correlations were found between time series of different skeletal metrics and ichthyosaur diversity. There is a significant geographical variation in completeness, with the well-studied northern hemisphere producing fossils of much higher quality than the southern hemisphere. Medium-sized ichthyosaurs are significantly more complete than small or large taxa: the incomplete-

ness of small specimens was expected, but it was a surprise that larger specimens were also relatively incomplete. Completeness varies greatly between facies, with fine-grained, siliciclastic sediments preserving the most complete specimens. These findings may explain why the ichthyosaur diversity record is low at times, corresponding to facies of poor preservation potential, such as in the Early Cretaceous. Unexpectedly, we find a strong negative correlation between skeletal completeness and sea level, meaning the most complete specimens occurred at times of global low sea level, and vice versa. Completeness metrics, however, do not replicate the sampling signal and have limited use as a global-scale sampling proxy.

Key words: completeness, Ichthyosauria, sampling bias, geological bias, sampling metrics.

PALAEONTOLOGISTS are keen to discover a reliable means to identify completeness of the fossil record. Suggested approaches include sampling standardization to equalize sample sizes, comparison and correction of fossil record data with proposed metrics of sampling such as formation or collection counts, identification of implied gaps (Lazarus gaps, ghost ranges) and consideration of specimen quality (reviewed in Smith 2007; Benton *et al.* 2011). In terms of specimen quality, it might be hypothesized that times of overall poor sampling should also correspond to times of poor specimen quality: incomplete or damaged specimens would be hard to identify and so diversity would be underestimated. Completeness metrics have been devised to document the preservation quality of taxa or individual specimens. These include taxon completeness scores that document whether species are represented by isolated bones, complete skulls or multiple skeletons (Fontaine *et al.* 2005; Benton 2008; Dyke *et al.* 2009), and completeness scores that document the percentage of the skeleton that is present (Mannion and Upchurch 2010; Beardmore *et al.* 2012; Brocklehurst *et al.* 2012).

The relationship between specimen completeness and diversity is unclear. One might expect that diversity would be highest when skeletons were most complete, and indeed, Brocklehurst *et al.* (2012) found a positive and statistically significant correlation between completeness and diversity for Mesozoic birds, and Mannion and Upchurch (2010) also found a correlation for sauropodomorph dinosaurs, but only for the Late Cretaceous. On the other hand, Brocklehurst and Fröbisch (2014) found a negative relationship between skeletal completeness and diversity for early synapsids, indicating a tendency among palaeontologists to name many species based on incomplete material.

Equally interesting is to assess whether skeletal completeness is a predictor of sampling more generally. Initial studies using completeness scores on terrestrial animals including sauropodomorph dinosaurs (Mannion and Upchurch 2010), birds (Brocklehurst *et al.* 2012) and non-mammalian synapsids (Brocklehurst *et al.* 2013; Walther and Fröbisch 2013; Brocklehurst and Fröbisch 2014) did not find any relationship between times when skeletal completeness was low and times of poor overall sampling

(i.e. low numbers of species, low numbers of fossiliferous formations). If anything, some times of apparently poor overall sampling corresponded to high overall skeletal completeness values, based on small numbers of sites of exceptional preservation. This could reflect some particular aspects of the sporadic nature of preservation of terrestrial fossil deposits and terrestrial tetrapods, so we chose to explore a group that is marine and apparently has a rich fossil record (McGowan and Motani 2003), the ichthyosaurs.

Ichthyosaurs were highly successful pelagic predators with a temporal range from the Early Triassic to the early Late Cretaceous (Motani 2009). They have an abundant fossil record for a large proportion of this time and have been intensely studied since the early nineteenth century. Many researchers have examined the diversity of ichthyosaurs as part of studies of all Mesozoic marine reptiles or for particular Mesozoic stages. While some consider potential biases affecting the fossil record (Benson *et al.* 2010; Benson and Butler 2011; Benton *et al.* 2013; Kelley *et al.* 2014), others only briefly mention (Thorne *et al.* 2011; Fischer *et al.* 2012) or do not consider (Zammit 2012) how this might affect observed diversity.

Mesozoic marine reptiles, including ichthyosaurs, have figured prominently in recent debates about the quality of the fossil record. In an initial study of the marine reptile record (Benson *et al.* 2010), strong correlations were found between apparent diversity and numbers of fossil reptile-bearing formations, and this was taken as evidence of prevalent bias. In a further study by Benson and Butler (2011), the ranking of rock volume and apparent diversity was found to indicate a biased record for pelagic taxa, but the correlations between formations and diversity for shelf taxa were ascribed by them to a 'common cause' (Peters 2005), namely sea level change and the resultant areas of continental flooding. This example illustrates how the commonly found covariation between fossil diversity and fossil-bearing formations could result from one of three causes, namely bias (Barrett *et al.* 2009; Benson *et al.* 2010), common cause (Peters 2005) or redundancy (Dunhill *et al.* 2014a), and all three should be considered as potential explanations (Benton *et al.* 2011; Upchurch *et al.* 2011).

Here, we explore the completeness of ichthyosaur specimens through their entire temporal and geographical distributions and investigate relationships with palaeodiversity, the rock record and sea level. We seek to identify times of low preservation quality, when a paucity of well-preserved fossils could increase the difficulty of identifying species. We explore host facies and completeness, as original deposition conditions can greatly affect preservation. We also compare records from the northern and southern hemispheres, as a preliminary test for any geographical variation in specimen completeness.

METHODS

Data

We constructed a matrix of 351 specimens, representing all 102 currently valid ichthyosaur species (Cleary *et al.* 2015, appendix 1, sheets 1–3). Up to ten specimens were scored from each species (range 1–10, mean = 3.44 specimens per species), and information was drawn primarily from the literature, in papers containing good images or detailed descriptions of specimens (or a combination of both). TJC also visited Bristol City Museum and Art Gallery and the Natural History Museum in London to study otherwise inaccessible specimens, test the coding methods on actual fossils and check aspects of ichthyosaur anatomy.

Decisions on which taxa to include and exclude from this study were made using the most recent taxonomic literature (McGowan and Motani 2003; Maisch 2010). If a species was considered a *nomen dubium*, it was excluded, except in cases where taxonomic validity was debated, for example the Cretaceous genus *Platypterygius*. Here, for completeness, we chose to retain species whose status is debated (Zammit 2012) as the study is based on individual specimens, and records of stratigraphic age, geographical location and overall size are unaffected.

A wealth of information was collected for each specimen (Cleary *et al.* 2015, appendix 1, sheets 1–3, 15), including geographical locality (modern coordinates), age (stratigraphic stage), body size (based on the length of the humerus when available) and geological setting (facies, divided into fine and coarse siliciclastic and carbonate categories, or a combination of the two).

Completeness metrics

We used two completeness metrics, the Skeletal Completeness Metric (SCM) and Beardmore's Skeletal Completeness Metric (BSCM). The SCM was devised by Mannion and Upchurch (2010) to document the skeletal completeness of sauropodomorph dinosaurs, and we adapted it for use with ichthyosaurs. The premise is to separate the skeleton into regions and then assign each region a percentage based on how much of the total skeleton that region represents. For ichthyosaurs, we divided the body into the skull, cervical + dorsal vertebrae, caudal vertebrae, pectoral girdle and forelimb, and pelvic girdle and hindlimb (Fig. 1A). We altered the proportions assigned to each skeletal division between Triassic and Jurassic/Cretaceous ichthyosaurs, as their body structure changed through time. As an example, the skull is rated at 20% in Triassic ichthyosaurs, but 30% in Jurassic and Cretaceous forms because it accounts for relatively more

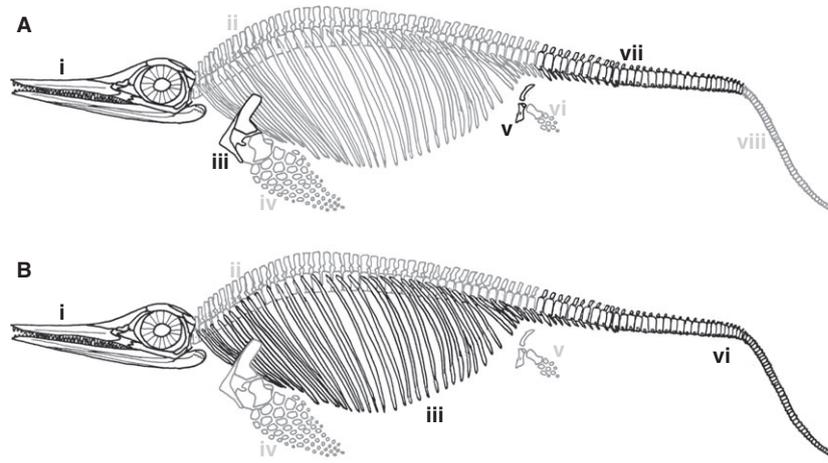


FIG. 1. Divisions of the ichthysaur skeleton for completeness metrics. A, Skeletal Completeness Metric region divisions and percentages, regions with different scores for the Triassic (Tr) and Jurassic–Cretaceous (J/K) are indicated: i, skull (Tr = 20%; J/K = 30%); ii, dorsal vertebrae and ribs (Tr = 40%; J/K = 30%); iii, pectoral girdle (Tr = 5%; J/K = 8%); iv, forelimbs (Tr = 10%; J/K = 10%); v, pelvic girdle (Tr = 3%; J/K = 3%); vi, hindlimbs (Tr = 7%; J/K = 4%); vii, tail axis (Tr = 10%; J/K = 10%); viii, tail fluke (Tr = 5%; J/K = 5%). B, Beardmore’s Skeletal Completeness Metric region divisions are scored up to a maximum of 4: i, skull; ii, dorsal vertebrae; iii, ribs; iv, forelimb and girdle; v, hindlimb and girdle; vi, caudal vertebrae. Outline adapted from Kirton (1983).

distinctive characters in later forms (Fig. 1A). Note that some regions are further subdivided. For example, in the Jurassic–Cretaceous, a preserved forelimb represents 10% of the total body and comprises the humerus (4%), radius (1%), ulna (1%) and phalanges (4%) (Fig. 1A). All divisions and subdivisions are listed in Figure 1A and Cleary *et al.* (2015, appendix 1, sheet 4). The sum of percentages from each area preserved gives a total SCM score.

The SCM had to be further adapted because ichthyosaurs are usually preserved in a lateral orientation, with only one side visible. Therefore, we report the completeness of one side of each specimen only. Where ichthyosaurs are preserved in the rarer dorsoventral orientation, we chose the best preserved side of the two. The skull must also be included as a whole entity, rather than as its individual components, as the compression of carcasses

often eliminates cranial sutures (McGowan and Motani 2003). Two SCM values were recorded for each species: the SCM1 was based on the most complete specimen from each species; and the SCM2 was a composite of the SCM1 value plus any missing parts added from other specimens.

The BSCM was designed for use with marine crocodylians (Beardmore *et al.* 2012), but we modified it for use with the ichthysaur body plan. The skeleton is divided into areas (Fig. 1B), and the completeness of each region is assessed according to a simplified scale, with a value between 0 (absent) and 4 (mostly/totally complete). For example, if approximately 40% of the dorsal vertebrae are present, then the dorsal section will score 2 (25–50% complete). The criteria for each numbered category can be found in Figure 1B and Table 1 (see also Cleary *et al.* 2015, appendix 1, sheet 2). The sections are totalled and

TABLE 1. Categorical completeness measures used here, termed the Beardmore’s Skeletal Completeness Metric (BSCM), in which portions of the skeleton are assigned to classes depending on a visual assessment of completeness.

BSCM	0	1	2	3	4
Skull	No skull	1–3 elements remain; limited preservation	1/2 elements	1–3 elements missing; skull shape recognizable	Complete
Dorsal vertebrae	0%	1–25%	25–50%	50–75%	75–100%
Forelimb + girdle	Absent	2/8 elements present	4/8 elements	6/8 elements	All present
Ribs	0%	Extensive loss (1–25%)	Moderate loss (25–50%)	Majority present (50–75%)	75–100%
Hindlimb +girdle	Absent	1/7 elements remain	3/7 elements	5/7 elements	All present
Caudal vertebrae	0%	1–25%	25–50%	50–75%	75–100%

For hindlimb measures, between-category designations are decided by the relative sizes of remaining elements vs. what is missing; for example, two elements but only pubo-ischial bones would be category 1. The BSCM total is (Total score/24)*100. Modified from Beardmore *et al.* (2012).

divided by the total possible score (24) to give a BSCM score, which is then multiplied by 100 to obtain a percentage, for better comparability to the SCM. As with the SCM, only one side of each specimen is measured. Furthermore, the cervical vertebrae are amalgamated with the dorsal vertebrae, as it can be hard to determine the division between these two areas in some taxa (Fig. 1B). We also integrated the pelvic and pectoral girdles into the limb categories, as they were not included in the original (Fig. 1B). Two versions of the BSCM are given, BSCM1 for the single best preserved specimen of each species, and BSCM2 for a composite comprising the best individual plus others that provide information on elements missing in the best specimen, to provide the most complete value possible.

Comparative time series

Several time series of physical environmental variables and potential sampling metrics were compiled and divided into time bins equivalent to Mesozoic stratigraphic stages. There are no values for the Bathonian or Valanginian stages because these stages have not yielded ichthyosaurs identified to species level. This may affect correlation strength and significance, but omitting these stages might have removed a genuine signal of non-preservation and so the decision was made to run the analyses twice, both retaining and removing the zero-value data. Mean completeness values were calculated for each time bin from the sum of all SCM and BSCM values from ichthyosaur species included in that time bin. For each time bin, we also recorded ichthyosaur diversity (number of species, from our data), the number of all fossiliferous marine formations (FMFs) and ichthyosaur collections, taken from the Paleobiology Database (PaleoDB; <http://fossilworks.org>; <http://www.paleobiodb.org/>).

Sea level data were taken from the standard summaries by Haq *et al.* (1987) and Miller *et al.* (2005), for which Butler *et al.* (2011, supplementary information) created equally spaced interpolations at 0.1-Myr intervals, enabling us to fit sea level data to our stage bins. We used the most recent *Geologic Time Scale* (Gradstein *et al.* 2012) to set dates for stage boundaries. As these data extend back only to the Ladinian, correlations with sea level exclude the Olenekian and Anisian stages. Research was carried out prior to the recent revision of the Rhaetian (Wotzlaw *et al.* 2014) and thus incorporates a longer stage duration; this does not affect the placement of formations in time bins, however, because our time bins are stages, not time increments.

Additional data recorded for each individual specimen included body size. This was assessed in classes, based on the length of the humerus, as small (<6 cm), medium

(6–14 cm) and large (>14 cm) categories. Exact body sizes were not estimated, because the humerus is easy to measure accurately and is proportional to total body size in any taxon (Maxwell 2012; Martin *et al.* in press), and we were interested simply in broad patterns of skeletal completeness in size classes. Ichthyosaur specimens were further categorized as coming from the modern northern and southern hemispheres, as a means of assessing evenness of collecting across the globe. Sedimentary facies for each specimen were also noted, as predominantly siliclastic or carbonate, based mainly on categories given in the PaleoDB. For these additional data, we grouped all individual specimens into categories, rather than using the ‘best specimen’ and ‘composite’ metrics.

In our study, we did not distinguish Lagerstätten from other deposits for statistical comparison, as the distinction is not clear for ichthyosaurs, and perhaps also for other marine reptiles, especially when compared to pterosaurs and birds (e.g. Brocklehurst *et al.* 2012). An easy solution would have been to choose only those geological formations that are traditionally called Lagerstätten (e.g. Guanling, Holzmaden, Solnhofen) and compare them with the rest. However, there is a sliding scale of ichthyosaur completeness between these, and other units of excellent preservation that are only sometimes called Lagerstätten (e.g. Lias of Dorset, Oxford Clay). Drawing the line would be arbitrary.

Relationships between pairs of time series were assessed using pairwise Spearman rank correlation tests and multiple regression models following the methods of Benson and Butler (2011), Benton *et al.* (2013) and Dunhill *et al.* (2014b). Time series were detrended using generalized differencing prior to correlation tests (with the *gen.diff* function of G. Lloyd: <http://www.graemetlloyd.com/meth-gd.html>; Cleary *et al.* 2015, supplement, appendix 6). False discovery rate (FDR) corrections were applied to families of associated correlation tests using the method of Benjamini and Hochberg (1995) to reduce the chance of acquiring type I statistical errors. Both linear modelling (‘lm’ and ‘step’ functions in R) and generalized least squares models (GLS; nlme and qpcR programs in R, ‘gls’ and ‘AICc’ functions) were applied. The linear models allowed sequential removal and addition of time series to seek the model that best explained the completeness metrics. GLS models take account of autocorrelation, and the GLS estimator is unbiased, consistent, efficient and asymptotically normal. We used the first-order autoregressive (AR(1)) correlation model, which has the property of seeking autocorrelation at up to one lag in either direction, and of minimizing the error term (Box *et al.* 1994). The quality of fit of models can be estimated using AIC and BIC values given by the GLS output, but these may not provide the best results for small sample sizes, as we have here. Therefore, we used the Akaike’s second-

order corrected information criterion (AICc command in qpcR program in R). We do not provide correlation coefficients. We do not compute R-squared ('pseudo-R-squared'), F-value or p-value for the GLS models as the merits of such estimators are currently debated (e.g. Freese and Long, 2006).

The aim was to determine whether any of the various metrics might be a reliable indicator of sampling quality, and also why some time bins might be better or worse represented by fossil specimens, and whether this might be associated with differences in specimen size or sedimentary facies available. Differences in completeness were assessed using Kruskal–Wallis tests. All analyses were carried out in R (v. 3.1.1), and we give code for the functions we used (Cleary *et al.* 2015, supplement, appendix 6).

RESULTS

Time series of completeness and sampling proxies

Both measures of skeletal completeness, SCM and BSCM, follow an almost identical pattern (Fig. 2; Cleary *et al.* 2015, appendix 1, sheets 4–9) and correlate strongly with each other, in both the 'best specimen' (SCM1, BSCM1) and 'composite' (SCM2, BSCM2) variants (Table 2; Cleary *et al.* 2015, supplement, appendix 2). There are significant differences between the SCM scores of each stage (Kruskal–Wallis: $\chi^2 = 87.329$, $df = 21$; $p < 0.001$).

During the Triassic, completeness is lowest during the Ladinian for all metrics (Fig. 2). This dip reflects the limited geographical range of sampling: only two species are known from one area of British Columbia. The rise in the Carnian after this low represents the Chinese Guanling Lagerstätte, whereas during the Norian there are numerous specimens, but poor completeness. Most of the Norian specimens are from a small area of Canada, and there is evidence for a marine transgression during this stage (Edwards *et al.* 1994), which may have led to a lack of restricted basinal facies that are associated with exceptional preservation.

Completeness varies throughout the Jurassic (Fig. 2), with the first peak in the Sinemurian, corresponding to the heavily sampled Blue Lias and Charmouth Mudstone formations (Dunhill *et al.* 2012), which have yielded many excellent, complete specimens of ichthyosaurs since the early 1800s. Completeness falls during the Middle Jurassic (Fig. 2), reflecting a paucity of localities that only produce a sparse assemblage of incomplete specimens. The Callovian peak in completeness reflects the geographically restricted collections from the Oxford Clay Formation that yield exquisite, and occasionally mostly complete ichthyosaur specimens (Martill 1986). There is a dramatic drop in

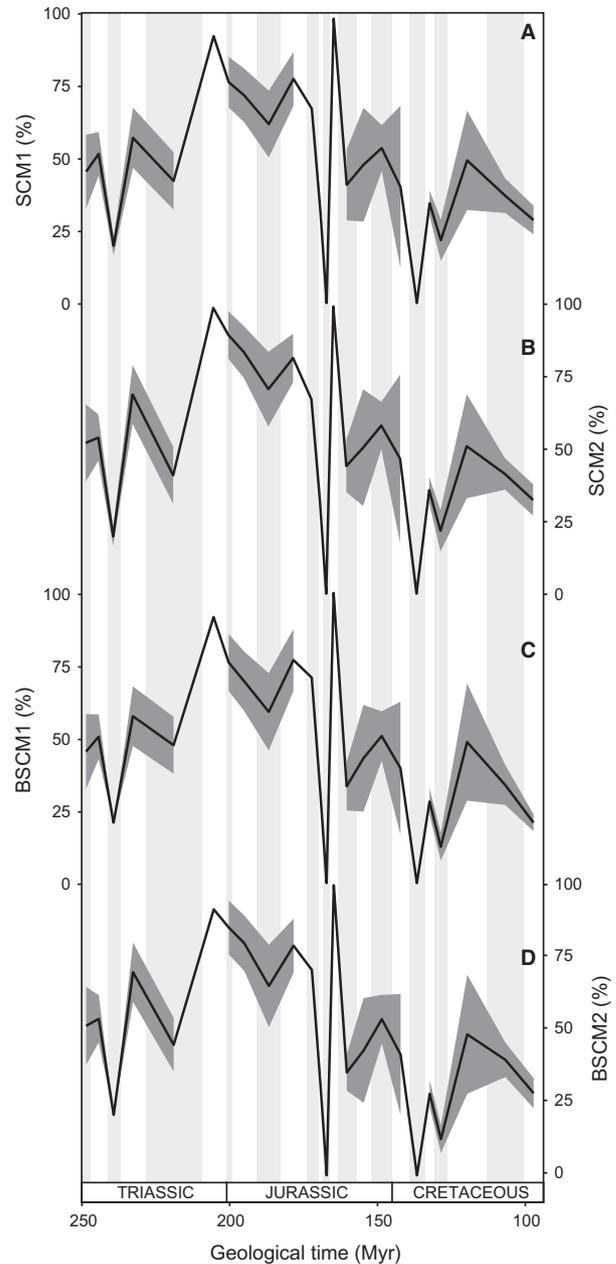


FIG. 2. Completeness metrics through the Mesozoic range of ichthyosaurs. A, SCM1; B, SCM2; C, BSCM1; D, BSCM2. Grey areas surrounding Skeletal Completeness Metric and Beardmore's Skeletal Completeness Metric represent 95% confidence intervals.

preservation quality across the Jurassic–Cretaceous boundary (Fig. 2), and it has long been debated whether this represents an extinction event or simply a major facies change, from marine to continental deposits, across Europe. In fact, the extinction rate of ichthyosaurs across the J/K boundary appears no higher than the background rate (Fischer *et al.* 2012; Zammit 2012), despite claims of an apparent mass extinction event at that time (Bambach

TABLE 2. Spearman's rank correlation coefficients (r_s) between time series data for all time bins, Triassic–Jurassic time bins, and Cretaceous time bins when time bins with no ichthyosaurs (Bathonian and Valanginian) are removed from the analysis.

	All time bins	Triassic–Jurassic	Cretaceous
SCM1 and BSCM1	0.98**	0.98**	1*
SCM2 and BSCM2	0.99**	0.99**	0.9
Diversity and collections	0.49*	0.35	1*
Diversity and FMFs	0.35	−0.08	0.8
Diversity and sea level	−0.15	−0.32	−0.1
Diversity and SCM1	0.18	0.26	0.9
Diversity and SCM2	0.15	0.23	0.7
Diversity and BSCM1	0.12	0.14	0.9
Diversity and BSCM2	0.2	0.23	0.9
Collections and FMFs	0.45*	0.21	0.8
Collections and sea level	−0.03	0.14	−0.1
Collections and SCM1	0.34	0.2	0.9
Collections and SCM2	0.3	0.19	0.7
Collections and BSCM1	0.32	0.19	0.9
Collections and BSCM2	0.33	0.18	0.9
FMFs and sea level	0.21	0.5	0.4
FMFs and SCM1	−0.12	−0.34	0.6
FMFs and SCM2	−0.13	−0.4	0.5
FMFs and BSCM1	−0.16	−0.31	0.6
FMFs and BSCM2	−0.13	−0.38	0.6
Sea level and SCM1	−0.68**	−0.78*	−0.3
Sea level and SCM2	−0.69**	−0.82**	−0.4
Sea level and BSCM1	−0.63*	−0.77*	−0.3
Sea level and BSCM2	−0.69**	−0.77*	−0.3

*p significant at 0.05.

**p significant after false discovery rate corrections (Benjamini and Hochberg 1995).

BSCM, Beardmore's Skeletal Completeness Metric; FMF, fossiliferous marine formation; SCM, Skeletal Completeness Metric.

2006). Completeness remains relatively low throughout the Cretaceous (Fig. 2), apart from a spike during the Albian, although is lower than the periods of best preservation in the Jurassic.

Our plot of ichthyosaur diversity through time (Fig. 3A) shows peaks in the Early and Middle Triassic, Early and Middle Jurassic, latest Triassic (Tithonian), and in the early Late Cretaceous. This diversity time series represents counts from the taxa we assessed, so is not complete, but it shows the same pattern as seen in previous, comprehensive compilations (e.g. Benson and Butler 2011, fig. 3), except for our J/K peak. The peaks in many cases represent Lagerstätten, sites of exceptional fossil preservation.

We compared the various completeness metrics with a number of sampling proxies (Figs 3–5). The results with

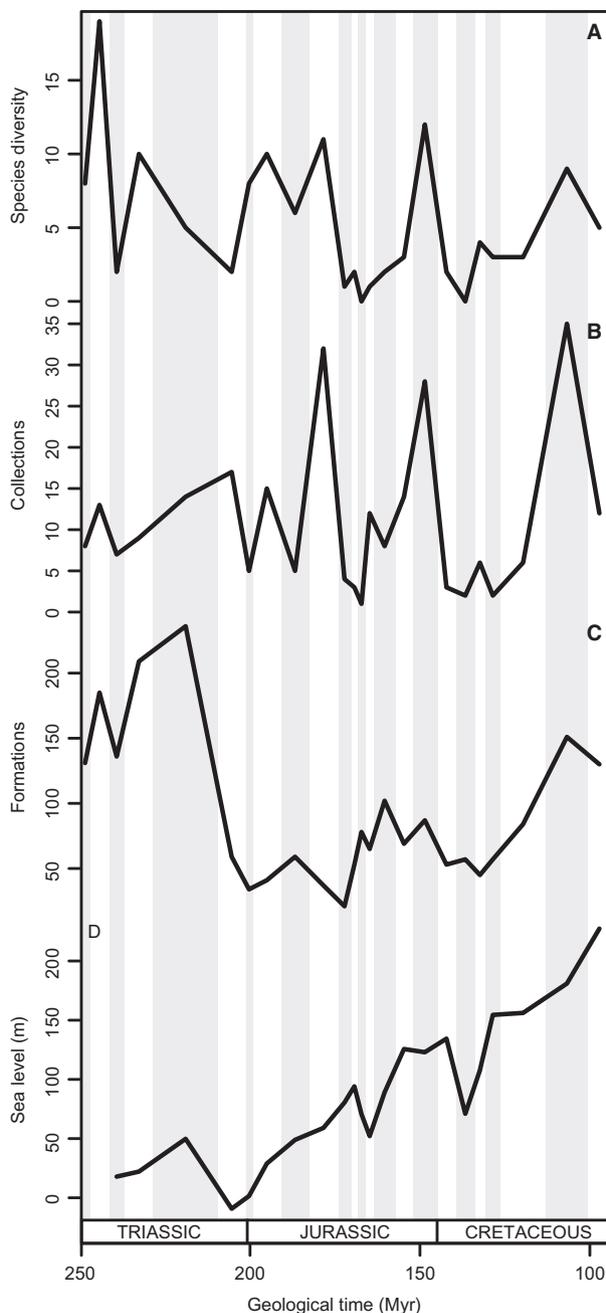


FIG. 3. Sampling proxies through the Mesozoic range of the ichthyosaurs. A, ichthyosaur species diversity; B, number of ichthyosaur-bearing collections; C, number of fossiliferous marine formations (FMFs); D, sea level. Sea level data from Butler *et al.* (2011); there are no sea level data for stages prior to the Ladinian.

and without zero–zero data are broadly similar (Table 2; Cleary *et al.* 2015, supplement, appendix 2), although the removal of the zero–zero data highlights the relationships between sea level and specimen completeness. All the results discussed further in this study refer to the data set with the zero–zero Bathonian and Valanginian data removed.

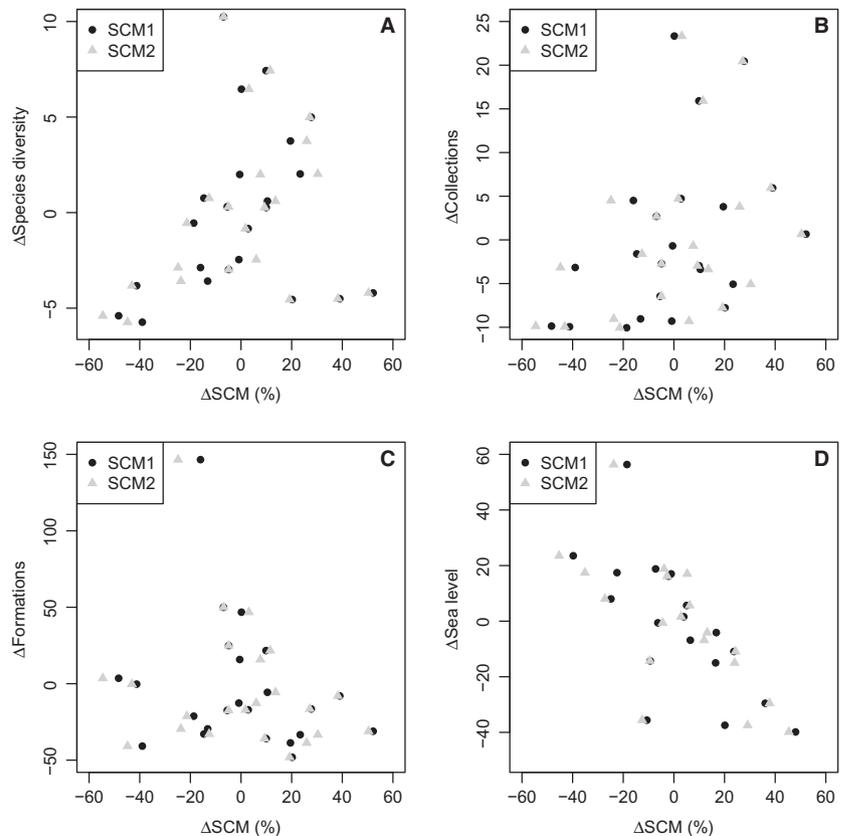
Ichthyosaur diversity correlates significantly with collection count, and collection and formation counts correlate significantly before FDR correction (Table 2). This could indicate a sampling bias or, more likely, may relate to the relative rarity of ichthyosaur fossils, compared to other fossil groups, and thus redundancy between diversity and collections metrics (Dunhill *et al.* 2014a). The non-correlation between raw diversity and completeness metrics, however, confirms that these metrics have no relationship to diversity, and that their use as a sampling proxy is limited. Ichthyosaur collections show no correlation with completeness metrics (Table 2; Figs 4–5). This suggests that there is no link between time bins, the abundance of ichthyosaur specimens and specimen completeness. Fossiliferous marine formation counts (FMFs) show no correlation with diversity, sea level or the completeness metrics (Table 2; Figs 4–5). One would expect rising sea level to increase formation count, because most marine formations are from the continental shelf, and rising sea level expands the area of continental shelf, but it appears not to be the case in this study. Sea level also does not correlate with any of the other proxies. However, sea level does correlate negatively and significantly with all the specimen completeness metrics (Table 2; Figs 4–5), and all but the correlation between sea level and BSCM1 survive FDR correction (Table 2). This shows that ichthyosaur

specimen completeness is highest during times of low sea level and deteriorates as sea levels rise.

As the completeness of ichthyosaur specimens seems to vary considerably between the Triassic–Jurassic time bins and the Cretaceous time bins, with an apparent marked dip in completeness across the Jurassic–Cretaceous boundary (Fig. 2), all correlations were run again for the Triassic–Jurassic and Cretaceous separately (Table 2). The results for the Triassic–Jurassic data were very similar to the total data set, albeit with stronger negative correlations between sea level and completeness, and non-significant correlations and non-significant results between diversity and collections, and collections and formations (Table 2). The Cretaceous data consist of fewer time bins, and therefore, the analysis lacks sufficient statistical power to make any conclusions.

The model fitting procedures provide rather different results. Multiple regressions highlight combinations of sea level, formations, collections and time period as the best predictors of specimen completeness (Table 3). As with the correlation results, the relationship between completeness and both sea level and formations is negative, suggesting that lower sea levels and fewer sampled formations result in specimens of higher completeness. The relationship between time period and completeness is also negative (as the coding refers to Triassic–Jurassic = 1,

FIG. 4. Correlation plots showing the relationships between completeness metrics and sampling proxies. Δ indicates that data have undergone generalized differencing prior to the application of Spearman rank correlation tests. A, SCM1/2 and diversity; B, SCM1/2 and collections; C, SCM1/2 and fossiliferous marine formations (FMFs); D, SCM1/2 and sea level. See Table 2 for correlation coefficients and p-values.



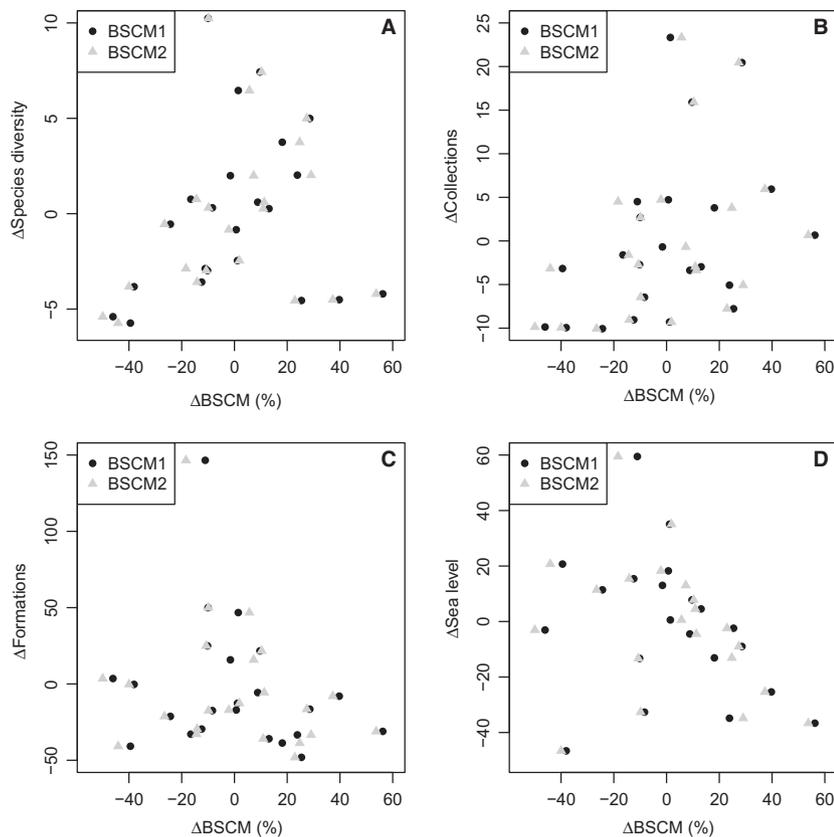


FIG. 5. Correlation plots showing the relationships between completeness metrics and sampling proxies. Δ indicates that data has undergone generalized differencing prior to the application of Spearman rank correlation tests. A, BSCM1/2 and diversity; B, BSCM1/2 and collections; C, BSCM1/2 and fossiliferous marine formations (FMFs); D, SCM1/2 and sea level. See Table 2 for correlation coefficients and p-values.

and Cretaceous = 2), confirming that Triassic–Jurassic specimens are, on average, more complete than Cretaceous specimens. The only independent variable that does not feature in any of the best fitting models is diversity, providing further evidence that recorded ichthyosaur diversity is not linked to specimen completeness.

Generalized least squares models do not eliminate diversity as a part of the best models for predicting specimen completeness (Table 4). Ranked by AICc value, the SCM1 is best explained by the model comprising collections, formations and sea level, and worst by the model comprising diversity, collections and formations. All five time series are roughly equally distributed between the best 16 models and the poorest 16 models, although, of single-factor models, time period performed best, and sea level, collections, diversity and formations were progressively poorer and poorer correlates of the SCM1 time series. The ‘top five’ models all contain collections and sea level as parameters, while the ‘bottom five’ do not all contain any one parameter, but formations occur in four of the five. All four SCMs showed similar best and poorest models (Cleary *et al.* 2015, supplement, appendix 4): the best models were 14 and 23 in all cases, with 7, 3, 16 and 9 always within the top five. The poorest five models were generally some mix of 25, 30, 17, 18 and 8, with 19, 22 and 28 featuring once. The GLS results then are equiv-

ocal, and do not confirm the exclusion of diversity as in some way related to specimen completeness.

Variation in completeness with body size, geography and lithology

Completeness varies with size: medium-sized ichthyosaurs were significantly more complete than smaller or larger ichthyosaurs (BSCM; Kruskal–Wallis: $\chi^2 = 10.578$, $df = 2$; $p = 0.005$). Small and large ichthyosaurs had very similar median completeness (Fig. 6; Cleary *et al.* 2015, appendix 1, sheet 10). This is surprising, because the null expectation was that larger ichthyosaurs would be more completely preserved than smaller ones, given the robustness of larger bones and their increased resistance to disarticulation and decay. There is a large range of completeness in each category (Fig. 6), however, which may be attributed to other factors such as geographical location and facies. Note that for the statistics in this section, SCM and BSCM were so similar that only one set of results is mentioned for size, hemisphere and geology comparisons (see Figs 6–8).

Northern hemisphere ichthyosaurs tend to be much more complete than southern hemisphere specimens (Fig. 7; Cleary *et al.* 2015, appendix 1, sheet 11; for SCM,

TABLE 3. Multiple regression results showing sampling proxy effects on completeness metrics.

Dependent	Full				Best			
	Independents	R ²	p	AIC	Independents	R ²	p	AIC
SCM1	Diversity + collections + formations + sea level + time period	0.45	0.02	-38.89	Collections + formations + sea level	0.48	0.004	-41.18
SCM2	Diversity + collections + formations + sea level + time period	0.37	0.04	-31.9	Collections + formations + sea level	0.43	0.007	-35.25
BSCM1	Diversity + collections + formations + sea level + time period	0.45	0.02	-31.24	Collections + formations + sea level + time period	0.48	0.006	-33.19
BSCM2	Diversity + collections + formations + sea level + time period	0.41	0.03	-29.86	Collections + formations + sea level + time period	0.45	0.01	-31.81

Details of full and best models and model selection process can be found in Cleary *et al.* (2015, supplement, appendix 4).

Kruskal–Wallis: $\chi^2 = 8.745$, $df = 1$; $p = 0.003$). While materials from the northern hemisphere show a large amount of variation at individual localities, southern hemisphere ichthyosaurs consistently show low completeness values, with the exception of two specimens from Argentina that are reasonably complete (Fig. 7).

When comparing the completeness of specimens recovered from different facies, we found no detectable difference in completeness between those recovered from coarse- vs. fine-grained lithologies (Cleary *et al.* 2015, appendix 1, sheet 12; for SCM, Kruskal–Wallis: $\chi^2 = 2.374$, $df = 1$; $p = 0.1$). However, ichthyosaur specimens in (coarser-grained) sandstones generally showed lower completeness scores than those in (finer-grained) mudstones in the original data, indicating that grain size should have an effect on completeness, but that a combination of facies factors (grain size and composition) is more important in preservation. A key example of these factors is whether each sediment is primarily siliciclastic or carbonate in its underlying lithology. There is a significant difference (Cleary *et al.* 2015, appendix 1, sheet 13; for SCM, Kruskal–Wallis: $\chi^2 = 8.840$, $df = 2$; $p = 0.01$) in completeness scores for specimens preserved in different lithologies of differing composition. Ichthyosaurs from predominantly siliciclastic deposits were best preserved, followed by those from mixed siliciclastic/carbonate facies, with the worst preserved recovered from predominantly carbonate units. When lithological categories are combined to reflect both composition and grain size, the five categories (Fig. 8; Cleary *et al.* 2015, appendix 1, sheet 14) show significant differences in completeness (for SCM, Kruskal–Wallis: $\chi^2 = 17.474$, $df = 4$; $p = 0.002$). Coarse siliciclastic and fine carbonate sediments appear to be associated with a poor level of fossil completeness, while fine siliciclastic sediments consistently yield the most complete specimens (Fig. 8). However, we do see a high variance of completeness values, especially among the finer-grained lithologies and mixed facies (Fig. 8).

DISCUSSION

Comparison of completeness metrics

The very close correlation between the SCM and BSCM was surprising, as they had been expected to differ. SCM assigns completeness based on the amount each region contributes to the overall skeleton, but BSCM counts all regions as having the same relative weighting. This means that SCM accounts for the higher preservation potential of some parts over others, while BSCM does not. However, the nearly uniform very highly significant correlation between the two (Table 2; Cleary *et al.* 2015, supplement, appendix 2) shows that such differences presumably do

TABLE 4. Statistical comparison of possible explanatory models for diversity of ichthyosaurs through the Mesozoic (Ladinian–Cenomanian interval).

Model	Parameters	df	Weighting	AICc	AIC	BIC	logLik
14	CFS	6	0.3612	202.3731	200.0202	206.5664	-94.01009
23	CS	5	0.1255	204.488	203.1547	208.6099	-96.57733
7	CFST	7	0.0967	205.0077	201.2577	208.895	-93.62887
3	DCFS	7	0.0667	205.7513	202.0014	209.6386	-94.00067
16	CST	6	0.0410	206.7245	204.3716	210.9178	-96.18579
15	CFT	6	0.0358	206.9954	204.6425	211.1887	-96.32123
24	CT	5	0.0316	207.2483	205.9149	211.3702	-97.95747
9	DCS	6	0.0283	207.4652	205.1122	211.6585	-96.55611
21	DT	5	0.0245	207.7519	206.4186	211.8738	-98.2093
12	DFT	6	0.0210	208.0632	205.7102	212.2565	-96.85512
32	T	4	0.0183	208.3366	207.705	212.0692	-99.85249
20	DS	5	0.0168	208.508	207.1747	212.6299	-98.58734
2	DCFST	8	0.0148	208.7609	203.1609	211.8893	-93.58047
11	DFS	6	0.0121	209.1612	206.8083	213.3545	-97.40413
4	DCFT	7	0.0101	209.5195	205.7695	213.4068	-95.88473
26	FT	5	0.0096	209.6287	208.2954	213.7506	-99.1477
31	S	4	0.0096	209.6339	209.0023	213.3664	-100.50114
10	DCT	6	0.0094	209.6683	207.3153	213.8616	-97.65767
13	DST	6	0.0088	209.8051	207.4522	213.9984	-97.72609
5	DCST	7	0.0080	210.0045	206.2545	213.8918	-96.12726
27	ST	5	0.0062	210.4988	209.1654	214.6206	-99.58271
6	DFST	7	0.0061	210.526	206.776	214.4133	-96.38801
22	CF	5	0.0057	210.6726	209.3392	214.7945	-99.66962
29	C	4	0.0055	210.7263	210.0947	214.4589	-101.04734
28	D	4	0.0047	211.0436	210.412	214.7762	-101.20602
19	DF	5	0.0045	211.1392	209.8059	215.2611	-99.90294
1	Null	3	0.0044	211.2081	211.0082	214.2813	-102.50407
25	FS	5	0.0042	211.2833	209.9499	215.4051	-99.97496
30	F	4	0.0027	212.1848	211.5533	215.9174	-101.77663
17	FST	6	0.0026	212.2211	209.8682	216.4144	-98.9341
18	DC	5	0.0018	213.0098	211.6765	217.1317	-100.83824
8	DCF	6	0.0018	213.0284	210.6755	217.2217	-99.33775

Models 1–31 comprise all different combinations of the diversity, collections, formations, sea level and time period (i.e. Triassic–Jurassic vs. Cretaceous) time series as parameters that might explain the skeletal completeness metric. As an example, the SCM1 metric is documented here, and the other metrics (SCM2, BSC1 and BSC2) are given in Cleary *et al.* (2015, supplement, appendix 5). Models 1–31, and the null model, are ranked in order of explanatory power, according to the Akaike's second-order corrected information criterion (AICc), where the smaller the value, the better the fit. In addition, the log-likelihood and the AIC and BIC of the best fitting model are given. The correlation structure is ARMA(1,0) in each case. *Abbreviations of parameters:* C, collections; D, diversity; F, formations; P, period; S, sea level; T, geological time.

not matter. Perhaps also this might suggest that either metric would be equally useful in studies of overall skeletal completeness such as this; the BSCM (Beardmore *et al.* 2012) is more rapid to assess than the SCM (Mannion and Upchurch 2010).

Drivers of diversity and fossil quality

Diversity and collections correlate, albeit only before FDR correction (Table 2). In general, any single ichthyosaur species may be a part of many collections (as listed by the

PaleoDB). This could be read as a simple metric of sampling – the more collections that are made (reflecting a combination of rock availability and collecting effort), the more ichthyosaur species are identified. Equally, this could be an indicator of the 'bonanza effect' (Raup 1977): time bins containing abundant fossils are much visited and much collected, so many ichthyosaur taxa are identified (Raup 1977; Dunhill *et al.* 2014b). Brocklehurst *et al.* (2013) identified that this may have been the case in their study of synapsid diversity, in which they found a similar significant correlation. Do collections drive diversity in this case (evidence of bias) or does diversity

FIG. 6. The effect of body size on ichthyosaur completeness metrics. A, Skeletal Completeness Metric; and B, Beardmore's Skeletal Completeness Metric. Body size is represented by humerus length, with small (< 6 cm), medium (6–14 cm) and large (> 14 cm) categories.

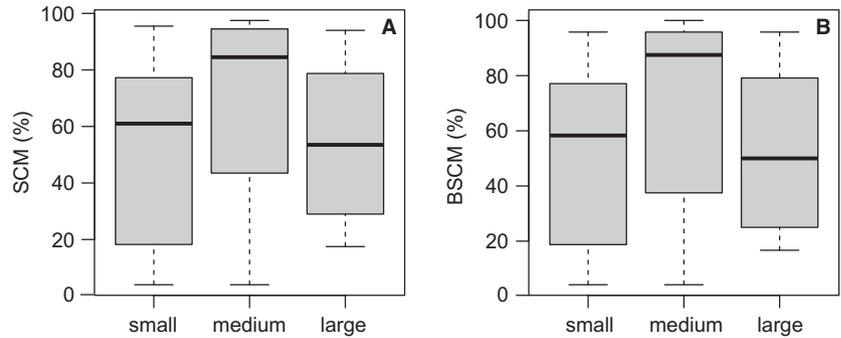


FIG. 7. The effect of geographical distribution on ichthyosaur completeness metrics. A, Skeletal Completeness Metric; and B, Beardmore's Skeletal Completeness Metric scores for ichthyosaurs from the northern and southern hemispheres.

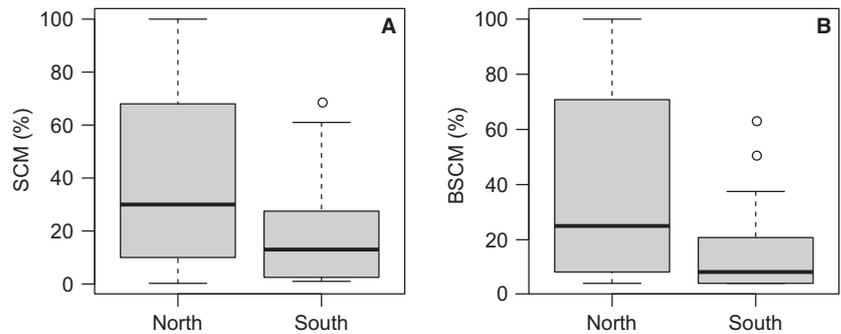
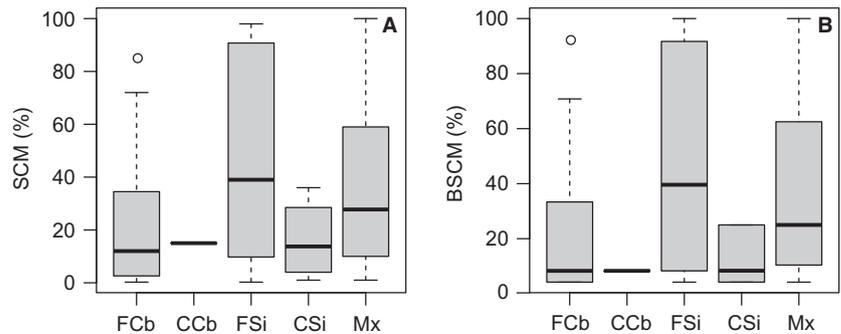


FIG. 8. The effect of lithology on ichthyosaur completeness metrics. A, Skeletal Completeness Metric; and B, Beardmore's Skeletal Completeness Metric. Completeness scores are included from all specimens where locality geology is known. FCb, fine carbonate; CCb, coarse carbonate; FSi, fine siliciclastic; CSi, coarse siliciclastic; Mx, mix of carbonate and siliciclastic facies.



(= fossil availability) drive collecting? The answer is probably both, and it is hard to discriminate whether time bins with high collection and high diversity values reflect sampling bias or genuinely high diversity, an example of redundancy (Benton *et al.* 2011; Dunhill *et al.* 2014a).

Similarly, low collection counts show some correspondence with low ichthyosaur diversity, reflecting an absence of ichthyosaur materials. It is unclear whether this means that ichthyosaurs were rare or absent in life (biological signal), were not preserved (preservation bias; geological signal) or were there and in the rocks, but just have not been collected (sampling bias). Dunhill *et al.* (2014a) found that the two variables drove each other equally in the fossil record of Great Britain, suggesting redundancy between the two signals. It is therefore not a given that the rarity or abundance of specimens or

collections is a metric simply of sampling; it could reflect reality.

There are exceptions to the correlation between diversity and collections. The Albian, for example, has the highest number of ichthyosaur collections but only nine recognized species (the Anisian holds the record, with 19 species). Here, other factors come into play. The Albian shows generally low values of specimen completeness, and this compromises the ability of palaeontologists to identify ichthyosaur collections, and a lack of collections generally hinders the identification of new species. Further, a mix of siliciclastic and carbonate facies is associated with lower completeness values. We do not have independent evidence, but it could also be that Albian ichthyosaur localities have been less intensively studied than those from some other stages.

There was no correlation between diversity and any of the four completeness metrics (Table 2; Fig. 4A–B). It was predicted that high completeness of specimens ought to enable more to be identified to species level and thus should enhance the reported diversity. Instead, specimen completeness appears to have no bearing on diversity and is therefore a poor proxy for global-scale sampling. It does not take into account other confounding factors that can affect how completeness varies between time bins. For example, one time bin may have beautiful, near-complete fossils, but be poorly sampled, while another may be heavily sampled but only produce an abundance of scrappy fossils. A case in point is the Anisian, which shows moderate mean completeness values, but high diversity, arising from large numbers of formations that show wide variation in completeness scores, including high values in some Lagerstätten.

Other studies have found a variety of results for this relationship. Brocklehurst *et al.* (2012) found a significant positive correlation between diversity and completeness for Mesozoic birds. Perhaps completeness provides a better proxy for sampling with terrestrial species, or for the avian fossil record in particular, which is notoriously patchy. Mannion and Upchurch (2010), however, demonstrated a lack of correlation between SCM and diversity in their sauropodomorph study. We found that medium-sized ichthyosaurs had higher preservation values than small or large specimens; there may be an upper limit on large size and preservation beyond which completeness begins to decline. It is possible that diversity has been inflated in some places because of the habit of naming new species from poor fossil remains; this may also apply to our study. A similar explanation is offered by Brocklehurst and Fröbisch (2014), who noted poor taxonomic practices in the mid-twentieth century in naming pelycosaurian-grade synapsids. Completeness metrics are useful to elucidate certain aspects of bias in the fossil record, as Benton *et al.* (2013) noted, but they cannot capture the entirety of the sampling biases affecting the fossil record.

The correlation of FMFs with collections (Table 2) is in line with earlier studies (Benson *et al.* 2010; Benson and Butler 2011), which suggests that both reflect some aspect of sampling. However, contrary to these studies, we found no correlation between formations and diversity. Benson and Butler (2011) regarded the formations–diversity relationship as key evidence for a rock record bias mechanism driving the record of open-ocean, pelagic marine reptiles. Here, without any significant correlation, we can only conclude that the FMF metric is not a good sampling proxy or, if it is a good proxy, we are not observing any significant sampling bias. FMFs also did not show any correlation with sea level.

Arguably our most striking result is the strong, but negative, correlation between sea level and all variants of

the completeness metrics (Table 2). Oddly, Mannion and Upchurch (2010) also found a negative correlation between skeletal completeness of sauropodomorph dinosaurs and sea level, primarily in the Late Jurassic and Early Cretaceous, which was hard to explain. They suggested that high sea levels might decrease the availability of land area, and so in some way diminish the quality of preservation of sauropod skeletons. In our case, the finding that ichthyosaurs are better preserved at times of low sea level and more poorly preserved at times of high sea level could indicate something about their habitats and eventual death locations. Some classic Lagerstätten, such as Solnhofen, correspond to shallow settings, but at a time of high sea level globally, whereas others, such as Holzmaden and the Oxford Clay, correspond to deeper water settings at times of high global sea level.

Similar patterns are present in invertebrate species on a more local geographical and temporal scale (Smith *et al.* 2001). In this case, the culprit appears to be the lack of suitable taphonomic settings: repeated transgressions in the Cretaceous created new areas of onshore, moderate depth (20–50 m) deposits in which skeletal remains were best preserved (Kidwell and Baumiller 1990). However, these were removed in the following regression by erosion of this part of the sequence, building a sequence of deeper water, less well preserving facies. Many ichthyosaur specimens are found in these shallower water settings (Martill 1986). The effect of increasing sea level through the Cretaceous does have a negative correlation with completeness (Table 2), but this is minor compared to what is found in the Triassic–Jurassic, and in the complete data.

Size, geographical location and geology

It was expected that larger ichthyosaurs would be better preserved. This is the norm for most fossil groups, including marine invertebrates (Cooper *et al.* 2006; Sessa *et al.* 2009) and some dinosaurs (Brown *et al.* 2013). Unexpectedly, we found that medium-sized ichthyosaurs had a higher median completeness than small or large, although there is much variation in each category (Fig. 6). It was expected that smaller specimens would not preserve as well, because of the lower robustness of smaller carcasses (Brown *et al.* 2013), but this is confounded by Lagerstätten that can preserve small forms in excellent detail. The largest ichthyosaur specimens might have been expected to be the best preserved, but this category contains many of the incomplete ichthyosaurs from poorly sampled areas such as Argentina and Russia. Mannion and Upchurch (2010) found a low completeness for sauropodomorphs despite their size, indicating that other factors may be at play. On the other hand, they noted that basal sauropodomorphs were the most complete and

titanosaurs, the least complete, perhaps reflecting the giant size of the latter. Is there an upper limit of size for good preservation quality, and could this apply to ichthyosaurs? It may also be that there are geographical or sampling biases, which are further described below.

Ichthyosaur specimens from the southern hemisphere tend to be much less complete than those from the northern hemisphere (Fig. 7). This could arise either from low preservation potential of facies in the south, or, most likely, from a lack of study. Argentinian ichthyosaur specimens show consistently low completeness scores, for example, but the vast majority of southern hemisphere ichthyosaurs originate from mudstones or even black shales, which should be associated with high completeness. This finding can be seen as only provisional, however, because the southern hemisphere sample size is very small, with only 21 specimens. The vast majority of localities in the northern hemisphere have been studied for a long time and have yielded dozens or hundreds of specimens. Argentina, on the other hand, continues to produce new species of ichthyosaur (Fernandez and Maxwell 2012), and there are still many countries that have barely been explored by palaeontologists. Brocklehurst *et al.* (2012) found a similar scenario between north and south for Mesozoic birds, with the majority of specimens originating from modern 30–60°N latitudes.

Our finding that sedimentary grain size does not affect fossil completeness (Fig. 8) was unexpected. It is predicted that fine-grained sediments are likely to host higher quality fossils than coarse-grained rocks. This is because sediment permeability is affected by the way grains fit together: the finer the grains, the less space there is between them, and the lower the permeability. Finer grains restrict ion transport in pore waters, impeding decay-causing bacteria (Allison 1988a), so such sediments tend to produce fossils of a much higher completeness than coarse-grained sediments. It is likely that the controls on preservation quality are more varied than simply grain size. Each sedimentary lithology represents a different depositional environment that can exert a variety of controls on preservation potential and thus specimen completeness. For example, black shales deposited in deep, anoxic waters represent ideal conditions for excellent preservation of marine fossils (Allison 1988a, b), but other kinds of mudstones, such as those deposited at delta fronts, lack such properties.

The higher chance of good preservation of ichthyosaurs in siliciclastic than carbonate sediments (Fig. 8) may be because benthic organisms may be more abundant in marine carbonates than in muds and sands, especially in coral reefs. Benthic organisms often scavenge carcasses on the seafloor, which can scatter parts and reduce their completeness. An oxygenated water column is needed for these organisms to thrive; the most complete ichthyosaur fossils often come from areas of anoxia, which cannot

support benthic organisms, and these are generally black shales, as noted above. Best and Kidwell (2000) found higher quality preservation in siliciclastic sediments than in carbonate or a siliciclastic–carbonate mix.

Some fine-grained siliciclastic rocks in this study do not yield ichthyosaur fossils of high completeness, whether for reasons of the original environment of deposition, or because sites may today lie in inaccessible or remote areas (e.g. Spitsbergen, Norway) where it is hard to recover complete skeletons, or where they are prone to intense weathering. There is also the issue of immature sampling in particular geographical areas, as mentioned above. More data are required from localities that have not been intensely sampled to identify areas of particularly poor preservation and elucidate causal mechanisms.

Implications for palaeodiversity studies

This study is the first to examine specimen completeness in a group of marine vertebrates, and it has revealed that ichthyosaur fossil completeness varies greatly through the Mesozoic, and how skeletal completeness relates to diversity and various sampling proxies. It is widely agreed that the vertebrate fossil record is incomplete and poorly sampled (Benson *et al.* 2010; Mannion and Upchurch 2010; Benson and Butler 2011; Benton *et al.* 2011, 2013), but determining the amount of error is extremely difficult. Widespread covariation of rock record and palaeodiversity signals has frequently been interpreted simply as evidence of bias, but it could equally be explained by a common cause model, or by varying degrees of redundancy between palaeodiversity and sampling proxy signals (especially counts of formations or collections for sparsely sampled taxa).

In the case of ichthyosaurs, it is evident that sampling worldwide has been extremely uneven, with long histories of collecting in western Europe, but relatively limited collecting in many other parts of the world. With greater effort devoted to collecting in the southern hemisphere, some of the inequalities of human sampling effort could be mitigated.

The absence of a relationship between numbers of fossiliferous formations and apparent diversity suggests that unevenness in knowledge of ichthyosaurs may result from other factors. Dunhill *et al.* (2014a) showed that formation count may fail as a sampling signal because it is redundant with recorded diversity. Clearly there are also geographical inequalities in sampling, and these may reflect differences in human effort devoted to particular time bins and to northern, rather than southern, continents. Most important was the evidence we have identified for selectivity in skeletal completeness scores relating to specimen size and facies. The most favourable

conditions for high skeletal completeness scores were for medium-sized specimens preserved in fine-grained siliciclastic rocks (particularly in black shales).

There is no reason that SCMs should correlate with other sampling metrics that attempt to quantify rock volume or rock availability, such as formation counts (Mannion and Upchurch 2010). Completeness metrics, however, have a direct causal link to observed diversity (Benton *et al.* 2013), as taxonomic identification requires a certain level of completeness. We therefore expected that we would find a correlation between completeness and diversity, but this was not the case. It has yet to be determined whether ichthyosaur specimen completeness correlates with the sampling-corrected diversity signal, as Brocklehurst *et al.* (2012) found for Mesozoic birds.

CONCLUSIONS

1. A study of 351 specimens belonging to 102 ichthyosaur species from the Olenekian to the Cenomanian shows that skeletal completeness does not correlate with diversity and thus is likely not a good global-scale sampling proxy. Completeness does have a relationship with collection counts, but the weakness of the correlation means only tenuous conclusions can be drawn.
2. Completeness fluctuated throughout the Mesozoic, with times of high quality often marked by localities of exceptional preservation (Lagerstätten). Times of low completeness were also identified: the Ladinian, much of the Middle Jurassic (Aalenian–Bathonian) and the majority of the Early Cretaceous. These times of low data quality should be taken into account when examining apparent ichthyosaur diversity.
3. Completeness is affected by ichthyosaur body size. Fossils of medium body size are the best preserved, but small sample size may account for the apparent poorer preservation potential seen in larger ichthyosaurs, as we would expect their robust bodies to be more resistant over time.
4. Ichthyosaurs from the northern hemisphere are much more complete than those from the South; it is unclear whether this is due to sampling or geological biases. The prevalence of fine-grained siliciclastic formations suggests the latter is more likely.
5. Facies composition has a significant effect on fossil completeness, with fine siliciclastic sediments showing the highest preservation potential, particularly if associated with anoxia, as in black shales. Coarse carbonate and coarse siliciclastic sediments appear to have the poorest preservation potential, perhaps because of their increased permeability to pore waters with oxygen and decay bacteria.

6. Skeletal completeness varies negatively with global sea level, which relates to the availability of suitable facies for preservation of fossils.
7. Completeness metrics are an effective proxy to highlight preservational bias in the fossil record. However, they do not capture the entirety of the bias signal and cannot explain ichthyosaur diversity patterns. Further study is required to obtain a larger picture of ichthyosaur fossil completeness for the Mesozoic, and to understand how this may affect observed diversity and the perceived evolution of the ichthyosaurs.

Acknowledgements. We thank Sandra Chapman at the NHMUK for her help in gaining access to the ichthyosaur collections. AMD is supported by a Royal Commission for the Exhibition of 1851 research fellowship. This work was partly supported by NERC doctoral training grant NE/H525111/1 to AMD.

DATA ARCHIVING STATEMENT

Data for this study are available in the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.657cc>.

Editor. Marcello Ruta

REFERENCES

- ALLISON, P. A. 1988a. Konservat-lagerstätten: cause and classification. *Paleobiology*, **14**, 331–344.
- 1988b. The role of anoxia in the decay and mineralization of proteinaceous macrofossils. *Paleobiology*, **14**, 139–154.
- BAMBACH, R. K. 2006. Phanerozoic biodiversity mass extinctions. *Annual Review of Earth and Planetary Sciences*, **34**, 127–155.
- BARRETT, P. M., MCGOWAN, A. J. and PAGE, V. 2009. Dinosaur diversity and the rock record. *Proceedings of the Royal Society B*, **276**, 2667–2674.
- BEARDMORE, S. R., ORR, P. J., MANZOCCHI, T. and FURRER, H. 2012. Float or sink: modelling the taphonomic pathway of marine crocodiles (Mesoeucrocodylia, Thalattosuchia) during the death–burial interval. *Palaeobiodiversity and Palaeoenvironments*, **92**, 83–98.
- 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, Series B (Methodological)*, **57**, 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, London, Special Publications*, **358**, 191–208.
- — LINDGREN, J. and SMITH, A. S. 2010. Mesozoic marine tetrapod diversity: mass extinctions and temporal heterogeneity in geological megabiases affecting vertebrates. *Proceedings of the Royal Society B*, **277**, 829–834.
- BENTON, M. J. 2008. Fossil quality and naming dinosaurs. *Biology Letters*, **4**, 729–732.

- DUNHILL, A. M., LLOYD, G. T. and MARX, F. G. 2011. Assessing the quality of the fossil record: insights from vertebrates. *Geological Society, London, Special Publications*, **358**, 63–94.
- RUTA, M., DUNHILL, A. M. and SAKAMOTO, M. 2013. The first half of tetrapod evolution, sampling proxies, and fossil record quality. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **372**, 18–41.
- BEST, M. M. R. and KIDWELL, S. M. 2000. Bivalve taphonomy in tropical mixed siliciclastic-carbonate settings. I. Environmental variation in shell condition. *Paleobiology*, **26**, 80–102.
- BOX, G. E. P., JENKINS, G. M. and REINSEL, G. C. 1994. *Time series analysis: forecasting and control*, Third edition. Prentice-Hall, Englewood Cliffs, NJ, 592 pp.
- BROCKLEHURST, N. 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.
- KAMMERER, C. F. and FRÖBISCH, J. 2013. The early evolution of synapsids, and the influence of sampling on their fossil record. *Paleobiology*, **39**, 470–490.
- 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 (Cretaceous, Alberta), a model Mesozoic terrestrial alluvial-paralic system. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **372**, 108–122.
- BUTLER, R. J., BENSON, R. B. J., CARRANO, M. T., MANNION, P. D. and UPCHURCH, P. 2011. Sea level, dinosaur diversity and sampling biases: investigating the “common cause” hypothesis in the terrestrial realm. *Proceedings of the Royal Society B*, **278**, 1165–1170.
- CLEARY, T. J., MOON, B. C., DUNHILL, A. M. and BENTON, M. J. 2015. Data from: the fossil record of ichthyosaurs, completeness metrics, and sampling biases. *Dryad Digital Repository*, doi:10.5061/dryad.657cc.
- COOPER, R. A., MAXWELL, P. A., CRAMPTON, J. S., BEU, A. G., JONES, C. M. and MARSHALL, B. A. 2006. Completeness of the fossil record: estimating losses due to small body size. *Geology*, **34**, 241.
- DUNHILL, A. M., BENTON, M. J., TWITCHETT, R. J. and NEWELL, A. J. 2012. Completeness of the fossil record and the validity of sampling proxies at outcrop level. *Palaeontology*, **55**, 1155–1175.
- HANNISDAL, B. and BENTON, M. J. 2014a. Disentangling rock record bias and common-cause from redundancy in the British fossil record. *Nature Communications*, **5**, 4818.
- BENTON, M. J., TWITCHETT, R. J. and NEWELL, A. J. 2014b. Testing the fossil record: Sampling proxies and scaling in the British Triassic-Jurassic. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **404**, 1–11.
- DYKE, G. J., MCGOWAN, A., NUDDS, R. and SMITH, D. 2009. The shape of pterosaur evolution: evidence from the fossil record. *Journal of Evolutionary Biology*, **22**, 890–898.
- EDWARDS, D., BARCLAY, J., GIBSON, D., KVILL, G. and HALTON, E. 1994. Triassic strata of the Western Canada sedimentary basin. 257–275. In MOSSOP, G. and SHETSON, I. (eds) *Geological atlas of the Western Canada sedimentary basin*. Canadian Society of Petroleum Geologists & Alberta Research Council, 510 pp.
- FERNANDEZ, M. S. and MAXWELL, E. E. 2012. The genus *Arthropterygius* Maxwell (Ichthyosauria: Ophthalmosauridae) in the Late Jurassic of the Neuquen Basin, Argentina. *Geobios*, **45**, 535–540.
- FISCHER, V., MAISCH, M. W., NAISH, D., KOSMA, R., LISTON, J., JOGER, U., KRÜGER, F. J., PÉREZ, J. P., 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.
- FOUNTAIN, T. M. R., BENTON, M. J., NUDDS, R. L. and DYKE, G. J. 2005. The quality of the fossil record of Mesozoic birds. *Proceedings of the Royal Society B*, **272**, 289–294.
- FREESE, J. and LONG, J. S. 2006. *Regression models for categorical dependent variables using Stata*. Second edition. Stata Press, College Station, 368 pp.
- GRADSTEIN, F. M., OGG, G. and SCHMITZ, M. 2012. *The geologic time scale 2012*. Elsevier, Amsterdam, 1176 pp.
- HAQ, B. U., HARDENBOL, J. and VAIL, P. R. 1987. Chronology of fluctuating sea levels since the Triassic. *Science*, **235**, 1156–1167.
- KELLEY, N. P., MOTANI, R., JIANG, D., RIEPPEL, O. and SCHMITZ, L. 2014. Selective extinction of Triassic marine reptiles during long-term sea-level changes illuminated by seawater strontium isotopes. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **400**, 9–16.
- KIDWELL, S. M. and BAUMILLER, T. 1990. Experimental disintegration of regular echinoids: roles of temperature, oxygen, and decay thresholds. *Paleobiology*, **16**, 247–272.
- KIRTON, A. M. 1983. A review of British Upper Jurassic ichthyosaurs. Unpublished PhD thesis, University of Newcastle upon Tyne, 235 pp.
- MAISCH, M. W. 2010. Phylogeny, systematics, and origin of the Ichthyosauria – the state of the art. *Palaeodiversity*, **3**, 151–214.
- 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.
- MARTILL, D. M. 1986. The stratigraphic distribution and preservation of fossil vertebrates in the Oxford Clay of England. *The Mercian Geologist*, **10**, 161–188.
- MARTIN, J., VINCENT, P., SUAN, G., SHARPE, T., HODGES, P., WILLIAMS, M., HOWELLS, C. and FISCHER, V. in press. A mysterious giant ichthyosaur from the lowermost Jurassic of Wales. *Acta Palaeontologica Polonica*, doi:10.4202/app.00062.2014.
- MAXWELL, E. E. 2012. New metrics to differentiate species of *Stenopterygius* (Reptilia: Ichthyosauria) from the Lower Jurassic of southwestern Germany. *Journal of Paleontology*, **86**, 105–115.
- MCGOWAN, C. and MOTANI, R. 2003. *Ichthyopterygia*. 1–175. In SUES, H. D. (ed.) *Handbook of Paleoherpology, Part 8*. Verlag Dr Friedrich Pfeil, Munich.

- MILLER, K. G., KOMINZ, M. A., BROWNING, J. V., WRIGHT, J. D. and MOUNTAIN, G. S. 2005. The Phanerozoic record of global sea-level change. *Science*, **310**, 1293–1298.
- MOTANI, R. 2009. The evolution of marine reptiles. *Evolution: Education and Outreach*, **2**, 224–235.
- PETERS, S. E. 2005. Geologic constraints on the macroevolutionary history of marine animals. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 12326–12331.
- RAUP, D. M. 1977. Systematists follow the fossils. *Paleobiology*, **3**, 328–329.
- SESSA, J. A., PATZKOWSKY, M. E. and BRALOWER, T. J. 2009. The impact of lithification on the diversity, size distribution, and recovery dynamics of marine invertebrate assemblages. *Geology*, **37**, 115–118.
- SMITH, A. B. 2007. Marine diversity through the Phanerozoic: problems and prospects. *Journal of the Geological Society*, **164**, 731–745.
- GALE, A. S. and MONKS, N. E. A. 2001. Sea-level change and the rock-record bias in the Cretaceous: a problem for extinction and biodiversity studies. *Palaebiologia*, **27**, 241–253.
- THORNE, P. M., RUTA, M. and BENTON, M. J. 2011. Resetting the evolution of marine reptiles at the Triassic-Jurassic boundary. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 8339–8344.
- UPCHURCH, P., MANNION, P. D., BUTLER, R. J., BENSON, R. B. J. and CARRANO, M. T. 2011. Geological and anthropogenic controls on the sampling of the terrestrial fossil record: a case study from the Dinosauria. *Geological Society, London, Special Publications*, **358**, 209–240.
- WALTHER, M. and FRÖBISCH, J. 2013. The quality of the fossil record of anomodonts (Synapsida, Therapsida). *Comptes Rendus Palevol*, **12**, 495–504.
- WOTZLAW, J.-F., GUEX, J., BARTOLINI, A., GALLET, Y., KRISTYN, L., MCROBERTS, C. A., TAYLOR, D., SCHOENE, B. and SCHALTEGGER, U. 2014. Towards accurate numerical calibration of the Late Triassic: high-precision U–Pb geochronology constraints on the duration of the Rhaetian. *Geology*, **42**, 571–574.
- ZAMMIT, M. 2012. Cretaceous ichthyosaurs: dwindling diversity, or the empire strikes back? *Geosciences*, **2**, 11–24.