Quantifying the completeness of the bat fossil record

Brown, Emily; Cashmore, Daniel; Simmons, Nancy; Butler, Richard

DOI:
10.1111/pala.12426

QUANTIFYING THE COMPLETENESS OF THE BAT FOSSIL RECORD

EMILY E. BROWN[1], DANIEL D. CASHMORE[1], NANCY B. SIMMONS[2], RICHARD J. BUTLER[1]

[1]School of Geography, Earth and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

[2]Department of Mammalogy, Division of Vertebrate Zoology, American Museum of Natural History, New York, New York, 10024, USA

Corresponding authors:

Emily E. Brown (emilyb62@hotmail.com), Daniel D. Cashmore (ddcashmore13@gmail.com)

Abstract: Bats (Chiroptera) are one of the most successful extant mammalian orders, uniquely capable of powered flight and laryngeal echolocation. The timing and evidence for evolution of their novel adaptations have been difficult to ascertain from the fossil record due to chronological gaps and the fragmentary nature of most fossil bat material. Here, we quantify the quality of the bat fossil record using skeletal and character completeness metrics, which respectively document for each taxon what proportion of a complete skeleton is preserved, and the proportion of phylogenetic characters that can be scored. Completeness scores were collected for 441 valid fossil bat species in 167 genera from the Eocene to Pleistocene. All metrics record similar temporal patterns: peak completeness in the Lutetian stage reflects the presence of Lagerstätten, while subsequent stages have very low completeness, excepting an Aquitanian high and a Pleistocene peak in skeletal completeness. Bat completeness is not correlated with intensity of sampling through geological time but has a weak negative correlation with publication date. There is no correlation between taxonomic richness and completeness, as the bat record predominately consists of diagnostic but isolated teeth. Consequently, bat skeletal completeness is the lowest of any previously assessed tetrapod group, but character completeness is similar to parareptiles and birds. Bats have significantly higher character completeness in the northern hemisphere, likely due to heightened historical interest and presence of Lagerstätten. Taxa derived from caves are more complete than those from fluviolacustrine and marine deposits, but do not preserve highly complete specimens.

KEYWORDS: Chiroptera, completeness metrics, Cenozoic, caves, Lagerstätten, sampling bias

INTRODUCTION

Bats (Chiroptera) are one of the most successful and diverse orders of extant mammals. They have a nearly global distribution, a varied range of body sizes, are known to have insectivorous, carnivorous, frugivorous, nectarivorous and hematophagous diets, and with nearly 1400 species, bats account for ~20% of all extant mammalian species (Wilson and Reeder 2005; Shi and Rabosky 2015; Tsang *et al.* 2015; Simmons and Cirranello 2018). Bats are the only living mammals, and one of only three vertebrate groups (along with pterosaurs and birds), that are capable of true self-powered flight (Padian 1985; Norberg 1990; Templin 2000; Dudley *et al.* 2007; Swartz *et al.* 2012). The extreme elongation of their forelimb digits plus development of a thin interconnected membrane (patagium) gives bats a unique body plan (Swartz *et al.* 1992; Swartz and Middleton 2008). They are also the only mammals capable of sophisticated laryngeal echolocation, in which they emit high-frequency sounds and interpret returning echoes to perceive their environment, enabling 'visualisation' in complete darkness (Arita and Fenton 1997; Teeling *et al.* 2012). Echolocation is the dominant mode of sensory perception in all bat families, with the exception of many of the fruit bats (Pteropodidae), which instead rely primarily on vision and olfaction (Arita and Fenton 1997; Boonman *et al.* 2013). Understanding the timing of and mechanisms underpinning the evolution of the unique adaptations of bats is an important goal of evolutionary biologists.

Monophyly of Chiroptera is consistently strongly supported by morphological and molecular analyses (e.g. Springer *et al.* 2001; Teeling *et al.* 2005; Song *et al.* 2012; Meredith *et al.* 2011; O'Leary *et al.* 2013; Tsagkogeorga *et al.* 2013; Lei and Dong 2016). The relative position of Chiroptera within the placental mammal tree has been the subject of debate, but it

is now uniformly placed within Laurasiatheria (Springer *et al.* 2001; Teeling *et al.* 2002; Song *et al.* 2012; Meredith *et al.* 2013; O'Leary *et al.* 2013; Tsagkogeorga *et al.* 2013; Halliday *et al.* 2017). Chiroptera is divided into two suborders, Yinpterochiroptera and Yangochiroptera, based on morphological, molecular, and genomic evidence (Teeling *et al.* 2002; Meredith *et al.* 2013; Tsagkogeorga *et al.* 2013; Lei and Dong 2016). These suborders are not distinguished by echolocation ability (both clades contain laryngeal echolocators), and differ from the traditional division of bats into Megachiroptera and Microchiroptera (Teeling *et al.* 2005; O'Leary *et al.* 2013). Time-calibrated molecular and genomic phylogenies indicate that bats evolved shortly after the Cretaceous-Palaeogene (K-Pg) boundary, with the split of Yinpterochiroptera and Yangochiroptera occurring around 63 million years ago (Ma) (Teeling *et al.* 2005; Lei and Dong 2016). Analyses combining morphological and molecular data similarly indicate an early origin around the K-Pg boundary, but suggest that the bat crown group diversified later, perhaps 57 Ma (O'Leary *et al.* 2013). Fossils could potentially provide the critical direct evidence of these relationships and divergence dates. However, the fossil record does not shed light on earliest history of bat evolution as the first unambiguous bats occur only in the early Eocene, leaving a 10-million-year gap between the apparent origin of the lineage and the first recognizable bat fossils (Gunnell and Simmons 2005).

Despite existence of several well-known complete skeletons, the fossil record of bats has generally been regarded as being exceptionally poor (Gunnell and Simmons 2005; Teeling *et al.* 2005; Eiting and Gunnell 2009). There are large chronological gaps, and most taxa are known from isolated teeth, jaw fragments, or postcranial elements. Only a small number of Eocene taxa and one Miocene taxon are known from Lagerstätten. Consequently, our understanding of the evolution of early bats, including their biology and functional morphology, comes predominantly from a few complete or near complete fossils recovered from Lagerstätten. The fragmentary nature of the bat record has almost exclusively been

proposed to result from the fragility of their skeletons (Gunnell and Simmons 2005). It is very important to take into account the limitations of a fossil record when making inferences about the evolution of a group. The quality of the bat fossil record has been previously quantified twice. Teeling *et al.* (2005) concluded that 61% of the bat fossil record is missing based on the extent to which the fossil record underestimates first appearances derived from molecular phylogenies. Eiting and Gunnell (2009) calculated that only 12% of bat fossil occurrence ranges represent 'true' temporal ranges, based on the phylogenetically independent extinction rates and preservation potential (Foote and Raup 1996). Both of these assessments analysed the absolute completeness of the fossil record - out of all bats that have ever existed, the relative percentage that are actually represented in the fossil record.

Completeness of the fossil record, and the impact(s) that missing data may have on interpretations of modes and patterns of evolution, have been a source of discussion for several decades (e.g. Dingus 1984; Foote *et al.* 1999; Benton *et al.* 2000; Smith 2001; Cooper *et al.* 2006). Recently, numerous studies have assessed completeness of the fossil record based on specimen-level data (Benton *et al.* 2004; Fountaine *et al.* 2005; Smith 2007; Dyke *et al.* 2009; Benton 2010; Mannion and Upchurch 2010; Brocklehurst *et al* 2012; Walther and Fröbisch 2013; Brocklehurst and Fröbisch 2014; Cleary *et al.* 2015; Dean *et al.* 2016; Verrière *et al.* 2016; Davies *et al.* 2017). These studies aim to assess variation in information content of fossil taxa and the entire record of a group, rather than trying to quantify temporal gaps in the record or stratigraphic fit. A high-quality record defined using this approach would be one that contains many highly complete taxa. Meaningful comparisons can subsequently be drawn to various sampling biases that can influence the macroevolutionary understanding of a group. Specimen quality of a group or time bin can theoretically be influenced by environmental and geological parameters, such as the types of depositional regimes or the number of localities preserved in the record (Dingus 1984; Retallack 1984). A

high number of localities from depositional settings with higher quality preservation could lead to increased specimen completeness. Specimen quality can be associated with body size or the robustness of skeletons (Cooper *et al.* 2006; Brown *et al.* 2013). The anthropogenic sampling of a group or time bin can also potentially influence the quality of the fossil record, as variation in historical or geographical sampling by palaeontologists could reduce or increase taxon completeness. Incomplete specimens may also be hard to identify taxonomically, either reducing estimates of diversity for a group or time bin or, conversely, increasing diversity as a result of taxonomic oversplitting (Brocklehurst and Fröbisch 2014) based on partial skeletons.

A number of completeness metrics have been conceived to accurately quantify the specimen completeness of different groups (Mannion and Upchurch 2010). Previous studies have found varying correlations between these completeness metrics and changes in diversity and fossil record sampling metrics through time (Mannion and Upchurch 2010; Brocklehurst *et al* 2012; Walther and Fröbisch 2013; Brocklehurst and Fröbisch 2014; Cleary *et al.* 2015; Dean *et al.* 2016; Verrière *et al.* 2016; Davies *et al.* 2017), thus highlighting the various biases acting upon different fossil records.

Here, we quantitatively assess the fossil record of bats using several completeness metrics originally developed by Mannion and Upchurch (2010) and Beardmore *et al.* (2012). We statistically compare the relationships between completeness and temporal and geographical distributions, depositional environments, rock record and taxonomic diversity changes through geological time, and human research effort through historical time. We aimed to understand the extent of the environmental and geological controls acting on the bat fossil record, and to compare bat completeness to other tetrapod groups.

MATERIALS AND METHODS

Dataset

Detailed records of the preserved osteological elements from over 4000 specimens of fossil bat were compiled from photographs, illustrations, and descriptions from published literature and museum catalogues. A list of genera from Eiting and Gunnell (2009) was used as the basis for taxon inclusions, but any genera named since Eiting and Gunnell (2009) were also collected from the literature. The Paleobiology Database (PaleoDB, www. paleobiodb.org) was used to determine the validity of each taxon as of July 2018. We gathered completeness scores for most specimens of each species and also for material indeterminate at species level (e.g. *Archaeonycteris* sp.). Specimens taxonomically assigned using the *cf.* connotation were treated as true representatives of those taxa. To avoid duplication of effort, for each individual species additional referred specimens were only added to the dataset if the skeletal elements represented were not already present in the material scored for that species (for example, if the first specimen scored for a species included a complete right mandible, further specimens of right mandibles for that species were not added to the dataset as they do not add to the scores for the completeness metrics used). Only specimens from intervals prior to the Holocene were recorded due to the vast number of specimens in the most recent deposits. "Quaternary" or "sub-Recent" deposits were therefore only included if explicitly cited as Pleistocene or dated older than 11,700 years old. In addition to preservation quality, information on depositional environment (either 'fluvio-marine', defined as fluviolacustrine and marine sedimentary regimes, or 'caves', here encompassing depositional settings defined as caves, karstic fissure fills and sinkholes in the literature; Jass and George 2010) and geographic locality were also recorded. Specimen records that were indeterminate at the

species-level were excluded from species-level analyses, but were included when assessing generic-level completeness. The final dataset consists of skeletal and character completeness data for 441 fossil species in 167 genera and is up to date as of July 2018 (see Brown *et al.* 2019).

Completeness metrics

Fossil record quality has been examined previously using a wide range of methods. Individual taxon completeness has been increasingly used to assess variation in fossil record quality over the last decade, and applied to a variety of taxonomic groups. Initial methods for quantifying taxon completeness were relatively crude and somewhat subjective. Benton *et al.* (2004), Fountaine *et al.* (2005) and Benton (2008) scored the completeness of known fossil material for a species by splitting preservation quality into between four and five categories (e.g. 'one bone', '> one bone', 'one specimen', 'more than one specimen'; Fountaine *et al.* 2005). Similarly, Beardmore *et al.* (2012) assessed species completeness using five categories (0-4) of preservation quality, but calculated the preservation quality separately for six skeletal regions.

Two new completeness metrics, the skeletal completeness metric (SCM) and the character completeness metric (CCM), were devised by Mannion and Upchurch (2010) in an attempt to quantify specimen completeness more objectively and in more detail. SCM is a measure of the absolute proportion of the skeleton that is preserved for a species, whereas CCM measures the proportion of phylogenetically informative characters preserved. Each has two implementations, respectively designated as SCM1/SCM2 and CCM1/CCM2. The first implementation (SCM1 and CCM1) determines completeness for a species based on its most completely known specimen, while the second uses all known specimens of a species in

calculating completeness. The second implementation has been generally favoured in previous studies (Brocklehurst *et al.* 2012; Dean *et al.* 2016; Verrière *et al.* 2016) and is also preferred here. These metrics have been widely used in recent completeness studies on a variety of fossil groups (Brocklehurst *et al.* 2012; Walther and Fröbisch 2013; Brocklehurst and Fröbisch 2014; Cleary *et al.* 2015; Dean *et al.* 2016; Verrière *et al.* 2016; Davies *et al.* 2017; Tutin and Butler 2017).

For the SCM, relative proportions of skeletal elements need to be determined. Different methods for calculating the relative proportions of skeletal regions have been developed: Brocklehurst and Fröbisch (2014) modelled different pelycosaur-grade synapsid bones as simple 3-dimensional shapes (e.g. cones, cylinders, etc.) to work out the relative volumes of skeletal regions. Verrière *et al.* (2016) revised this method to increase accuracy for use with parareptiles by modelling elements with truncated shapes to work out the surface area (instead of the volume) of different skeletal regions. Here the skeletal proportions were determined from: (1) the surface area of bones based on a 2D scientifically-informed skeletal reconstruction (Fig. 1A-B), and (2) from bone volume of a CT-scanned 3D skeleton (Fig. 1C). These were used to generate "SCM2 2D" and "SCM2 3D" scores. Lateral view skeletal reconstructions of *Icaronycteris index* (an Eocene fossil bat) and *Myotis myotis* (an extant vespertilionid bat) were redrawn from Wimsatt (2012) in Adobe Illustrator (CS5, www.adobe.com). The surface area of all skeletal elements was calculated for these two species using a free Illustrator plug-in, Patharea Filter (http://telegraphics.com.au/sw/product/patharea). The relative weighting of different skeletal regions was determined respectively in both species. The relative skeletal proportions of *Icaronycteris index* were attributed to all species belonging to extinct genera, while the proportions of *Myotis myotis* were used for members of extant genera. The use of both Ypresian and extant taxa – species that bracket the majority of the phylogenetic tree of bats

(Simmons and Giesler 1998; Simmons *et al.* 2008) – accounts for some of the major morphological changes in bat evolution; however, there are limitations as many families or even genera exhibit large morphological disparity (Freeman 2000; Shi *et al.* 2018) not captured by simple segregations of taxa as either extinct or extant.

The 3D method utilised a CT scan of a preserved extant specimen of *Myotis daubentonii* (AMNH 218932). AMNH 218932 was scanned by NBS at The University of Texas High-Resolution X-ray CT Facility as 1595 coronal slices (1024 x 1024 pixels, 0.036 voxels). Skeletal elements were virtually segmented using the 'Masks' and 'Curves' tools in SPIERSedit (2.20, www.spiers-software.org). The 'Measure volumes' option provided the volume (in voxels) of each element which was used to determine the relative weighting of skeletal regions. This 3D approach may be the most accurate representation of relative proportions used yet in the literature. The calcar was removed from the 3D skeletal body proportions for extinct taxa as the 2D proportions for extinct taxa derived from analysis of *Icaronycteris index* lacked the element. 2D SCM2 and 3D SCM2 scores of extant taxa were statistically compared to directly test the validity of using 2D scientifically informative diagrams to estimate relative skeletal proportions.

In addition to the methods described above, we also used a modified version of SCM2. Traditional SCM2 assesses completeness of both the left and right sides of the skeleton separately even though this is likely not necessary as tetrapods are bilaterally symmetrical and only one whole side of the skeleton is potentially needed to understand the complete anatomy of a taxon. For this reason, we also calculated the proportional completeness of each taxon in relation to only one bilateral side of a full skeleton, but utilising the composite material available from both sides. Cleary *et al.* (2015) first employed a similar method because of the compressed nature of most ichthyosaur fossil specimens, but it is used here to test whether traditional SCM significantly underestimates our understanding

of fossil bats. This new metric does not provide an absolute measure of the preserved skeleton like the standard SCM2, but instead more closely assesses the proportion of preserved anatomical information, bridging the gap between SCM and CCM. This new approach was used to calculate two additional versions of SCM2, CSCM2 2D and CSCM2 3D ('Composite' SCM2).

Beardmore's Skeletal Completeness Metric (BSCM) was initially developed to assess the geologically localised taphonomy of marine crocodylomorphs (Beardmore *et al.* 2012), but has also since been modified and applied to assess the global fossil record of ichthyosaurs (Cleary *et al.* 2015). The BSCM is a simpler metric that involves assigning a value between 0 (absent) and 4 (mostly/fully complete) different regions of the skeleton (in this case six regions, combining neck, dorsal and lumbar vertebral sections). The values from these regions are then totalled and transformed into a skeletal completeness percentage. Here BSCM is incorporated to assess whether a less refined but more easily implemented metric might serve to represent the completeness of the bat fossil record just as well as other metrics.

To assess completeness with respect to characters we used traditional CCM2 (CCM2a, Mannion and Upchurch 2010) where skeletal regions are weighted by assigned numbers of characters, and the physical completeness of individual elements are used as a guide for the presence of characters within each element. For example, if a femur is complete it is assumed all the characters of that femur are obtainable. However, this might not necessarily always be the case as bones frequently suffer surface damage and erosion. We used an extensive osteological character list (699 distinct characters) provided by NBS and specifically designed to include fossil chiropterans (see Brown *et al.* 2019). The number of characters per skeletal region (e.g. skull, mandible, axial skeleton etc.) were counted and used to determine the relative character proportions assigned to each (Brown *et al.* 2019: Table S1). These relative proportions were used to score completeness values per taxon. The

individual completeness of elements was based on the composite completeness of the elements derived from both left and right bilateral sides of the skeleton, as the distinction between left and right sides is irrelevant for phylogenetic character scoring in symmetrical animals such as bats. Characters pertaining to the calcar were removed from the proportions for extinct taxa to keep consistent body proportions between metrics, as mentioned above. Character 698 and 699, were omitted from the analyses as they pertain to the baculum, a gender-specific element. Characters that were not discrete to a specific element were assigned a proportion of the character (e.g. character X references both the humerus and femur: humerus and femur were assigned 50% of a character for determination of relative proportions).

Time bins and comparative data

For analyses over geological time, stage-level time bins were used between the Ypresian (early Eocene) and Messinian (late Miocene). Subsequently, epoch-level time bins were used due to the short durations of the Pliocene and Pleistocene. The completeness for each metric was calculated for each time bin as the mean and median of the completeness scores for all species and genera (separately) occurring within that time bin.

Two different estimates of fossil record sampling were used. All fossil records of Cenozoic tetrapods from terrestrial depositional settings were downloaded from the PaleoDB on 15th May 2018 (see Brown *et al.* 2019). Numbers of geological formations yielding tetrapod fossils (tetrapod-bearing formations, TBFs) and numbers of PaleoDB 'collections' (= localities) yielding tetrapod fossils (tetrapod-bearing collections, TBCs) were calculated from this download. TBFs and TBCs were used instead of counts of numbers of formations yielding bats, in order to provide a more global estimate of sampling, reducing the possibility

of signal "redundancy" (Benton *et al.* 2011). It is worth noting, however, that recent simulation studies have indicated that using a wider clade does not necessarily remove the issue of redundancy (Brocklehurst 2015, Dunhill *et al.* 2018).

Bat completeness metrics were also compared with those from other tetrapod groups for which completeness data has been compiled. SCM2 data were compared to values for sauropodomorphs (Mannion and Upchurch 2010), pelycosaurs (Brocklehurst and Fröbisch 2014), ichthyosaurs (Cleary *et al.* 2015), parareptiles (Verrière *et al.* 2016), and plesiosaurs (Tutin and Butler 2017). CCM data were compared to those for sauropodomorphs (Mannion and Upchurch 2010), birds (Brocklehurst *et al.* 2012), anomodonts (Walther and Fröbisch 2013) pelycosaurs (Brocklehurst and Fröbisch 2014), pterosaurs (Dean *et al.* 2016), parareptiles (Verrière *et al.* 2016), plesiosaurs (Tutin and Butler 2017), and Cretaceous and Paleogene eutherian mammals (Davies *et al.* 2017). Mannion and Upchurch (2010) presented their sauropodomorph data in substage time bins, but this was recalculated into stage-level time bins by Dean *et al.* (2016).

SCM2 and CCM2 results were further compared between five major subgroups of Chiroptera (*sensu* Teeling *et al.* 2012): the family Pteropodidae, the superfamilies Rhinolophoidea, Emballonuroidea, Noctilionoidea, Vespertilionoidea and superfamily-level *incertae sedis* taxa, mostly consisting of basal chiropterans.

Previous studies (Brocklehurst *et al.* 2012; Dean *et al.* 2016) of completeness metrics have attempted to parse out the influence of sites of exceptional preservation (Lagerstätten). In this study only Konservat-Lagerstätten were considered, which are here defined as deposits that preserve articulated fossils and associated soft tissues under unique depositional circumstances. Three Lagerstätten relevant to bats were recognised: Green River Formation, USA (Ypresian, Jepsen 1966; Simmons *et al.* 2008); Messel Shale, Germany (Lutetian,

Russell and Sige 1970; Simmons and Geisler 1998); and the Shanwang Formation, China (Burdigalian, Yang and Yang 1994). Completeness scores of species from these three Lagerstätten were excluded from some analyses as noted below.

Unlike some previous analyses of completeness (e.g. Brocklehurst *et al.* 2012), we do not attempt to make comparisons between completeness metrics and estimates of species richness that attempt to account for variable spatiotemporal sampling of the fossil record. This is because sampling of the bat record is insufficient to allow diversity to be estimated using the most rigorous and appropriate diversity estimators (e.g. shareholder quorum subsampling; Alroy 2010; Close *et al.* 2018). We do, however, make comparisons between sampled (i.e. uncorrected) species counts and completeness.

Statistical analyses

All statistical analyses were conducted in R. Time series plots were produced using the package strap (Bell and Lloyd 2015). Non-temporal pairwise comparisons of populations of completeness values for bats and other tetrapod groups were made using non-parametric Mann-Whitney-Wilcoxon tests, which assess differences in the population medians and distribution. Mann-Whitney-Wilcoxon tests were also used to compare the validity of different completeness metrics, completeness between the major subgroups of Chiroptera, across northern and southern hemispheres, different major continents, and between caves and normal fluvio-marine deposits (for all time bins). False discovery rate (FDR; Benjamini and Hochberg 1995) corrections were used where appropriate to correct for multiple comparisons. For comparisons of more than two datasets (e.g. comparing populations of values for different continental regions), Kruskal-Wallis tests were used.

Time series comparisons compared changes in bat completeness metrics through time to one another, between species and generic levels, as well as to temporal variation in observed species richness, fossil record sampling (TBFs and TBCs), and time bin length. We used generalized least-squares regression (GLS) with a first order autoregressive model (corARMA) fitted to the data using the function gls() in the R package nlme v. 3.1–137 (Pinheiro *et al.* 2018). GLS reduces the chance of overestimating statistical significance of regression lines due to serial correlation. Time series were ln-transformed prior to analysis to ensure normality and homoskedasticity of residuals. We calculated likelihood-ratio based pseudo-R^2 values using the function r.squaredLR() of the R package MuMIn (Bartoń 2018). We also tested whether combinations of completeness scores, fossil record sampling and time bin length provided significant explanations of observed species richness, by fitting GLS autoregressive models to combinations of potential explanatory variables. Results were compared using Akaike's information criterion for small sample sizes (AICc) and Akaike weights were calculated to identify the best combination of explanatory variables from those tested. AICc was calculated using the function AICc() of the R package qpcR (Spiess 2018), and Akaike weights calculated using the aic.w() function of the R package phytools (Revell 2017).

Ordinary least squares (OLS) linear regressions were used to test for trends in overall patterns of completeness, observed taxonomic diversity and sampling metrics when compared to geological time, as well as to test the relationship between species completeness and the year in which that species was named.

RESULTS

Changes in mean and median completeness through time.

All SCM2 metrics show highly consistent patterns through time (Fig. 2A), with the modified SCM2 metrics (CSCM2 2D and 3D) being consistently slightly offset towards higher values than the standard SCM2 metrics. Mean completeness is initially comparatively high (c. 7–8%) in the Ypresian and rises to a peak in the Lutetian (c. 38–39%). Subsequently, completeness is very low (c. 2–7%) and almost flat from the Bartonian to the Pliocene, with just a single peak of moderately higher values (c. 7–9%) in the Aquitanian. The Pleistocene has SCM2 values (c. 11–16%) that are substantially higher than other post-Lutetian intervals.

When Lagerstätten taxa are removed, there is little variation in SCM2 completeness values prior to the Pleistocene (Fig. 2B), with the previously comparatively high Ypresian and Lutetian values being greatly diminished. There is still a moderate peak in completeness in the Aquitanian, and a more substantial peak in the Pleistocene. Median completeness values (including Lagerstätten taxa; Fig. 2C) are similar to the mean values excluding Lagerstätten taxa: they are consistently very low (c. 1–5%), with the exception of higher values in the Aquitanian (c. 6–10%) and the Pleistocene (c. 7–11%).

CCM2 patterns (Fig. 2A) are broadly similar to, although always higher than, those for SCM2, in that they show high values in the Lutetian (c. 49%), after which there is relatively little variation in completeness (c. 22–29%). The Pleistocene does not form a distinct peak, in contrast to the SCM2 data. The Lutetian peak disappears when Lagerstätten taxa are excluded (Fig. 2B).

BSCM2 is high in the Ypresian (c. 13%) and Lutetian (c. 40%) when Lagerstätten taxa are included (Fig. 2A), and then virtually flat (c. 8–10%) subsequently, with no peaks in the Aquitanian or Pleistocene. When Lagerstätten taxa are excluded then BSCM2 is virtually flat throughout the Cenozoic (Fig. 2B), and median BSCM2 values form a perfectly straight line (Fig. 2C).

No significant correlation is recovered when all data points for SCM2 (p = 0.306; R^2 = 0.0001) are compared to geological time using OLS, whereas CCM2 has a very weak trend towards decreasing values towards the present (p = 0.02; R^2 = 0.01) (Brown *et al.* 2019: Fig. S1). When Lagerstätten taxa are excluded then there is a significant trend towards increasing SCM2 values towards the Recent (p = 7.17e-12; R^2 = 0.10), but not for CCM2 values (p = 0.135; R^2 = 0.003). The significant trend for SCM2 data is lost when Pleistocene data are excluded (p = 0.433; R^2 = 0.03).

Generic level completeness metrics show significantly different temporal patterns to species level metrics (Brown *et al.* 2019: Fig. S2). In most time bins completeness is higher for genera, but generic completeness drops below species completeness in the Lutetian skeletal metrics. All generic time series have higher completeness in the Ypresian (20% SCM2 2D, 34.2% CCM2) and Bartonian (19.8% SCM2 2D, 43.8% CCM2), and peaks in the Chattian (20.1% SCM2 2D, 46.4% CCM2) and Aquitanian (20.7% SCM2 2D, 44.3% CCM2) which are not revealed in the species level data.

Differences between completeness metrics

Non-temporal comparisons using Mann-Whitney-Wilcox tests show that there are no significant differences between SCM2 metrics generated using either 2D or 3D approaches (Table 1). The modified SCM2 approach (CSCM2), which does not consider left and right sides of the body separately, has significantly higher median completeness values than the standard SCM2 approach (adjusted p = 0.0002–0.0003; Table 1). CCM2 median completeness is significantly higher than that of all SCM2 (adjusted p = 1.05E-72 to 3.91E-61) and BSCM2 (adjusted p = 1.42E-47) approaches (Table 1). Indeed, median completeness for CCM2 (21.14%) is nearly an order of magnitude greater than median completeness for

SCM2 (2.8%, using the 2D approach), and more than four times greater than median completeness for CSCM2 (4.48%, using the 2D approach). BSCM2 shows a tiny interquartile range, indicating very limited variation in BSCM2 values. Violin plots emphasise the bottom-heavy nature of the distributions of completeness values (Brown *et al.* 2019: Fig. S3).

Time series comparisons of mean values using GLS find significant relationships between all completeness metrics (Table 2; Fig. 2A). The strongest relationships are between different SCM2 metrics (pseudo-R^2 = 0. 979–0.996). CCM2 is strongly correlated with SCM2 metrics (pseudo-R^2 = 0.564–0.61), and both SCM2 and CCM2 are significantly correlated with BSCM2 (pseudo-R^2 = 0.673–0.735). However, scatterplots of data for SCM2 versus CCM2, SCM2 versus BSCM2 and CCM2 versus BSCM2 suggest that these relationships are potentially influenced by a single outlier, the Lutetian, for which completeness values are anomalously high due to the presence of the Messel Lagerstätte. When Lagerstätten taxa are excluded from the time series (Fig. 2B), the relationships between SCM2 metrics remain strongly significant (Table 3). However, the relationships between SCM2 metrics and CCM2 are non-significant, as are those between CCM2 and BSCM2 and SCM2 and BSCM2 (Table 3).

All generic and species level time series lack significant correlation apart from BSCM2 when Lagerstätten taxa are included (Brown *et al.* 2019: Table S2). Excluding Lagerstätten taxa removes any correlation between the two taxonomic resolutions (Brown *et al.* 2019: Table S3).

Given the strength of correlation between temporal changes in the different SCM2 metrics, subsequent analyses and comparisons are focused on the following datasets only: SCM2 (2D approach), CSCM2 (2D approach) and CCM2. We do not focus further on

BSCM2 given the highly limited variation present within taxon scores for this metric (see Discussion).

Comparisons with taxonomic diversity and sampling

Raw taxonomic diversity (Fig. 2D) fluctuates from the Ypresian to the Miocene, with peaks in the Priabonian (40 species), Rupelian (35 species), Burdigalian (52 species) and Serravallian (38 species), and troughs in the Bartonian (10 species), Aquitanian (8 species), Langhian (17 species) and Tortonian–Messinian (17 and 13 species, respectively). Diversity then begins to increase in the Pliocene (68 species), and Pleistocene diversity (165 species) far exceeds that of all earlier intervals. There is no significant overall trend in diversity through time ($p = 0.127$; $R^2 = 0.1153$). Excluding the small number of Lagerstätten has no substantial impact on perceived taxonomic diversity patterns beyond substantially reducing taxon counts for the Lutetian.

TBFs show a general trend of increases through time ($p = 0.0003$; $R^2 = 0.65$), with lower numbers of formations from the Ypresian to the Rupelian, followed by considerably higher values from the Chattian to the Pleistocene (Brown *et al.* 2019: Fig. S4). A broadly similar but marginally non-significant pattern is seen in TBCs ($p = 0.06$; $R^2 = 0.196$), although TBCs are marked by an enormous peak in numbers during the Pleistocene, with values 2–3 times higher than in preceding stages.

The best models for observed taxonomic diversity were those including either only CCM2 or only SCM2 as explanatory variables, followed by the models involving only time bin length as an explanatory variable (Brown *et al.* 2019: Table S4). However, all of these models were non-significant with negligible R^2 values.

Comparisons between taxonomic groups

For comparisons of SCM2, bats show significantly lower median completeness values than all taxonomic groups to which they were compared (Table 4; Fig. 3). Indeed, the upper quartile of the bat distribution falls below the lower quartile of the distributions of all of the other groups, with the exception of pelycosaurs. The distribution of SCM2 values for bats is very bottom-heavy (Fig. 3), and this is more extreme than in other groups that also show bottom-heavy SCM2 distributions, such as sauropodomorphs and pelycosaurs.

As noted above, CCM2 values for bats are higher than SCM2 values, but they are still significantly lower than most other groups to which they were compared (Table 5; Fig. 4), with the exception of Mesozoic birds and pelycosaurs. Median CCM2 values for bats (21.14%) are moderately lower than for pelycosaurs (25.6%) but higher than Mesozoic birds (15.57%). Bird data have a substantially greater interquartile range than that of bats (61.6% versus 21.9%) and a more bimodal distribution.

There is no statistical difference in the distribution of SCM2 and CCM2 values between any bat subgroup (Brown *et al.* 2019: Table S5 and S6). Kruskal-Wallis tests further reveal no significant difference between the distribution of values in either SCM2 ($p=0.2312$) or CCM2 ($p=0.5497$). Pteropodidae and Noctilionoidea are the only groups that are not represented by complete and near complete taxa. All groups share a very similar bottom-heavy distribution of completeness values except Pteropodidae, which has a strongly bimodal distribution with a much larger interquartile range (Brown *et al.* 2019: Fig. S5 and S6).

Spatial and environmental comparisons

Mann-Whitney-Wilcox tests indicate that bat specimens recovered solely from Lagerstätten are considerably more complete than those derived from either non-Lagerstätten fluvio-

marine settings (SCM2, p = 3.736e-08; CCM2, p = 4.189e-08) or caves (SCM2, p = 2.612e-08; CCM2, p = 5.72e-08). However, specimens derived from cave deposits are significantly more complete than specimens discovered from fluviolacustrine and marine regimes, even when included Lagerstätten (SCM2, p = 2.2E-16; CCM2, p = 9.5E-11; Fig. 5). For northern versus southern hemisphere comparisons, results are non-significant for SCM2 (p = 0.343), but significant for CCM2 (p = 0.001), with completeness higher in the northern hemisphere (Fig. 6). Kruskal-Wallis tests indicate significant differences occur between median completeness values for different continental regions for both SCM2 (p = 1.29e-13) and CCM2 (1.55e-09). Most notably, median SCM2 values for North America are significantly higher than those for all other continental regions with the exception of Australasia, whereas for CCM2, median values for Europe are significantly higher than those for all other continental regions with the exception of Australasia (Brown *et al.* 2019: Table S7 and S8; Brown *et al.* 2019: Fig. S7 and S8).

Completeness and date of discovery

OLS recovered a significant relationship between date of discovery and SCM2, with more recently discovered species being known from less complete material, but the explanatory power is very low (p = 2.40E-06; adjusted R^2 = 0.047). This relationship is significant even when Lagerstätten taxa are excluded (p = 1.11e-09; adjusted R^2 = 0.08). The same result is recovered for comparisons of date of discovery and CCM2 (p = 4.89e-08; adjusted R^2 = 0.06), also when Lagerstätten taxa are excluded (p = 1.11e-09; adjusted R^2 = 0.08) (Fig. 7A-B). A collector's curve shows few discoveries from the early part of the 19th century up until the 1950s, followed by a steep rise in the rate of discoveries in the second half of the 20th

century. There is no sign of any decline in the rates of discovery towards the present day (Fig. 7C).

DISCUSSION

Comparison of completeness metrics

There is no significant difference between the time series or non-temporal distributions of SCM2 values for bats derived from either 2D or 3D methods (Table 1-2). This suggests that completeness estimates generated by simpler (e.g. 2D skeletal drawings) methods of calculating body proportions are statistically indistinguishable from those generated by more detailed methods. Efforts to develop highly accurate means of calculating body proportions (e.g. Brocklehurst and Fröbisch 2014; Verrière *et al.* 2016; our use of 3D CT scans) may be unnecessarily precise and time consuming given these results.

Using CSCM2 significantly increases the median and distribution of completeness scores in comparison to the original SCM2 (Table 1, Fig S3). Calculating SCM2 by separating the left and right sides of the skeleton therefore significantly underestimates the completeness of the bat fossil record. SCM2 2D seems to be most strongly altered when using the composite metric. However, time series comparisons show that mean and median CSCM2 scores are only minimally higher than SCM2 per time bin, and the original and composite metrics show the same temporal trends.

BSCM2 scores are consistently higher than other skeletal completeness metrics. Its temporal patterns lack the resolution of the other metrics (Fig. 2A-C), and their distribution is unlike the other skeletal metrics (Brown *et al.* 2019: Fig. S3), with an unrepresentative absence of taxon completeness scores below ~4%. Unlike with ichthyosaurs (Cleary *et al.* 2015), it is likely this metric is too coarse to accurately estimate the completeness of the bat

fossil record, and is possibly inadequate for use with groups known from very limited fossil material.

CCM2 has higher mean and median time series, and significantly higher non-temporal median and distribution of completeness values in comparison to the skeletal completeness metrics (Table 1-3; Fig. 2A-C). The large differences between the character- and skeletal-based metrics is likely because teeth are the most common element recovered in the bat fossil record and because of the high proportion of dental characters (42% of all phylogenetic characters) compared to the skeletal mass of the dentition (1.92% of total skeleton). This highlights the different aspects of fossil record quality that are assessed by character- and skeletal-based metrics. Even though skeletal material is sparse in the bat fossil record, it is still relatively rich in character information.

All mean time series are significantly correlated with each other, suggesting that the same temporal signal is recovered regardless of the metric used. However, removing Lagerstätten taxa from the time series renders non-significant the correlations between BSCM2 and SCM2 and CCM2 and SCM2 (Table 3), suggesting that Lagerstätten can have a strong influence on metric results and our interpretations of the bat fossil record.

The significant changes and lack of correlation between the species and generic time series suggests that the quality of the fossil record can strongly differ depending on the taxonomic resolution used. The higher generic compared to species completeness for most time bins is expected since most genera are known from multiple species and therefore incorporate many more specimens. The drop in mean generic completeness in the Lutetian below species completeness levels is surprising considering that the stage contains many Lagerstätten taxa. This, however, is likely due to the addition of more fragmentary taxonomic records to the analyses that are indeterminate at species level.

Teeling *et al.* (2005) used ghost lineages from a molecular phylogeny representing 30 genera to estimate that 61% of the bat fossil record is missing and that first appearances are underestimated by 73%. Eiting and Gunnell (2009) later concluded the generic bat fossil record was 12% complete using phylogenetically independent extinction rate and preservation potential to assess fossil occurrence ranges. These studies assessed very different concepts of 'completeness' (Benton and Storrs 1994; Foote and Sepkoski 1999; Benton *et al.* 2000; Benton *et al.* 2011) to the metrics used here. However, it is interesting that the mean generic character (27.8% CCM2, ~70% incomplete) and skeletal completeness (12.6% SCM2 2D, 14% SCM2 3D, 14.6% CSCM2 2D, 16.8% CSCM2 3D and 13.8% BSCM2) converge on relatively similar values to the previous studies, respectively. This suggests the overall signals of a comparatively poor bat fossil record are consistent regardless of the approach used.

Taxonomic comparisons

Bats seem to have the least complete skeletal fossil record of any tetrapod group previously studied using the skeletal completeness metric, but their character completeness is more similar to some other groups (Fig. 3-4). The extremely poor skeletal record is most likely due to the fragile nature of bat skeletons, and the very common preservation and reporting of isolated teeth relative to other skeletal elements. Bat teeth, like those of most fossil mammals, are highly character-rich and therefore allow for the recognition of new species from very limited fossil material. Even with this in mind, bat CCM2 is comparatively high, with a non-temporal median score higher than that of Mesozoic birds. However, the similarity between CCM2 scores for bats and birds may not be that surprising considering their similar ecologies (e.g. flight capability) and the likely similar taphonomic biases acting upon their records

(Brocklehurst *et al.* 2012; Gardner *et al.* 2016) (see below). There are large differences apparent visually between the two CCM2 distributions, but the median values are similar. The much larger interquartile range and more bimodal distribution of Mesozoic birds reflects their more common preservation in Lagerstätten deposits (34.7% and 2.5% in Mesozoic birds and fossil bats, respectively [Brocklehurst *et al.* 2012; this study]).

Although bats and early eutherians both represent subgroups within mammals, early eutherian CCM2 values (from Davies *et al.* 2017) are significantly higher than those of bats. This may in part reflect differential character scoring between the groups. Early eutherians share similar dental character proportions (37%; Davies *et al.* 2017) to bats (42%), but the proportion of cranial characters (34%) used in phylogenetic studies of early eutherians is much greater than in bats (5.6%), whereas the proportion of postcranial characters is substantially reduced (26% in early eutherians, 52% in bats). This means the bat character completeness is more reliant on infrequently preserved (or identified/reported) postcranial material than the early eutherian record. In future, it would be interesting to see how skeletal completeness metrics for other eutherian groups compare to the extremely poor skeletal record of bats.

In contrast, the statistical and visual similarity between the distributions of CCM2 scores in the bat and pelycosaur records makes little morphological, ecological and taphonomic sense. Brocklehurst and Fröbisch (2014) concluded that before the widespread use of cladistics, many pelycosaur species were named on the basis of limited material, leading to taxonomic "over-splitting" and therefore artificially reducing the quality of the pelycosaur fossil record. This could explain its similarity with the bat fossil record which by its nature is poor but taxonomically informative.

No significant differences in the completeness of the six major chiropteran subgroups were found in our study, as all share similar median SCM2 and CCM2 values. Pteropodidae, however, have an unusual bimodal distribution that differs visually from that of the other subgroups. This record is likely not statistically different from the other subgroups because of the extended interquartile range and the small sample size (n=7). This, like previous assessments (Gunnell and Simmons 2005; Teeling *et al.* 2005; Eiting and Gunnell 2009), emphasises how comparatively poor the pteropodid fossil record is. The low completeness of pteropodid specimens is perhaps unexpected given that they are the first major group to diverge (Simmons 2005) and also have the largest and thus most robust skeletons (Norberg and Norberg 2012; Brown *et al.* 2013), possibly giving them a greater opportunity to be represented in the fossil record. The lack of a well-sampled fossil record for this important group is troubling for understanding the evolutionary development of all bats, but may potentially stem from aspects of pteropodid distribution and ecology (e.g. their tendency to occupy tropical wooded environments and avoidance of caves by many species), as well as potential origin of this clade in tropical Australasia (Almeida *et al.* 2016) (see below).

Sampling biases

Palaeoecology and Palaeoenvironment. The generally poor fossil record of bats is not surprising considering their small size, delicate skeletons, and their preference for environments rarely preserved in the fossil record (e.g. forests, caves). 58% of all species included in our dataset are solely derived from cave deposits, whereas only 34% are solely derived from non-Lagerstätten fluvio-marine settings, and 2.5% from exceptional Lagerstätten deposits.

Lagerstätten understandably recover the most complete bat specimens and therefore represent our best chances to enhance our knowledge of early bat evolution, handicapped only by their temporal and spatial restrictions. The much more widespread cave, fissure fill and sinkhole deposits on average preserve significantly more complete (for both SCM2 and CCM2) bat specimens than fluvio-marine deposits, but as yet have not produced even a single published complete bat skeleton. Though they represent more frequent sources of bat remains, the absence of complete material could well be a limiting factor in their usefulness for understanding bat evolution.

The Pleistocene sees a large increase in species richness and a peak in non-Lagerstätten influenced skeletal completeness. The Pleistocene is the only stage that has individual taxa known from more than 50% of the skeleton, with the exclusion of those stages with Lagerstätten deposits (Brown *et al.* 2019: Fig. S1). 88% of all Pleistocene species are known from caves, and 81% are known solely from them. Caves are known to preserve highly useful fossil records for mammals, having often very large sample sizes (perhaps tens of thousands of fossils) and preserving behavioural processes and palaeoecological signals (Jass and George 2010). Jass and George (2010) found that in the USA, cave deposits account for a high proportion (62%) of individual late Pleistocene mammal species records in comparison to non-cave deposits. They further showed that late Pleistocene bats have a comparatively high occurrence rate in caves. As some bat cave remains are autochthonous, sampling the actual caves that were previously occupied by fossil bats greatly increases the likelihood of preserving more complete skeletal material. Furthermore, Pleistocene deposits have likely undergone the least taphonomic and diagenetic processes of any time bin and a number may be unlithified, which have been demonstrated to better preserve specimens (Kowalewski *et al.* 2006; Hendy 2009). The heightened preservation of cave deposits in the Pleistocene could explain the increase in completeness of this time period.

Geographic and human biases. Bats have an almost global distribution. The generally more complete remains known from the northern hemisphere therefore most likely reflect heightened historical collection effort and fortunate preservation regimes. Europe has the highest median CCM2 of any continental region, likely resulting from long historical interest, the presence of large spatial and temporal fissure fill deposits such as the Quercy Phosphorites (France) and the exceptional preservation of the Messel Oil Shales (Germany). North America has the highest SCM2 values, likely because 83% of its taxa are known from cave deposits, 72% solely so. Asia, by contrast, has a comparatively limited record, with only 42 occurrences and only 33% of taxa known from caves. The southern hemisphere has a significantly lower CCM2 in comparison to the northern hemisphere. This likely reflects a relatively limited record and the lack of any complete taxa due to the absence of Lagerstätten. South America has a better SCM2 record than Europe, possibly because 55% of taxa are derived from cave deposits (all Pleistocene), but constitutes one of the poorer CCM2 records. Australasia seems to have relatively high completeness scores because of the small sample size (24 taxa) and the fact that 50% of known taxa are derived from cave deposits. Africa has moderate CCM2 scores but one of the poorest SCM2 records, possibly because only 22% are known from caves, but 93% of taxa have at least one occurrence in fluvio-marine formations. Africa, Asia, Australasia and South America potentially hold a wealth of unexplored and undiscovered information on bat fossils, and more effort may be required to sample the bat record in new geographical localities.

The negative correlation between completeness and discovery date suggests researchers have named more bat species based on less complete material as time has passed. However, the correlations are very weak and explain only small amounts of variation. The patterns observed likely reflect the extreme paucity of the bat fossil record during the 19th

century in comparison to the major increase in recognition of new bat species in recent decades (Fig. 7) since an explosion of finds in the 1950s and the improvement of techniques for fossil discovery (Morgan and Czaplewski 2012).

Macroevolutionary understanding

The vast majority of the bat fossil record is known from isolated teeth and jaw fragments, which fortunately are diagnostic enough to identify many different species. As such, there is no correlation between taxon completeness and changes in bat diversity through time. However, such a record only provides limited insight. Near complete specimens from Lagerstätten are often the only way to answer important biological and evolutionary questions about the early history of the group. Such questions include how body sizes in the group evolved (Giannini *et al.* 2012), or whether laryngeal echolocation evolved multiple times convergently, or evolved just once and was lost in pteropodids (Teeling *et al.* 2012; Teeling *et al.* 2016). Even then there are fundamental disagreements as to what morphological markers are actually informative for identifying such features and behavioural traits in fossils (Simmons *et al.* 2008; 2010; Veselka *et al.* 2010; Teeling *et al.* 2012). Given the paucity of fossil data, much research attention has focused on genomic, molecular and morphological studies of extant taxa, which have made great strides in deciphering bat relationships and answering some of these important questions (Springer *et al.* 2001; Teeling *et al.* 2002; Pederson and Timm 2012; Tsagkogeorga *et al.* 2013; Wang *et al.* 2017). A more even representation of different skeletal elements, such as seen in other tetrapod groups, would provide a more enriched understanding of the evolution of different aspects of bat biology and functional morphology. However, the Pleistocene sees an increase in skeletal completeness and diversity, but lacks a peak in CCM2. As explained earlier, this may be a

consequence of non-dental elements having comparatively little morphological character weight and therefore little impact on average phylogenetic understanding. Complete skeletons from Lagerstätten deposits are necessary to significantly enhance the morphology-based phylogenetic datasets.

With relatively little fluctuation except for the Lagerstätten peak in the Lutetian, bat completeness is fairly constant through time and does not correlate with geological age, changes in tetrapod bearing formations, collections or taxon richness. This suggests the general poor quality of the bat fossil record has been unaffected by large scale evolutionary and environmental changes through the Cenozoic, or by the geological outcrop availability and the number of sampled localities. Moreover, there are many chronological and geographical gaps in the record (Teeling *et al.* 2005; Eiting and Gunnell 2009; Morgan and Czaplewski 2012). As previously discussed (Gunnell and Simmons 2005), the lack of unambiguous Paleocene bat fossils is a major barrier to understanding the origins of the group. Molecular and combined-data phylogenies (Teeling *et al.* 2005; Miller-Butterworth *et al.* 2007; Meredith *et al.* 2013; O'Leary *et al.* 2013; Lei and Dong 2016) have estimated that bats first evolved just after the Cretaceous-Palaeogene boundary, meaning that early bats existed during almost the entirety of the Paleocene without leaving a discernible fossil record. This leaves a ~10 Myr gap (Miller-Butterworth *et al.* 2007) between their origin and the first unambiguous bat fossils. The lack of any taxonomically useful Paleocene fossils is puzzling. Bats have recently been hypothesized to have evolved within forested environments (Giannini 2012). The relatively poor preservation of these environments in the fossil record (Newell 1959; Raup 1988) and potential geographic isolation of early bats (Smith *et al.* 2012) may explain the lack of a Paleocene record. The metrics we employ here do not take into consideration this large chronological gap in the bat fossil record and can possibly present a false sense of completeness, as this gap is removed from analyses. This must be considered

when interpreting our findings, and future studies must address the absent Paleocene bat record. The quality of early Eocene bat record is a dramatic improvement, and has been hypothesized to relate to an insect radiation at the PETM fuelling diversification and radiation of five bat superfamilies within five million years (Teeling *et al.* 2005). Bats exploiting a more varied tapestry of environments and niches in the early Eocene potentially enhanced the likelihood of individuals being preserved in the fossil record. Though convincing arguments have been presented (Giannini 2012), without Paleocene fossils researchers can only theorise as to how and why ancestral bats transitioned to flying locomotion and the unique bat morphology, as the first bats known from complete specimens have all the unique anatomical features of extant forms (Gunnell and Simmons 2005).

To try and address some of the temporal and spatial gaps in the bat fossil record highlighted by this study, our results might inform future exploration, and be utilised in combination with ancestral area reconstructions (Ruedi *et al.* 2012; Ruedi *et al.* 2013; Almeida *et al.* 2016) and innovative predictive modelling techniques (Anemone *et al.* 2011; Conroy *et al.* 2012; Wills *et al.* 2018), to locate new and productive fossil localities.

CONCLUSIONS

Quantitative analysis of the bat fossil record reveals it to be comparatively distinct and derived almost entirely from either isolated teeth, fragmentary elements or complete skeletons.

- Bats have the poorest skeletal completeness of any previously assessed tetrapod group, but considerably higher character completeness, similar to that previously found for pelycosaurs and birds.

- There are no major differences in the completeness of the six major bat subgroups, but Pteropodidae is notable for its extremely limited record and highly bimodal distribution of scores.
- There is little fluctuation in completeness through time for any metric, except for a peak in the Lutetian, and a skeletal high in the Pleistocene. Temporal completeness patterns reveal significant differences depending on the taxonomic resolution used. Observed bat diversity is not correlated with completeness metrics, likely because highly diagnostic teeth allow for low level taxonomic assignments.
- All complete and near complete bat specimens are derived from Lagerstätten.
- Bat fossils from caves are generally more complete than those from fluvio-marine deposits. However, the former environments have limited use as no complete specimens have ever been described from a cave deposit.
- The heightened historical research interest and presence of Lagerstätten potentially explain the significantly higher character completeness in the northern hemisphere in comparison to the southern hemisphere.
- Simple 2D measures of skeletal body proportions reveal the same completeness signals as more accurate 3D models. BSCM2 is possibly too coarse a metric to accurately represent the quality and changes in the bat fossil record.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Jessie Maisano (University of Texas), DigiMorph.org, and NSF grants DEB-9873663 and IIS-9874781 for CT data used in this study. We thank all contributors to the Paleobiology Database: this is Paleobiology Database official publication XXX. This research was originally completed as EEB's BSci research in Palaeobiology &

Palaeoenvironments at the University of Birmingham. Review comments from Neil Brocklehurst, Sally Thomas and Nicholas Czaplewski improved the final manuscript. This research was partially funded by the European Union's Horizon 2020 research and innovation programme under grant agreement 637483 (ERC Starting Grant TERRA to R.J.B.).

REFERENCES

ALMEIDA, F. C., GIANNINI, N. P. and SIMMONS, N. B. 2016. The evolutionary history of the African fruit bats (Chiroptera: Pteropodidae). *Acta Chiropterologica*, **18**, 73-108.

ALROY, J. 2010. Geographical, environmental and intrinsic biotic controls on Phanerozoic marine diversification. *Palaeontology*, **53**, 1211-1235.

ANEMONE, R. L., EMERSON, C. W. and CONROY, G. C. 2011. Finding fossils in new ways: An artificial neural network approach to predicting the location of productive fossil localities. E*volutionary Anthropology: Issues, News, and Reviews*, **20**, 169-180.

ARITA, H. T. and FENTON, M. B. 1997. Flight and echolocation in the ecology and evolution of bats. *Trends in Ecology & Evolution*, **12**, 53-58.

BARTOŃ, K. 2018. MuMIn: Multi-model inference. R package version 1.42.1.

BEARDMORE, S. R., ORR, P. J., MANZOCCHI, T., FURRER, H. and JOHNSON, C. 2012. Death, decay and disarticulation: modelling the skeletal taphonomy of marine reptiles demonstrated using *Serpianosaurus* (Reptilia; Sauropterygia). *Palaeogeography, Palaeoclimatology, Palaeoecology*, **337**, 1-13.

BELL, M. A. and LLOYD, G. T. 2015. strap: an R package for plotting phylogenies against stratigraphy and assessing their stratigraphic congruence. *Palaeontology*, **58**, 379-389.

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)*, 289-300.

BENTON, M. J. and STORRS, G. W. 1994. Testing the quality of the fossil record: paleontological knowledge is improving. *Geology*, **22**, 111-114.

— WILLS, M. A. and HITCHIN, R. 2000. Quality of the fossil record through time. *Nature*, **403**, 534-537.

— TVERDOKHLEBOV, V. P. and SURKOV, M. V. 2004. Ecosystem remodelling among vertebrates at the Permian–Triassic boundary in Russia. *Nature*, **432**, 97-100.

— 2008. Fossil quality and naming dinosaurs. *Biology Letters*, **4**, 729-732.

— 2010. Naming dinosaur species: the performance of prolific authors. *Journal of Vertebrate Paleontology*, **30**, 1478-1485.

— 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.

BOONMAN, A., BAR-ON, Y. and YOVEL, Y. 2013. It's not black or white—on the range of vision and echolocation in echolocating bats. *Frontiers in physiology*, **4**, 1-12.

BROCKLEHURST, N., 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.

— FRÖBISCH, J. 2014. Current and historical perspectives on the completeness of the fossil record of pelycosaurian-grade synapsids. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **399**, 114–126.

— 2015. A simulation-based examination of residual diversity estimates as a method of correcting for sampling bias. *Palaeontologia Electronica*, **18**, 1-15.

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.

BROWN, E. E., CASHMORE, D. D., SIMMONS, N. B. and BUTLER, R. J. 2019. Data from: Quantifying the completeness of the bat fossil record. Dryad Digital Repository. https://doi.org/10.5061/dryad.hp250fb

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.

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

CONROY, G. C., EMERSON, C. W., ANEMONE, R. L. and TOWNSEND, K. B. 2012. Let your fingers do the walking: A simple spectral signature model for "remote" fossil prospecting. *Journal of Human Evolution*, **63**, 79-84.

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**, 241-244.

DAVIES, T. W., BELL, M. A., GOSWAMI, A. and HALLIDAY, T. J. 2017. Completeness of the eutherian mammal fossil record and implications for reconstructing mammal evolution through the Cretaceous/Paleogene mass extinction. *Paleobiology*, **43**, 521-536.

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

DINGUS, L. 1984. Effects of stratigraphic completeness on interpretations of extinction rates across the Cretaceous-Tertiary boundary. *Paleobiology*, **10**, 420-438.

DUDLEY, R., BYRNES, G., YANOVIAK, S. P., BORRELL, B., BROWN, R. M. and MCGUIRE, J. A. 2007. Gliding and the functional origins of flight: Biomechanical novelty or necessity? *Annual Review of Ecology, Evolution, and Systematics*, **38**, 179-201.

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

DYKE, G. J., MCGOWAN, A. J., NUDDS, R. L. and SMITH, D. 2009. The shape of pterosaur evolution: evidence from the fossil record. *Journal of Evolutionary Biology*, **22**, 890-898.

EITING, T. P. and GUNNELL, G. F. 2009. Global completeness of the bat fossil record. *Journal of Mammalian Evolution*, **16**, 151-173.

FOOTE, M. and RAUP, D. M. 1996. Fossil preservation and the stratigraphic ranges of taxa. *Paleobiology*, **22**, 121-140.

— and SEPKOSKI, J. J. Jr 1999. Absolute measures of the completeness of the fossil record. *Nature*, **398**, 415-417.

FOUNTAINE, T. M., 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 of London B: Biological Sciences*, **272**, 289-294.

FREEMAN, P. W. 2000. Macroevolution in Microchiroptera: recoupling morphology and ecology with phylogeny. *Evolutionary Ecology Research,* **2**: 317–335.

GARDNER, E. E., WALKER, S. E. and GARDNER, L. I. Palaeoclimate, environmental factors, and bird body size: A multivariable analysis of avian fossil preservation. *Earth-Science Reviews*, **162**, 177-197.

GIANNINI, N. P. 2012. Towards an integrative theory on the origin of bat flight. 353-384. *In* GUNNELL, G. F. and SIMMONS, N. B. (eds). *Evolutionary History of Bats: Fossils, Molecules and Morphology*. Cambridge University Press, 560 pp.

— GUNNELL, G. F., HABERSETZER, J. and SIMMONS, N. B. 2012. Early evolution of body size in bats. 530-555. *In* GUNNELL, G. F. and SIMMONS, N. B. (eds). *Evolutionary History of Bats: Fossils, Molecules and Morphology*. Cambridge University Press, 560 pp.

GUNNELL, G. F. and SIMMONS, N. B. 2005. Fossil evidence and the origin of bats. *Journal of Mammalian Evolution*, **12**, 209-246.

HALLIDAY, T. J., UPCHURCH, P. and GOSWAMI, A. 2017. Resolving the relationships of Paleocene placental mammals. *Biological Reviews*, **92**, 521-550.

HENDY, A. J. W. 2009. The influence of lithification of Cenozoic marine biodiversity trends. *Paleobiology*, **35**, 51-62.

JASS, C. N. and GEORGE, C. O. 2010. An assessment of the contribution of fossil cave deposits to the Quaternary paleontological record. *Quaternary International*, **217**, 105-116.

JEPSEN, G. L. 1966. Early Eocene bat from Wyoming. *Science*, **154**, 1333-1339.

KOWALEWSKI, M., KIESSLING, W., ABERHAN, M., FUERSICH, F. T., SCARPONI, D., BARBOUR WOOD, S. L. and HOFFMEISTER, A. P. 2006. Ecological, taxonomic, and taphonomic components of the post-Paleozoic increase in sample-level species diversity of marine benthos. *Paleobiology*, **32**, 553-561.

LEI, M. and DONG, D. 2016. Phylogenomic analyses of bat subordinal relationships based on transcriptome data. *Scientific Reports*, **6**, 27726.

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.

MEREDITH, R. W., JANECKA, J. E., GATESY, J., RYDER, O. A., FISHER, C. A., TEELING, E. C., GOODBLA, A., EIZIRIK, E., SIMÃO, T. L. L., STADLER, T., RABOSKY, D. L., HONEYCUTT, R. L., FLYNN, J. J., INGRAM, C. M., STEINER, C., WILLIAMS, T. L., ROBINSON, T. J., BURK-HERRICK, A., WESTERMAN, M., AYOUB, N. A., SPRINGER, M. S. and MURPHY, W. J. 2011. Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science*, 1211028.

MILLER-BUTTERWORTH, C. M., MURPHY, W. J., O'BRIEN, S. J., JACOBS, D. S., SPRINGER, M. S. and TEELING, E. C. 2007. A family matter: conclusive resolution of the taxonomic position of the long-fingered bats, *Miniopterus*. *Molecular Biology and Evolution*, **24**, 1553-1561.

MORGAN, G. S. and CZAPLEWSKI, N. 2012. Evolutionary history of the Neotropical Chiroptera: the fossil record. 105-161. *In* GUNNELL, G. F. and SIMMONS, N. B. (eds). *Evolutionary History of Bats: Fossils, Molecules and Morphology*. Cambridge University Press, 560 pp.

NEWELL, N. D. 1959. The nature of the fossil record. *Proceedings of the American Philosophical Society*, **103**, 264-285.

NORBERG, U. M. 1990. Vertebrate flight: mechanics, physiology, morphology, ecology and evolution. *Zoophysiology*, **27,** Springer-Verlag Berlin Heidelberg, pp 291.

— and NORBERG, R. Å. 2012. Scaling of wingbeat frequency with body mass in bats and limits to maximum bat size. *Journal of Experimental Biology*, **215**, 711-722.

O'LEARY, M. A., BLOCH, J. I., FLYNN, J. J., GAUDIN, T. J., GIALLOMBARDO, A., GIANNINI, N. P., GOLDBERG, S. L., KRAATZ, B. P., LUO, Z., MENG, J., NI, X., NOVACEK, M. J., PERINI, F. A., RANDALL, Z. S., ROUGIER, G. W., SARGIS, E. J., SILCOX, M. T., SIMMONS, N. B., SPAULDING, M., VELAZCO, P. M., WEKSLER, M., WIBLE, J. R. and CIRRANELLO, A. L. 2013. The placental mammal ancestor and the post–K-Pg radiation of placentals. *Science*, **339**, 662-667.

PADIAN, K. 1985. The origins and aerodynamics of flight in extinct vertebrates. *Palaeontology*, **28**, 413-433.

PEDERSON, S. C. and TIMM, D. W. 2012. Echolocation, evo-devo and the evolution of bat crania. 470-499. *In* GUNNELL, G. F. and SIMMONS, N. B. (eds). *Evolutionary History of Bats: Fossils, Molecules and Morphology*. Cambridge University Press, 560 pp.

PINHEIRO, J., BATES, D., DEBROY, S. and SARKAR, D. 2018. nlme: linear and nonlinear mixed effects models, R package version 3.1-137.

RAUP, D. M. 1988. Diversity crisis in the geological past. 51-57. *In* WILSON, E. O. (eds). *Biodiversity*. National Academy Press, 538 pp.

RETALLACK, G. 1984. Completeness of the rock and fossil record: some estimates using fossil soils. *Paleobiology*, **10**, 59-78.

REVELL, L. J. 2017. phytools: Phylogenetic Tools for Comparative Biology (and Other Things). R package version 0.6-60.

RUEDI, M., FRIEDLI-WEYENTH, N., TEELING, E. C., PUECHMAILLE, S. J. and GOODMAN, S. M. 2012. Biogeography of Old World emballonurine bats (Chiroptera: Emballonuridae) inferred with mitochondrial and nuclear DNA. *Molecular Phylogenetics and Evolution*, **64**, 204-211.

RUEDI, M., STADELMANN, B., GAGER, Y., DOUZERY, E. J., FRANCIS, C. M., LIN, L. K., GUILLÉN-SERVENT, A. and CIBOIS, A. 2013. Molecular phylogenetic reconstructions identify East Asia as the cradle for the evolution of the cosmopolitan genus *Myotis* (Mammalia, Chiroptera). *Molecular Phylogenetics and Evolution*, **69**, 437-449.

RUSSELL, D. E. and B. SIGE. 1970. Révision des chiroptères lutétien de Messel (Hesse, Allemagne). *Palaeovertebrata*, Montpellier, **3**, 83-182.

SHI, J. J. and RABOSKY, D. L. 2015. Speciation dynamics during the global radiation of extant bats. *Evolution*, **69**, 1528–1545.

— WESTEEN, E. P. and RABOSKY, D. L. 2018. Digitizing extant bat diversity: An open-access repository of 3D μCT-scanned skulls for research and education. *PLOS One*, **13**, e0203022.

SIMMONS, N. B. and GEISLER, J. H. 1998. Phylogenetic relationships of *Icaronycteris*, *Archaeonycteris*, *Hassianycteris*, and *Palaeochiropteryx* to extant bat lineages, with comments on the evolution of echolocation and foraging strategies in Microchiroptera. *Bulletin of the American Museum of Natural History*, **235**, 1-182.

— 2005. An Eocene big bang for bats. *Science*, **307**, 527-528.

— SEYMOUR, K. L., HABERSETZER, J. and GUNNELL, G. F. 2008. Primitive Early Eocene bat from Wyoming and the evolution of flight and echolocation. *Nature*, **451**, 818-822.

— — — — 2010. Inferring echolocation in ancient bats. *Nature*, **466**, E8.

— and CIRRANELLO, A. L. 2018. Bat species of the world: A taxonomic and geographic database. http://batnames.org. Accessed on 29/10/2018.

SMITH, A. B. 2001. Large–scale heterogeneity of the fossil record: implications for Phanerozoic biodiversity studies. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **356**, 351-367.

— 2007. Intrinsic versus 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.

SMITH, T., HABERSETZER, J., SIMMONS, N. B. and GUNNELL, G. 2012. Systematics and paleobiogeography of early bats. 23-66. *In* GUNNELL, G. F. and SIMMONS, N. B. (eds). *Evolutionary History of Bats: Fossils, Molecules and Morphology*. Cambridge University Press, 560 pp.

SONG, S., LIU, L., EDWARDS, S. V. and WU, S. 2012. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proceedings of the National Academy of Sciences of the USA*, **109**, 14942-14947.

SPEISS, A-N. 2018. QpcR: Modelling and Analysis of Real-Time PCR Data. R package version 1.4-1.

SPRINGER, M. S., TEELING, E. C., MADSEN, O., STANHOPE, M. J. and DE JONG, W. W. 2001. Integrated fossil and molecular data reconstruct bat echolocation. *Proceedings of the National Academy of Sciences of the USA*, **98**, 6241-6246.

SWARTZ, S. M., BENNETT, M. B. and CARRIER, D. R. 1992. Wing bone stresses in free flying bats and the evolution of skeletal design for flight. *Nature*, **359**, 726-729.

— and MIDDLETON, K. M. 2008. Biomechanics of the bat limb skeleton: scaling, material properties and mechanics. *Cells Tissues Organs*, **187**, 59-84.

— IRIARTE-DÍAZ, J., RISKIN, D. K. and BREUER, K. S. 2012. A bird? A plane? No, it's a bat: an introduction to the biomechanics of bat flight. 317-352. *In* GUNNELL, G. F. and SIMMONS, N. B. (eds). *Evolutionary History of Bats: Fossils, Molecules and Morphology*. Cambridge University Press, 560 pp.

TEELING, E. C., SPRINGER, M. S., MADSEN, O., BATES, P., O'BRIEN, S. J. and MURPHY, W. J. 2005. A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science*, **307**, 580-584.

— MADSEN, O., VAN DEN BUSSCHE, R. A., DE JONG, W. W., STANHOPE, M. J. and SPRINGER, M. S. 2002. Microbat paraphyly and the convergent evolution of a key innovation in Old World rhinolophoid microbats. *Proceedings of the National Academy of Sciences of the USA*, **99**, 1431-1436.

— DOOL, S. and SPRINGER, M. S. 2012. Phylogenies, fossils and functional genes: the evolution of echolocation in bats. 1-22. *In* GUNNELL, G. F. and SIMMONS, N. B. (eds). *Evolutionary History of Bats: Fossils, Molecules and Morphology*. Cambridge University Press, 560 pp.

— JONES, G. and ROSSITER, S. J. 2016. Phylogeny, genes, and hearing: implications for the evolution of echolocation in bats. 25-54. *In* FENTON, M. B., GRINNELL, A. D., POPPER, A. N. and FAY, R. R. (eds). *Bat Bioacoustics.* Springer, 304 pp.

TEMPLIN, R. J. 2000. The spectrum of animal flight: insects to pterosaurs. *Progress in Aerospace Sciences*, **36**, 393-436.

TSAGKOGEORGA, G., PARKER, J., STUPKA, E., COTTON, J. A. and ROSSITER, S. J. 2013. Phylogenomic analyses elucidate the evolutionary relationships of bats. *Current Biology*, **23**, 2262-2267.

TSANG, S. M., CIRRANELLO, A. L., BATES, P. J. J. and SIMMONS N. B. 2015. The roles of taxonomy and systematics in bat conservation. 503-538. *In* KINGSON, T. and VOIGHT, C. (ed.) *Bats in the Anthropocene: Conservation of Bats in a Changing World*. Springer International, 606 pp.

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.

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.

VESELKA, N., MCERLAIN, D. D., HOLDSWORTH, D. W., EGER, J. L., CHHEM, R. K., MASON, M. J., BRAIN, K. L., FAURE, P. A. and FENTON, M. B. 2010. A bony connection signals laryngeal echolocation in bats. *Nature*, **463**, 939.

WALTHER, M. and FRÖBISCH, J. 2013. The quality of the fossil record of anomodonts (Synapsida, Therapsida). *Comptes Rendus Palevol*, **12**, 495-504.

WANG, Z., ZHU, T., XUE, H., FANG, N., ZHANG, J., ZHANG, L., PANG, J., TEELING, E. and ZHANG, S. 2017. Prenatal development supports a single origin of laryngeal echolocation in bats. *Nature Ecology & Evolution*, **1**, 0021.

WILLS, S., CHOINIERE, J. N. and BARRETT, P. M. 2018. Predictive modelling of fossil-bearing locality distributions in the Elliot Formation (Upper Triassic–Lower Jurassic), South Africa, using a combined multivariate and spatial statistical analyses of present-day environmental data. *Palaeogeography, Palaeoclimatology, Palaeoecology,* **489**, 186-197.

WILSON, D. E. and REEDER, D. M. 2005. *Mammal Species of the World. A Taxonomic and Geographic Reference (3rd ed)*. Johns Hopkins University Press, 2142 pp.

WIMSATT, W. 2012. *Biology of bats*. Elsevier, 420 pp.

YANG, H. and YANG, S. 1994. The Shanwang fossil biota in eastern China: a Miocene Konservat-Lagerstätte in lacustrine deposits. *Lethaia*, **27**, 345-354.

FIGURE CAPTIONS

FIG 1. Scientifically-informed skeletal reconstructions of (a) *Icaronycteris index* and (b) *Myotis myotis* modified from Wimsatt (2012), used to calculate the 2D weighted skeletal proportions for ‘extinct’ and ‘extant’ taxa respectively; and (c) a CT-scanned *Myotis daubentonii* specimen, used to calculate the 3D weighted skeletal proportions. For detailed breakdown of weightings see Brown *et al.* (2018).

FIG 2. Changes in bat completeness and raw diversity through time: (a) mean scores, based on all taxa and data; (b) mean scores, excluding Lagerstätten taxa; (c) median scores, based on all taxa and data; and (d) raw species richness.

FIG 3. Distribution of bat SCM2 scores in comparison to other tetrapod groups. Comparative taxa: synapsid-grade pelycosaurs (Brocklehurst and Fröbisch 2014); parareptiles (Verrière *et al.* 2016); ichthyosaurs (Cleary *et al.* 2015); plesiosaurs (Tutin and Butler 2017); and sauropodomorphs (Mannion and Upchurch 2010).

FIG 4. Distribution of bat CCM2 scores in comparison to other tetrapod groups. Comparative taxa: eutherian mammals (Davies *et al.* 2017); synapsid-grade pelycosaurs (Brocklehurst and Fröbisch 2014); parareptiles (Verrière *et al.* 2016); anomodonts (Walther and Fröbisch 2013); plesiosaurs (Tutin and Butler 2017); pterosaurs (Dean *et al.* 2016); sauropodomorphs (Mannion and Upchurch 2010); and birds (Brocklehurst *et al.* 2012).

FIG 5. Distribution of bat SCM2 (light grey) and CCM2 (dark grey) scores between 'caves' (caves, fissure fills, sink holes) and fluvio-marine deposits, including Lagerstätten taxa.

FIG 6. Distribution of bat SCM2 (light grey) and CCM2 (dark grey) scores between the northern and southern hemispheres.

FIG 7. Changes in taxonomic completeness and collection rate over historical time. Publication date in correlation with (a) SCM2, (b) CCM2, and in comparison, to (c) cumulative species count.

TABLE CAPTIONS

TABLE 1. Results of comparisons of the population median and distribution of completeness values for different completeness metrics using Mann-Whitney-Wilcoxon tests. Statistically significant results indicated in bold.

Dataset 1	Dataset 2	Test statistic (W)	p-value	p-value following FDR corrections
SCM2 2D	SCM2 3D	93615	0.338	0.338
CSCM2 2D	CSCM2 3D	93587	0.334	0.338
SCM2 2D	CSCM2 2D	83036	**0.0002**	**0.0003**
SCM2 3D	CSCM2 3D	83419	**0.0003**	**0.0003**
SCM2 2D	CCM2	28584	**1.31E-73**	**1.05E-72**
CSCM2 2D	CCM2	34521	**9.78E-62**	**3.91E-61**
SCM2 2D	BSCM2	47996	**9.34E-40**	**3.74E-39**
CCM2	BSCM2	151565	**5.34E-48**	**1.42E-47**

TABLE 2. Results of pairwise comparisons between time series of completeness values using GLS. Statistically significant results indicated in bold.

Comparison	Slope	*t*-value	*p*-value	R^2
SCM2 2D ~ SCM2 3D	1.038	60.27	**<0.0001**	0.996
CSCM2 2D ~ CSCM2 3D	1.029	32.27	**<0.0001**	0.989
SCM2 2D ~ CSCM2 2D	1.191	31.86	**<0.0001**	0.988
SCM2 3D ~ CSCM2 3D	1.194	34.21	**<0.0001**	0.989
SCM2 2D ~ CCM2	2.709	4.14	**0.0014**	0.564
CSCM2 2D ~ CCM2	2.374	4.57	**0.0006**	0.61
SCM2 2D ~ BSCM2	1.349	5.81	**0.0001**	0.735
BSCM2 ~ CCM2	1.538	4.93	**0.0004**	0.673

TABLE 3. Results of pairwise comparisons between time series of completeness values using GLS and following removal of Lagerstätten taxa. Statistically significant results indicated in bold.

Comparison	Slope	*t*-value	*p*-value	R^2
SCM2 2D ~ SCM2 3D	1.012	45.19	**<0.0001**	0.994
CSCM2 2D ~ CSCM2 3D	0.949	27.76	**<0.0001**	0.985

SCM2 2D ~ CSCM2 2D	1.233	20.98	**<0.0001**	0.973
SCM2 3D ~ CSCM2 3D	1.174	23.57	**<0.0001**	0.978
SCM2 2D ~ CCM2	1.542	1.25	0.2366	0.106
CSCM2 2D ~ CCM2	1.278	1.33	0.2090	0.114
SCM2 2D ~ BSCM2	1.527	1.49	0.1612	0.161
BSCM2 ~ CCM2	-0.394	-1.49	0.1627	0.166

1061

1062 **TABLE 4**. Results of comparisons of the population median and distribution of SCM2 values
1063 for bats (2D approach) to other taxonomic groups using Mann-Whitney-Wilcoxon tests.
1064 Statistically significant results indicated in bold.

Dataset 1	Dataset 2	Test statistic (W)	p-value	p-value following FDR corrections
Bat SCM2	Ichthyosaur SCM2	2808	**3.19E-43**	**1.60E-42**
Bat SCM2	Plesiosaur SCM2	4456	**7.91E-42**	**1.98E-41**
Bat SCM2	Sauropodomorpha SCM2	12438	**5.25E-39**	**8.74E-39**
Bat SCM2	Parareptilia SCM2	4253	**2.57E-25**	**3.22E-25**
Bat SCM2	Pelycosauria SCM2	8249	**6.77E-21**	**6.77E-21**

1065

1066 **TABLE 5**. Results of comparisons of the population median and distribution of CCM2
1067 values for bats to other taxonomic groups using Mann-Whitney-Wilcoxon tests. Statistically
1068 significant results indicated in bold.

Dataset 1	Dataset 2	Test statistic (W)	p-value	p-value following FDR corrections
Bat CCM2	Bird CCM2	28427	0.902	0.902
Bat CCM2	Pterosaur CCM2	27999	**4.69E-07**	**6.25E-07**
Bat CCM2	Plesiosaur CCM2	11098	**3.64E-20**	**7.28E-20**
Bat CCM2	Sauropodomorpha. CCM2	27465	**3.96E-08**	**6.33E-08**
Bat CCM2	Anomodontia CCM2	2134	**9.90E-40**	**3.96E-39**
Bat CCM2	Parareptilia CCM2	5447	**5.28E-21**	**1.41E-20**
Bat CCM2	Pelycosauria CCM2	18747.5	0.079	0.090

Bat CCM2	Early Eutheria CCM2	29122.5	**2.36E-73**	**1.89E-72**	

1069

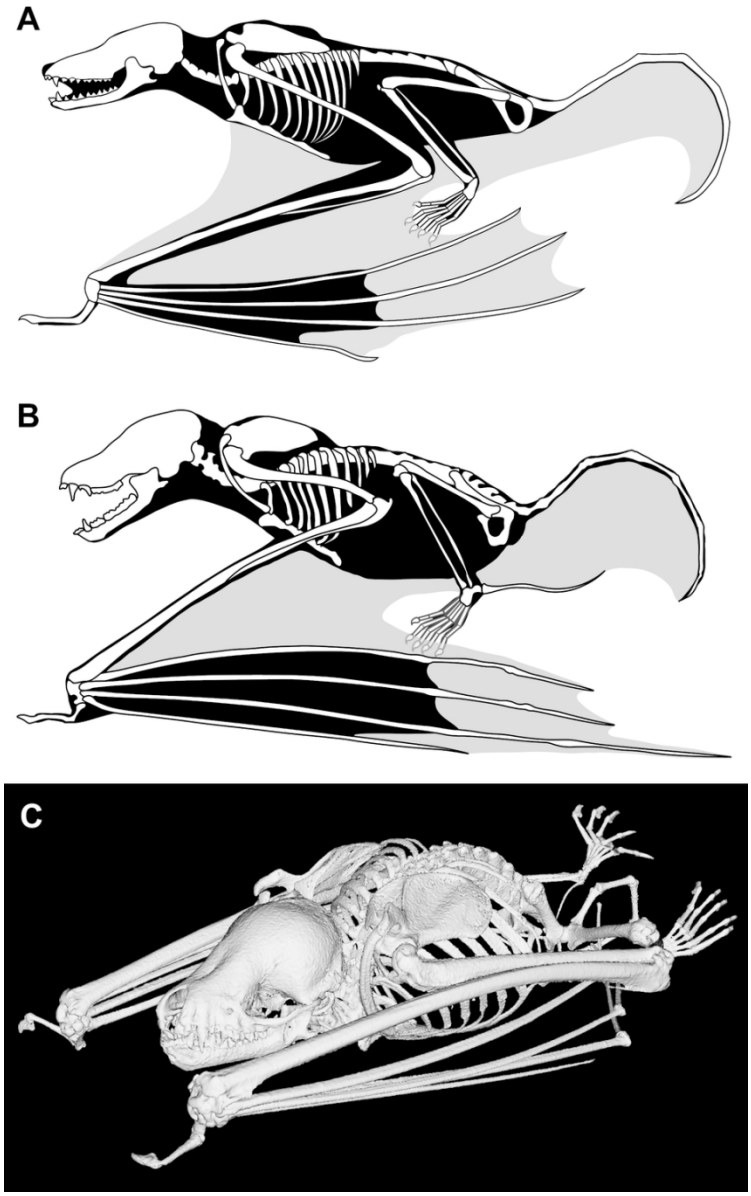

FIG 1. Scientifically-informed skeletal reconstructions of (a) Icaronycteris index and (b) Myotis myotis modified from Wimsatt (2012), used to calculate the 2D weighted skeletal proportions for 'extinct' and 'extant' taxa respectively; and (c) a CT-scanned Myotis daubentonii specimen, used to calculate the 3D weighted skeletal proportions. For detailed breakdown of weightings see Brown et al. (2018).

80x129mm (300 x 300 DPI)

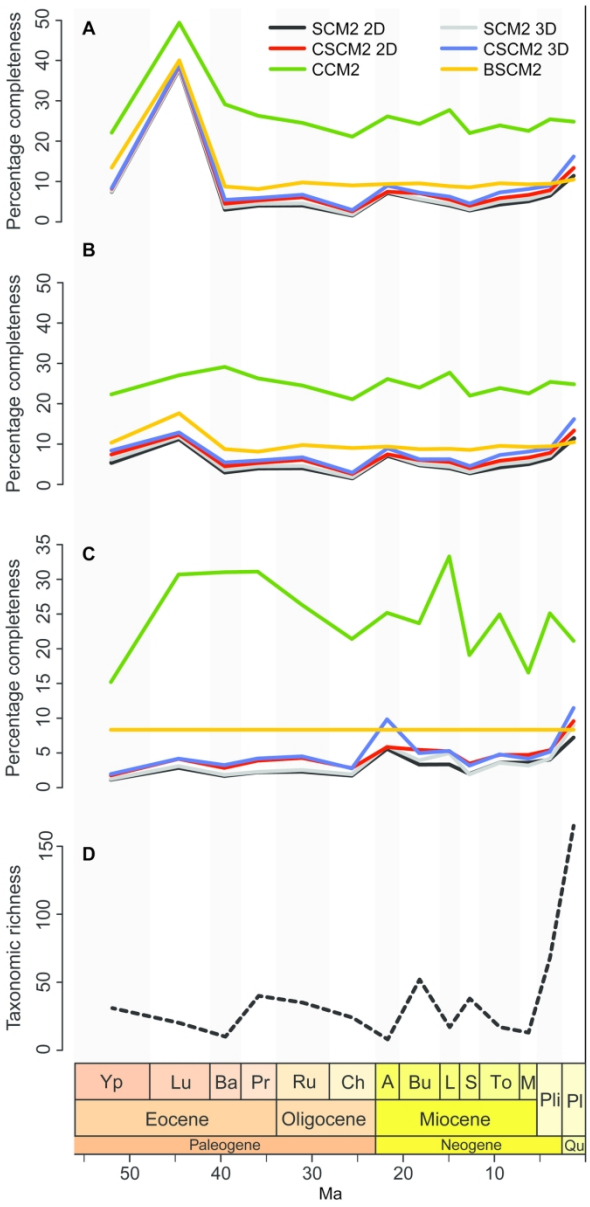

FIG 2. Changes in bat completeness and raw diversity through time: (a) mean scores, based on all taxa and data; (b) mean scores, excluding Lagerstätten taxa; (c) median scores, based on all taxa and data; and (d) raw species richness.

116x224mm (300 x 300 DPI)

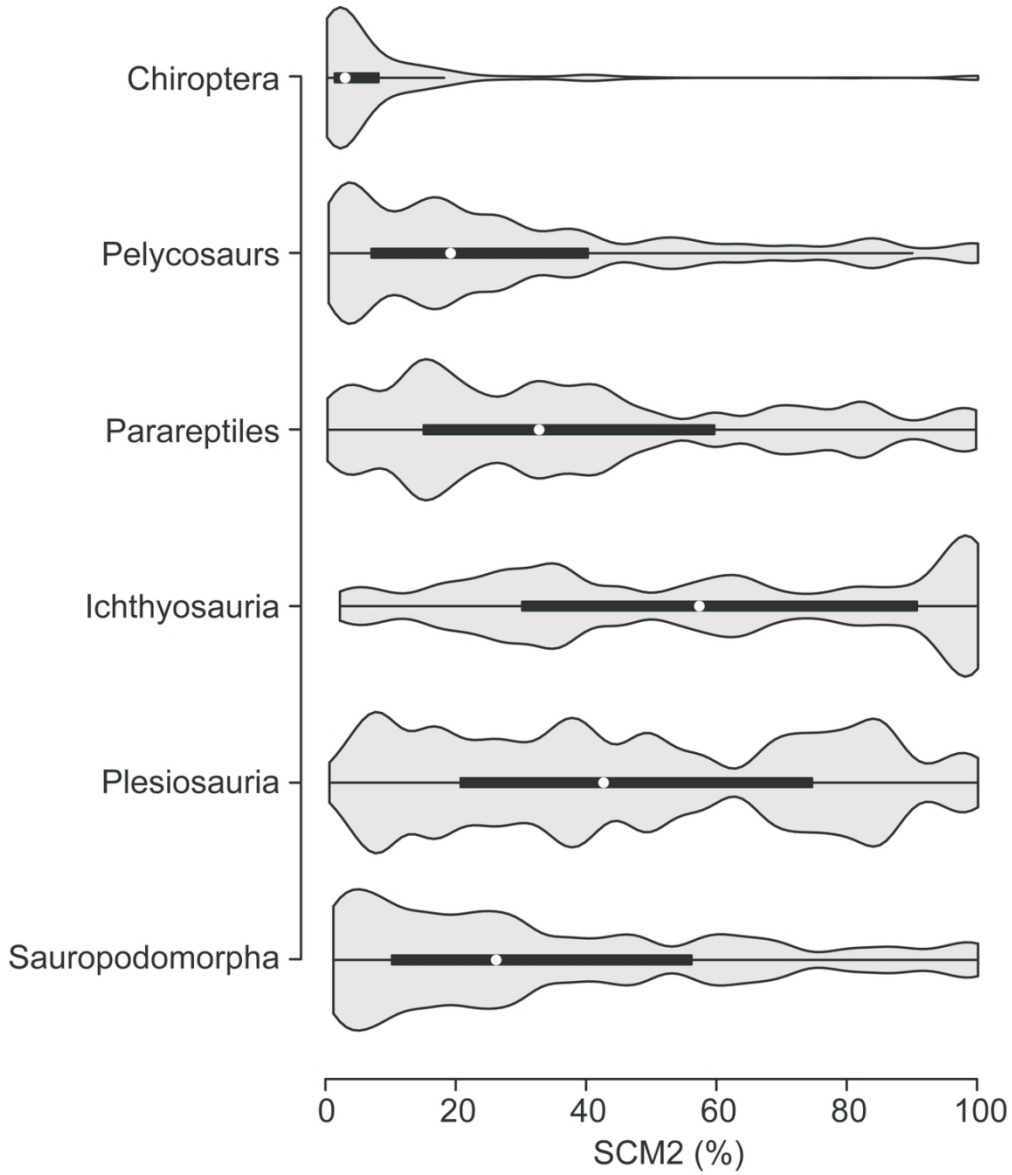

FIG 3. Distribution of bat SCM2 scores in comparison to other tetrapod groups. Comparative taxa: synapsid-grade pelycosaurs (Brocklehurst and Fröbisch 2014); parareptiles (Verrière et al. 2016); ichthyosaurs (Cleary et al. 2015); plesiosaurs (Tutin and Butler 2017); and sauropodomorphs (Mannion and Upchurch 2010).

107x125mm (300 x 300 DPI)

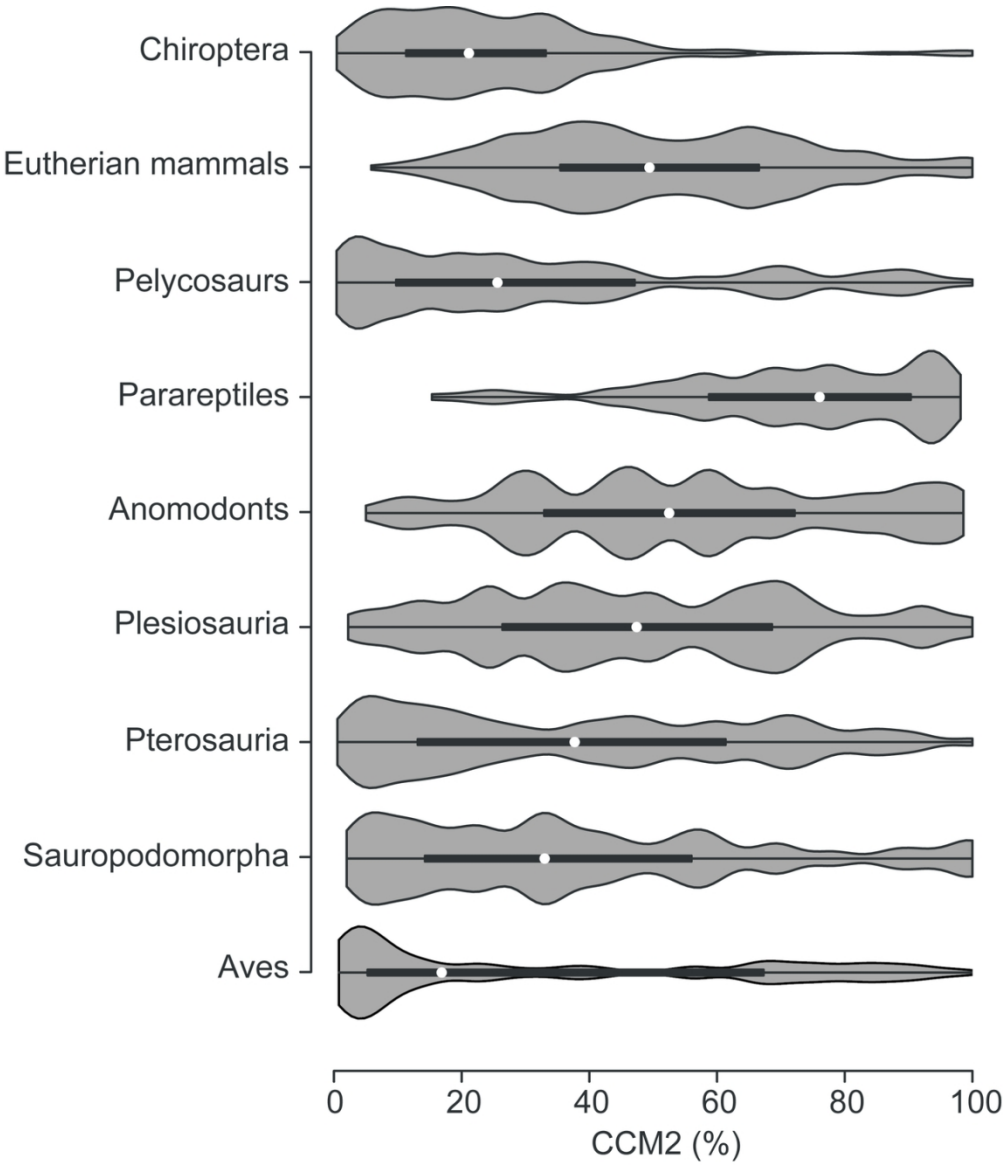

FIG 4. Distribution of bat CCM2 scores in comparison to other tetrapod groups. Comparative taxa: eutherian mammals (Davies et al. 2017); synapsid-grade pelycosaurs (Brocklehurst and Fröbisch 2014); parareptiles (Verrière et al. 2016); anomodonts (Walther and Fröbisch 2013); plesiosaurs (Tutin and Butler 2017); pterosaurs (Dean et al. 2016); sauropodomorphs (Mannion and Upchurch 2010); and birds (Brocklehurst et al. 2012).

109x126mm (300 x 300 DPI)

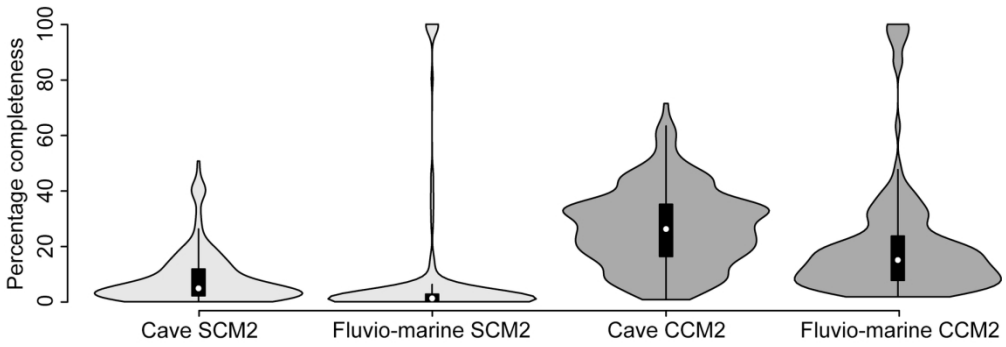

FIG 5. Distribution of bat SCM2 (light grey) and CCM2 (dark grey) scores between 'caves' (caves, fissure fills, sink holes) and fluvio-marine deposits, including Lagerstätten taxa.

161x54mm (300 x 300 DPI)

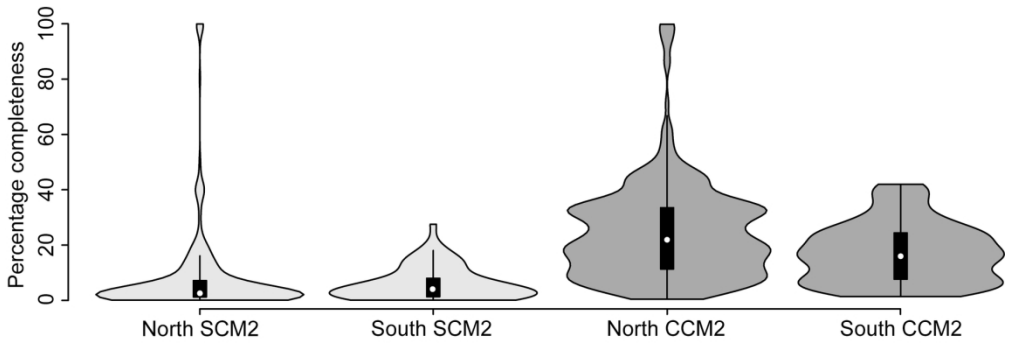

FIG 6. Distribution of bat SCM2 (light grey) and CCM2 (dark grey) scores between the northern and southern hemispheres.

164x55mm (300 x 300 DPI)

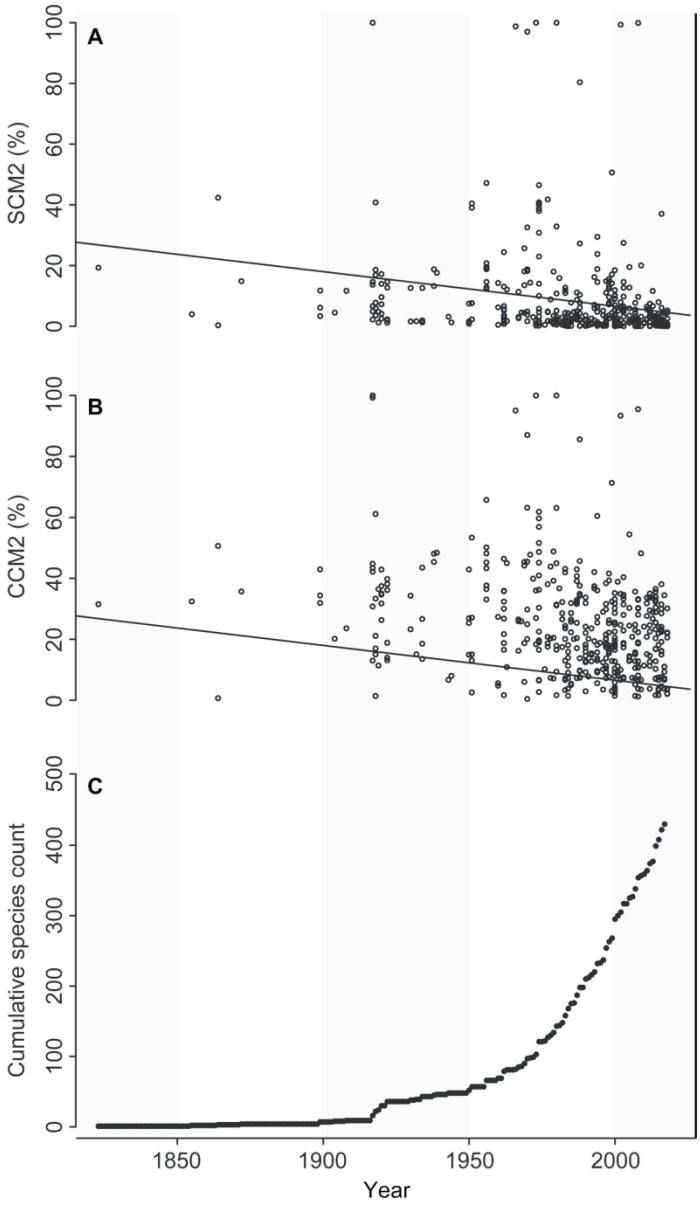

FIG 7. Changes in taxonomic completeness and collection rate over historical time. Publication date in correlation with (a) SCM2, (b) CCM2, and in comparison, to (c) cumulative species count.

109x189mm (300 x 300 DPI)