The accuracy and precision of body mass estimation in non-avian dinosaurs

Nicolás E. Campione[1]* and David C. Evans[2,3]

[1]*Palaeoscience Research Centre, University of New England, Armidale, New South Wales, 2351, Australia*
[2]*Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks St, Toronto, Ontario, M5S 3B2, Canada*
[3]*Department of Natural History, Royal Ontario Museum, 100 Queen's Park, Toronto, Ontario, M5S 2C6, Canada*

ABSTRACT

Inferring the body mass of fossil taxa, such as non-avian dinosaurs, provides a powerful tool for interpreting physiological and ecological properties, as well as the ability to study these traits through deep time and within a macroevolutionary context. As a result, over the past 100 years a number of studies advanced methods for estimating mass in dinosaurs and other extinct taxa. These methods can be categorized into two major approaches: volumetric-density (VD) and extant-scaling (ES). The former receives the most attention in non-avian dinosaurs and advanced appreciably over the last century: from initial physical scale models to three-dimensional (3D) virtual techniques that utilize scanned data obtained from entire skeletons. The ES approach is most commonly applied to extinct members of crown clades but some equations are proposed and utilized in non-avian dinosaurs. Because both approaches share a common goal, they are often viewed in opposition to one another. However, current palaeobiological research problems are often approach specific and, therefore, the decision to utilize a VD or ES approach is largely question dependent. In general, biomechanical and physiological studies benefit from the full-body reconstruction provided through a VD approach, whereas large-scale evolutionary and ecological studies require the extensive data sets afforded by an ES approach. This study summarizes both approaches to body mass estimation in stem-group taxa, specifically non-avian dinosaurs, and provides a comparative quantitative framework to reciprocally illuminate and corroborate VD and ES approaches. The results indicate that mass estimates are largely consistent between approaches: 73% of VD reconstructions occur within the expected 95% prediction intervals of the ES relationship. However, almost three quarters of outliers occur below the lower 95% prediction interval, indicating that VD mass estimates are, on average, lower than would be expected given their stylopodial circumferences. Inconsistencies (high residual and per cent prediction deviation values) are recovered to a varying degree among all major dinosaurian clades along with an overall tendency for larger deviations between approaches among small-bodied taxa. Nonetheless, our results indicate a strong corroboration between recent iterations of the VD approach based on 3D specimen scans suggesting that our current understanding of size in dinosaurs, and hence its biological correlates, has improved over time. We advance that VD and ES approaches have fundamentally (metrically) different advantages and, hence, the comparative framework used and advocated here combines the accuracy afforded by ES with the precision provided by VD and permits the rapid identification of discrepancies with the potential to open new areas of discussion.

Key words: body mass, estimation, accuracy, precision, linear predictive models, statistical error, 3D visualisation, Dinosauria

CONTENTS

* Author for correspondence (Tel: +61 2 6773 1906; E-mail: ncampion@une.edu.au).

I. INTRODUCTION

Body size of extinct terrestrial vertebrates represents one of the most important and useful proxies for studying numerous biological properties through deep time (e.g. Bakker, 1972; Farlow, 1976; Erickson, Rogers & Yerby, 2001; Henderson, 2004; Laurin, 2004; Gillooly, Allen & Charnov, 2006; Finarelli, 2008; Pontzer, Allen & Hutchinson, 2009; Benson et al., 2011; Sookias, Benson & Butler, 2012a; Sookias, Butler & Benson, 2012b; Benson et al., 2014, 2018). Its utility stems from decades of research in extant taxa where it became evident that body size and, more specifically, body mass is related to almost all aspects of organismal biology including physiology (e.g. metabolism, growth, and fecundity) and ecology (e.g. population density, land area, and extinction risk) (e.g., Kleiber, 1947; Hemmingsen, 1960; Jerison, 1969; Peters, 1983; Brown & Maurer, 1986; Brown, Marquet & Taper, 1993; Burness, Diamond & Flannery, 2001; Gillooly et al., 2001; Gillooly et al., 2002; Capellini & Gosling, 2007; McClain & Boyer, 2009). Since body mass cannot be directly measured in the fossil record, numerous studies over the past century

proposed methods to estimate body mass from skeletal fossil remains, in particular as they relate to dinosaurs (Alexander, 1985; Anderson, Hall-Martin & Russell, 1985; Bates et al., 2009b; Campbell Jr & Marcus, 1992; Campione, 2017; Campione & Evans, 2012; Campione et al., 2014; Colbert, 1962; Erickson & Tumanova, 2000; Gregory, 1905; Gunga et al., 1995; Henderson, 1999; Hurlburt, 1999; Mallison, 2010; Mazzetta, Christiansen & Fariña, 2004; O'Gorman & Hone, 2012; Paul, 1997; Seebacher, 2001; Sellers et al., 2012). These methods can be categorized into two principal approaches that are fundamentally different in their methodology and assumptions. These include: (i) the volumetric-density (VD) approach, which incorporates the greatest amount of information about a skeleton; and (ii) the extant-scaling (ES) approach, which integrates empirical knowledge of the relationship between bone dimensions and body mass in extant analogues. Within the last two decades, VD and ES approaches have been extensively applied to investigate a suite of biological properties in non-avian dinosaurs, including metabolism (Seebacher, 2003; Gillooly et al., 2006; Pontzer et al., 2009; Grady et al., 2014), growth patterns and rates (Erickson et al., 2001; Hutchinson et al., 2011; Myhrvold, 2013;

Otero *et al.*, 2019), locomotion (Alexander, 1985; Christiansen, 1997; Hutchinson *et al.*, 2011; Sellers *et al.*, 2013), estimation and implications of the centre of mass (Henderson, 1999, 2006; Allen *et al.*, 2013; Maidment, Henderson & Barrett, 2014), aquatic abilities (Henderson, 2004, 2014, 2018), defence capabilities (Mallison, 2011*a*), relative sizes of muscles/organs (Gunga *et al.*, 1995; Franz *et al.*, 2009), macroevolutionary dynamics (Allen *et al.*, 2013; Dececchi & Larsson, 2013; Benson *et al.*, 2014, 2018), and both palaeoecology and taphonomy (O'Gorman & Hone, 2012; Codron *et al.*, 2012*b*; Brown *et al.*, 2013).

Because both approaches share a common goal, they are often viewed in opposition (e.g. Campione & Evans, 2012; Sellers *et al.*, 2012; Bates *et al.*, 2015; Brassey, Maidment & Barrett, 2015). However, given the variety of research programmes in dinosaur palaeobiology that use body size, the approaches are often question specific. For instance, biomechanical and physiological studies, which are often investigated through a few and largely complete specimens, are best addressed through a VD approach, as it integrates a more holistic assessment of a specimen and permits interpretations of body proportions (e.g. muscle size, centre of mass, moments of inertia, and surface area) and volumetric information (Henderson, 2006; Hutchinson, Ng-Thow-Hing & Anderson, 2007; Pontzer *et al.*, 2009; Hutchinson *et al.*, 2011; Mallison, 2011*a*; Allen *et al.*, 2013; Maidment *et al.*, 2014). By contrast, studies reconstructing large-scale patterns of evolution prefer an ES approach because of its suitability for assembling large data sets (Benson *et al.*, 2014, 2018). However, it should be noted that these are recent applications of ES approaches, previous studies looking at body size evolution opted instead for hybrid approaches that utilized VD-based estimates to generate general regression equations (such as those derived by Christiansen & Fariña, 2004; O'Gorman & Hone, 2012; Seebacher, 2001) or skeletal proxies (e.g. femoral length) without estimating body mass (Hone *et al.*, 2005; Carrano, 2006; Turner *et al.*, 2007; Hone *et al.*, 2008; Ősi *et al.*, 2012; Sookias *et al.*, 2012*a*, 2012*b*; Lee *et al.*, 2014; Puttick, Thomas & Benton, 2014).

Recent computational advances and innovations in three-dimensional (3D) scanning technologies have led to a broader application of VD approaches to non-avian dinosaurs relative to those of ES (but see Campione, 2017; Campione & Evans, 2012; Campione *et al.*, 2014). As a result, VD approaches transitioned greatly over research history from physical, statuesque reconstructions (e.g, Colbert, 1962; Alexander, 1985), to reconstructions based directly on two-dimensional (2D) skeletal restorations (i.e. illustrations) – either through physical (e.g. Paul, 1988, 1997, 2010) or mathematical (e.g. Henderson, 1999; Seebacher, 2001) means – and, finally, to 3D-based virtual models (e.g. Bates *et al.*, 2009*b*; Mallison, 2010; Hutchinson *et al.*, 2011). The general principle behind VD approaches, however, remains the same. To estimate mass (*M*), the VD approach requires: (*i*) a life reconstruction of the fossil taxon; (*ii*) a method with which to calculate the volume (*V*) of the reconstruction; and (*iii*) an estimate of body density (*D*). The body mass of an animal is then estimated using the standard density equation, solved for *M* ($M = D \times V$). In comparison, the ES approach utilizes skeletal measurements taken from living animals that are then regressed against body mass to derive a predictive equation. Assuming that the estimated animal follows the same scaling pattern as the extant exemplars on which the regression model is based, the resulting predictive equation and its associated error can be used as an accurate approximation of body mass. Both approaches can include a variety of different parameters and can be based on different data sets, both have advantages and disadvantages, and both make unavoidable methodological assumptions, resulting in a somewhat dichotomous outlook among some researchers on the capacity of these approaches to estimate body mass accurately, with criticisms forwarded against both camps (Paul, 1997; Carrano, 2001; Gunga *et al.*, 2002; Hutchinson *et al.*, 2011; Campione & Evans, 2012; Sellers *et al.*, 2012). However, both can, and should, be used as reciprocals of each other to better describe the sizes of extinct taxa and to achieve a better understanding of their palaeobiology.

In this review we provide a synopsis of the principal methods used to estimate body masses of non-avian dinosaurs. We focus on this group because non-avian dinosaurs receive a great deal of attention by both VD and ES approaches and provide the best comparative model. However, both approaches are applied to extinct mammals (e.g. Romer & Price, 1940; Damuth & MacFadden, 1990; Gingerich, 1990; Fortelius & Kappelman, 1993; Fariña, Vizcaíno & Bargo, 1998; De Esteban-Trivigno, Mendoza & De Renzi, 2008; Rinderknecht & Blanco, 2008; Millien & Bovy, 2010; De Esteban-Trivigno & Köhler, 2011; Brassey & Sellers, 2014; Larramendi & Palombo, 2015; Basu, Falkingham & Hutchinson, 2016; Cuff, Goswami & Hutchinson, 2017; Hopkins, 2018), non-mammalian synapsids (Romer & Price, 1940; Blob, 2001; Romano & Manucci, 2019; Romano & Rubidge, 2019), and birds (Campbell Jr & Marcus, 1992; Field *et al.*, 2013; Brassey & Sellers, 2014; Serrano, Palmqvist & Sanz, 2015) and see Hopkins (2018) for a recent review of body mass estimation techniques in fossil mammals. We discuss the assumptions inherent to VD and ES approaches, as well as their advantages and disadvantages. We compile a data set of mass estimates derived using VD approaches and quantitatively compare these estimates to the ES relationships derived in Campione & Evans (2012) and Campione *et al.* (2014). Finally, we forward our comparative context as an option that combines the precision and accuracy of VD and ES approaches, respectively.

II. BODY MASS INFERENCE IN THE FOSSIL RECORD

(1) Volumetric-density approach

The volumetric-density approach is the older of the two approaches, at least as it applies to dinosaurs. It was employed for the first time over a century ago to estimate the body mass of *Brontosaurus* (AMNH 460; Gregory, 1905),

most recently considered an indeterminate apatosaurine (Tschopp, Mateus & Benson, 2015). Given its long history, various permutations of the approach have been proposed. In this section, we present a synopsis of the chief VD rendition, however, the original author(s) should be referred to for more detailed explanations.

(a) Approaches

i Physical scale models. The original VD approach for estimating body mass is based on the physical construction of scale models. Gregory (1905) was the first to provide a body mass estimate for a dinosaur based on a reconstructed model by Charles M. Knight of what was then considered *Brontosaurus excelsus* (Fig. 1). The author created plaster replicas, one of which he segmented into several pieces, and ascertained the volume of the model by submerging the segments in water following Archimedes' Principle: the volume of fluid displaced by a submerged body will be equal to the volume of the body. Gregory initially assumed a total body density of water (approximately 1 g/cm^3), and then applied a scaling factor of 16 to obtain a mass of 31 t. Finally, he added 10% to the mass based on the common supposition, at the time, that sauropods were aquatic and, as a result, must have had a density greater than that of water, for a final mass estimate of 34.1 t.

The method outlined by Gregory (1905) was expanded upon by Colbert (1962) who presented the first data set of body sizes in non-avian dinosaurs, including exemplars from major branches of the dinosaurian evolutionary tree (e.g. sauropods, theropods, ornithopods, ceratopsians, and thyreophorans; Fig. 2). Colbert (1962) also used plaster reconstructions but, unlike Gregory (1905), the models were kept in one piece and submerged in sand, rather than water. The mathematical calculation to determine body mass was the same as that presented by Gregory (1905), with the main exception that Colbert (1962) did not agree with the assumption that non-avian dinosaurs were denser than water. Based on measurements taken from a specimen of *Alligator* and *Heloderma*, Colbert (1962) assumed a body density of 0.9 g/cm^3 for non-avian dinosaurs.

Fig 2. Historical plaster reconstructions used by, and originally published in, Colbert (1962). The volumes of these reconstructions were determined by volumetric displacement in sand and the associated body masses represent the first body mass data set ($N = 15$) generated for dinosaurs. Images (#328210, #328211, #328212 from top to bottom, respectively) reproduced with permission from the American Museum of Natural History Library.
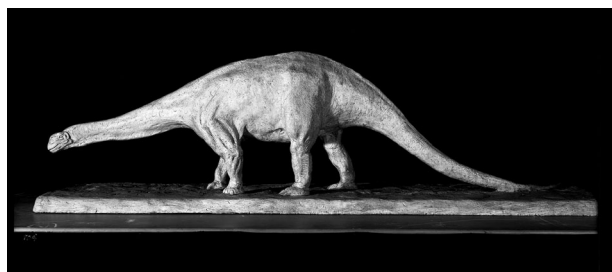

Fig 1. Replica of Charles M. Knight's historical reconstruction of what was then considered *Brontosaurus*; the reconstruction of which was used by William K. Gregory to generate the first body mass estimate of a dinosaur in 1905. Image (#35742) reproduced with permission from the American Museum of Natural History Library.

The application of Archimedes's Principle to body mass estimation from a scale model was refined by Alexander (1985) using solid, plastic, 1/40 scale toy models then sold at the British Museum of Natural History. Instead of measuring the volume of the fluid displaced by the scale model, Alexander used a scale, which at one end hung the model within a beaker and, at the other, a set of weights that kept the scale at equilibrium. The beaker was subsequently filled with water, decreasing the weight of the model. The difference between the weight of the model in water and in air is equal to the volume of the model. This technique is less prone to measurement error as that applied by the previous authors, but still requires the use of a body density, which Alexander (1985) assumed to be equal to that of water (i.e. 1 g/cm^3).

The most recent applications of physical reconstructions to estimate body mass were by Paul (1988, 1997, 1998, 2010, 2019), Christiansen (1997, 1998), and Mazzetta, Christiansen & Fariña (2004), who created clay models. Unlike the previous studies, which used artistic renditions, these models were generated based on 2D skeletal reconstructions that incorporated measurements taken from specimens. The volumes of these models were determined using water displacement methods similar to that used by Gregory (1905), which were then multiplied by a density parameter varying from 0.85 g/cm^3 to 0.95 g/cm^3.

ii Two-dimensional mathematical models. An alternative method for obtaining the volume of an animal is to generate a reconstruction in two dimensions and then mathematically integrate it into the third dimension. Although this method is somewhat simpler than those described in the previous section, it generally assumes a circular or oval cross-sectional shape to the body. Originally developed by Jerison (1969) to calculate the volume of brain endocasts, Graphic Double Integration (GDI) was adapted by Hurlburt (1999) as a viable method for calculating the volume of a 2D reconstruction. A similar approach may have been first used by Mazzetta, Fariña & Vizcaíno (1998) to estimate the mass of *Carnotaurus*, although the details of their approach were unclear. Graphic double integration is based on the notion that the volume of a complex 3D shape can be modelled as an elliptical cylinder. Two different reconstructed views are needed that are at right angles to each other (e.g. lateral and dorsal). Transverse measurements are then taken at equal intervals in both views and the mean of these measurements in each view is then used as the principal axes (*a* and *b*) of an ellipse. The length of the reconstruction is used as the height (*h*) of the elliptical cylinder and hence the volume of the model is calculated by using the standard equation for calculating the volume of an elliptical cylinder ($V = \pi abh$). Jerison (1969) showed that, for his brain endocast data set, GDI calculated volumes that were within 5% of that obtained using a water displacement approach, but see Motani (2001) for issues associated with assuming elliptical cross sections. Hurlburt (1999) calculated the body volume of *Edaphosaurus boanerges*, a Permian synapsid, by performing separate GDIs for the head, trunk, tail, as well as the proximal and distal portion of the fore and hind limbs. The volumes of each section were added together to obtain the total body volume and, similar to previous studies, Hurlburt (1999) assumed a body density equal to that of water.

Seebacher (2001) developed a method known as the 'polynomial' technique. As in GDI, it takes length (or depth) measurements at standard intervals along the length of a 2D reconstruction of the body outline based on a skeletal schematic. Unlike GDI, this method only requires use of a lateral-view reconstruction. The depth measurements are then plotted against their position along the length of the body, and the points fitted using an eighth-order polynomial. The polynomial function is then integrated over the length of the body, which provides the volume of a circular body shape. Given that the latter is not a good approximation for

true body shape, Seebacher (2001) developed a correction factor based on the relationship between body depth and width.

An alternative method for constructing 3D volumetric reconstructions from 2D images was presented by Henderson (1999). The basis, with some variation, is essentially a computational translation of the 2D method presented by Jerison (1969) and Seebacher (2001). 3D computational reconstructions are derived from 2D reconstructions of lateral and dorsal views of an animal, organized on a 3D Cartesian grid (*x*-, *y*-, and *z*-axes). The outlines are then sliced into segments. The intersections between the slices and the outline of the reconstruction represent a set of coordinates that can be inputted into a drafting software package. The distances between the *x*–*y* and *x*–*z* coordinates represent the main axes of a series of ellipses along the body that correspond to the initial slices. Taken together, sets of adjacent slices form a series of three-dimensional 'slabs'. Using linear algebra and integration techniques (see Henderson, 1999 for full details), the volume of each slab is calculated and the total volume is the sum of the slab volumes. The volume of the limbs, horns, frills, and plates were estimated separately using the same method. Body density was assumed to be 1 g/cm^3 for all dinosaurs, however, in the case of *Triceratops*, Henderson (1999) assumed a density of 1.7 g/cm^3 for the horns and frill and, for sauropod heads and necks, Henderson (2006) assumed a density of 0.3 g/cm^3, both based on extant comparisons. In addition, a percentage of lung volume (e.g. 10–15%) can be incorporated to adjust the trunk volume.

iii Three-dimensional virtual models. One of the first attempts to reconstruct a dinosaur in three dimensions was based on the principle of photogrammetry (Gunga *et al.*, 1995, 2002; Gunga *et al.*, 1999), reviewed by Stoinski, Suthau & Gunga (2011). The approach uses 2D photographs of specimens (generally mounts) from three or more different angles that complement each other to create a 3D image through the use of photogrammetric instruments and, more recently, software (Mallison & Wings, 2014). The 3D depiction of a mount is converted into a series of digital points, or a point cloud, representing particular regions of the skeleton (i.e. skull, neck, thorax, etc.) or a complete skeletal mount. These points can be inputted into drafting software packages to create a wireframe reconstruction that, depending on the anatomical region, can be modelled as standard geometric shapes (e.g. cylinders, spherical caps, and truncated cones) from which volumes can be determined using standard geometric volumetric equations. This method was used extensively by German research groups interested in the size and correlated physiological attributes of *Giraffatitan brancai*, *Dicraeosaurus hansmanni*, and *Diplodocus carnegiei* (Gunga *et al.*, 1995, 2002; Gunga *et al.*, 1999; Wiedemann, Suthau & Albertz, 1999). In practice, photogrammetry continues to be a viable source of 3D data and, given adequate photographic resolution and specimen coverage, should provide as detailed a 3D model as a 3D scanner, although processing times can be substantially higher.

Following technological advances in 3D scanning and 3D imaging software, reconstructions can be obtained *via* complete 3D scans of museum mounts and specimens (Gunga *et al.*, 2007, 2008; Bates *et al.*, 2009*a*, 2009*b*; Mallison, 2010, 2011*a*, 2011*b*; Stoinski *et al.*, 2011; Sellers *et al.*, 2012; Brassey *et al.*, 2015). Scanners can rapidly generate a full-size virtual 3D skeleton that can be imported into a drafting software package where the body outline can be created and volumes calculated. Scans can be obtained by various instruments [e.g. surface scanners: Mensi S25 (Gunga *et al.*, 2007, 2008) and RIEGL LMS-Z420i (Bates *et al.*, 2009*b*; Hutchinson *et al.*, 2011), or CT scanners (Mallison, 2010; Hutchinson *et al.*, 2011)], but the general principle is the same. Scans are taken of museum mounts from various positions around the skeleton, the number and actual position of which will depend on the complexity of the skeleton and mount. These scans produce virtual images that are accurate to the sub-centimetre scale (e.g. Gunga *et al.*, 2007; Bates *et al.*, 2009*b*; Stoinski *et al.*, 2011). The resulting scan is in actuality a series of points in virtual 3D space that combine to create a point cloud or mesh representing the different scans taken from the skeleton. The clouds or meshes can be manipulated and sampled by various programs (e.g. PolyWorks; Bates *et al.*, 2009*b*) where they are stitched together to form a single virtual depiction of the mounted skeleton. Once in virtual space, the scan can either be treated as a single entity or subdivided into various body segments (e.g. skull, thorax, tail, etc.). This segmentation technique is commonly used in VD approaches to reduce complex body outlines into simpler shapes, but more importantly, to determine the mass properties of particular anatomical regions, and their relative proportions (Gunga *et al.*, 1995, 2002; Gunga *et al.*, 1999; Henderson, 1999; Gunga *et al.*, 2007; Gunga *et al.*, 2008; Bates *et al.*, 2009*b*, 2016; Mallison, 2010; Hutchinson *et al.*, 2011; Allen *et al.*, 2013).

The body outline can be modelled by various drafting packages (e.g. AutoCad and Maya) and although original studies modelled outlines based on a series of rotational solids for distinct regions along the body (e.g. Gunga *et al.*, 2007), later studies advocated the use of non-uniform rational B (Bézier)-splines (NURBs) (Hutchinson *et al.*, 2007; Gunga *et al.*, 2008; Bates *et al.*, 2009*b*; Stoinski *et al.*, 2011), which mathematically incorporate information from 2D objects (e.g. lines, circles, and arcs) to create a 3D object. These are easily manipulated by most CAD packages to create a body outline of a full-scale skeleton from which a volume can be calculated. As is the case for all VD approaches, an estimate of density is still required, which, in the studies cited above, ranges from 0.8 to 1 g/cm^3, depending on the body region and anatomy of the specific dinosaur being modelled.

iv Minimum convex hulling. Sellers *et al.* (2012) presented the most recent rendition of the VD approach, building upon original work done using LiDAR scanning techniques (Bates *et al.*, 2009*a*, 2009*b*). The approach differs from previous 3D modelling VD approaches in that it attempts to better integrate extant data into the generation of a body outline. The methodological foundation incorporates the original scan data obtained from a mount (i.e. the point cloud or mesh)

but, instead of reconstructing the body outline based on assumptions of soft tissue, the authors separate the scan into various anatomical segments for which 3D minimum envelopes are calculated *via* 3D minimum convex hulls (MCHs). As MCHs are derived from the skeletal data, they are an objective representation of the minimum volume needed to envelop the skeleton. The volume of the resultant MCHs is calculated and then multiplied by an assumed body (or segment) density to obtain a mass estimate. However, it is obvious that a body mass determined *via* a minimum convex hull of a skeleton will always underestimate the true mass, as these hulls are not real, but mathematical representations of body outlines (Sellers *et al.*, 2012). Accordingly, the authors applied this technique to a suite of large-bodied extant mammals that ranged from approximately 90 kg to 3000 kg and compared masses determined *via* the VD$_{MCH}$ method against estimates obtained from the literature of extant exemplars. The VD$_{MCH}$ method was found to consistently underestimate the body mass of their extant sample by approximately 21%; an expansion factor that could be applied to MCHs derived from skeletons of extinct taxa, such as non-avian dinosaurs. The extant sample also allowed for the generation of 95% confidence intervals, forming the basis for body mass estimation error ranges (Sellers *et al.*, 2012).

(b) Criteria

i Assumptions. By definition, the reconstruction of an extinct animal, which is known from taphonomically biased material (e.g. only the skeleton), is subjective. A reconstruction can be conservatively rooted by empirical data (such as from extant animals), but it still incurs important assumptions with regards to soft tissue reconstruction. The various permutations of the VD approach, with the partial exception of the Sellers *et al.* (2012) VD$_{MCH}$ model, make two important suppositions: (*i*) an accurate body outline upon which to base volume calculations, and (*ii*) an estimate(s) of body (or body segment) density in order to covert the volume into a mass.

The reconstruction of a body outline is one of the most subjective aspects of the VD approach. It requires appreciable assumptions to be made regarding soft-body anatomy and, in particular, relies heavily on the completeness of the skeleton on which the model is based. In this regard, reconstructions of taxa such as *Tyrannosaurus rex* (Hutchinson *et al.*, 2011) and *Stegosaurus* (Brassey *et al.*, 2015) can be viewed with a certain level of scientific rigour, whereas those based on highly incomplete specimens, such as *Argentinosaurus* (Sellers *et al.*, 2013), should be viewed with a healthy level of scepticism. Only a few non-avian dinosaur fossils preserve complete tails and, when they do, their lengths are remarkably variable, even among closely related genera (Hone, 2012). This is particularly problematic for the application of VD methods to non-avian dinosaurs, as they retained the plesiomorphic reptilian condition in which major locomotor muscles (such as the *M. caudofemoralis*) originate on the tail (Gatesy, 1990, 1991; Persons IV &

Currie, 2011). Such factors will increase the uncertainty and error associated with the rather large proportion of mass held within the tail of a non-avian dinosaur. In general, the nature of variation in the proportions and anatomical properties of tissue types (e.g. skin, organs, and muscle) and their relationship with body size remain poorly understood. The proportions of mass appropriated by different tissues can vary significantly among mammals of constant body mass, especially when muscles are linked to specific locomotor patterns (Grand, 1990). Determining the sizes of organs in extinct taxa is equally problematic as non-avian reptiles, mammals, and birds exhibit differential size-dependent scaling patterns, complicating the extrapolation of organ size in non-avian dinosaurs (Franz *et al.*, 2009). Other soft-tissue-related assumptions include the amount of interosseous spacing associated with cartilage, the variation of which is only appreciated in a few extant forms (e.g. Bonnan *et al.*, 2010; Tsai & Holliday, 2015).

The unavoidable assumption made by all VD approaches is that of body density. Initial VD approaches implemented a constant value for the entire body based on hypotheses and observations held at respective times. Gregory (1905) assumed a density greater than water (about 1.1 g/cm^3), due to the hypothesis held during the 19th and early 20th centuries on the aquatic lifestyle of sauropods. Later studies reduced this number to ≤ 1 g/cm^3. Only Colbert (1962) based his assumption (0.9 g/cm^3) on empirical extant data collected from an alligator and Gila monster (*Heloderma horridum*). Modern approaches can now virtually segment body reconstructions so that densities can be applied differentially along the body (Henderson, 1999; Gunga *et al.*, 2008; Hutchinson *et al.*, 2011). However, empirical studies of body density in extant taxa suggest that this property may vary wildly, especially among taxa with highly pneumatic bodies (Colbert, 1962; Hazlehurst & Rayner, 1992; Hurlburt, 1999). That the relationship between density and body mass in birds is negative (Brassey & Sellers, 2014), suggests that larger pneumatic taxa may have relatively lower body densities. Density assumptions, therefore, need careful consideration when applying VD approaches to saurischian dinosaurs given the weak, but significant positive trends between pneumaticity and overall size in theropods (Benson *et al.*, 2011). At this time, however, the interplay between pneumaticity, density, and body size requires substantially more data to understand the variation therein and empirically ground assumptions of density in VD approaches.

Both body envelope and density assumptions are unavoidable but can have an appreciable effect on the final mass estimate. As a case in point, body mass estimates by the same research group between 1995 and 2008 for the *Giraffatitan brancai* specimen mounted at the Humboldt Museum in Berlin, Germany, differed widely. Originally, Gunga *et al.* (1995) presented a body mass of 74.4 tonnes based on photogrammetric techniques, which modelled the body outlines as a series of standard shapes (e.g. truncated cones and cylinders) and an assumed body density of 1 g/cm^3. By contrast, later boyd mass estimates incorporated lower estimates of body density (0.8 g/cm^3) and more carefully modelled body outline using NURBs,

to obtain a final mass estimate of 38 tonnes, an almost 50% difference from the original estimate (Gunga *et al.*, 2008). This by no means reflects inaptitude in earlier studies, but rather reflects shifts in palaeobiological interpretations and in our understanding of the anatomical properties of fossils (e.g. skeletal pneumaticity; Wedel, 2003), along with technological and computational improvements.

ii Advantages. Volumetric-density approaches represent the most adequate technique with which to address functional (e.g. locomotion and defence) and physiological (e.g. metabolism, body temperature, organ size, and circulation) hypotheses in the fossil record (Christiansen, 1997, 1998; Gunga *et al.*, 1999, 2007, 2008; Henderson, 2004, 2006, 2010, 2014; Henderson & Snively, 2004; Hutchinson *et al.*, 2007; Franz *et al.*, 2009; Pontzer *et al.*, 2009; Mallison, 2010, 2011a, 2011b, 2011c; Hutchinson *et al.*, 2011; Allen *et al.*, 2013). In terms of body mass estimation, however, VD approaches have an advantage in that they take a more holistic view by incorporating information about the entire skeleton, rather than relying on individual skeletal measurements (as in ES approaches). Such a specimen-focused approach means that it can be applied to any specimen, regardless of ontogenetic state, without the need for additional scaling corrections [such as developmental mass extrapolation (DME): Brassey *et al.*, 2015; Erickson & Tumanova, 2000]. Current growth curves rely on DME to generate body masses of juvenile specimens, which means that juvenile body mass is dependent on the mass estimate of the adult, thereby violating the assumptions of downstream statistical analysis that generally assume independence between data points.

Body masses based on VD approaches were traditionally reported as point estimates, due to the time-consuming nature of generating a reconstruction (e.g. Colbert, 1962). However, modern virtual techniques permit for the broader incorporation of sensitivity analyses to provide a range of possible mass estimates (Hutchinson, 2012). Advantageously, the holistic perspective taken by VD approaches should generate more precise mass estimates (i.e. small possible ranges), given a range of sensitivity analyses (Bates *et al.*, 2009b; Mallison, 2010; Bates *et al.*, 2015).

Of all the VD approaches reviewed here, VD$_{MCH}$ has a major advantage over the others in that it directly integrates information from extant taxa, through the use of an expansion factor (21%; Sellers *et al.*, 2012). Importantly, the extant data on which the expansion factor is based can be used to generate an empirically derived 95% confidence interval of the expansion factor, much like those generated *via* ES approaches. This makes VD$_{MCH}$ the first volumetric approach to incorporate advantages of both ES (extant-based error bounds) and VD (full-skeleton data) into a single approach.

iii Disadvantages. Despite the advantage of precision, generating realistic, non-arbitrary error margins is a challenge and, indeed, initial VD approaches provided no error margins (Colbert, 1962; Alexander, 1989; Paul, 1997; Seebacher, 2001). Modern virtual methods allow for flexibility in

model production and reconstructions can be manipulated to give upper and lower bounds. However, this approach is often arbitrary (e.g. Gunga *et al.*, 2007; Bates *et al.*, 2009*b*). Determination of the preferred model and the amount of additional upper and lower error (e.g. ±50%) remains at the discretion of the researcher(s) and does not represent a true measure of sensitivity or its impact on reliability (as defined by Hutchinson, 2012). Nevertheless, studies using VD approaches have become increasingly more interested in determining the sensitivity of volumetric calculations to differences in tissue densities, assumed lung volume (and air sacs), and missing body parts, which should provide a better approximation of possible body mass ranges (Henderson, 1999; Hutchinson *et al.*, 2007; Bates *et al.*, 2015, 2016). Furthermore, extant data can be used empirically to inform and constrain the volume of soft tissue needed (e.g. Allen, Paxton & Hutchinson, 2009; Sellers *et al.*, 2012), thus reducing the arbitrary nature of error bounds or, as in the case of VD$_{MCH}$, to evaluate statistical uncertainty associated with the expansion factor (Sellers *et al.*, 2012). To this end, 3D-scanning techniques, in particular VD$_{MCH}$, have revealed some of the lowest differences when compared to stylopodial circumference (see Section (II.2)) and, given that they are constructed in virtual space, represent the best method for incorporating sources of uncertainty into reconstructions and assessing their effects on resultant estimates (e.g. Hutchinson *et al.*, 2007; Allen *et al.*, 2009; Brassey & Sellers, 2014; Bates *et al.*, 2015).

The holistic, specimen-focused approach taken by VD models is a distinct advantage although, in practice, it also presents a limitation. Skeletons of non-avian dinosaurs, and all extinct vertebrates, are generally incomplete (Benton, 2008; Mannion & Upchurch, 2010; Brown *et al.*, 2013), requiring the approximation of missing components. This adds to the subjectivity of the reconstruction, unless adequate efforts are made to assess the sensitivity of mass estimates to missing data (e.g. Bates *et al.*, 2015). As a result, precise and reliable VD models can only be generated for a select few taxa. This limitation, compounded with the time-consuming nature of VD methodologies (compared to ES approaches), precludes its application to large-scale macroevolutionary studies. To date, only a single large-scale study ($\mathcal{N} > 100$) of body size in the fossil record is based on estimates from VD approaches (Codron *et al.*, 2012*b*) and its data set was heavily criticized for its biased taxonomic representation and peculiar mass estimates [see Brown *et al.*, 2012 and Codron *et al.*, 2012*a* for full discussion; along with Section IV herein]. Due to this limitation, studies interested in reconstructing macroevolutionary patterns of *Baupläne* use unavoidably small sample sizes (e.g. $\mathcal{N} = 17$ in Allen *et al.*, 2013; $\mathcal{N} = 22$ in Bates *et al.*, 2016).

(2) Extant scaling approach

Extant scaling approaches are almost universally applied to estimate body mass in extinct members of crown clades, such as Mammalia and Aves (Greenewalt, 1975; Campbell Jr &

Tonni, 1983; Damuth & MacFadden, 1990; Gingerich, 1990; Campbell Jr & Marcus, 1992; Finarelli & Flynn, 2006; Butler & Goswami, 2008; De Esteban-Trivigno *et al.*, 2008; Hone *et al.*, 2008; Rinderknecht & Blanco, 2008; Millien & Bovy, 2010; De Esteban-Trivigno & Köhler, 2011; Vizcaino *et al.*, 2011; Field *et al.*, 2013; Ghizzoni, 2014). These studies use standardised measurements regressed against body mass, including various teeth measurements, skull length, total body length, and various appendicular measurements analysed either within a bivariate or multivariate framework. Dental and cranial characters, however, can be highly variable, especially when considering diverse phylogenetic histories, and can lead to biases in final mass estimates if interspecific scaling patterns are not properly assessed (Damuth & MacFadden, 1990; Millien, 2008; Rinderknecht & Blanco, 2008; Millien & Bovy, 2010). By contrast, limb measurements play a key role in weight support and are, therefore, regarded to be a better proxy for mass when considering a broader phylogenetic scope. Among limb bones, proximal elements are found to correlate most strongly with body mass (Damuth & MacFadden, 1990; Campione & Evans, 2012; Field *et al.*, 2013; Campione *et al.*, 2014). Here, we only describe ES approaches as they relate to non-avian dinosaurs, but these methods have clear implications for estimating body mass in other stem terrestrial vertebrate groups. Despite the large amount of literature on ES approaches for estimating body mass in stem taxa, only four separate, extant-based, ES approaches are proposed for estimating body mass in non-avian dinosaurs.

(a) Equations for quadrupeds

Quadrupedality is the plesiomorphic posture for all tetrapods and, as a result, extant taxa offer a disparate phylogenetic assemblage from which to sample and generate ES equations. The first quadrupedal-based equation aimed at non-avian dinosaurs was advanced by Anderson *et al.* (Anderson *et al.*, 1985, their Fig. 8A), based on a regression equation between the combined humeral and femoral circumference and body mass (*BM*) in a sample of 33 extant mammals (ranging from 0.47 kg to 5900 kg):

$$BM = 0.084 \cdot C_{\text{humerus+femur}}^{2.73} \tag{1}$$

The equation above differs slightly from that published by Anderson *et al.* (1985) as Alexander (1989) noted a minor error in its original calculation (Fig. 3). The original equation was published with an intercept coefficient of 0.078. Anderson *et al.* (1985) remains one of the most highly cited papers in dinosaur biology (over 400 times according to *Google Scholar*). Despite its criticisms (see next paragraph) it remains one of the simplest methods with which to estimate body mass in a large sample of non-avian dinosaurs and, as indicated by the results of Campione & Evans (2012), Equation 1 can likely be extended to all terrestrial quadrupeds.

Campione & Evans (2012) revisited the scaling model developed by Anderson *et al.* (1985) between combined
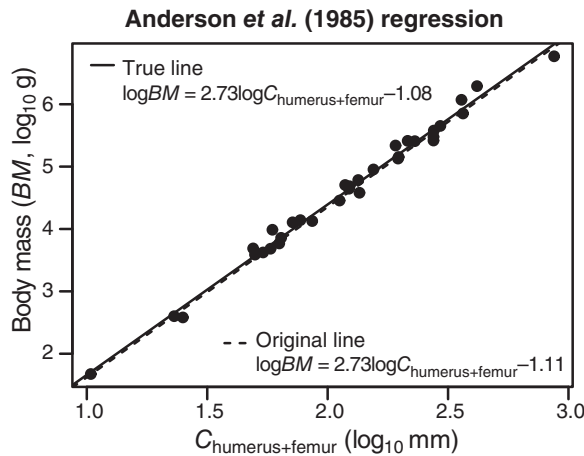
Anderson *et al.* (1985) regression

Fig 3. The regression model presented by Anderson *et al.* (1985) based on a sample of 33 extant mammals. The different line types reflect the true relationship for the 33 specimens (solid line) and the original Anderson *et al.* (1985) line (dashed line), which erred on the intercept.

stylopodial circumference and body mass, noting that the major criticisms forwarded against Anderson *et al.* (1985) had yet to be tested. These include: (*i*) differential scaling patterns between major taxonomic groups (e.g. ungulates *versus* carnivorans); (*ii*) differential scaling patterns associated with differences in gait and posture (e.g. mammals *versus* reptiles); and (*iii*) the effects of outlier data points. We tested those within the context of a large limb-measurement data set of terrestrial quadrupeds (200 mammals and 47 reptiles), the body mass of which was obtained prior to death or skeletonization and found negligible support for the aforementioned criticisms. Differences between extant quadrupeds, whether due to gait, limb posture, and/or phylogenetic history, affected scaling relationships between stylopodial length and body mass or proportional scaling patterns (i.e. length *versus* circumference), but the relationship between stylopodial circumference and body mass was highly conserved, even when comparing such disparate quadrupeds as mammals and reptiles. The relationship was particularly conserved when humeral and femoral circumferences were combined, thereby rejecting the main criticisms forwarded against Anderson *et al.*'s (1985) quadruped model. The updated equation provides a quick, consistent, and accurate method to estimate body mass in a wide range of terrestrial quadrupeds. The revised Anderson *et al.* (1985) equation from Campione & Evans (2012) is:

$$\log_{10}BM_{\text{quad}} = 2.749 \cdot \log_{10}C_{\text{humerus+femur}} - 1.104 \quad (2)$$

Anderson *et al.* (1985) and Campione & Evans (2012) tested whether a multiple regression approach, in which coefficients for humeral and femoral circumferences were calculated separately, outperformed the combined approach. These result in the following two equations, respectively:

$$\log_{10}BM_{\text{quad}} = 1.28 \cdot \log_{10}C_{\text{humerus}} + 1.46 \cdot \log_{10}C_{\text{femur}} - 0.18 \quad (3)$$

$$\log_{10}BM_{\text{quad}} = 1.78 \cdot \log_{10}C_{\text{humerus}} + 0.939 \cdot \log_{10}C_{\text{femur}} - 0.215 \quad (4)$$

Anderson *et al.* (1985) argue for the use of the combined approach, stating that the residual variance is greater when using a multiple regression approach, whereas Campione & Evans (2012) more conservatively stated that, statistically, neither approach was preferable. However, as discussed in Section II.2c.*iii*, the multiple regression approach is likely to be more affected by proportional differences between humeral and femoral circumferences than the combined approach.

Most recently, Campione (2017) presented a non-linear (quadratic) equation for the extant combined circumferences to body mass data:

$$\log_{10}BM_{\text{quad}} = -0.049 \cdot \log_{10}C_{\text{humerus+femur}}^{2} + 2.923 \cdot \log_{10}C_{\text{humerus+femur}} - 1.25 \quad (5)$$

Statistically, the quadratic equation cannot be differentiated from the linear equation (Equation 2) and it received lower likelihood support values. However, it generates body mass estimates that are more consistent with those based on VD approaches, when extrapolated to some of the largest sauropods (Bates *et al.*, 2015; Campione, 2017).

(b) Equations for bipeds

In addition to their widely cited quadrupedal equation, Anderson *et al.* (1985) present a second equation intended for bipeds. It assumes a constant scaling coefficient between bipeds and quadrupeds, where the intercept coefficient was adjusted based on the femoral circumference and body mass derived from a life-size reconstruction of *Stenonychosaurus inequalis*. This model was estimated at 41.9 kg based on water displacement methods and an assumed density of 0.9 g/cm^3. The resultant bipedal equation is:

$$BM = 0.16 \cdot C_{\text{femur}}^{2.73} \quad (6)$$

In an attempt to provide an entirely empirically based equation for estimating the body mass of extinct bipeds (including dinosaurs), Campbell Jr & Marcus (1992) published an equation based on the femoral circumference to body mass relationship in extant birds. They presented two equations, one based on an extensive specimen data set ($N = 795$):

$$\log_{10}BM = 2.411 \cdot \log_{10}C_{\text{femur}} - 0.065 \quad (7)$$

and another based on a pruned data set in which each species was represented by their mean value, weighted relative to its sample size ($N = 391$):

$$\log_{10} BM = 2.414 \cdot \log_{10} C_{\text{femur}} - 0.069 \qquad (8)$$

Given the low scaling coefficients in Equations 7 and 8 these equations result in remarkably low body mass estimates for large-bodied non-avian bipedal dinosaurs, such as 3244 kg for *Tyrannosaurus rex* compared to more traditional estimates of ~6000 kg (reviewed in Campione *et al.*, 2014).

Campione *et al.* (2014) used Equation 2, which is strictly based on quadrupeds, as the basis for a model to estimate bipedal body masses. The novel bipedal model is derived from the geometric difference in the circumference to area relationships of two circles (representing quadrupeds) to that of one circle (representing bipeds). The differences between these purely mathematical relationships are systemic and are thus used to generate a constant (α) that, when integrated into the quadrupedal formula, permits the mass estimation of bipeds from their femoral circumferences alone. In its initial derivation, assuming limb circularity and an approximate equal circumference between the humerus and femur, the correction factor was calculated at $\alpha = \sqrt{2}$. To implement the correction factor, either the femoral circumference of the biped can by multiplied by α and the product inserted into Equation 2, or through the direct modification of the elevation, resulting in:

$$\log BM_{\text{bip}} = 2.754 \cdot \log C_{\text{femur}} - 0.683 \qquad (9)$$

(c) Criteria

i Assumptions. ES approaches make one important assumption: the relationship between a skeletal measurement (e.g. femoral circumference) and body mass is assumed to be consistent between the extant taxa used to generate the model and the extinct taxa for which a body mass estimate is desired. Ideally, a proxy for body mass should be intimately related to weight support. For this reason, limb measurements are generally recognized as a better representation of size than other measurements (e.g. from teeth or crania; see Damuth & MacFadden, 1990 and chapters therein) and, of the limb bones, the stylopodium is more strongly correlated with mass than the zeugopodium (Field *et al.*, 2013; Campione *et al.*, 2014). In flying birds, however, Field *et al.* (2013) recovered a much stronger relationship between the maximum diameter of the coracoid articular facet and mass than for any other limb measurement, and Serrano *et al.* (2015) revealed the potential of multiple regression models when applying them to stem birds.

The assumption of scaling consistency between extant and extinct taxa can never be tested in the fossil record as true body mass cannot be determined. As a result, this presents an important limitation for the application of ES approaches to fossil (stem) taxa and the main reason that ES approaches are most prevalent for estimating body mass in extinct members of crown clades. Furthermore, the phylogenetic, morphological, and behavioural variation of the tetrapod skeleton means that it is unlikely that numerous ES proxies will be found that transcend all tetrapods. However, as shown by us, a comprehensive analysis comparing scaling patterns across a broad range of extant groups (including amphibians, reptiles, and mammals) can serve to justify a particular scaling pattern as consistent and, therefore, applicable to disparate *Baupläne* (Campione & Evans, 2012). A better understanding of interspecific scaling [such as in studies conducted on limb bones (Campbell Jr & Marcus, 1992; Christiansen, 1999*b*; Blob, 2000; Carrano, 2001; Campione & Evans, 2012; Doube *et al.*, 2012; Field *et al.*, 2013; Serrano *et al.*, 2015)] can, at the very least, provide insights into the strength of proxies and provide a realistic depiction of error (i.e. a standard of accuracy) when applying ES approaches to extinct taxa outside of crown clades, such as non-avian dinosaurs.

ii Advantages. ES approaches or the direct use of a proxy offer the most efficient method with which to generate large data sets of body mass (or size) for reconstructing large-scale patterns of evolution (Laurin, 2004; Carrano, 2006; Finarelli & Flynn, 2006; Hone *et al.*, 2008; Sookias *et al.*, 2012*a*, 2012*b*; Benson *et al.*, 2014, 2018; Puttick *et al.*, 2014; Kubo & Kubo, 2016) and, by extension, offer the ability to study its physiological and ecological correlates on a macroevolutionary scale and within the context of deep time and changing environments (e.g. Smith *et al.*, 2010; Sookias *et al.*, 2012*a*).

Mass estimates of fossil taxa are replete with uncertainty, questioning the accuracy of any point estimate. Accordingly, a strong estimation approach should incorporate some measure of error. In this regard, extant scaling approaches are ideal, as statistical error around scaling estimates can be easily calculated (Smith, 1980), applied to mass predictions (Campione & Evans, 2012; Serrano *et al.*, 2015), and then incorporated into subsequent analyses (Benson *et al.*, 2014, 2018). There are many statistics that can be used to represent the statistical error of a scaling equation, including the standard approach of presenting confidence intervals about the coefficients of the equation. When predicting a *y*-value, however, there are three principal equations one can use to generate intervals about an estimate. If one is interested in an average estimate of body mass at a given *x*-value (e.g. average femoral circumferences of a species), the confidence intervals (CIs) can be calculated as:

$$\log BM \pm t_{\alpha, n-2} \, s_e \sqrt{\frac{1}{n} + \frac{(x - \overline{x})^2}{(n-1) s_x^2}} \qquad (10)$$

where x is the predicting variable, $\overline{x}$ is the mean of x, n is the sample size, s_x is the standard deviation of x, s_e is the standard deviation of the residuals derived from the scaling equation, and t is the Student's t critical value given a significance value (α), and the degrees of freedom (n–2). If, however, one is interested in the confidence about a specific point estimate (based on a single measured individual), then the prediction intervals (PIs) can be calculated as:

$$\log BM \pm t_{\alpha, n-2}\, s_e \sqrt{1 + \frac{1}{n} + \frac{(x - \overline{x})^2}{(n-1)s_x^2}} \qquad (11)$$

The latter are the broadest of the intervals, as they take into consideration more extreme residual values that stray from the mean. In practice, PIs result in very broad ranges for large dinosaurs such as sauropods (Bates *et al.*, 2015; Campione, 2017; Carballido *et al.*, 2017; Benson *et al.*, 2018).

It should be noted that the vast majority of mass-estimating scaling equations (Anderson *et al.*, 1985; Christiansen & Fariña, 2004; Mazzetta *et al.*, 2004; Campione & Evans, 2012; Campione *et al.*, 2014) calculate CIs and PIs on a logarithmic scale, resulting in asymmetrical intervals when converted back to the antilog. Accordingly, Smith (1980) proposed an alternative approach based on (*i*) back-transforming all values to their antilog scale, and (*ii*) calculating a scaled residual called the per cent prediction error:

$$\%PE = \left(\frac{(\text{observed} - \text{predicted})}{\text{predicted}}\right) \times 100 \qquad (12)$$

Equation 12 can be applied across the data set, and the absolute values can be used to calculate the mean per cent prediction error for the scaling equation:

$$\overline{|\%PE|} = \frac{\sum |\%PE|}{n} \qquad (13)$$

and used to generate an expected average error range for a body mass estimate:

$$BM \pm \frac{BM \times \overline{|\%PE|}}{100} \qquad (14)$$

In practice, the error intervals generated by $\overline{|\%PE|}$ are not as inclusive as PIs, but more conservative than CIs. The principal advantage of these error ranges, and ES approaches in general, is that as long as intervals/errors are based on tested and consistent scaling equations, they are accurate, regardless of their magnitude.

iii Disadvantages. ES approaches tend to incorporate a few measurements (or even a single measurement) that can lead to biased results, especially if the assumption of scaling consistency is not met. Such issues were debated in the palaeomammalogy literature with regard to the advantages and disadvantages of craniodental and limb measurements as size proxies (see also Damuth & MacFadden, 1990 and several chapters therein; Fariña *et al.*, 1998; Millien, 2008; Millien & Bovy, 2010; Rinderknecht & Blanco, 2008). In relation to dinosaurs, femoral length is a commonly used proxy for size (Carrano, 2006; O'Gorman & Hone, 2012; Sookias *et al.*, 2012b; Puttick *et al.*, 2014). However, a detailed look at scaling patterns between femoral length and mass indicates significant differences between extant tetrapods and overall high prediction errors for femoral length (Campione &

Evans, 2012; Doube *et al.*, 2012; Field *et al.*, 2013; Campione *et al.*, 2014). Such high error values advocate caution when applying such a proxy across highly disparate *Baupläne*, such as those of dinosaurs. Proxies should always be ground-truthed and not assumed to be valid, even if commonly used and/or readily available.

The incorporation of multiple measurements within an extant-scaling framework can be achieved through multiple regression or principal component approaches (e.g. Campione & Evans, 2012; Serrano *et al.*, 2015). Serrano *et al.* (2015) successfully showed the potential of multiple regressions to lessen the effects of non-size related aspects on the proxy-to-size signal (e.g. locomotor abilities). However, caution is advised when using a multiple regression approach as a predictive model. Multiple regression models are often contingent on the taxa chosen to produce them and cannot always be generalized (see Smith, 2002 and references therin). Indeed, the latter review emphasizes the importance of applying an appropriate validation test when dealing with multiple predicting variables. The problem of non-generalizability stems, in part, from the redundancy of predictor variables (or multicollinearity) that can be addressed statistically (Serrano *et al.*, 2015). However, we emphasize a more profound issue resulting from variations in the proportional differences between the chosen predictor variables in the extant and extinct sample. For instance, unlike the vast majority of extant mammals and reptiles, non-avian dinosaurs have differentially sized forelimbs relative to the hind limb. Campione *et al.* (2014) briefly described this discrepancy within the context of generating the bipedal correction factor and showed that extant taxa (with the exception of Marsupialia and Euarchontoglires) have femur circumference to humerus circumference ratios close to one, whereas dinosaurs have comparatively much larger femora. Scaling of femur-to-humerus circumferences in quadrupeds demonstrates the disparity between quadrupedal dinosaurs (in particular ornithischians) and most extant forms (Fig. 4). The hindlimb-driven locomotion of many quadrupedal dinosaurs means that the importance of their hind limbs, in supporting their weight, will be reduced when applying a multiple regression model based on extant taxa, resulting in significantly lower body mass estimates overall. The combined use of both humeral and femoral circumferences as a single proxy, as proposed by Anderson *et al.* (1985), circumvents this issue by assuming that the relative importance of the fore and hind limb in supporting the weight of the animal is dictated by their relative contribution/size to their combined total. Application of the combined proxy revealed low per cent prediction errors for marsupial taxa (Campione *et al.*, 2014), many of which are equally hindlimb-driven and occur within the dinosaurian range of variation in circumferential scaling (Fig. 4), supporting the use of a combined proxy for dinosaurs, rather than multiple regression. However, future work could investigate the nature of the difference in fore to hind limb ratios in extant animals and test whether differences in limb sizes reflect variation in the distribution of weight in the body. Such a study
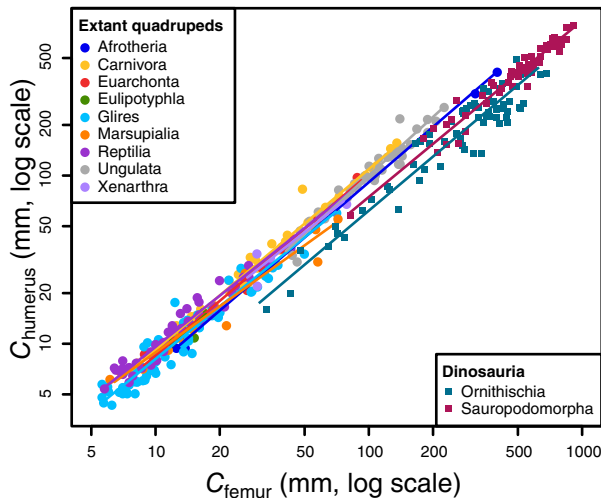
Fig 4. Differential scaling between humeral and femoral circumferences among and between extant and dinosaurian quadrupeds. Extant data taken from Campione & Evans (2012) and quadrupedal dinosaurs from Benson *et al.* (2018).

could have major implications for generating rapid estimates of the centre of mass of extinct taxa, which are currently dependent on VD approaches (Henderson, 1999; Allen *et al.*, 2013; Bates *et al.*, 2016; Macaulay, Hutchinson & Bates, 2017).

Although statistical errors can be easily calculated for ES-based body mass estimates and the non-arbitrary nature of their calculation offers an accurate representation of possible mass estimates, ES-based mass estimates tend to lack precision (Bates *et al.*, 2015). This limits the utility of ES-based approaches to define the specific mass of a dinosaur, especially when interpreting mass-specific patterns at the scale of a single specimen. However, this limitation does not encumber our ability to reconstruct long-terms patterns of evolution (Benson *et al.*, 2014, 2018) and lack of precision should not be used to question accuracy. As previously demonstrated, such issues are particularly prevalent when estimating masses of large animals >10000 kg, which occur outside the range of the extant data set, such as sauropod dinosaurs (Bates *et al.*, 2015; Campione, 2017; Carballido *et al.*, 2017; Benson *et al.*, 2018).

(3) Hybrid approaches

In an attempt to generate large body mass data sets through the use of VD approaches (e.g. Turner *et al.*, 2007), several studies proposed a suite of linear bivariate/multivariate equations that describe the relationship between skeletal measurements and estimated body masses in a (small) sample of dinosaurs (Seebacher, 2001; Christiansen & Fariña, 2004; Mazzetta *et al.*, 2004). These are here termed hybrid approaches, as they attempt to combine VD and ES approaches. The VD_{MCH} approach could also be considered

a hybrid approach, given that it integrates extant data into its conceptual framework. However, it is fundamentally different from those identified here and, hence, is described and discussed elsewhere. Seebacher (2001) provided a series of equations between total body length and body mass (estimated using the 'polynomial' method) for seven dinosaurian clades. Christiansen & Fariña (2004) and Mazzetta *et al.* (2004) presented numerous bivariate and multivariate regressions between various limb measurements and body mass in a sample of 16 theropods and 13 sauropods, respectively. Measurements in both studies include the length, circumference (or perimeter), and diameters of the femur, tibia, and fibula. These measurements were scaled against body mass estimates obtained from physical VD models (Christiansen, 1997, 1998). Most recently, O'Gorman & Hone (2012) presented equations between femoral length and body mass in a sample of 19 ornithischians, 27 sauropodomorphs, and 31 theropods in order to generate a large body mass data set to investigate body size distributions. The authors obtained body masses from various literature sources that used either ES or VD approaches.

Although hybrid approaches permit the rapid estimation of body mass of incomplete specimens (an advantage of ES approaches) based on more holistic body mass estimates (an advantage of VD approaches), the end result will propagate the errors (disadvantages) of both approaches. For instance, slight differences in the VD-based data sets used to generate the prediction models can result in major differences in body mass estimates (Campione *et al.*, 2014) and, as shown in Section (IV.3), the application of these hybrid models can lead to highly differential circumference–body mass relationships than those observed among extant non-avian tetrapods (Campione & Evans, 2012). This suggests that future research on hybrid approaches should focus on testing the consistency of scaling coefficients across taxonomic and methodological data sets, similar to that done to validate ES approaches, to better understand their utility.

III. A UNIFIED FRAMEWORK FOR DINOSAUR MASS ESTIMATION

A review of the criteria (assumptions, advantages, and disadvantages) needed to generate body mass estimates *via* VD and ES approaches highlights the fundamental differences between these techniques but does not suggest that they need to be regarded in opposition. In fact, a unified framework presents the possibility of reciprocal illumination, combining the precision of VD with the accuracy of ES. To become a standard for accuracy, ES approaches must be validated through adequate sampling and statistical testing (e.g. Campione & Evans, 2012; Field *et al.*, 2013; Serrano *et al.*, 2015) and, even so, the taxonomic scope to which an ES approach can serve as a standard for accuracy must be considered critically. For example, mammalian tooth-based ES approaches cannot be used as a standard of accuracy

for dinosaurs, given the apomorphic morphology and life history of mammalian teeth.

VD approaches (with the exception of VD$_{MCH}$) are generally validated *via* the application of a preferred reconstruction method to a suite of extant animals (e.g. Paul, 1997; Henderson, 1999; Seebacher, 2001; Gunga *et al.*, 2007; Bates *et al.*, 2009*b*). The resultant mass is then compared to either actual mass, which are unfortunately generally not available, or published averages for the particular extant species. If both masses approximate each other, the VD approach is considered validated. As identified by Bates *et al.* (2009*b*), however, the ability to model an extant taxon accurately does not inform the accuracy of the fossil reconstruction. In other words, it does not test the entire three-stage process needed to complete a volumetric reconstruction: (*i*) model creation, (*ii*) volume calculation, and (*iii*) density application ($M = DV$). When applied to an extant animal, its 'reconstruction' is not based on assumptions of soft tissue structures and density (stages 1 and 3), but rather actual knowledge of the soft-tissue proportions in these taxa, as already known by the researcher making the reconstruction. This regularly used validation technique can provide some context as to whether the approach can produce a viable result and, in the case of centre of mass estimation, can be used to assess where the centre of mass occurs in a natural position. However, faithful mass estimates of extant taxa should not be considered as a test of the soft tissue and body density assumptions made for an extinct animal and, hence, are not true validations of the accuracy of a VD estimation approach (Bates *et al.*, 2009*b*).

Alternatively, the final body mass estimates themselves could be constrained *via* an independent skeletal proxy that, following careful scrutiny and testing, offers a standard of accuracy upon which to assess a VD estimate. The use of an independent proxy to validate volumetric estimates was first presented by Motani (2001), who used a similar VD approach to that of Henderson (1999) to estimate the body masses of marine reptiles. To test the estimates, measurements of fork length (measured from the tip of the snout to the middle of the fluke), which are strongly associated with body mass in extant cetaceans, were used. Using this method, Motani (2001) was able to show that his estimates of marine reptiles occurred within the 95% prediction intervals of extant cetaceans. By contrast, other mass estimates (Massare, 1988) fell outside of these intervals.

Despite being often pitted against each other, VD and ES mass estimates are generally carried out in isolation, with little effort placed on quantifying their differences and, when compared, they are generally done so on a very limited sample (e.g. in a single table). As a result, do estimates truly differ to such a degree that one must choose between VD and ES approaches? Furthermore, confusion often exists because different specimens are used to generate masses for a particular species and, in some cases, studies are unclear as to the specimen(s) used to generate the reconstruction (e.g. some of the estimates presented by Christiansen, 1998 and Seebacher, 2001). Here we use the conserved relationship between combined humeral and femoral circumference and body mass in extant quadrupeds (Campione & Evans, 2012), the bipedal correction factor developed by Campione *et al.* (2014), and the non-linear model (Campione, 2017) as the ES standard of accuracy against which we can quantify the differences between VD and ES approaches through reciprocal illumination. We ask: do VD- and ES-based mass estimates corroborate each other? How different are they? As terrestrial tetrapods, non-avian dinosaurs are expected to follow the universal combined stylopodial circumference-to-body mass pattern and expect that, given a measurement of total stylopodial circumference, 95% of the body masses generated by a VD approach should occur within the wide 95% prediction intervals of the extant data set (in log space) and, on average, within the ±25% mean per cent prediction error (in antilog space), as determined by the extant data (Campione & Evans, 2012).

(1) Methods

(a) *Database construction*

We amassed a database of body masses generated for dinosaurs between 1905 and 2020 through various VD approaches (Fig. 5; see online Supporting information, Appendix S1). To these data, we added humeral and femoral circumferential measurements that, for the most part, correspond to the specimens used to create the volumetric reconstructions. The database consists of 626 entries, 447 of which are viable as they include both VD mass estimates and associated limb circumferential data. Of the 447 entries, we omitted the VD$_{MCH}$ minimal mass estimates (Bates *et al.*, 2016; Carballido *et al.*, 2017); these represent minimal convex hull models but, although they serve a descriptive purpose, they are not real minimal mass estimates. The final 432 entries span Ornithischia ($N = 123$), Sauropodomorpha ($N = 120$), and Theropoda ($N = 189$), most of which can then be further subdivided into 18 lower taxonomic units (Fig. 5A). Mass estimates were based on seven VD approaches, including: physical scale models (VD$_{PS}$), Graphic Double Integration (VD$_{GDI}$), polynomial (VD$_{PY}$), mathematical slicing (VD$_{MS}$), minimum convex hull (VD$_{MCH}$), photogrammetry (VD$_{PG}$), and 3D scanning (VD$_{3D}$) (Fig. 5B), and were collected from 44 studies (Fig. 5C). Note that both VD$_{PG}$ and VD$_{3D}$ generate 3D virtual models that should, theoretically, produce identical skeletal models; they differ only in the assumptions used to generate the body volumes and are separated here to reflect the research gap between these approaches (VD$_{PG}$: 1995–2002; VD$_{3D}$: 2007–2013; Fig. 5D). For 284 entries (~66% of the usable data; Fig. 5E), the femoral and humeral circumference measurements were taken from the same specimen used to generate the reconstruction (referred to as 'Perfect match'). Measurements could not be obtained for 148 specimens and, therefore, limb measurements correspond to a specimen approximately matching in size and phylogeny. All analyses were implemented with and without
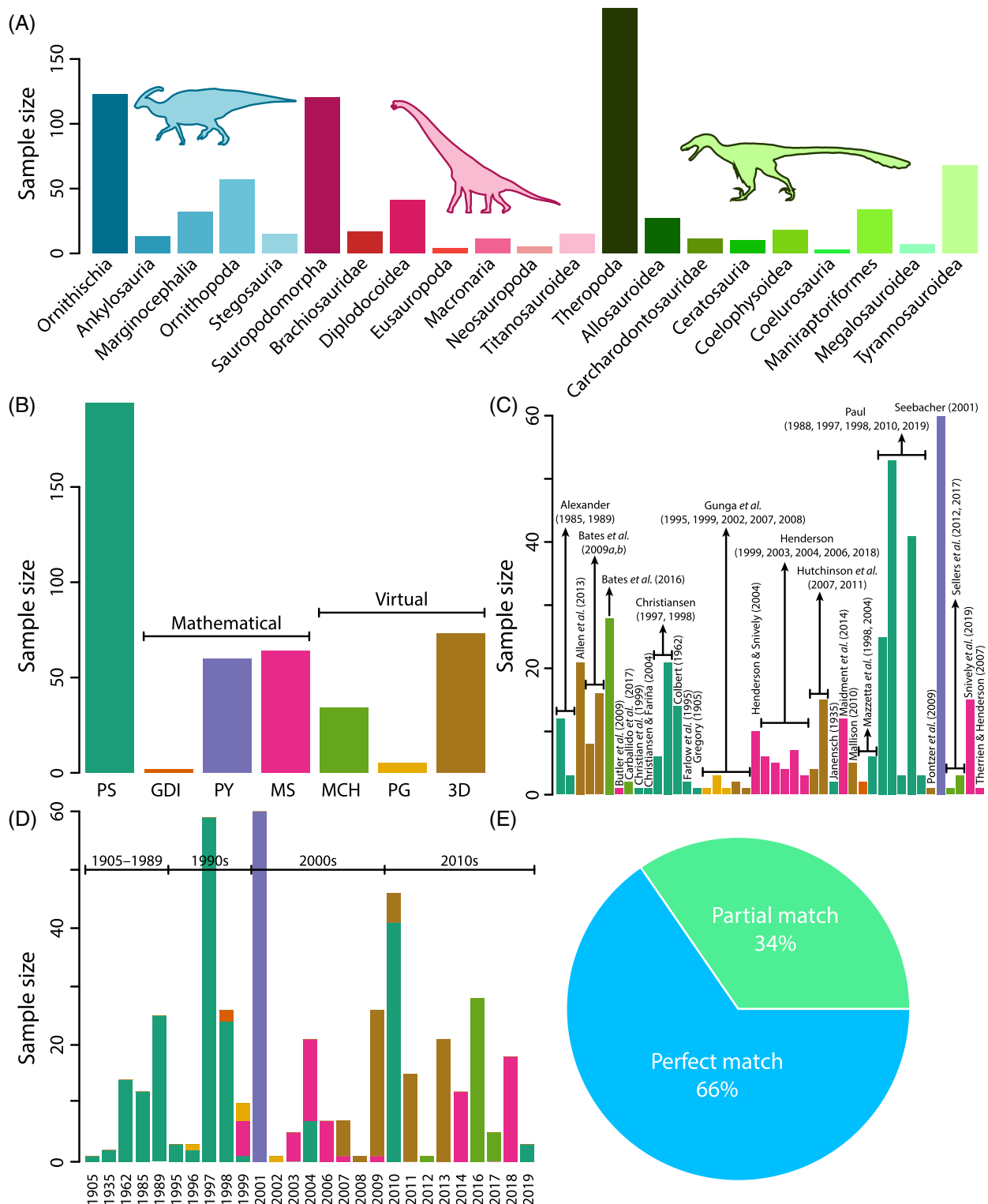
Fig 5. Plots exploring the structure of the volumetric-density (VD) database ($N = 432$). (A) Taxonomic composition. Ornithischia, Sauropodomorpha, and Theropoda represent the grand total for each major group, followed by their subdivisions. (B) Tally of the VD approaches, as applied to dinosaurs and summarized in the text (Section II.1a). PS, physical scale model; GDI, graphic double integration; PY, polynomial; MS, mathematical slicing; MCH, minimum convex hull; PG, photogrammetry; 3D, 3D scanning. (C) List and sample size provided by the studies compiled in the database. Colour scheme represents the VD approach used as in B. (D) Summary of VD-based mass estimates since 1905. Colour scheme represents the VD approach used as in B. (E) Proportion of the database for which a perfect match (blue) was obtained between the specimens used to generate the VD mass estimate and those from which the circumferential limb data were measured. Silhouettes generated by Scott Hartman and reused under Creative Commons licenses: BY-NC-SA 3.0 (*Parasaurolophus* and *Velociraptor*) or BY 3.0 (*Giraffatitian*).

the partial matches to detect if error associated with these inconsistencies affected the final interpretations.

(b) Analyses

Dinosaur limb measurements and VD masses were log-transformed and projected onto the plot of combined humeral and femoral circumference to body mass in extant quadrupeds (Campione & Evans, 2012). The extant regression and 95% prediction intervals were based on the raw, non-phylogenetically corrected equation (Campione & Evans, 2012; Equation 2). We opted for the raw regression because (*i*) the non-avian dinosaur taxa included here are not part of the phylogenetic variance–covariance matrix used to generate the phylogenetic generalized least-squares regression (Garland & Ives, 2000), and (*ii*) Akaike information criterion (AIC) values indicate that incorporating phylogeny into the equation does not increase the strength of the model ($\text{AIC}_{\text{OLS}} = -282$; $\text{AIC}_{\text{PGLS}} = -232$; Campione & Evans, 2012). Circumferential data of quadrupedal taxa are plotted in their raw form but, to visualize the bipedal taxa onto the same plot, circumference measurements were adjusted by multiplying them by a correction factor ($\alpha = \sqrt{2}$; Campione *et al.*, 2014). The deviations between the VD-based mass estimates and those expected using the ES model were measured as: (*i*) residuals (*e*), calculated as the difference between the $\log_{10}$ model mass derived from the VD approaches and the $\log_{10}$ predicted mass derived using stylopodial circumference(s),

$$e = \log_{10} BM_{\text{VD}} - \log_{10} BM_{\text{ES}} \tag{15}$$

and (*ii*) absolute per cent prediction error ($|\%PE|$), which provides a measure of error (in absolute terms) of the actual, rather than log-transformed body masses, calculated as:

$$|\%PE| = \left(\frac{|BM_{\text{VD}} - BM_{\text{ES}}|}{BM_{\text{ES}}} \right) \times 100 \tag{16}$$

Deviations were grouped and compared at a taxonomic level, subdivided into three major and 18 minor taxonomic units (Fig. 5A), and at the VD approach level (Fig. 5B). Systematic deviations were tested using both residual and %PE values, against the expected values of $\bar{e} = 0$ and $\overline{|\%PE|} = 25.6$, respectively. In addition, deviations were tested against: (*i*) size, in this case measured as the fitted values derived from the combined $\log_{10}$ stylopodial circumferences for quadrupeds and femoral circumference for bipeds and tested using a Spearman's ρ, and (*ii*) year of publication for the VD reconstruction. Approximately 87% of VD mass estimates were generated in the last three decades (Fig. 5D), preventing parametric testing. As a result, the time axis was measured as the $\log_{10}$ difference (in years) between the date of model publication and the present (i.e. 2020). Finally, the overall factors potentially driving the deviations, such as taxonomy, approach, partial matches, gait, year of

publication, and size were tested through a type-II non-parametric two-way analysis of variance (ANOVA), implemented using a Procrustes ANOVA in geomorph. ANOVAs were carried out on the standard and absolute residuals values.

All data were compiled in Excel, and plots/analyses were largely conducted using standard packages in R v4.0.1 (R Core Team, 2020), with additional functions provided by geomorph v3.3.0 (Adams, Collyer & Kaliontzopoulou, 2020), MASSTIMATE v2.0 (Campione *et al.*, 2014; Campione, 2020), smatr v3.4-8 (Warton *et al.*, 2012), and car v3.0-8 (Fox & Weisberg, 2019). MASSTIMATE v2.0 now includes a function that projects VD mass estimates onto the ES standard of accuracy (called 'MASSunite'), along with additional hybrid models not included in previous versions. All analyses and additional results are available through the R source code (see online Supporting information, Appendix S2).

(2) Results

Comparisons between the full sample ($N = 432$) and the perfect match sample ($N = 284$) reveal few differences, other than a reduction in statistical power (Table 1). Notable differences in statistical results between samples occurred in 'basal' Macronaria, where mean residual deviations lose significance in the perfect match sample, Allosauroidea, where residual and per cent prediction error deviations generate opposing results depending on the sample used, and Maniraptoriformes, where deviation values lose significance in the perfect match sample. On a broad scale, residual deviations are significantly related to the sample used, whether perfect match or the approximate samples (Actual *e*: $F_{\text{Sample}} = 8.131$, $P \approx 0.005$; Table 2). However, mean residuals for the perfect match and approximate subsamples of the data set are -0.064 and -0.041, respectively, suggesting that approximate samples do not intensify the observed deviations of VD mass estimates from that expected by stylopodial circumferences. Furthermore, significance is lost once absolute residual values are considered (Absolute *e*: $F_{\text{Sample}} = 0.503$, $P \approx 0.442$; Table 2). Combined with their associated low coefficients of determination (Actual *e*: $r^2_{\text{Sample}} = 0.012$ and Absolute *e*: $r^2_{\text{Sample}} = 0.001$), these results indicate that the source of the sample has low power to explain the deviations observed in our data set, especially when compared to taxonomy and approach (Table 2). Interpretations and discussions will henceforth utilize the complete data set, which provides a greater level of statistical power.

Of the 432 VD mass estimates tested here, approximately 73% occur within the conservative 95% prediction intervals. This suggests that there is a general pattern of consistency between stylopodial circumferential measurements and VD mass estimates (e.g. Fig. 6). Overall residual deviations are close to zero ($\bar{e}_{\text{All}} \approx -0.056 \pm 0.249$), although its effect size, compared to an expected value of 0, is large enough to reach significance (Table 1). Accordingly, VD mass estimates are, on average, significantly lower than would be expected given

Table 1. Deviations between volumetric-density and extant-scaling mass estimates, measured as the mean of the residuals ($\bar{e}$), the mean absolute per cent prediction error ($\overline{|\%PE|}$), and associated standard deviations (σ)

	Total					Perfect match				
	N	$\bar{e}$	$\sigma_{\bar{e}}$	$\overline{\|\%PE\|}$	$\sigma_{\overline{\|\%PE\|}}$	N	$\bar{e}$	$\sigma_{\bar{e}}$	$\overline{\|\%PE\|}$	$\sigma_{\overline{\|\%PE\|}}$
All	432	−0.056***	0.249	41.58***	37.5	284	−0.064***	0.237	38.21***	27.95
Taxonomy										
Ornithischia	123	−0.203***	0.274	44.64***	33.29	69	−0.214***	0.252	41.95***	24.53
Ankylosauria	13	−0.023	0.276	50.2	45.68	3	0.132[NA]	0.227	60.44[NA]	62.11
Marginocephalia	32	−0.132**	0.234	39.77*	30.14	14	−0.213**	0.234	45.1*	26.34
Ornithopoda	57	−0.234***	0.259	39.34***	20.97	41	−0.225***	0.255	39.71***	19.9
Stegosauria	15	−0.398***	0.31	75.16**	50.96	5	−0.367*	0.241	51.63	25.55
Sauropodomorpha	120	−0.084***	0.23	38.81***	28.79	74	−0.156***	0.23	39.63***	23.18
Brachiosauridae	17	−0.023	0.161	31.33	29.71	16	−0.015	0.163	31.59	30.67
Diplodocoidea	41	−0.111**	0.217	37.98**	26.98	27	−0.193***	0.21	44.22***	19.69
Eusauropoda	4	0.119[NA]	0.09	33.69[NA]	27.39	4	0.119[NA]	0.09	33.69[NA]	27.39
Macronaria	11	−0.122*	0.147	31.23	14.86	6	−0.127	0.146	32.7	13.16
Neosauropoda	5	−0.063	0.07	14	11.9	0				
Titanosauroidea	15	−0.204***	0.167	38.58*	17.83	12	−0.22***	0.161	40.75**	16.61
Theropoda	189	0.057***	0.182	41.36***	44.38	141	0.057***	0.158	35.64**	31.49
Allosauroidea	27	0.031	0.169	36.52*	24.53	11	0.141***	0.099	43.4	27.82
Carcharodontosauridae	11	0.193***	0.103	59.9*	37.49	11	0.193***	0.103	59.9*	37.49
Ceratosauria	10	−0.011	0.272	58.41	65.75	8	−0.058	0.247	47.49	54.39
Coelophysoidea	18	0.134***	0.123	42.98	39.39	16	0.111**	0.109	34.78	33.09
Coelurosauria	3	−0.25[NA]	0.061	43.32[NA]	8.15	2	−0.284[NA]	0.021	47.91[NA]	2.54
Maniraptoriformes	34	0.084*	0.23	58.75*	72.27	15	0.049	0.135	29.75	26.44
Megalosauroidea	7	−0.175*	0.166	35.8	16.01	7	−0.175*	0.166	35.8	16.01
Tyrannosauroidea	68	0.077***	0.11	29.23	29.29	63	0.081***	0.113	30.82	29.81
VD Approach										
PS	194	−0.104***	0.202	34.31***	26.45	122	−0.131***	0.196	33.85***	20.24
GDI	2	−0.132[NA]	0.131	24.45[NA]	22.48	2	−0.132[NA]	0.131	24.45[NA]	22.48
PY	60	−0.161**	0.361	52.6***	44.52	36	−0.135*	0.36	47.75**	38.36
MS	64	−0.085**	0.227	36.04**	26.12	47	−0.067*	0.217	34.87**	21.18
MCH	34	−0.051	0.162	27.69	20.96	24	−0.056	0.18	30.75	21.95
PG	5	0.217*	0.15	72.5	56.78	3	0.168[NA]	0.167	54.7[NA]	59.54
3D	73	0.164***	0.162	61.55***	55.62	50	0.137***	0.134	48.28***	36.95

Subdivision of residuals and absolute per cent prediction errors based on major taxonomic groupings of non-avian dinosaurs (Taxonomic; note that not all entries could be given a lower taxonomic classification, and thus some were only included in the higher taxonomic groupings) and the VD approach used to reconstruct the skeleton (VD Approach).
PS, physical scale model; GDI, graphic double integration; PY, polynomial; MS, mathematical slicing; MCH, minimum convex hull; PG, photogrammetry; 3D, 3D scanning.
Two-tailed t-tests were conducted on both measures of deviation against expected values of $\bar{e} \approx 0$ and $\overline{|\%PE|} \approx 25\%$, obtained from the extant terrestrial quadruped data set (Campione & Evans, 2012). Statistical tests were run at a minimum sample size of five (below which the value is considered non-applicable, or NA) and are reported as $* = 0.05 > P > 0.01$, $** = 0.01 > P > 0.001$, and $*** = P < 0.001$.

the relationship between stylopodial circumference and body mass in extant terrestrial vertebrates. Approximately 18% of data points occur outside the lower 95% prediction interval compared to 9% above (Fig. 6). Mean absolute per cent prediction errors ($\overline{|\%PE|}_{All} \approx 42\% \pm 38\%$) are also significantly above the 25.6% average calculated for extant terrestrial quadrupeds (Table 1) (Campione & Evans, 2012).

Residual deviations between stylopodial circumferential expectations and VD-based mass estimates, both in terms of their actual and absolute values, are largely related to taxonomy, followed by VD approach (Table 2). Significant differences in residual deviations are recovered between quadrupeds and bipeds where, on average, bipedal mass estimates are more consistent with the ES expectations ($\bar{e}_{bipeds} = 0.03$, $P = 0.045$; $\bar{e}_{quadrupeds} = −0.166$,

$P < 0.001$). Within the context of the general ANOVA, no significant variation in deviation values is explained by year of publication, however, there is a significant relationship with size (i.e. ES fitted values; Table 2) when absolute residuals are considered.

(a) By taxonomy

Categorizing residuals and absolute per cent prediction errors by major taxonomic groups (Fig. 6; Table 1: Taxonomy) reveals that, on average, VD-based estimates generate significantly lower body mass estimates than would be predicted by limb scaling among ornithischians (except ankylosaurs) and some sauropods. By contrast, significantly higher

Table 2. Results of the non-parametric two-way analyses of variance for actual and absolute residual deviations between stylopodial circumferential expectations and volumetric-density (VD)-based mass estimates

Source	d.f.	Sums of squares	Mean square	r^2	F-value	z-scores
Actual e values						
Taxonomy	20	3.784**	0.189	0.141	4.918	4.548
Approach	6	2.184**	0.364	0.082	9.463	3.59
Sample	1	0.313**	0.313	0.012	8.131	1.555
Gait	1	0.484**	0.483	0.018	12.568	1.728
Year	1	0.072	0.072	0.003	1.859	0.835
Size	1	0.029	0.029	0.001	0.756	0.444
Residuals	401	15.425	0.038	0.576		
Total	431	26.771				
Absolute e values						
Taxonomy	20	1.539**	0.077	0.128	3.569	3.71
Approach	6	0.67**	0.112	0.055	5.176	2.747
Sample	1	0.011	0.011	0.001	0.503	0.31
Gait	1	0.401**	0.401	0.033	18.584	1.862
Year	1	0.006	0.006	0.0005	0.26	0.05
Size	1	0.25**	0.25	0.021	11.599	1.677
Residuals	401	8.648	0.022	0.716		
Total	431	12.072				

Taxonomy is subdivided by the categories identified in the text.
Approach represents the various VD techniques.
Sample reflects the subdivisions of the data set into data for which a perfect match was obtained between the measured and estimated specimen or those where an approximate specimen was used.
Gait denotes whether the taxon was treated as quadrupedal or bipedal.
Year designates the year in which the VD reconstruction was published.
Size represents the fitted values predicted by stylopodial circumference(s).
Significance tests were run at 999 iterations and are reported as: * = $0.05 > P > 0.01$, ** = $0.01 > P > 0.001$, and *** = $P < 0.001$.

mass estimates are recovered for theropods, specifically carcharodontosaurids, coelophysoids, and tyrannosauroids. The largest deviations between VD-based mass estimates and those expected based on stylopodial circumferences were recovered for stegosaurs ($\bar{e}_{\text{Stegosauria}} \approx -0.398 \pm 0.31$; $\overline{|\%PE|}_{\text{Stegosauria}} \approx 75\% \pm 51\%$), with most estimates falling below the 95% prediction intervals and above the 25.6% extant-based average absolute per cent prediction error (Fig. 6B, C). Highly consistent results were obtained for brachiosaurids ($\bar{e}_{\text{Brachiosauridae}} \approx -0.023 \pm 0.161$; $\overline{|\%PE|}_{\text{Brachiosauridae}} \approx 31\% \pm 30\%$), with almost all estimates occurring with the 95% prediction intervals (Fig. 6B). Similarly, all non-titanosauroid/diplodocoid neosauropods and non-neosauropod eusauropods occur within the 95% prediction intervals, despite their small sample sizes.

(b) By approach

Subdivision of residuals and absolute per cent prediction errors by the type of VD approach employed (Fig. 7; Table 1: VD Approach) indicates that traditional physical reconstructions and mathematical approaches (VD_{PS}, VD_{PY}, VD_{MS}, and qualitatively VD_{GDI} in Table 1) generate significantly lower body mass estimates than that expected by stylopodial circumferences. By contrast, initial attempts at virtual reconstruction (VD_{PG} and VD_{3D}) generated, on average, higher estimates for given stylopodial circumferences. The largest deviations were recovered for VD_{PY}, VD_{PG}, and VD_{3D}, which exhibited the highest prediction

errors (approximately 53%, 73%, and 62%, respectively); VD_{PY} generated some of the most extreme deviations, which are higher at smaller body sizes ($\rho_{\overline{|\%PE|}:\text{PY}} = -0.258$, $P = 0.047$; Table 3; Fig. 8B). Notable outliers were also generated by certain VD_{PS} and VD_{MS} mass estimates (Fig. 7) and their residual values were significantly negatively related to the fitted values of the ES model (Table 3).

Of all the methodologies, the VD_{MCH} approach generated the lowest deviations relative to that expected from stylopodial circumferences ($\bar{e}_{\text{MCH}} \approx -0.051 \pm 0.162$; $\overline{|\%PE|}_{\text{MCH}} \approx 28\% \pm 21\%$; Table 1). Almost all mass estimates derived using the VD_{MCH} approach on dinosaurs fall within the 95% prediction intervals and both residual and absolute per cent prediction errors cannot be differentiated from those derived from the extant data. Importantly, this consistency is achieved by omitting the estimates derived by the minimum envelope of the VD_{MCH} approach. When the minimum convex hull mass estimate is included, the average VD_{MCH} residual deviation is significantly below that of extant taxa ($\bar{e}_{\text{MCH w/minimal}} \approx -0.109 \pm 0.178$, $P \ll 0.001$) and the absolute per cent prediction error is just shy of significance ($\overline{|\%PE|}_{\text{MCH w/minimal}} \approx 32\% \pm 21\%$, $P = 0.073$).

(c) Relationship to size and year

Deviations as a function of size indicate some significant negative relationships between both deviation metrics and size (Table 3). These trends were particularly prevalent among
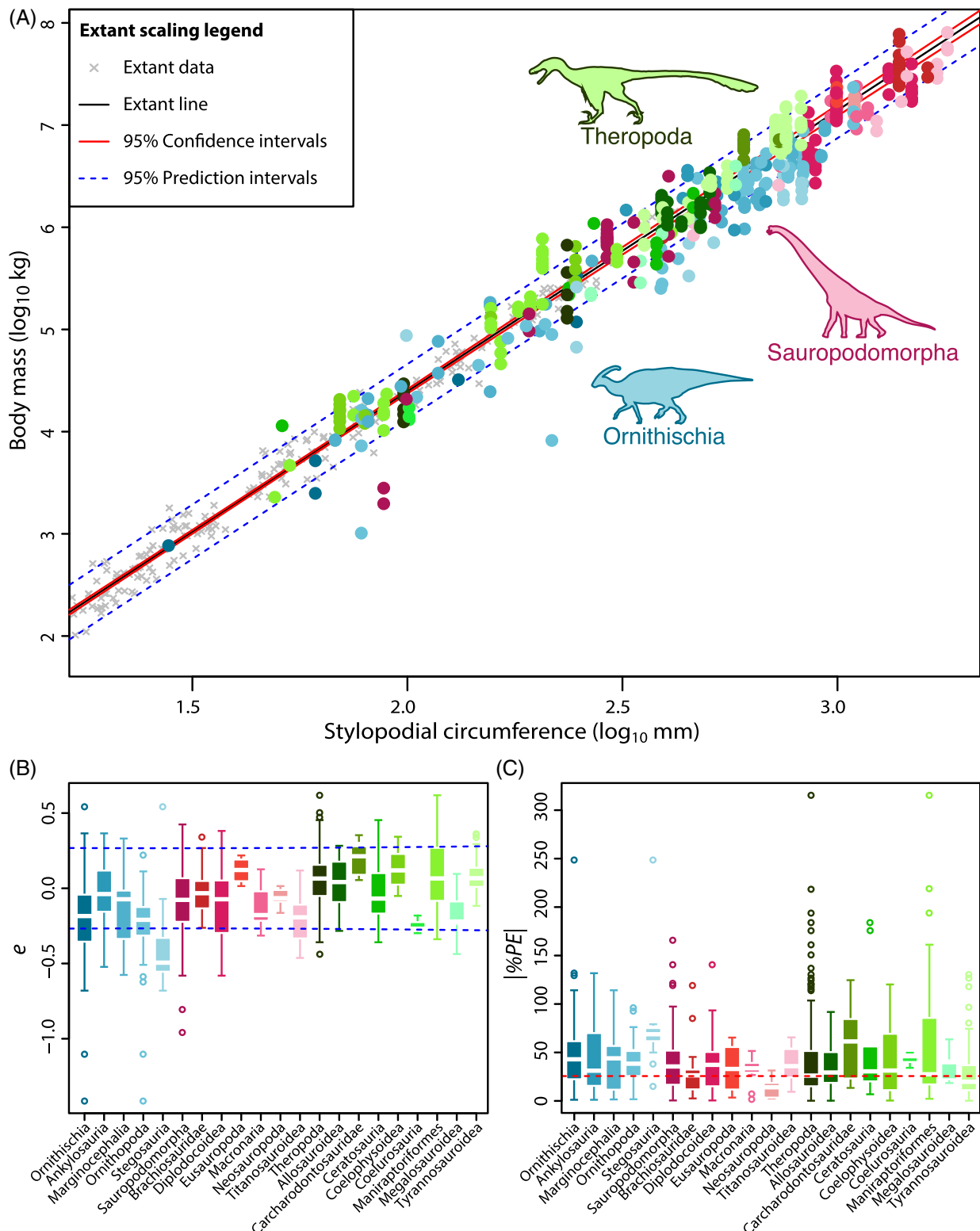
Fig 6. Stylopodial circumferences and volumetric-density (VD) mass estimates subdivided by major taxonomic designations. The *x*-axis represents stylopodial circumference, which refers to the combined humeral and femoral circumferences for assumed quadrupeds and femoral circumference for assumed bipeds (see Appendix S1 for full list of gait designations). (B, C) Deviations calculated from A represented as the residuals (*e*; B) and absolute per cent prediction errors ($\overline{|\%PE|}$; C) along with associated 95% prediction errors (blue dashed lines) and the mean per cent prediction error (red dashed line) calculated from extant data. Most VD mass estimates (~73%) occur within the 95% prediction intervals (A and B). Silhouettes generated by Scott Hartman and reused under Creative Commons licenses: BY-NC-SA 3.0 (*Parasaurolophus* and *Velociraptor*) or BY 3.0 (*Giraffatitian*).
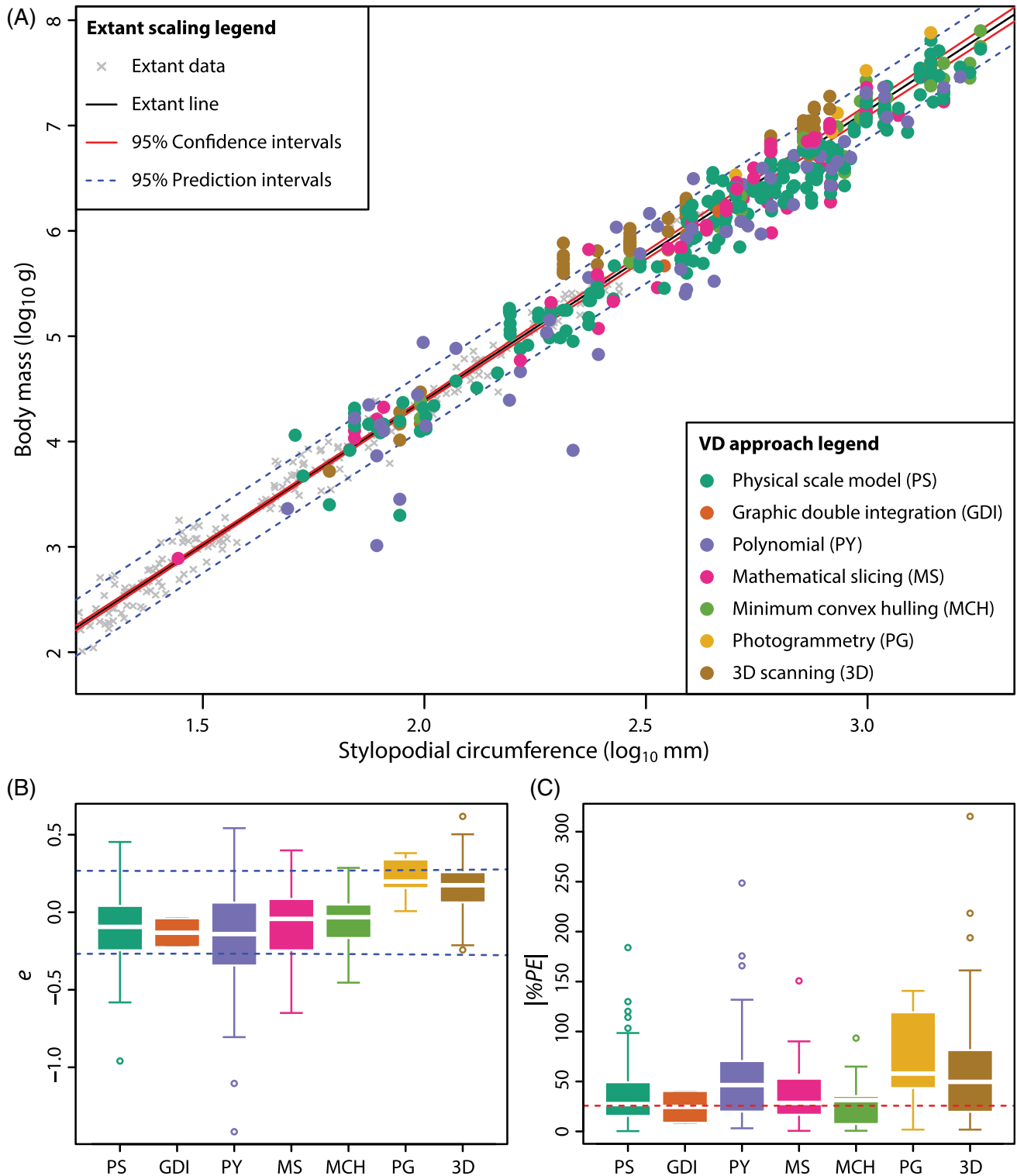
Fig 7. Stylopodial circumferences and volumetric-density (VD) mass estimates subdivided by VD approach. Other details as in Fig. 6. Elevated deviation values calculated from absolute per cent prediction errors are present in the PY and PS approaches, especially at small sizes, whereas low deviation values were recovered for the VD$_{MCH}$ approach.

traditional physical reconstructions ($\rho_{\bar{e}:PS} = -0.194$, $P = 0.007$) and mathematical approaches ($\rho_{\bar{e}:MS} = -0.309$, $P = 0.013$; $\rho_{\overline{|\%PE|}:PY} = -0.258$, $P = 0.047$). In general, these

approaches present larger deviations among smaller taxa (~10 kg or less), rather than larger taxa, as shown in Fig. 8A, B. Virtual-based reconstruction approaches revealed no

Table 3. Non-parametric (Spearman's ρ) statistics summarizing the relationship between deviations metrics and size (fitted values of the extant scaling models) or year of volumetric-density publication (VD$_{Annum}$) (see also Fig. 7)

Source	$\mathcal{N}$	$\rho_{\bar{e}}$	$\rho_{\overline{|\%PE|}}$
versus fitted			
All	432	−0.17***	−0.073
PS	194	−0.194**	−0.024
PY	60	−0.068	−0.258*
MS	64	−0.309*	0.188
MCH	34	−0.193	−0.015
PG	5	0.6	0.600
3D	73	−0.06	−0.179
versus VD$_{Annum}$			
All	432	0.155**	0.029

PS, physical scale model; PY, polynomial; MS, mathematical slicing; MCH, minimum convex hull; PG, photogrammetry; 3D, 3D scanning. Graphic double integration (GDI) was not included due to low sample size ($N_{GDI} = 2$).
Statistical tests are reported as: * = $0.05 > P > 0.01$, ** = $0.01 > P > 0.001$, and ***$P < 0.001$.

significant trends with size, although it is worth noting that only a few taxa <10 kg have been estimated through these, more modern, approaches.

A significant trend was noted between residuals and year of VD publication ($\rho_{\bar{e}} = 0.155$, $P = 0.001$; Fig. 8C; Table 3) suggesting a general trend from lower body mass estimates generated by early VD attempts to more recent attempts generating higher estimates relative to stylopodial circumferences. However, when absolute per cent prediction errors are used, the relationship is not significant ($\rho_{\overline{|\%PE|}} = 0.029$, $P = 0.547$) indicating that overall deviations, relative to stylopodial circumferences, have remained largely static over the 115 years sampled in this study. The range in deviation values fluctuated appreciably through time (Fig. 8C, D), with the largest deviations occurring in 1997, 2001, 2009–2011, and 2013.

IV. DISCUSSION

(1) On accuracy and precision

The biological importance of body size (and by extension mass) has led to over 100 years of scientific interest in generating realistic body mass estimates of fossil vertebrates, in particular, for Mesozoic dinosaurs. As a predicted parameter, the scientific concepts of accuracy and precision are highly relevant to the discussion of body mass estimation and both are referred to frequently in related literature (e.g. Bates *et al.*, 2009*a*; Hutchinson *et al.*, 2011; Campione & Evans, 2012; Hutchinson, 2012; Sellers *et al.*, 2012; Paul, 2019). It is, therefore, important to consider and review the distinction between these concepts, as expressed by the Joint Committee for Guides in Metrology (JCGM), but see also Hopkins (2018).

Accuracy is defined as "Closeness of agreement between a measured quantity value and a true quantity value of a measurand." (JCGM 200:2012, section 2.13).

With regards to body mass estimation, the *measurand* is body mass, of which the *true quantity* is unknown in the fossil record. Accuracy must, therefore, be interpreted and approximated through external means. Applying a VD approach to an extant animal is not a test of accuracy to estimate a body mass generated *via* a reconstructed fossil, as it does not test an investigator's ability to generate a reconstruction from limited fossil material (Bates *et al.*, 2009*b*). Rather, the application of VD approaches to extant organisms only validates the accuracy with which a method can calculate the volume of a complex object. In comparison, following careful validation (e.g. Campione & Evans, 2012; Field *et al.*, 2013; Serrano *et al.*, 2015), the accuracy of an ES approach to estimate body mass can be determined as it makes a direct comparison between a measured quantity (e.g. stylopodial circumferences) and the measurand (body mass). A well-validated linear model (i.e. one of minor/insignificant differential scaling) provides accuracy, regardless of how broad a prediction error it may generate. Whether the accuracy of such a linear model can be extended to the fossil record, however, must be carefully interpreted; hence the importance of validating ES relationships across a wide range of *Baupläne* and phylogenetic histories.

Precision is defined as "Closeness of agreement between indications or measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions." (JCGM 200:2012, section 2.15).

In comparison to accuracy, precision, or rather imprecision, which can be measured by typical measures of variance, standard deviations, and confidence intervals (JCGM 200:2012), can be high among ES approaches (Campione & Evans, 2012; Bates *et al.*, 2015; Campione, 2017; Carballido *et al.*, 2017; Benson *et al.*, 2018). This is particularly true of the non-linear antilog values that increase in variance as a function of size (Cawley & Janacek, 2009) and lead to proportionally higher error ranges at larger body masses than at lower body masses (Benson *et al.*, 2018; Campione, 2017). In comparison, VD mass estimates, which ideally incorporate a greater extent of observational data on the anatomy of a reconstructed fossil animal, have the advantage that (*i*) they can accommodate variation in reconstruction parameters (e.g. Hutchinson *et al.*, 2007), a measure of sensitivity akin to replication, and (*ii*) they should generate a more precise body mass estimate range, albeit of unknown accuracy. Precision, as defined here, is quantitative and, therefore, differs from the concept of reliability, which relates to the qualitative resilience of a conclusion to sensitivity analyses (Hutchinson, 2012), although they are evidently related.

It is important that these concepts, especially those of accuracy and precision, are not confounded; ES approaches (specifically that based on stylopodial circumferences) provide an accurate error range within which more precise VD-generated mass estimates should occur. The realization that these approaches have different conceptual strengths
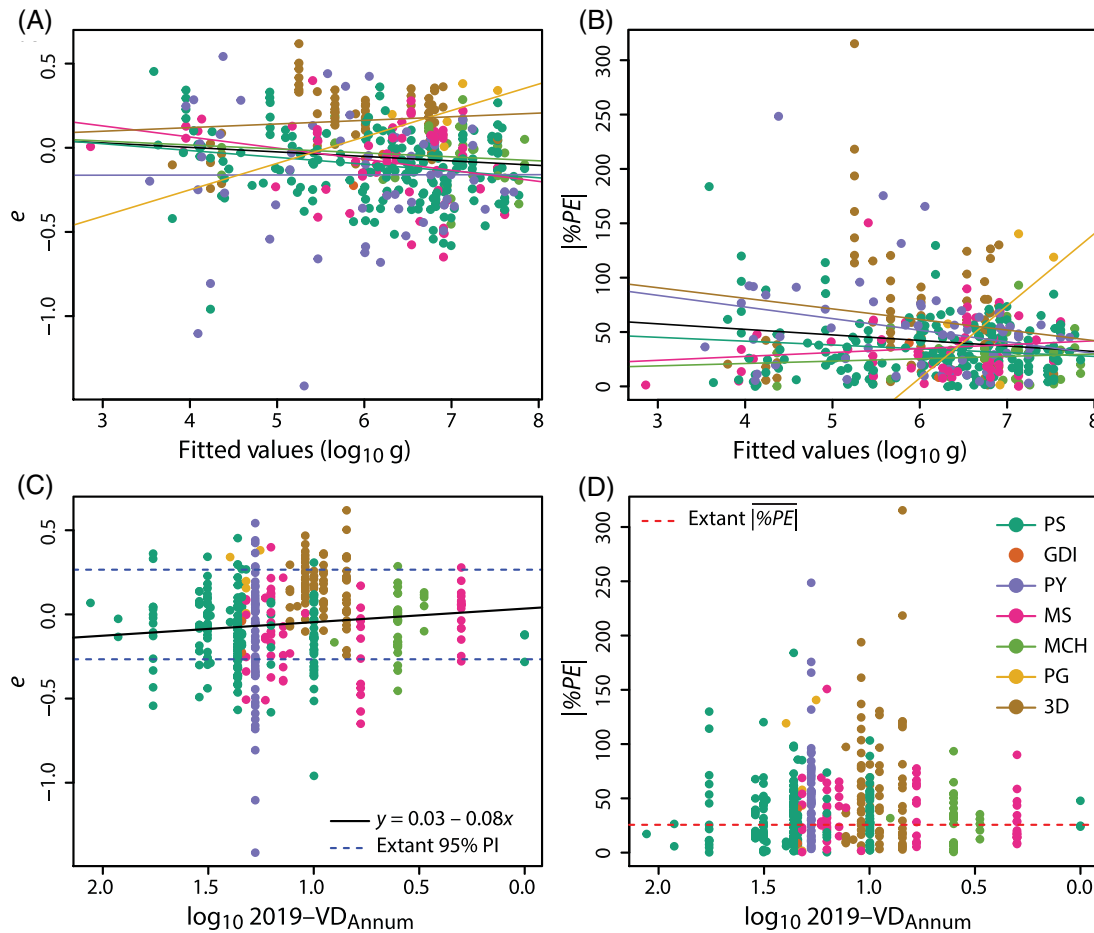
Fig 8. Deviations as a function of size and year of volumetric-density (VD) publication. Data are colour-coded according to VD approach: PS, physical scale model; GDI, graphic double integration; PY, polynomial; MS, mathematical slicing; MCH, minimum convex hull; PG, photogrammetry; 3D, 3D scanning. (A) Residual deviations against fitted values of the extant-scaling (ES) models. (B) Absolute per cent prediction error against fitted values of the ES models. Note the greater incidence of larger deviations (extreme outliers) at small sizes. (C) Residual deviations and (D) per cent prediction error against the $\log_{10}$ difference between 2020 and the year the VD reconstruction was published (VD$_{Annum}$). Red dashed line represents the ~25% mean per cent prediction error and the blue dashed lines represent the 95% prediction intervals, both recovered from the extant quadrupedal data set. Fitted lines represent the parametric ordinary least-squares line for visualization purposes only, statistics were determined through non-parametric means (Table 3).

underscores the need for them to be used in tandem, such as through the unified framework presented here, rather than in opposition.

(2) The relationship between VD and ES

Our study reviewed the various body mass estimation methods applied to non-avian dinosaurs. Through a series of quantitative comparisons, we propose an integrated methodological framework that uses the highly conserved relationship between combined stylopodial circumferences and body mass in extant terrestrial tetrapods as an accurate representation of error, within which more precise mass estimates obtained *via* VD approaches should occur. Consistency should be reached between the methods; however, should inconsistencies arise, they will require *ad hoc*

interpretations of apomorphic limb scaling patterns or a re-evaluation of the VD reconstruction assumptions. Achieving consistency, or justifying the lack thereof, is particularly important given the subsequent impact of body mass estimates to address a wide range of fundamental biological questions in the fossil record (e.g. Pontzer *et al.*, 2009; Hutchinson *et al.*, 2011; O'Gorman & Hone, 2012; Benson *et al.*, 2014, 2018; Bates *et al.*, 2016; Henderson, 2018).

The unified framework is based on the assumption that non-avian dinosaurs follow the plesiomorphic stylopodial circumference–body mass pattern observed in living non-avian tetrapods (Campione & Evans, 2012). Although this assumption is challenged by apomorphic and differential scaling patterns in birds (Campbell Jr & Marcus, 1992; Campione *et al.*, 2014) and we do not yet understand the femoral circumference–body mass scaling relationship in non-

avian avialians, the deeply nested plesiomorphic tetrapod scaling pattern (Campione & Evans, 2012), the observations that avian femora are biomechanically distinct from those of non-avian dinosaurs, including from those of some of the earliest birds (Carrano, 1998; Carrano & Biewener, 1999; de Margerie *et al.*, 2005), and that the derived cross-sectional distribution of bone tissues lies away from the centre of the cross section (Doube *et al.*, 2012), all support the hypothesis that the limbs of birds and non-avian dinosaurs are likely to be under different evolutionary constraints. Accordingly, given a particular stylopodial circumference, 95% of the body masses obtained *via* VD reconstructions of, at least most, non-avian dinosaurs should occur within the 95% prediction intervals presented in Campione & Evans (2012) and Campione *et al.* (2014).

The results presented here indicate that the majority of VD-based mass estimates (~73%) are consistent with those expected by the preferred ES approach. The incidence of approximately 27% data points outside the 95% prediction intervals suggests that some reconstructions are significantly inconsistent relative to the sizes of their limbs, requiring *ad hoc* hypotheses. These could include the revision of VD-reconstruction assumptions, the discussion of derived biological factors leading to apomorphic limb proportions, or a combination thereof. For instance, Campione & Evans (2012) noted potential apomorphic scaling patterns among tetrapods with extreme burrowing behaviours, such as moles (Talpidae), which occurred outside the 95% prediction intervals. Such ecologies were proposed for certain dinosaurs (Varricchio, Martin & Katsura, 2007; Longrich, 2010) and may well explain some of the negative deviations noted among small taxa, especially among ornithopods and possibly marginocephalians.

Of the dinosaur VD estimates recovered outside the 95% prediction intervals (115 in total; see Appendix S1: outliers; top-10 shown in Table 4), 11 could be explained by the original use of juvenile specimens to generate the VD-based mass estimates. ES approaches represent interspecific patterns expressed by skeletally mature individuals (i.e. adults), which need not be followed ontogenetically (e.g. Erickson & Tumanova, 2000; Kilbourne & Makovicky, 2010). Given a more isometric ontogenetic trajectory between stylopodial circumferences and body mass, the interspecific relationship used here may over-estimate the body mass of immature individuals; although, it should be noted that its use may provide a rough range estimate for the typical mass of an associated adult (Campione, 2017). However, the extent of the remaining deviations, many of which exceed those of talpids (Campione & Evans, 2012), and the observation that over half of the outliers relate to specific VD approaches – either the polynomial approach of Seebacher (2001) or the physical reconstructions generated by Paul (1997) – suggests that the VD reconstructions need re-examination. For instance, Paul (1988, p. 235) acknowledges that his models were built in a "lean condition", which likely contributes to the significantly lower residual deviations recovered among physical scale models (Table 1). In general, the polynomial approach

exhibited some of the highest deviations and seven of the top-10 most extreme outliers (Tables 1 and 4; Fig. 7), which may reflect the over-generalised form in which the approach corrects for the width of an animal (see Section II.1*a*; Seebacher, 2001). The polynomial approach suggests particularly low mass estimates for small ornithopods (e.g. *Thescelosaurus neglectus*, *Gasparinisaura cincosaltensis*, *Tenontosaurus tilletti*, and *Cumnoria prestwichii*; Table 4). The overall trend of certain VD approaches to exhibit higher deviations than expected at small body sizes (Table 3) also requires further scrutiny, since a disparity between smaller and larger body sizes can lead to seemingly amplified differences in the subsequently inferred biology of these animals, including their community structure (Codron *et al.*, 2012*b*, who relied heavily on estimates derived from the polynomial approach) and metabolic rates (Seebacher, 2003; Pontzer *et al.*, 2009; Grady *et al.*, 2014).

The observation of large deviations in Stegosauria, with virtually all VD estimates occurring below the 95% prediction limits (Fig. 6), remains unclear. Notable extreme examples include: *Kentrosaurus aethiopicus*, *Scelidosaurus harrisonii* (a non-eurypodan thyreophoran), and *Stegosaurus stenops* (Table 4). Bates *et al.* (2015), Brassey *et al.* (2015) and Campione (2017) all emphasized the importance of understanding the ontogenetic status of the specimen when assessing discrepancies between VD- and ES-based approaches. However, the reconstructions of all three taxa were based on skeletally mature specimens, suggesting that ontogeny cannot explain the large deviations. The lack of specimen consistency between the VD and ES approach could be a factor, given that only five of the 15 VD mass estimates could be perfectly matched, although even the perfectly matched specimens exhibit significantly high deviation values (Table 1). Further research could focus on the sensitivity of VD-based mass estimates of stegosaurs to inherent assumptions about body proportions and investigate the possible impact of armour on limb robusticity. However, although armour was hypothesised as a factor driving differential limb scaling in stegosaurs (Maidment *et al.*, 2012), it cannot solely be the cause of the observed body mass deviations because armoured ankylosaurs exhibited comparatively low deviation values (Fig. 6; Table 1). It is therefore possible that, unlike ankylosaurs whose body envelopes can be adequately constrained by their heavy armament (e.g. Brown *et al.*, 2017), low body mass estimates among stegosaurs using VD approaches reflect a lack of understanding of the quantity of soft tissue needed for their reconstruction.

Theropods exhibited the lowest overall errors, although still significantly different from extant tetrapods (Table 1). Certain data points exhibit particularly high deviations above the 95% prediction intervals, including: *Deltadromeus agilis*, *Herrerasaurus ischigualastensis*, *Struthiomimus sedens*, and *Masiakasaurus knopfleri* (see full list in Appendix S1). Deviations could be explained by the incomplete nature of the specimens used to model these taxa – specimens of *D. agilis* and *M. knopfleri* include partial limbs and tail but are missing data on the chest cavity and neck (Sereno *et al.*, 1996; Carrano,

Table 4. The most extreme (top 10) deviations between volumetric-density (VD)-based mass estimates and that expected by stylopodial circumferences (see Appendix S1: outliers, for full list)

Clade	Species	VD specimen	VD approach	VD estimate (kg)	ES specimen	ES estimate (kg)
Ornithopoda	*Thescelosaurus neglectus*	USNM 7757	PY	7.9	USNM 7757	152–258
Ornithopoda	*Gasparinisaura cincosaltensis*	MUCPv-208	PY	0.98	MUCPv-219	9.2–15.6
Sauropodomorpha	*Eoraptor lunensis*	PVSJ 512	PS	1.9	PVSJ 512	12.8–21.7
Sauropodomorpha	*Eoraptor lunensis*	PVSJ 512	PY	2.7	PVSJ 512	12.8–21.7
Stegosauria	*Kentrosaurus aethiopicus*	HMN Mount	PY	321	HMN Mount	1144–1932
Stegosauria*	*Scelidosaurus harrisonii*	Galton (1997)	PY	65	BMNH R1111	219–370
Stegosauria	*Stegosaurus stenops*	USNM 4934	MS	1821	AMNH Mount	6031–10188
Ornithopoda	*Tenontosaurus tilletti*	OU 11	PY	243	OU 11	758–1280
Maniraptoriformes	*Struthiomimus sedens*	BHI 1266	3D	741	AMNH 5375	133–224
Ornithopoda	*Cumnoria prestwichii*	OUM J.3303	PY	268	YPM 1880	771–1303

Extant scaling (ES) mass estimate range is based on the 25.6 mean per cent prediction error and reflects the expectation, on average, given the associated circumferential measurements.
*_Scelidosaurus harrisonii_ is a non-eurypodan thyreophoran, but was included within Stegosauria for the purposes of our analyses.
PY, polynomial; PS, physical scale model; MS, mathematical slicing; 3D, 3D scanning.

Loewen & Sertich, 2011) and *S. sedens* was based on a 50% complete specimen (Bates *et al.*, 2009*b*) – or on a chimeric reconstruction, as in the case of *H. ischigualastensis*). Overall, however, VD and ES mass estimates of theropods are highly consistent compared to sauropodomorphs (unless a quadratic ES model is assumed; see Section IV.4) and, especially, compared to ornithischians (Figs 6 and 9; Table 1).

Sauropodomorpha, specifically Sauropoda, include species that attained the largest body sizes of any terrestrial vertebrate and, unsurprisingly, attempts to infer their body masses extend back to the early 20th century, including the first ever dinosaur mass estimate (Gregory, 1905). However, many sauropod mass estimates necessitate extrapolations beyond extant observations (Campione, 2017). Obvious extrapolations are evident when using ES approaches (e.g. circumferences are outside the range of those used to build the ES model). Obscure extrapolations are those that result from our necessity to ground fossil reconstructions on testable observations made from the extant (validation; as per Hutchinson, 2012). In spite of these extrapolations, estimating the sizes of sauropods permits the investigation of physiological and life-history attributes within the context of gigantism and the documented upper limits of terrestrial size (see Klein *et al.*, 2011; Sander *et al.*, 2011; Benson, 2018; Benson *et al.*, 2018).

Our results suggest that VD-based mass estimates of sauropods are generally consistent with stylopodial circumference expectations, especially when adopting a non-linear quadratic function. Extreme deviations are noted in the early sauropodomorph, *Eoraptor lunensis* and *Amargasaurus cazaui*. The VD reconstructions of *E. lunensis* were based on a highly complete, putatively mature specimen (PVSJ 512; Sereno, Martínez & Alcober, 2012) and generated through either physical (Paul, 1997) or mathematical (polynomial; Seebacher, 2001) means. Both approaches tend to generate

higher deviations at smaller size (Table 3), which may explain the observed deviations. Four of the five VD mass estimates of *A. cazaui* occur outside the 95% prediction intervals. These estimates were generated through either physical (Paul, 1997; Mazzetta *et al.*, 2004) or minimum convex hulling means (Bates *et al.*, 2016) based on a partial skeleton of unknown growth stage (MACN-N 15). Although the reason for the deviation remains unclear, the unique morphology of the cervical vertebrae in *A. cazaui* suggests that reconstructions should incorporate a broader range of sensitivity analysis.

Of the most commonly estimated gigantic sauropods, our results support lower estimates (25257–42665 kg; Fig. 9E) for *Giraffatitan brancai* (Anderson *et al.*, 1985; Alexander, 1989; Paul, 1997; Henderson, 1999; Seebacher, 2001; Gunga *et al.*, 2008; Sellers *et al.*, 2012; Bates *et al.*, 2015) over gigantic historical estimates (~70 tonnes; Christian, Heinrich & Golder, 1999; Colbert, 1962; Gunga *et al.*, 1995). By contrast, we support higher estimates for *Apatosaurus louisae* (CM 3018; 30370–51303 kg; Fig. 9D), more consistent with recent estimates generated *via* convex hulling (Bates *et al.*, 2016), but substantially higher than many estimates generated through physical or mathematical means (between 15 and 20 tonnes; Christiansen, 1997; Henderson, 2006; Henderson & Snively, 2004; Mazzetta *et al.*, 2004; Paul, 1997).

(3) Using hybrid approaches

In an effort to generalise VD-based estimates to the much larger sample of incompletely known dinosaurs, several studies proposed scaling approaches that describe the relationship between skeletal measurements and VD-based body masses (Seebacher, 2001; Christiansen & Fariña, 2004; Mazzetta *et al.*, 2004; Therrien & Henderson, 2007; O'Gorman &
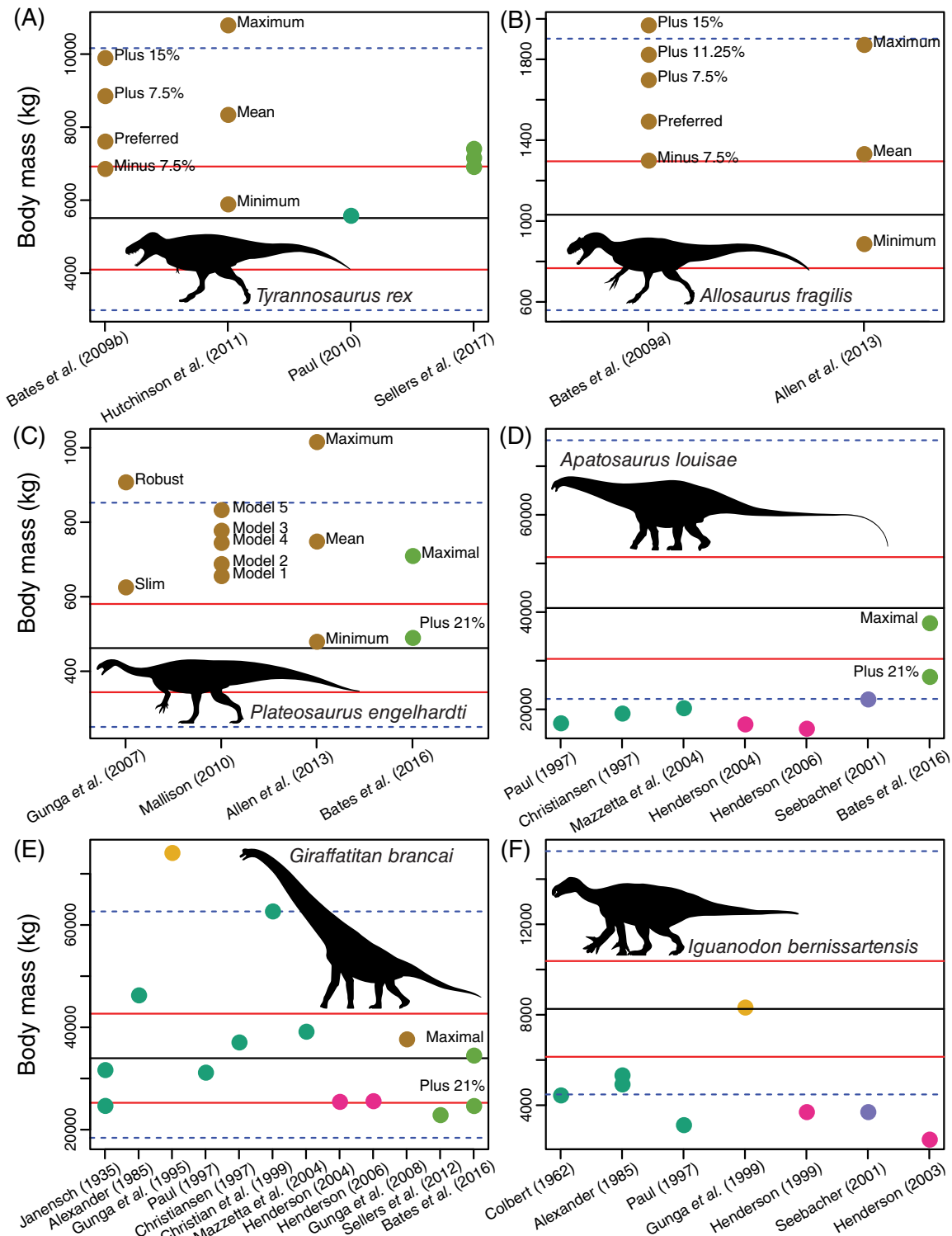
Fig 9. Body mass estimates of commonly modelled dinosaurs. Horizontal lines depict our extant-scaling (ES) standard of accuracy, based on the stylopodial circumference(s), given as the points estimate (black line), mean per cent prediction error (red lines), and 95% prediction intervals (blue dashed lines). Volumetric-density approaches are colour-coded as in Fig. 5. (A) *Tyrannosaurus rex* (ES estimate based on BHI 3033), (B) *Allosaurus fragilis* (ES estimate based on MOR 693), *Plateosaurus engelhardti* (ES estimate based on AMNH 2107), *Apatosaurus louisae* (ES estimate based on CM 3018), *Giraffatitan brancai* (ES estimate based on HMN SKII), and *Iguanodon bernissartensis* (ES estimate based on IRSNB R51). Silhouettes generated by Scott Hartman and reused under Creative Commons licenses: BY-NC-SA 3.0 (A–D) or BY 3.0 (E, F).

Hone, 2012). These equations may be applied to large-scale evolutionary studies (e.g. Turner *et al.*, 2007; Dececchi & Larsson, 2013) but, as for ES approaches, their generalizability relies on the baseline mass-estimate data set on which they were derived. As previously demonstrated, the polynomial method of Seebacher (2001) exhibits large deviations relative to those expected from limb circumferences when compared to other VD approaches and these deviations may be size dependent. It is, therefore, important that large deviations be considered on a case-by-case basis and differential scaling tested if VD mass estimates are to be used to generate general mass-estimation equations.

Here we examine four studies that proposed hybrid approaches within the comparative framework proposed herein. (*i*) Christiansen & Fariña (2004) proposed a set of 39 bivariate or multiple regression equations for theropods. The data included a broad range of hind-limb bone measurements and body masses generated through physical reconstructions (originally created by Christiansen, 1998). These equations were subsequently used to estimate the body masses of *Giganotosaurus* and *Carnotaurus* (Mazzetta *et al.*, 2004), along with numerous small theropods on the avian lineage (e.g. Turner *et al.*, 2007; Zhang *et al.*, 2008; Xu *et al.*, 2011). (*ii*) A set of similar bivariate equations for sauropods were proposed by Mazzetta *et al.* (2004). (*iii*) O'Gorman & Hone (2012) developed three femur length-to-body mass equations for theropods, sauropodomorphs, and ornithischians, with which to explore dinosaur body-size distributions. (*iv*) Finally, Therrien & Henderson (2007) proposed a skull length-to-body mass equation based on body masses estimated through mathematical slicing. To explore these approaches, the limb-based hybrid models (1–3) were applied to the dinosaur limb data of Benson *et al.* (2018) and the skull-based hybrid model (4) reused the data presented in Table 2 of Therrien & Henderson (2007).

The coherence between hybrid-based mass estimates and our comparative framework depends on two factors. The first is the consistency between the initial VD models on which the hybrid approach is based and the ES expectations. For instance, the physical reconstructions used by Christiansen & Fariña (2004) are highly consistent with the expectations of femoral circumference and, therefore, most of their hybrid approaches follow suit (Fig. 10A, B). In comparison, the physical reconstructions used by Mazzetta *et al.* (2004) are, on average, lower than expected given their stylopodial circumferences, resulting in similarly lower hybrid-based mass estimates (Fig. 10C). A curvilinear function can reconcile these lower estimates (Campione, 2017), although then extrapolation to smaller body sizes is not recommended as a linear extrapolation of Mazzetta *et al.*'s (2004) hybrid models to smaller body sizes will not accommodate for the non-linear model fit of the extant data. The second factor relates to the scaling pattern of the skeletal size-proxy used in the hybrid approach. For instance, the femoral mediolateral diameter equation of Christiansen & Fariña (2004) generates body mass estimates that imply an exceedingly inconsistent (systemically lower) scaling relationship between

stylopodial circumference and body mass in theropods compared with that expected by the extant data and those produced by other hybrid equations presented by the authors (Fig. 10A, B). Notably, this is despite the fact that the initial VD models occur well within the 95% prediction intervals of the extant data. In other cases, hybrid equations can generate estimates that are apparently consistent (within the 95% prediction intervals) but nonetheless imply a different scaling pattern when compared to the extant data or other hybrid approaches [e.g. femoral anteroposterior diameter (FAP) and tibial circumference (TC), Fig. 10A]. Hybrid equations implying differential scaling need to be considered carefully, given the consistency of stylopodial circumference(s) relative to body mass across taxonomically and functionally distinct extant terrestrial tetrapods. Equations such as those of FAP or TC (Christiansen & Fariña, 2004) reveal specific trends among the residual deviations, whereby smaller taxa will generate higher estimates than expected, with the reverse true for larger taxa. Despite these noted discrepancies, femoral length and circumference hybrid approaches were highly consistent with those expected by the ES approach and support the applicability of both Christiansen & Fariña's (2004) model and the bipedal correction factor (Campione *et al.*, 2014), along with the universality of load-bearing stylopodial circumferences in terrestrial tetrapods as accurate proxies for body mass (Anderson *et al.*, 1985; Campione & Evans, 2012).

Our comparative framework suggests some caution when adopting the use of skull length as a proxy for body mass as proposed by Therrien & Henderson (2007). The skull length-based hybrid approach infers a differential circumference–body mass scaling relationship, despite the fact that the VD reconstructions used by the authors to generate the hybrid model occur within the 95% prediction intervals (Fig. 11). We support Therrien & Henderson's (2007) acknowledgement that their mass estimates for both *Mei long* (erroneously identified as an overestimate) and *Carnotaurus sastrei* should be regarded with caution, given the juvenile nature of the former and the apomorphically short skull of the latter. However, in addition, our ES approach suggests a substantially higher estimate for *Sinornithoides youngi* and lower mass estimates for *Coelophysis bauri* and *Giganotosaurus carolinii*. In general, we recommend that skull length as a size proxy be considered with a certain level of scepticism (e.g. 11988–20888 kg estimate for *Spinosaurus aegypticus*) due to the allometric nature of the preorbital region in at least some dinosaurs (e.g. Campione & Evans, 2011) and the likelihood that skull length, and cranial properties in general, were under selection pressures other than body size, such as those imposed by diet (Zanno & Makovicky, 2011) or sexual selection and/or species recognition (Sampson, 1997; Padian & Horner, 2014).

The femoral length-to-body mass equations of O'Gorman & Hone (2012) are, in part, extensions of the equivalent equations published by Christiansen & Fariña (2004) and Mazzetta *et al.* (2004) for theropods and sauropods, respectively, with the addition of an ornithischian
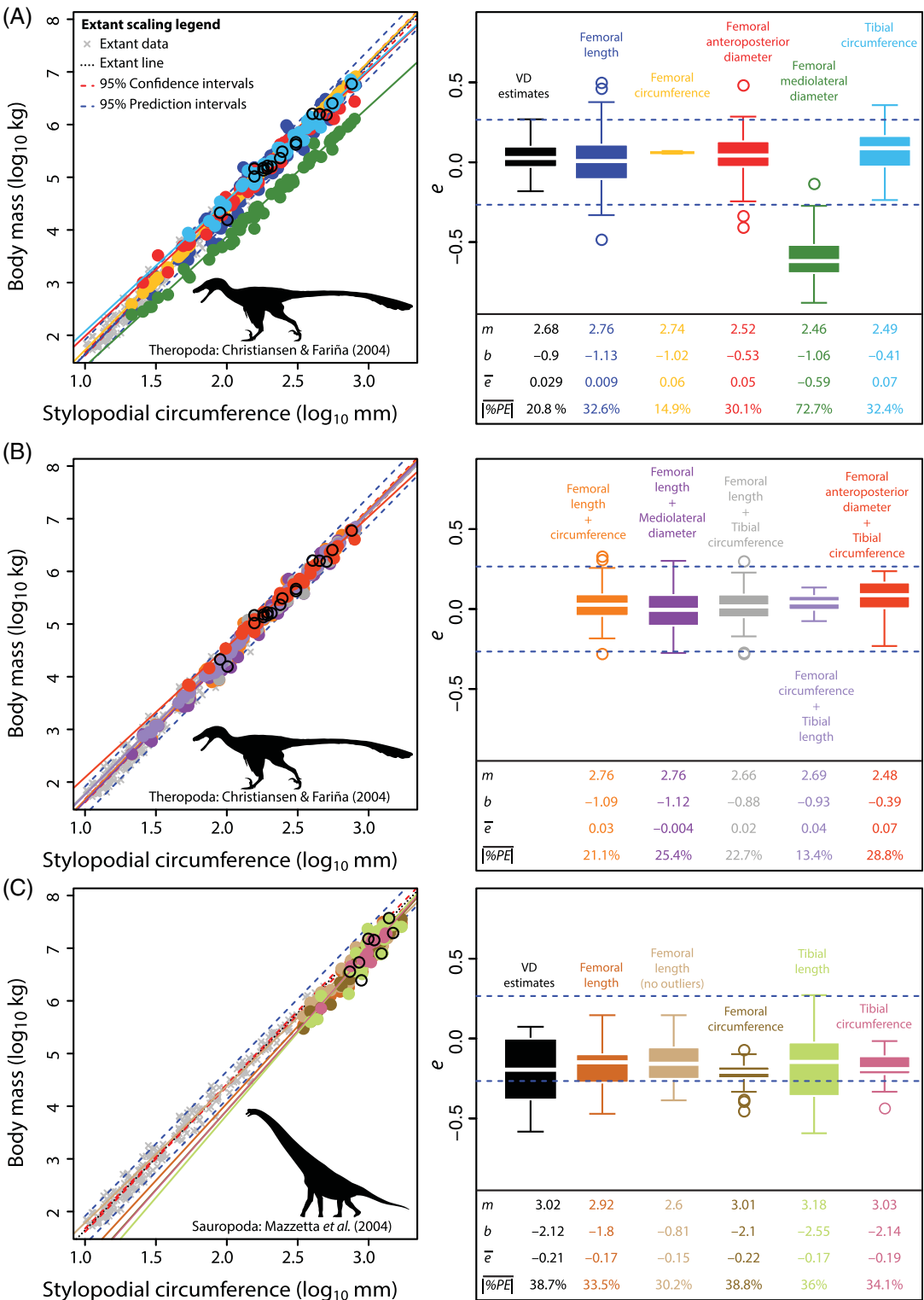
Fig 10. Application of the comparative framework to the hybrid approaches of Christiansen & Fariña (2004) and Mazzetta *et al.* (2004). Residual deviations (plots on right) are calculated in relation to the extant-scaling ordinary least-squares line and summarised in a boxplot. The tables below the residual plots provide the implied scaling line ($\log y = m\log x + b$) between the circumference(s) and body mass, along with the mean residual ($\bar{e}$) and mean absolute per cent prediction error ($\overline{|\%PE|}$) values. Scaling coefficients should be compared against those of the extant line (Campione & Evans, 2012; $m = 2.75$, $b = -1.1$). (A) Certain bivariate hybrid models of Christiansen & Fariña (2004), note the large deviations for femoral mediolateral diameter and lower scaling coefficients (m) for femoral anteroposterior diameter and tibial circumference. (B) Certain multivariate models of Christiansen & Fariña (2004). (C) Bivariate models of Mazzetta *et al.* (2004), note the overall lower estimates compared to the linear extant model. Silhouettes generated by Scott Hartman and reused under Creative Commons licenses: BY-NC-SA 3.0 (A, B) or BY 3.0 (C).
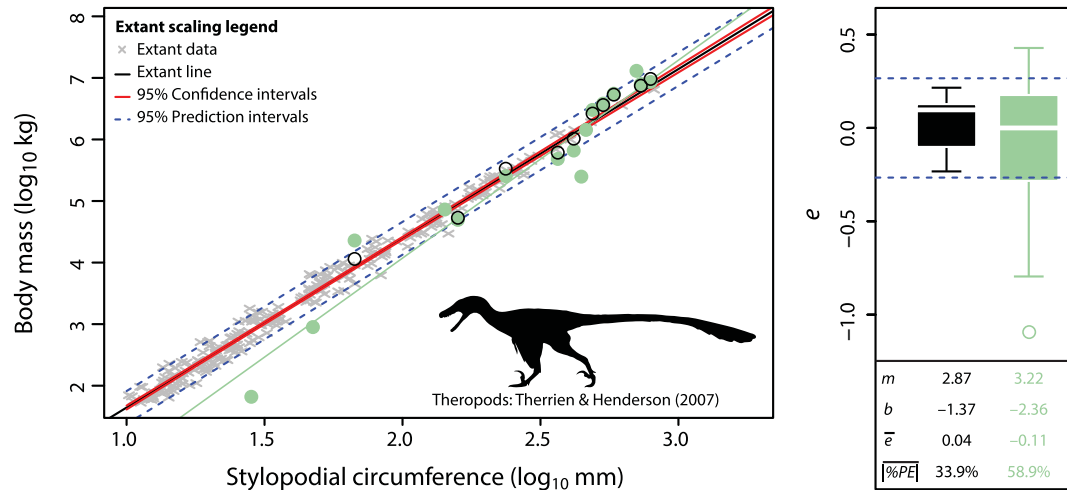
Fig 11. Application of the comparative framework to the skull length-based hybrid approach proposed by Therrien & Henderson (2007). Other details are as in Fig. 10. Note the high scaling coefficient (m) and the relatively high incidence of outliers, associated with elevated deviation values, despite high consistency between the original volumetric-density (VD) mass estimates and stylopodial circumference. Silhouette generated by Scott Hartman and reused under Creative Commons license BY-NC-SA 3.0.

equation. The theropod equation is highly consistent with the expectations of femoral circumference (Fig. 12A), both in terms of the VD and hybrid mass estimates, as was the case when assessing Christiansen & Fariña's (2004) equivalent equation (Fig. 10A). Furthermore, the inferred linear model is virtually identical to that of the extant data (Fig. 12A). In contrast, both the sauropod and ornithischian lines imply differential scaling patterns inconsistent with the ES relationship (Fig. 12B, C, respectively). In sauropods, the low scaling coefficients cross the upper 95% prediction interval at approximately 1000 kg, below which species will receive elevated mass estimates relative to that suggested by their stylopodial circumferences. Incidentally, a value of ~1 tonne is close to the modal body mass of dinosaurs, especially sauropods, and represents a 'taphonomic peak', below which dinosaurs are considered to be taphonomically filtered from the fossil record (Benson, 2018). Our results, therefore, indicate that hybrid estimation methods can intensify the seemingly negatively skewed body size distribution generally recovered among Mesozoic dinosaurs (O'Gorman & Hone, 2012; Codron *et al.*, 2012b), through systematically increasing the mass estimates of smaller taxa (Fig. 12B).

As is evident from the various hybrid approaches examined here, some models are highly consistent with each other and with the expectations of the ES relationship, while others are highly inconsistent, leading to knock-on effects on palaeobiological interpretations (e.g. body-size frequency distributions: O'Gorman & Hone, 2012). Notably, the somewhat unpredictable outcome of hybrid approaches is often independent of whether the VD estimates on which they are based were deemed accurate *a priori*. Hybrid approaches continue to be used (e.g. Wang *et al.*, 2019) and are undoubtedly valuable when circumferential or virtual data are lacking. As a result, we recommend that the application and future generation of hybrid approaches be considered in light of the

results presented here and explored through the use of our comparative framework. In particular, there is the potential that, if differential scaling patterns are recovered, they can be mathematically adjusted (e.g. Campione *et al.*, 2014).

(4) Quadratic *versus* linear ES models

The act of logarithmically transforming morphometric data generally serves to linearize allometric (i.e. non-linear) data to fit a standard linear model (i.e. $\log y = m\log x + b$). In addition, log-transforming can better model the relative errors across a large size spectrum (Cawley & Janacek, 2009). However, the large orders of magnitude exhibited by vertebrates and the anatomical alterations to their skeletons, as a result of size-dependent changes in biomechanical demands, can lead to non-linear (quadratic) relationships in log-space; larger taxa may have lower scaling coefficients compared to small taxa (Bertram & Biewener, 1990; Christiansen, 1999a; Carrano, 2001; Chan, 2017). Such differential scaling and a quadratic model were statistically rejected for the relationship between body mass and combined stylopodial circumferences (Campione & Evans, 2012; Campione, 2017), although the quadratic model received some support based on a small sample of sauropod dinosaurs (Campione, 2017). We extend these comparisons to the entire sample compiled in this study (Table 5) and, in general (compare deviations between Tables 1 and 5), find that a quadratic model does not result in dramatically different residual or per cent prediction deviations from those recovered by the standard linear model (Fig. 13). This supports previous assertions that the quadratic model is, statistically, linear (Campione, 2017). However, some differences are noted among sauropodomorphs, whereby residual deviations are no longer significantly different from 0, suggesting higher consistency between VD reconstructions and ES
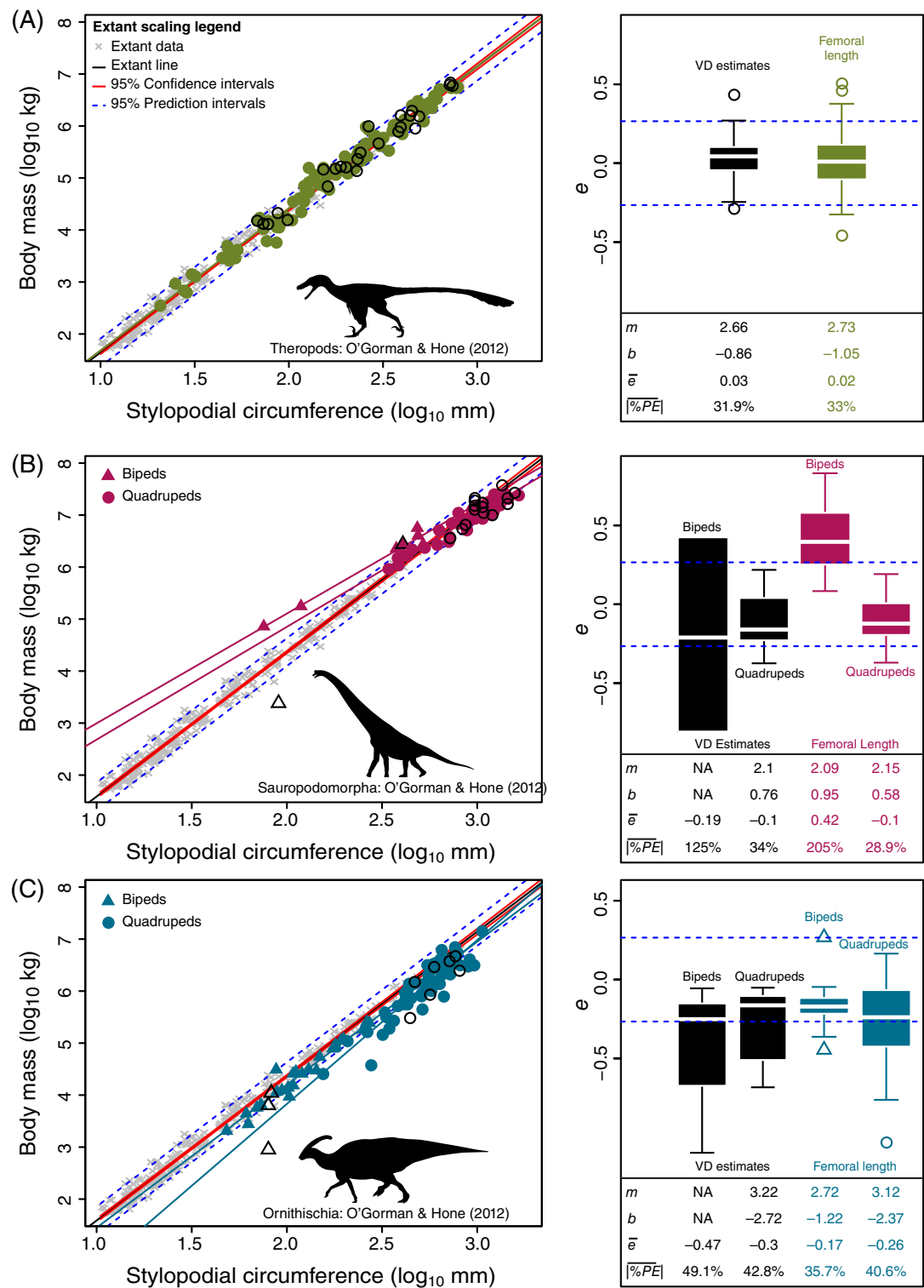
Fig 12. Application of the comparative framework to the clade-specific femoral length-based hybrid approaches proposed by O'Gorman & Hone (2012). (A) Theropoda, (B) Sauropodomorpha, and (C) Ornithischia. Other details are as in Fig. 10. Note the coherence between extant-scaling (ES), volumetric-density (VD), and hybrid approaches in theropods, but the stark deviations among sauropodomorphs and ornithischians. Silhouettes generated by Scott Hartman and reused under Creative Commons licenses: BY-NC-SA 3.0 (A, C) or BY 3.0 (B).

expectations under a quadratic ES model (Table 5; Fig. 13). Per cent prediction errors for sauropodomorphs, however, did not see an associated decrease and their deviations, in absolute terms, remain high despite more evenly distributed residuals about the quadratic line. In contrast to sauropodomorphs, residual deviations increased in Theropoda when using the quadratic model, suggesting that the potential benefits of the quadratic model noted by Campione (2017) cannot be extended to all dinosaurs.

Taken at face value, a potential decrease in scaling coefficient suggests that circumferences, relative to mass, increase at a faster rate in large sauropodomorphs ($\gtrsim$10000 kg) than in smaller forms. Such a phenomenon was observed at a smaller size threshold among mammals ($\gtrsim$100 kg), when comparing length and diameter measures, and hypothesised to reflect the inability of larger mammals to accommodate mass-dependent increases in strain through postural changes,

thereby requiring a shift in skeletal allometry (Bertram & Biewener, 1990). It is important to note, however, that differential scaling between small and large extant quadrupeds is more pronounced in the lengths of the stylopodia than in their associated circumferences, when explored in relation to body mass (see Table 4 in Campione & Evans, 2012 and Table 3 in Christiansen, 1999*b*). Thus, there is less statistical support for differential scaling in extant quadrupeds when considering the circumference–body mass ES model (Campione, 2017).

At present, the reconciliation between the two ES models amounts to empirical *versus* theoretical considerations that result from the need to extrapolate for very large dinosaurs. The linear ES model receives better empirical support and there is no significant size-dependent differential scaling when considering the relationship between stylopodial circumferences and body mass. Furthermore, circumference–body mass

Table 5. Deviations between volumetric-density and extant-scaling mass estimates using a quadratic prediction model (Campione, 2017), measured as the mean of the residuals ($\bar{e}$), the mean absolute per cent prediction error ($\overline{|\%\boldsymbol{PE}|}$), and associated standard deviations (σ) of the total data set

	N	$\bar{e}$	$\sigma_{\bar{e}}$	$\overline{|\%\boldsymbol{PE}|}$	$\sigma_{\overline{|\%\boldsymbol{PE}|}}$
All	432	−0.023*	0.247	43.42***	39.79
Taxonomy					
Ornithischia	123	−0.175***	0.271	42.51***	34.05
Ankylosauria	13	0.009	0.27	51.46	50.87
Marginocephalia	32	−0.109**	0.222	36.53	30.22
Ornithopoda	57	−0.202***	0.264	36.79***	21.19
Stegosauria	15	−0.363***	0.299	71.66**	50.37
Sauropodomorpha	120	−0.029	0.23	40.49***	33.85
Brachiosauridae	17	0.053	0.163	37.31	42.55
Diplodocoidea	41	−0.046	0.216	39.99**	32.47
Eusauropoda	4	0.18[NA]	0.09	53.64[NA]	31.48
Macronaria	11	−0.058	0.145	29.77	14.23
Neosauropoda	5	0.002	0.07	11.85*	8.96
Titanosauroidea	15	−0.133*	0.173	33.41	20.31
Theropoda	172	0.08***	0.189	47.54***	47.62
Allosauroidea	27	0.055	0.167	39.15*	28.08
Carcharodontosauridae	11	0.233***	0.1	74.93**	40.03
Ceratosauria	10	0.004	0.267	57.26	66.42
Coelophysoidea	18	0.131***	0.122	42.25	38.45
Coelurosauria	3	−0.258[NA]	0.061	44.36[NA]	8
Maniraptoriformes	34	0.084*	0.23	58.64*	73.21
Megalosauroidea	7	−0.154	0.174	34.58	16.88
Tyrannosauroidea	68	0.119***	0.113	38.99**	35.49
VD Approach					
PS	194	−0.072***	0.199	33.6***	27.07
GDI	2	−0.11[NA]	0.138	20.35[NA]	24.89
PY	60	−0.134**	0.362	53.29***	44.61
MS	64	−0.05	0.221	36.94**	27.54
MCH	34	0.004	0.16	30.54	24.58
PG	5	0.271*	0.16	96.87	69.67
3D	73	0.189***	0.164	70.07***	57.23

Subdivision of residuals and absolute per cent prediction errors and statistical protocols as in Table 1.
PS, physical scale model; GDI, graphic double integration; PY, polynomial; MS, mathematical slicing; MCH, minimum convex hull; PG, photogrammetry; 3D, 3D scanning.
Two-tailed *t*-tests were conducted on both measures of deviation against expected values of $\bar{e}=0$ and $\overline{|\%PE|}\approx25\%$, obtained from the extant terrestrial quadruped data set (Campione, 2017).
Statistical results are reported as: * = 0.05 > P > 0.01, ** = 0.01 > P > 0.001, and ***P < 0.001.
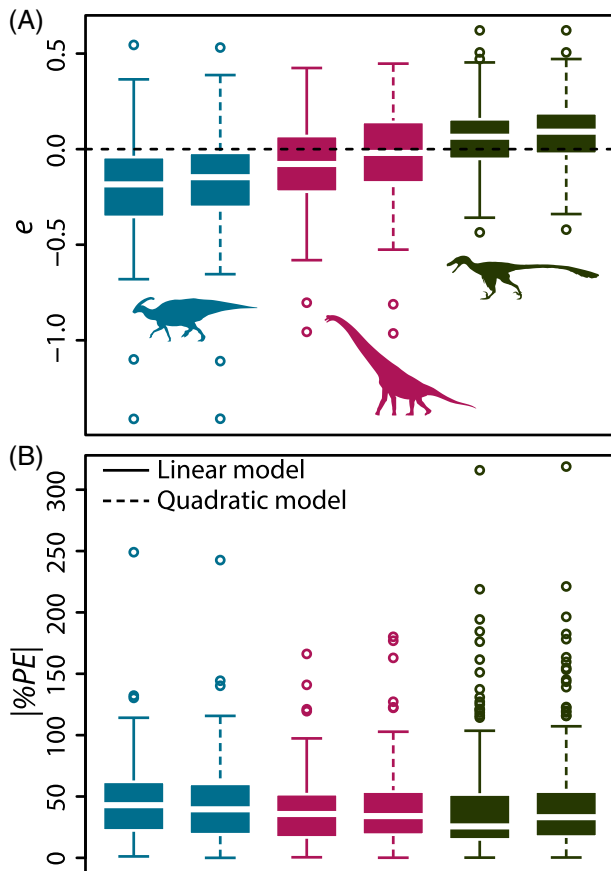
Fig 13. Distribution of the deviations between volumetric-density (VD) reconstructions and extant-scaling (ES) expectations in the three main dinosaurian clades given a linear or quadratic model. Deviations given as (A) residuals (e) and (B) per cent prediction error ($|\%PE|$). Results are, by in large, consistent between models, but note the improvement in median value ($\tilde{e} \approx 0$) in residual deviations for the quadratic models for Sauropodomorpha. Silhouettes generated by Scott Hartman and reused under Creative Commons licenses: BY-NC-SA 3.0 (*Parasaurolophus* and *Velociraptor*) or BY 3.0 (*Giraffatitian*).

scaling coefficients generally occur below geometric (isometric) expectations, following or approaching elastic similarity, which indicates that there is a natural tendency for quadrupeds to have relatively thicker stylopodia, independent of size class (Christiansen, 1999b; Campione & Evans, 2012). The theoretical possibility that the rate at which circumference increases relative to body mass is size dependent presently lies outside the size range of extant data and relies on the accuracy of sauropod VD reconstructions that, as discussed earlier, cannot be intrinsically assessed. Nevertheless, the very high degree of overlap between both ES models (see Fig. 2 in Campione, 2017) indicates that, for most taxa, either will generate similar results (see examples in Tables 6 and 7). Unsurprisingly, discrepancies are most prevalent at exceptionally large sizes (e.g. *Argentinosaurus*, *Patagotitan*, *Dreadnoughtus*, and *Brachiosaurus*; Table 6), where the difference in absolute fitted values can reach between 10000–20000 kg, although there is

substantial overlap in their errors. The difference between the fitted values of the ES models, however, decreases rapidly, with a ~ 6500 kg difference for *Futalognkosaurus*, 840 kg difference for *Tyrannosaurus*, and a seemingly negligible difference at small sizes (Tables 6 and 7). It is, therefore, only at very large body sizes ($\gtrsim$30000 kg) where choice of ES model needs careful consideration. Accordingly, we interpret the noted improvement in residual deviations in sauropodomorphs as support for the quadratic ES model in these taxa, with the caveat that it may not apply to all Dinosauria (Table 5 Fig. 13).

(5) Shifting paradigms and the 'evolution' of body mass estimates

Almost two centuries of research have been dedicated to the discovery, description, and interpretation of Mesozoic dinosaurs. Perhaps most vibrant are the long-running discussions and interpretations of dinosaurian biology (e.g. Dollo, 1883; Bakker, 1986; Sampson, 2009; Brusatte, 2012) borne out of innovations in technological and analytical tools and conceptual frameworks. These paradigm shifts in dinosaurian palaeobiology are built upon new fossil discoveries that reveal greater details about overall dinosaur appearance that should, ideally, lead to greater convergence between mass estimates generated *via* different approaches. Perhaps counterintuitively, our results reveal no significant trend in any direction over time (Fig. 8C, D) and, surprisingly, early estimates such as those of Gregory (1905), Janensch (1935), and certain models of Colbert (1962) exhibit very low deviations compared to ES expectations (Fig. 14). In comparison, some more recent estimates based on mathematical and updated physical reconstructions exhibited some of the highest deviations, many far outside the very liberal 95% prediction intervals of the ES model (Fig. 8C, D).

A non-significant trend in deviations through time does not necessarily indicate that early reconstructions were valid in terms of their general appearance given our current understanding of dinosaurian palaeobiology. Rather, it suggests that some of the assumptions inherent to early VD approaches (e.g. the amount of soft tissue reconstructed) should be reconsidered. It is possible that, in light of the 'Dinosaur Renaissance', reconstructions were pushed towards more emaciated proportions. Indeed, the lowest body mass estimates ever generated for dinosaurs are those that occured in the 90s and early 2000s ($\log_{10}2019-VD_{Annum} = 1$–$1.5$ in Fig. 8C), prior to the use of virtual 3D reconstruction approaches.

The more recently proposed VD_{MCH} approach generated the highest consistency between VD and ES expectations (Table 1; Figs 7 and 15). Of the 34 models explored, only four occurred outside the 95% prediction intervals: both models of *Amargasaurus* and the 21% model of *Dreadnoughtus* result in extreme negative residuals, whereas the maximal model of *Barosaurus* is higher than expected (Fig. 15). The variation in body mass estimates of *Dreadnoughtus* have been discussed in multiple studies (Lacovara *et al.*, 2014; Bates *et al.*, 2015; Campione, 2017; Carballido *et al.*, 2017; Paul, 2019).

Table 6. Mass comparisons between linear and quadratic extant-scaling (ES) equations for the top ten largest and smallest quadrupedal non-avian dinosaurs

	Clade	Subclade	Linear ES estimate (kg)	Quadratic ES estimate (kg)
Top 10 largest				
Argentinosaurus huinculensis	Sauropodomorpha	Titanosauria	94717 (70440–118993)	75077 (56027–94126)
Patagotitan mayorum	Sauropodomorpha	Titanosauria	690917 (51384–86800)	55681 (41553–69809)
Dreadnoughtus schrani	Sauropodomorpha	Titanosauria	59291 (44095–74487)	48158 (35939–60378)
Brachiosaurus altithorax	Sauropodomorpha	Titanosauriformes	57606 (42842–72371)	46858 (34969–58748)
Ruyangosaurus giganteus	Sauropodomorpha	Titanosauriformes	53927 (40106–67748)	44012 (32845–55179)
Turiasaurus riodevensis	Sauropodomorpha	Eusauropoda	50858 (37823–63893)	41630 (31067–52192)
Apatosaurus louisae	Sauropodomorpha	Diplodocoidea	41217 (30653–51781)	34090 (25440–42739)
Futalognkosaurus dukei	Sauropodomorpha	Titanosauria	38091 (28328–47853)	31625 (23601–39650)
Traukutitan eocaudata	Sauropodomorpha	Titanosauria	38081 (28321–47842)	31618 (23595–39640)
Top 10 smallest				
Protoceratops andrewsi	Ornithischia	Ceratopsia	82.6 (61.5–103.8)	83.1 (62–104.2)
Struthiosaurus transilvanicus	Ornithischia	Ankylosauria	128.1 (95.3–161)	128 (95.5–160.4)
Montanoceratops cerorynchus	Ornithischia	Ceratopsia	187.5 (139.4–235.5)	186 (138.8–233.2)
Leptoceratops gracilis	Ornithischia	Ceratopsia	415.9 (309.3–522.5)	405.8 (302.8–508.8)
Saichania chulsanensis	Ornithischia	Ankylosauria	611.5 (454.8–768.2)	591.1 (441.1–741.1)
Hungarosaurus tormai	Ornithischia	Ankylosauria	687.6 (511.4–863.9)	662.7 (494.6–830.9)
Magyarosaurus dacus	Sauropodomorpha	Titanosauria	745.5 (554.5–936.6)	717 (535.1–898.9)
Animantarx ramaljonesi	Ornithischia	Ankylosauria	763.5 (567.8–959.2)	733.8 (547.6–920)
Gargoyleosaurus parkpinorum	Ornithischia	Ankylosauria	890.2 (662.1–1118)	852.1 (635.9–1068)
Tenontosaurus tilletti	Ornithischia	Iguanodontia	1019 (757.7–1280)	971.5 (725–1218)

Limb measurement data from Benson *et al.* (2018). Range in parentheses represents the mean per cent prediction error based on the associated equation.

Although the specimen upon which *Dreadnoughtus* was reconstructed is incomplete (e.g. the neck is missing: Lacovara *et al.*, 2014) substantial effort was allocated to test the sensitivity of the VD$_{MCH}$ model to estimated missing data. These sensitivity tests, which should be standard practice for all volumetric reconstructions (e.g. Bates *et al.*, 2016), demonstrate that large increases in neck size do not necessarily translate to large increases in body volume (Bates *et al.*, 2015). As a result, it seems unlikely that the deviations noted here and elsewhere are due to incompleteness. It is possible, however, that lower VD reconstruction estimates compared to ES stem from the immature state of the hypodigm, despite its large size. This hypothesis is partly supported by the uniqueness of its elevated deviation, when compared to most other models

generated under the same VD$_{MCH}$ criteria, and by a similarly low mass estimate recently proposed for *Dreadnoughtus* based a physical reconstruction (31000 kg; Paul, 2019), suggesting consistency between VD approaches. Alternatively (but not mutually exclusively), the seemingly large *Dreadnoughtus* deviation is alleviated if one assumes a quadratic model for the extant data (Campione, 2017). Explanations for the high deviation values for *Barosaurus* and *Amargasaurus* are not evident, although it is possible that the latter stems from its unusual vertebral morphology and/or the incomplete nature of the thoracic cavity compared to its vertebral and appendicular skeleton (Salgado & Bonaparte, 1991).

In theory, VD$_{MCH}$ models have the potential to be more objective as they do not rely on the researcher to generate

Table 7. Mass comparisons between linear and quadratic extant-scaling (ES) equations for the top 10 largest and smallest bipedal non-avian dinosaurs

	Clade	Specimen	Linear ES estimate (kg)	Quadratic ES estimate (kg)
Top 10 largest				
Tyrannosaurus rex	Theropoda	Tyrannosauroidea	7685 (5716–9655)	6845 (5108–8582)
Deinocheirus mirificus	Theropoda	Ornithomimosauria	7320 (5444–9197)	6533 (4875–8190)
Giganotosaurus carolinii	Theropoda	Allosauroidea	6130 (4559–7702)	5508 (4111–6906)
Tyrannotitan chubutensis	Theropoda	Allosauroidea	5361 (3987–6735)	4841 (3613–6070)
Segnosaurus galbinensis	Theropoda	Therizinosauria	4595 (3417–5773)	4173 (3114–5232)
Mapusaurus roseae	Theropoda	Allosauroidea	4137 (3076–5197)	3771 (2814–4728)
Saurophaganax maximus	Theropoda	Allosauroidea	3772 (2806–4739)	3451 (2575–4326)
Chilantaisaurus tashuikouensis	Theropoda	Allosauroidea	3587 (2668–4506)	3287 (2453–4121)
Lametasaurus indicus	Theropoda	Ceratosauria	3555 (2644–4467)	3259 (2432–4086)
Top 10 smallest				
Iberomesornis romerali	Theropoda	Avialae	0.0133 (0.0099–0.0167)	0.0122 (0.0091–0.0153)
Paraprotopteryx gracilis	Theropoda	Avialae	0.016 (0.0119–0.0201)	0.0148 (0.011–0.0186)
Qiliania graffini	Theropoda	Avialae	0.0174 (0.0129–0.0219)	0.0162 (0.0121–0.0203)
Huoshanornis huji	Theropoda	Avialae	0.0223 (0.0166–0.028)	0.0208 (0.0155–0.0261)
Hongshanornis longicresta	Theropoda	Avialae	0.026 (0.0193–0.0327)	0.0244 (0.0182–0.0306)
Cathayornis yandica	Theropoda	Avialae	0.03 (0.0223–0.0377)	0.0283 (0.0211–0.0355)
Sinornis santensis	Theropoda	Avialae	0.0352 (0.0262–0.0442)	0.0334 (0.0249–0.0419)
Alexornis antecedens	Theropoda	Avialae	0.0375 (0.0279–0.0471)	0.0356 (0.0266–0.0446)
Concornis lacustris	Theropoda	Avialae	0.0393 (0.0292–0.0494)	0.0374 (0.0279–0.0469)
Longirostravis hani	Theropoda	Avialae	0.0408 (0.0303–0.0513)	0.0389 (0.029–0.0488)

Limb measurement data from Benson *et al.* (2018). Range in parentheses represents the mean per cent prediction error based on the associated equation.

the final reconstruction and, unlike other VD approaches, VD_{MCH} models can be properly constrained by extant data, through the use of expansion or scaling factors. Perhaps its greatest strength, however, is that extant-based confidence intervals can be generated about the expansion factor (Sellers *et al.*, 2012), thereby incorporating a measure of accuracy into the final body mass estimate. Given this fundamental strength, however, the 21% expansion factor identified by Sellers *et al.* (2012) between the predicted and true body mass in their extant sample ($N = 14$) requires the level of statistical scrutiny needed to substantiate an ES approach. Much like the now seemingly universal nature of the Anderson *et al.* (1985) equation, which was tested within the context of a large, disparate extant data set (Campione &

Evans, 2012), the universality and accuracy of the 21% expansion and associated confidence intervals depends on a broader context studied through a phylogenetically disparate extant sample. Such an investigation is particularly important as differential scaling patterns between VD_{MCH} and true body masses are observed between mammals (Sellers *et al.*, 2012) and birds (Brassey & Sellers, 2014). A larger sample of reptiles will be particularly relevant; mammalian *Baupläne* do not fully encompass those of dinosaurs, such as their large muscular tails (Gatesy, 1990). Finally, in addition to future validation, the presentation of VD_{MCH} results should emphasize the 21% and 'maximal' models, as was done here. Current use of the approach presents the 'minimal' model (i.e. that derived from the minimum convex hull)
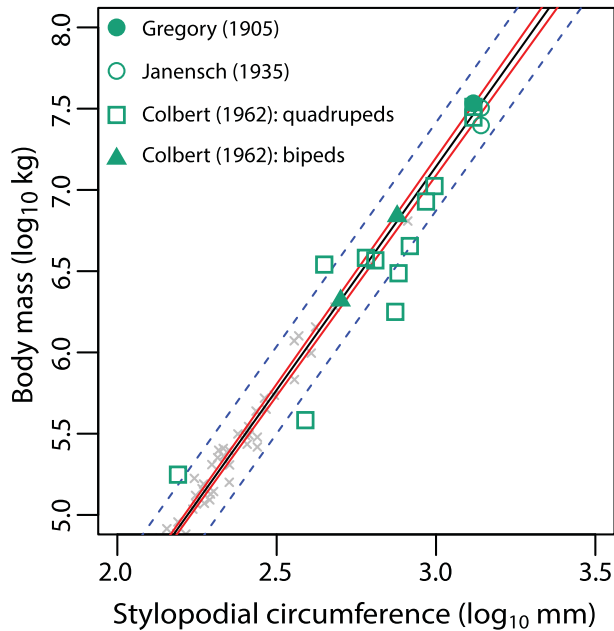
Fig 14. Historical body mass estimates compared to their respective stylopodial circumferences. Mass estimates are from Gregory (1905), Janensch (1935), and Colbert (1962). Note the high level of consistency between the circumferential data and Gregory's first estimate of *Brontosaurus* based on AMNH 460 (currently considered to be an indeterminate apatosaurine: Tschopp *et al.*, 2015).

reconstructions, but they are not realistic representations of the lower bounds of a possible mass estimate range. Minimal models are unrealistic mathematical envelopes and their inclusion within the possible range of mass estimates will inevitably lead to a tendency towards lower mass estimates.

V. FUTURE DIRECTIONS

The ecological and physiological implications of body mass are well recognized (e.g. Kleiber, 1947; Peters, 1983; Calder, 1984; Brown *et al.*, 1993; Burness *et al.*, 2001; Capellini & Gosling, 2007; McClain & Boyer, 2009) and the ability to reconstruct the mass of extinct forms, such as non-avian dinosaurs, offers the unique opportunity to study past life in a more biological context (Bakker, 1972; Farlow, 1976; Christiansen, 1999a; Henderson, 2004, 2006; Erickson, 2005; Varricchio *et al.*, 2008; Franz *et al.*, 2009; Pontzer *et al.*, 2009; Hutchinson *et al.*, 2011; Klein *et al.*, 2011; Sander *et al.*, 2011; Codron *et al.*, 2012b; Benson *et al.*, 2014, 2018; Bates *et al.*, 2016; Benson, 2018). Since body mass cannot be directly measured in the fossil record, several studies developed and advocate methods for estimating body mass that can be broadly categorised into volumetric-density or extant-scaling approaches. To date, no study has quantitatively compared these approaches on a large scale or assessed potential systemic differences between them. This study synthesizes previously proposed body mass estimation methods through the first extensive meta-analysis of dinosaur body masses. *Via* this undertaking, it is evident that a number of interesting questions remain about body mass properties in extinct animals.
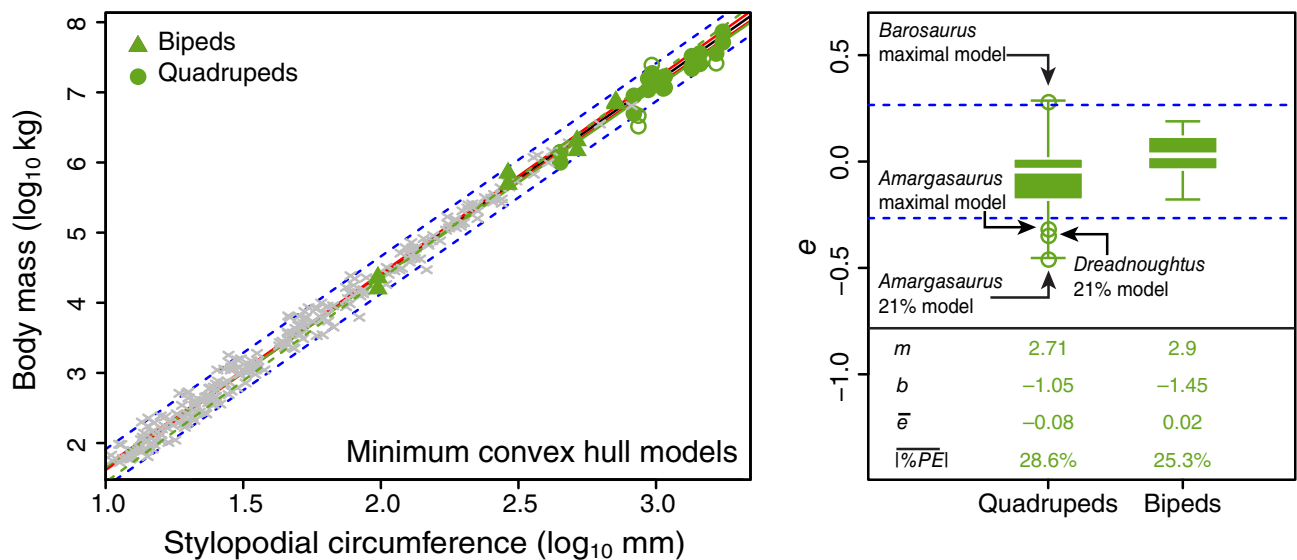
among the list of possible body mass estimates (Bates *et al.*, 2015, 2016; Carballido *et al.*, 2017). Depictions of the 'minimal' models are important to portray the virtual

Fig 15. Exploration of the minimum convex hull volumetric-density approach through the comparative framework. Plot details are as in Fig. 10. This approach is highly consistent with the body masses expected from stylopodial circumferences. Four models (open circles) occur outside the 95% prediction intervals (dashed blue lines), including the 21% model of *Dreadnoughtus* (Lacovara *et al.*, 2014; Bates *et al.*, 2015; Campione, 2017).

(1) Why are there general inconsistencies between VD and ES body mass estimates of ornithischians? All ornithischians, with the exception of ankylosaurs, exhibit high average deviation values within our comparative context. This suggests either long-standing misconceptions about the body proportions of these animals (e.g. size of chest cavity; Clauss *et al.*, 2017) and/or overall body density assumptions or, conversely, apomorphic limb scaling patterns between stylopodial circumferences and body mass in at least some ornithischians that require *ad hoc* explanations. For example, we previously identified that extreme burrowing behaviours (e.g. moles) are outliers of the stylopodial circumferences–body mass relationship (Campione & Evans, 2012). It is possible that such behaviours, hypothesized in ceratopsians and ornithopods (Varricchio *et al.*, 2007; Longrich, 2010; VanBuren, Campione & Evans, 2015), can be interpreted from comparing VD and ES mass estimates. However, the parameters used to generate ornithischian VD reconstructions require further exploration, perhaps through VD_{MCH} or VD_{3D} approaches, which are more broadly consistent with our ES expectations, but have yet to be applied to a range of ornithischians.

(2) How universal is the 21% expansion factor identified by Sellers *et al.* (2012) when using the VD_{MCH} technique? Brassey & Sellers (2014) recovered similar values in a limited sample of birds and primates (11–20%). However, they also noted potential differences between minimum convex hull volume-to-body mass scaling relationships in their extant sample, which can have dramatic effects when extrapolated to larger taxa (e.g. sauropods). Further applications of this approach to a large taxonomically inclusive sample of extant skeletons with known body masses will elucidate on this promising technique and help constrain realistic and non-arbitrary 'minimal' and 'maximal' models, respectively.

(3) Could density–size scaling provide a useful model for predicting body density from an *a priori* expectation of size? Brassey & Sellers (2014) identified a weak, but significant negative relationship between body density and mass in living birds, which is consistent with a weak and positive relationship between the level of skeletal pneumaticity and body size (measured as femoral length) in non-avian theropods (Benson *et al.*, 2011). These relationships and the high variance therein (i.e. low r^2 values) underscore our limited understanding of body density variation in living animals. Further exploration of such values using a systematic and broad taxonomic approach across living vertebrates would provide a useful avenue to constrain such values in the fossil record.

(4) Are relative limb circumferences a proxy for centre of mass? The centre of mass (CoM) of extinct tetrapods is currently estimated *via* VD reconstructions (e.g. Henderson, 1999; Allen *et al.*, 2013; Bates *et al.*, 2016). However, the time-consuming nature of this endeavour and the restriction of VD approaches to highly complete specimens limits the sample sizes available for macroevolutionary analysis (Allen *et al.*, 2013; Bates *et al.*, 2016). Variations in the relationship between humeral and femoral circumferences (Fig. 4) suggest that the relative circumferences of the humerus and femur may be linked to the proportion of body mass that they support. Indeed, the entire conceptual framework behind our ES relationship is predicated on such a link. Future studies could test for a relationship between previously estimated CoMs in dinosaurs and their relative limb sizes. A potential outcome is a new CoM estimation model that could be easily applied to a large sample of extinct taxa to study the evolution of *Baupläne* (as described by CoM) within the context of armour (Maidment *et al.*, 2014; Arbour & Zanno, 2018), gait transitions (e.g. to quadrupedality in ornithischians: Carrano, 2005; Maidment & Barrett, 2012), and the evolution of flight (Benson & Choiniere, 2013; Dececchi & Larsson, 2013).

VI. CONCLUSIONS

(1) Our study reviews the various iterations of VD and ES approaches as they apply to non-avian dinosaurs and compiles a large data set of VD-based mass estimates and the stylopodial circumferences of the specimens on which they were based. In so doing, we present the first analytical comparison between VD and ES approaches.

(2) Comparisons underscore that (*a*) overall, there is prevalent consistency between VD and ES approaches, and (*b*) the comparative framework presented herein is a powerful tool to illuminate, and corroborate, the unavoidable subjectivity inherent to VD reconstructions.

(3) Despite overall consistency, the comparative framework identified several inconsistencies and, in particular, suggests that some VD approaches (e.g. the polynomial method: Seebacher, 2001) may underestimate small-bodied taxa, in relation to the ES relationship, with important knock-on effects to subsequent biological interpretations [e.g. ontogenetic niche shifting (Codron *et al.*, 2012*b*); body-size distributions (O'Gorman & Hone, 2012)].

(4) Promisingly, there is great consistency between most models generated through three-dimensional scanning techniques of complete skeletons (Gunga *et al.*, 2007, 2008; Hutchinson *et al.*, 2007; Bates *et al.*, 2009*a*, 2009*b*, 2015, 2016; Mallison, 2010; Hutchinson *et al.*, 2011; Sellers *et al.*, 2012; Carballido *et al.*, 2017), especially those derived through minimum convex hulling. Although further validation of these approaches is still needed, virtual-based mass estimates were the most consistent with limb circumferential data providing strong evidence that these approaches, in particular VD_{MCH}, provide the most accurate VD approach with which to reconstruct extinct vertebrates.

(5) VD and ES approaches are fundamentally different and should not be viewed in opposition. ES approaches provide accuracy, whereas VD approaches should provide precision. As a result, differences between these approaches should be regarded as an opportunity to reassess assumptions – such as

body envelopes, density, etc. – or identify anatomical novelties – such as apomorphically robust/gracile limbs possibly associated with particular ecologies (e.g. Richards *et al.*, 2019).

(6) Emphatically, the results of mass estimation should never be presented as a point estimate. Errors, whether methodological or statistical need to be quantified, presented, and their knock-on effects considered.

ACKNOWLEDGMENTS

VII. REFERENCES

References marked with asterisks are cited within the supporting information.

Adams, D. C., Collyer, M. L. & Kaliontzopoulou, A. (2020). Geomorph: Software for geometric morphometric analyses *version 3.2.1*.

Alexander, R. M. (1985). Mechanics of posture and gait of some large dinosaurs. *Zoological Journal of the Linnean Society* 83, 1–25.

Alexander, R. M. (1989). *Dynamics of Dinosaurs and Other Extinct Giants*. Columbia University Press, New York.

Allen, V., Paxton, H. & Hutchinson, J. R. (2009). Variation in center of mass estimates for extant sauropsids and its importance for reconstructing inertial properties of extinct archosaurs. *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology* 292, 1442–1461.

Allen, V., Bates, K. T., Li, Z. & Hutchinson, J. R. (2013). Linking the evolution of body shape and locomotor biomechanics in bird-line archosaurs. *Nature* 497, 104–107.

Anderson, J. F., Hall-Martin, A. & Russell, D. A. (1985). Long-bone circumference and weight in mammals, birds and dinosaurs. *Journal of the Zoological Society of London A* 207, 53–61.

Arbour, V. M. & Zanno, L. E. (2018). The evolution of tail weaponization in amniotes. *Proceedings of the Royal Society B: Biological Sciences* 285, 20172299.

Bakker, R. T. (1972). Anatomical and ecological evidence of endothermy in dinosaurs. *Nature* 239, 81–85.

Bakker, R. T. (1986). *The Dinosaur Heresies*. Kensington Publishing Corp, New York.

Basu, C., Falkingham, P. L. & Hutchinson, J. R. (2016). The extinct, giant giraffid *Sivatherium giganteum*: skeletal reconstruction and body mass estimation. *Biology Letters* 12, 20150940.

Bates, K. T., Falkingham, P. L., Breithaupt, B. H., Hodgetts, D., Sellers, W. I. & Manning, P. L. (2009a). How big was 'Big Al'? Quantifying the effects of soft tissue and osteological unknowns on mass predictions for *Allosaurus* (Dinosauria: Theropoda). *Palaeontologia Electronica* 12, 33p.

Bates, K. T., Manning, P. L., Hodgetts, D. & Sellers, W. I. (2009b). Estimating mass properties of dinosaurs using laser imaging and 3D computer modelling. *PLoS One* 4, e4532.

Bates, K. T., Falkingham, P. L., Macaulay, S., Brassey, C. & Maidment, S. C. R. (2015). Downsizing a giant: re-evaluating *Dreadnoughtus* body mass. *Biology Letters* 11, 20150215.

Bates, K. T., Mannion, P. D., Falkingham, P. L., Brusatte, S. L., Hutchinson, J. R., Otero, A., Sellers, W. I., Sullivan, C., Stevens, K. A. & Allen, V. (2016). Temporal and phylogenetic evolution of the sauropod dinosaur body plan. *Royal Society Open Science* 3, 150636.

Benson, R. B. J. (2018). Dinosaur macroevolution and macroecology. *Annual Review of Ecology, Evolution, and Systematics* 49, 379–408.

Benson, R. B. J. & Choiniere, J. N. (2013). Rates of dinosaur limb evolution provide evidence for exceptional radiation in Mesozoic birds. *Proceedings of the Royal Society B* 280, 20131768.

Benson, R. B. J., Butler, R. J., Carrano, M. T. & O'Connor, P. M. (2011). Air-filled postcranial bones in theropod dinosaurs: physiological implications and the 'reptile'-bird transition. *Biological Reviews* 87, 168–193.

Benson, R. B. J., Campione, N. E., Carrano, M. T., Mannion, P. D., Sullivan, C., Upchurch, P. & Evans, D. C. (2014). Rates of dinosaur body mass evolution indicate 170 million years of sustained ecological innovation on the avian stem lineage. *PLoS Biology* 12, e1001853.

Benson, R. B. J., Hunt, G., Carrano, M. T. & Campione, N. (2018). Cope's rule and the adaptive landscape of dinosaur body size evolution. *Palaeontology* 61, 13–48.

Benton, M. J. (2008). How to find a dinosaur, and the role of synonymy in biodiversity studies. *Journal Information* 34, 516–533.

Bertram, J. E. A. & Biewener, A. A. (1990). Differential scaling of the long bones in the terrestrial Carnivora and other mammals. *Journal of Morphology* 204, 157–169.

Blob, R. W. (2000). Interspecific scaling of the hindlimb skeleton in lizards, crocodilians, felids and canids: does limb bone shape correlate with limb posture? *Journal of Zoology* 250, 507–531.

Blob, R. W. (2001). Evolution of hindlimb posture in nonmammalian therapsids: biomechanical tests of paleontological hypotheses. *Paleobiology* 27, 14–38.

Bonnan, M. F., Sandrik, J. L., Nishiwaki, T., Wilhite, D. R., Elsey, R. M. & Vittore, C. (2010). Calcified cartilage shape in archosaur long bones reflects overlying joint shape in stress-bearing elements: implications for nonavian dinosaur locomotion. *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology* 293, 2044–2055.

Brassey, C. A. & Sellers, W. I. (2014). Scaling of convex hull volume to body mass in modern Primates, non-primate mammals and birds. *PLoS One* 9, e91691.

Brassey, C. A., Maidment, S. C. R. & Barrett, P. M. (2015). Body mass estimates of an exceptionally complete *Stegosaurus* (Ornithischia: Thyreophora): comparing volumetric and linear bivariate mass estimation methods. *Biology Letters* 11, 20140984.

Brett-Surman, M. K. (1997). Ornithopods. In *The Complete Dinosaur* (eds J. O. Farlow and M. K. Brett-Surman), pp. 330–346. Indiana University Press, Indianapolis.

Brown, J. H. & Maurer, B. A. (1986). Body size, ecological dominance and Cope's rule. *Nature* 324, 248–250.

Brown, J. H., Marquet, P. A. & Taper, M. L. (1993). Evolution of body size: consequences of an energetic definition of fitness. *The American Naturalist* 142, 573–584.

Brown, C. M., Campione, N. E., Corrêa Giacomini, H., O'Brien, L. J., Vavrek, M. J. & Evans, D. C. (2012). Ecological modelling, size distributions and taphonomic size bias in dinosaur faunas: a comment on Codron et al. (2012). *Biology Letters* 9, 20120582.

Brown, C. M., Evans, D. C., Campione, N. E., O'Brien, L. J. & 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, C. M., Henderson, D. M., Vinther, J., Fletcher, I., Sistiaga, A., Herrera, J. & Summons, R. E. (2017). An exceptionally preserved three-dimensional armored dinosaur reveals insights into coloration and Cretaceous predator-prey dynamics. *Current Biology* 27, 2514–2521.e3.

Brusatte, S. L. (2012). *Dinosaur Paleobiology*. John Wiley and Sons Ltd, West Sussex.

*Brusatte, S. L., Carr, T. D., Erickson, G. M., Bever, G. S. & Norell, M. A. (2009). A long-snouted, multihorned tyrannosaurid from the Late Cretaceous of Mongolia. *Proceedings of the National Academy of Sciences* 106, 17261–17266.

Burness, G. P., Diamond, J. & Flannery, T. (2001). Dinosaurs, dragons, and dwarfs: the evolution of maximal body size. *Proceedings of the National Academy of Science* 98, 14518–14523.

BUTLER, R. J. & GOSWAMI, A. (2008). Body size evolution in Mesozoic birds: little evidence for Cope's rule. *Journal of Evolutionary Biology* **21**, 1673–1682.

BUTLER, R. J., GALTON, P. M., PORRO, L. B., CHIAPPE, L. M., HENDERSON, D. M. & ERICKSON, G. M. (2009). Lower limits of ornithischian dinosaur body size inferred from a new Upper Jurassic heterodontosaurid from North America. *Proceedings of the Royal Society B* **277**, 375–381.

CALDER, W. A. I. (1984). *Size, Function, and Life History*. Harvard University Press, Cambridge.

CAMPBELL, K. E. JR. & MARCUS, L. (1992). The relationships of hindlimb bone dimensions to body weight in birds. *Natural History Museum of Los Angeles County Science Series* **36**, 395–412.

CAMPBELL, K. E. JR. & TONNI, E. P. (1983). Size and locomotion in teratorns (Aves: Teratornithidae). *The Auk* **100**, 390–403.

CAMPIONE, N. E. (2017). Extrapolating body masses in large terrestrial vertebrates. *Paleobiology* **43**, 693–699.

CAMPIONE, N. E. (2020). MASSTIMATE: body mass estimation equations for vertebrates. *R package version 2.0*.

CAMPIONE, N. E. & EVANS, D. C. (2011). Cranial growth and variation in edmontosaurs (Dinosauria: Hadrosauridae): implications for latest Cretaceous megaherbivore diversity in North America. *PLoS One* **6**, e25186.

CAMPIONE, N. E. & EVANS, D. C. (2012). A universal scaling relationship between body mass and proximal limb bone dimensions in quadrupedal terrestrial tetrapods. *BMC Biology* **10**, 60.

CAMPIONE, N. E., EVANS, D. C., BROWN, C. M. & CARRANO, M. T. (2014). Body mass estimation in non-avian bipeds using a theoretical conversion to quadruped stylopodial proportions. *Methods in Ecology and Evolution* **5**, 913–923.

CAPELLINI, I. & GOSLING, L. M. (2007). Habitat primary production and the evolution of body size within the hartebeest clade. *Biological Journal of the Linnean Society* **92**, 431–440.

CARBALLIDO, J. L., POL, D., OTERO, A., CERDA, I. A., SALGADO, L., GARRIDO, A. C., RAMEZANI, J., CÚNEO, N. R. & KRAUSE, J. M. (2017). A new giant titanosaur sheds light on body mass evolution among sauropod dinosaurs. *Proceedings of the Royal Society B: Biological Sciences* **284**, 20171219.

CARPENTER, K. (1990). Pinacosaurus. In *Encyclopedia of Dinosaurs* (ed. P. DODSON), p. 208. Beekman House, New York.

CARRANO, M. T. (1998). Locomotion in non-avian dinosaurs: integrating data from hindlimb kinematics, in vivo strains, and bone morphology. *Paleobiology* **24**, 450–469.

CARRANO, M. T. (2001). Implications of limb bone scaling, curvature and eccentricity in mammals and non-avian dinosaurs. *Journal of Zoology* **254**, 41–55.

CARRANO, M. T. (2005). The evolution of sauropod locomotion: morphological diversity of a secondarily quadrupedal radiation. In *The Sauropods: Evolution and Paleobiology* (eds K. C. ROGERS and J. A. WILSON), pp. 229–251. University of California Press, Berkeley, CA.

CARRANO, M. T. (2006). Body-size evolution in the Dinosauria. In *Amniote Paleobiology: Perspectives on the Evolution of Mammals, Birds, and Reptiles* (eds M. T. CARRANO, R. W. BLOB, T. J. GAUDIN and J. R. WIBLE), pp. 225–268. University of Chicago Press, Chicago.

CARRANO, M. T. & BIEWENER, A. A. (1999). Experimental alteration of limb posture in the chicken (*Gallus gallus*) and its bearing on the use of birds as analogs for dinosaur locomotion. *Journal of Morphology* **240**, 237–249.

CARRANO, M. T., LOEWEN, M. A. & SERTICH, J. J. W. (2011). New material of *Masiakasaurus knopfleri* Sampson, Carrano, and Forster, 2001, and implications for the morphology of the Noasauridae (Theropoda: Ceratosaurinae). *Smithsonian Contributions to Paleobiology* **95**, 49.

CAWLEY, G. C. & JANACEK, G. J. (2009). On allometric equations for predicting body mass of dinosaurs. *Journal of Zoology* **280**, 355–361.

CHAN, N. R. (2017). Phylogenetic variation in hind-limb bone scaling of flightless theropods. *Paleobiology* **43**, 129–143.

*CHIAPPE, L. M., SHU'AN, J., QIANG, J. & NORELL, M. A. (1999). Anatomy and systematics of the Confuciusornithidae (Theropoda: Aves) from the late Mesozoic of northeastern China. *Bulletin of the American Museum of Natural History* **242**, 1–89.

CHRISTIAN, A., HEINRICH, W. D. & GOLDER, W. (1999). Posture and mechanics of the forelimbs of *Brachiosaurus brancai* (Dinosauria: Sauropoda). *Mitteilungen aus dem Museum für Naturkunde in Berlin, Geowissenschaftliche Reihe* **2**, 63–73.

CHRISTIANSEN, P. (1997). Locomotion in sauropod dinosaurs. *Gaia* **14**, 45–75.

CHRISTIANSEN, P. (1998). Strength indicator values of theropod long bones, with comments on limb proportions and cursorial potential. *Gaia* **15**, 241–255.

CHRISTIANSEN, P. (1999a). Long bone scaling and limb posture on non-avian theropods: evidence for differential allometry. *Journal of Vertebrate Paleontology* **19**, 666–680.

CHRISTIANSEN, P. (1999b). Scaling of the limb long bones to body mass in terrestrial mammals. *Journal of Morphology* **239**, 167–190.

CHRISTIANSEN, P. & FARIÑA, R. A. (2004). Mass prediction in theropod dinosaurs. *Historical Biology* **16**, 85–92.

CLARKE, J. A., ZHOU, Z. & ZHANG, F. (2006). Insight into the evolution of avian flight from a new clade of Early Cretaceous ornithurines from China and the morphology of *Yixianornis grabaui*. *Journal of Anatomy* **208**, 287–308.

CLAUSS, M., NURUTDINOVA, I., MELORO, C., GUNGA, H.-C., JIANG, D., KOLLER, J., HERKNER, B., SANDER, P. M. & HELLWICH, O. (2017). Reconstruction of body cavity volume in terrestrial tetrapods. *Journal of Anatomy* **230**, 325–336.

CODRON, D., CARBONE, C., MÜLLER, D. W. H. & CLAUSS, M. (2012a). Ecological modelling, size distributions and taphonomic size bias in dinosaur faunas: reply to Brown et al. *Biology Letters* **9**, 20120926.

CODRON, D., CARBONE, C., MÜLLER, D. W. H. & CLAUSS, M. (2012b). Ontogenetic niche shifts in dinosaurs influenced size, diversity and extinction in terrestrial vertebrates. *Biology Letters* **8**, 620–623.

COLBERT, E. H. (1962). The weights of dinosaurs. *American Museum Novitates* **2076**, 1–16.

COLBERT, E. H. (1989). The Triassic dinosaur *Coelophysis*. *Museum of Northern Arizona Bulletin* **57**, 1–160.

R CORE TEAM. (2020). R: a language and environment for statistical computing *version 4.0.1*.

CUFF, A. R., GOSWAMI, A. & HUTCHINSON, J. R. (2017). Reconstruction of the musculoskeletal system in an extinct lion. *Palaeontologia Electronica* **20.2** (23A), 1–25.

DAMUTH, J. & MACFADDEN, B. J. (1990). *Body Size in Mammalian Paleobiology: Estimation and Biological Implications*, p. 397. Cambridge University Press, Cambridge.

DE ESTEBAN-TRIVIGNO, S. & KÖHLER, M. (2011). New equations ofr body mass estimation in bovids: testing some procedures when constructing regression functions. *Mammalian Biology* **76**, 755–761.

DE ESTEBAN-TRIVIGNO, S., MENDOZA, M. & DE RENZI, M. (2008). Body mass estimation in Xenarthra: a predictive equation suitable for all quadrupedal terrestrial placentals? *Journal of Morphology* **269**, 1276–1293.

DECECCHI, T. A. & LARSSON, H. C. E. (2013). Body and limb size dissociation at the origin of birds: uncoupling allometric constraints across a macroevolutionary transition. *Evolution* **67**, 2741–2752.

DOLLO, L. (1883). Troisième note sur les dinosauriens de Bernissart [Third note on the dionosaurs of Bernissart]. *Bulletin de Musée Royal d'Histoire Naturelle de Belgique* **2**.

DOUBE, M., YEN, S. C. W., KLOSOWSKI, M. M., FARKE, A. A., HUTCHINSON, J. R. & SHEFELBINE, S. J. (2012). Whole-bone scaling of the avian pelvic limb. *Journal of Anatomy* **221**, 21–29.

ERICKSON, G. M. (2005). Assessing dinosaur growth patterns: a microscopic revolution. *Trends in Ecology and Evolution* **20**, 677–684.

ERICKSON, G. M. & TUMANOVA, T. A. (2000). Growth curve of *Psittacosaurus mongoliensis* Orborn (Ceratopsia; Psittacosauridae) inferred from long bone histology. *Zoological Journal of the Linnean Society* **130**, 551–566.

ERICKSON, G. M., ROGERS, K. C. & YERBY, S. A. (2001). Dinosaurian growth patterns and rapid avian growth rates. *Nature* **412**, 429–433.

FARIÑA, R. A., VIZCAÍNO, S. F. & BARGO, M. S. (1998). Body mass estimations in Lujanian (late Pleistocene-early Holocene of South America) mammal megafauna. *Mastozoología Neotropical* **5**, 87–108.

FARLOW, J. O. (1976). A consideration of the trophic dynamics of a Late Cretaceous large-dinosaur community (Oldman Formation). *Ecology* **57**, 841–857.

FARLOW, J. A., SMITH, M. B. & ROBINSON, J. M. (1995). Body mass, bone "strength indicator," and cursorial potential of *Tyrannosaurus rex*. *Journal of Vertebrate Paleontology* **15**, 713–725.

FIELD, D. J., LYNNER, C., BROWN, C. & DARROCH, S. A. F. (2013). Skeletal correlates for body mass estimation in modern and fossil flying birds. *PLoS One* **8**, e82000.

FINARELLI, J. A. (2008). Hierarchy and the reconstruction of evolutionary trends: evidence for constraints on the evolution of body size in terrestrial caniform carnivorans (Mammalia). *Paleobiology* **34**, 553–563.

FINARELLI, J. A. & FLYNN, J. J. (2006). Ancestral state reconstruction of body size in the Caniformia (Carnivora, Mammalia): the effects of incorporating data from the fossil record. *Systematic Biology* **55**, 301–313.

*FORSTER, C. A., SAMPSON, S. D., CHIAPPE, L. M. & KRAUSE, D. W. (1998). The theropod ancestry of birds: new evidence from the Late Cretaceous of Madagascar. *Science* **279**, 1915–1919.

FORTELIUS, M. & KAPPELMAN, J. (1993). The largest land mammal ever imagined. *Zoological Journal of the Linnean Society* **108**, 85–101.

FOX, J. & WEISBERG, S. (2019). *An R Companion to Applied Regression*, 3rd Edition (). Sage Publications, Inc., Thousand Oaks.

FRANZ, R., HUMMEL, J., KIENSLE, E., KÖLLE, P., GUNGA, H.-C. & CLAUSS, M. (2009). Allometry of visceral organs in living amniotes and its implications for sauropod dinosaurs. *Proceedings of the Royal Society B* **276**, 1731–1736.

GALTON, P. M. (1997). Stegosaurs. In *The Complete Dinosaur* (eds J. O. FARLOW and M. K. BRETT-SURMAN), pp. 291–306. Indiana University Press, Indianapolis.

GARLAND, T. J. & IVES, A. R. (2000). Using the past of predict the present: confidence intervals for regression equations in phylogenetic comparative methods. *The American Naturalist* **155**, 346–364.

GATESY, S. M. (1990). Caudofemoral musculature and the evolution of theropod locomotion. *Paleobiology* **16**, 170–186.

GATESY, S. M. (1991). Hind limb scaling in birds and other theropods: implications for terrestrial locomotion. *Journal of Morphology* **203**, 83–96.

GHIZZONI, M. (2014). Estimación de la masa corporal de un ejemplar cuaternario del carpincho extincto *Neochoerus* a través de medidas cráneo-dentales. *Revista Brasileira de Paleontologia* **17**, 83–90.

GILLOOLY, J. F., BROWN, J. H., WEST, G. B., SAVAGE, V. M. & CHARNOV, E. L. (2001). Effects of size and temperature on metabolic rate. *Science* **293**, 2248–2251.

GILLOOLY, J. F., CHARNOV, E. L., WEST, G. B., SAVAGE, V. M. & BROWN, J. H. (2002). Effects of size and temperature of developmental time. *Nature* **417**, 70–73.

GILLOOLY, J. F., ALLEN, A. P. & CHARNOV, E. L. (2006). Dinosaur fossils predict body temperatures. *PLoS Biology* **4**, e248.

*GILMORE, C. W. (1915). Osteology of *Thescelosaurus*, an orthopodous dinosaur from the Lance Formation of Wyoming. *Proceedings of the United States National Museum* **49**, 591–616.

GINGERICH, P. D. (1990). Prediction of body mass in mammalian species from long bone lengths and diameters. *Contributions from the Museum of Paleontology, University of Michigan* **28**, 79–92.

GRADY, J. M., ENQUIST, B. J., DETTWEILER-ROBINSON, E., WRIGHT, N. A. & SMITH, F. A. (2014). Evidence for mesothermy in dinosaurs. *Science* **344**, 1268–1272.

GRAND, T. I. (1990). The functional anatomy of Body mass. In *Body Size in Mammalian Paleobiology: Estimation and Biological Implications* (eds J. DAMUTH and B. J. MACFADDEN), pp. 39–47. Cambridge University Press, Cambridge.

GREENEWALT, C. H. (1975). The flight of birds: the significant dimensions, their departure from the requirements for dimensional similarity, and the effect on flight aerodynamics of that departure. *Transactions of the American Philosophical Society* **65**, 1–67.

GREGORY, W. K. (1905). The weight of *Brontosaurus*. *Science, New Series* **22**, 572.

GUNGA, H.-C., KIRSCH, K., BAARTZ, F. & RÖCKER, L. (1995). New data on the dimensions of *Brachiosaurus brancai* and their physiological implications. *Naturwissenschaften* **82**, 190–192.

GUNGA, H.-C., KIRSCH, K., RITTWEGER, J., RÖCKER, L., CLARKE, A., ALBERTZ, J., WIEDEMANN, A., MOKRY, S., SUTHAU, T., WEHR, A., HEINRICH, W.-D. & SCHULTZE, H.-P. (1999). Body size and body volume distribution in two sauropods from the Upper Jurassic of Tendaguru (Tanzania). *Mitteilungen aus dem Museum für Naturkunde der Humboldt-Universität Berlin, Geowissenschaftliche Reihe* **2**, 91–102.

GUNGA, H.-C., KIRSCH, K., RITTWEGER, J., CLARKE, A., ALBERTZ, J., WIEDEMANN, A., WEHR, A., HEINRICH, W.-D. & SCHULTZE, H.-P. (2002). Dimensions of *Brachiosaurus brancai*, *Dicraeosaurus hansemanni* and *Diplodocus carnegii* and their implications for gravitational physiology. *Adaptation Biology and Medicine* **3**, 156–169.

GUNGA, H.-C., SUTHAU, T., BELLMANN, A., ANDREAS, F., SCHWANEBECK, T., STOINSKI, S., TRIPPEL, T., KIRSCH, K. & HELLWICH, O. (2007). Body mass estimations for *Plateosaurus engelhardti* using laser scanning and 3D reconstriction methods. *Naturwissenschaften* **94**, 623–630.

GUNGA, H.-C., SUTHAU, T., BELLMANN, A., STOINSKI, S., FRIEDRICH, A., TRIPPEL, T., KIRSCH, K. & HELLWICH, O. (2008). A new body mass estimation of *Brachiosaurus brancai* Janensch, 1914 mounted and exhibited at the Museum of Natural History (Berlin, Germany). *Fossil Record* **11**, 33–38.

HAZLEHURST, G. A. & RAYNER, J. M. V. (1992). Flight characteristics of Triassic and Jurassic Pterosauria: an appraisal based on wing shape. *Paleobiology* **18**, 447–463.

VAN HEERDEN, J. (1997). Prosauropods. In *The Complete Dinosaur* (eds J. O. FARLOW and M. K. BRETT-SURMAN), pp. 242–263. Indiana University Press, Indianapolis.

HEMMINGSEN, A. M. (1960). Energy metabolism as related to body size and respiratory surfaces, and its evolution. *Steno Memorial Hospital and Nordinsk Insulin Laboratosium* **9**, 6–110.

HENDERSON, D. M. (1999). Estimating the masses and centers of mass of extinct animals by 3-D mathematical slicing. *Paleobiology* **25**, 88–106.

HENDERSON, D. (2003). Footprints, trackways, and hip heights of bipedal dinosaurs—testing hip height predictions with computer models. *Ichnos* **10**, 99–114.

HENDERSON, D. M. (2004). Tipsy punters: sauropod dinosaur pneumaticity, buoyancy and aquatic habits. *Biology Letters* **271**, S180–S183.

HENDERSON, D. M. (2006). Burly gaits: centers of mass, stability, and the trackways of sauropod dinosaurs. *Journal of Vertebrate Paleontology* **26**, 907–921.

HENDERSON, D. M. (2010). Pterosaur body mass estimates from three-dimensional mathematical slicing. *Journal of Vertebrate Paleontology* **30**, 768–785.

HENDERSON, D. M. (2014). Duck soup: the floating fates of hadrosaurs and ceratopsians at dinosaur Provincial Park. In *Hadrosaurs* (eds D. A. EBERTH and D. C. EVANS), pp. 459–463. Indiana University Press, Bloomington.

HENDERSON, D. M. (2018). A buoyancy, balance and stability challenge to the hypothesis of a semi-aquatic *Spinosaurus* Stromer, 1915 (Dinosauria: Theropoda). *PeerJ* **6**, e5409.

HENDERSON, D. M. & SNIVELY, E. (2004). *Tyrannosaurus* en pointe: allometry minimized rotational inertia of large carnivorous dinosaurs. *Biology Letters* **271**, S57–S60.

*HOLTZ, T. R. (1994). The phylogenetic position of the Tyrannosauridae: implications for theropod systematics. *Journal of Paleontology* **68**, 1100–1117.

HONE, D. W. E. (2012). Variation in the tail length of non-avian dinosaurs. *Journal of Vertebrate Paleontology* **32**, 1082–1089.

HONE, D. W. E., KEESEY, T. M., PISANI, D. & PURVIS, A. (2005). Macroevolutionary trends in the Dinosauria: Cope's rule. *Journal of Evolutionary Biology* **18**, 587–595.

HONE, D. W. E., DYKE, G. J., HADEN, M. & BENTON, M. J. (2008). Body size evolution in Mesozoic birds. *Journal of Evolutionary Biology* **21**, 618–624.

HOPKINS, S. S. B. (2018). Estimation of Body size in fossil mammals. In *Methods in Paleoecology: Reconstructing Cenozoic Terrestrial Environments and Ecological Communities* (eds D. A. CROFT, D. F. SU and S. W. SIMPSON), pp. 7–22. Springer International Publishing, Cham.

HURLBURT, G. (1999). Comparison of body mass estimation techniques, using recent reptiles and the pelycosaur *Edaphosaurus boanerges*. *Journal of Vertebrate Paleontology* **19**, 338–350.

HUTCHINSON, J. R. (2012). On the inference of function from structure using biomechanical modelling and simulation of extinct organisms. *Biology Letters* **8**, 115–118.

HUTCHINSON, J. R., NG-THOW-HING, V. & ANDERSON, F. C. (2007). A 3D interactive method for estimating body segmental parameters in animals: application to the turning and running performance of *Tyrannosaurus rex*. *Journal of Theoretical Biology* **246**, 660–680.

HUTCHINSON, J. R., BATES, K. T., MOLNAR, J., ALLEN, V. & MAKOVICKY, P. J. (2011). A computational analysis of limb and body dimensions in *Tyrannosaurus rex* with implications for locomotion, ontogeny, and growth. *PLoS One* **6**, e26037.

*HWANG, S. H., NORELL, M. A., QIANG, J. & KEQIN, G. (2002). New specimens of *Microraptor zhaoianus* (Theropoda: Dromaeosauridae) from Northeastern China. *American Museum Novitates* **2002**, 1–44.

JANENSCH, W. (1935). Die Schädel der Sauropoden *Brachiosaurus*, *Barosaurus* und *Dicraeosaurus* aus den Tendaguruschichten Deutsch-Ostafrikas. *Palaeontographica, Supplement* **7**, 147–298.

JERISON, H. J. (1969). Brain evolution and dinosaur brains. *The American Naturalist* **103**, 575–588.

*JI, Q., JI, S. A., LÜ, J. C., YOU, H.-L., CHEN, W., LIU, Y. Q. & LIU, Y. X. (2005). First avialian bird from China. *Geological Bulletin of China* **24**, 197–210.

Joint Committee for Guides in Metrology (JCGM 200:2012) (2012). *International Vocabulary of Metrology: Basic and General Concepts and Associated Terms*, 3rd Edition (VIM3), p. 91. Sèvres: Bureau International des Poids et Mesures.

KILBOURNE, B. M. & MAKOVICKY, P. J. (2010). Limb bone allometry during the postnatal ontogeny in non-avian dinosaurs. *Journal of Anatomy* **217**, 135–152.

KLEIBER, M. (1947). Body size and metabolic rate. *Physiological Reviews* **27**, 511–541.

KLEIN, N., REMES, K., GEE, C. T. & SANDER, P. M. (2011). *Biology of the Sauropod Dinosaurs: Understanding the Life of Giants*, p. 331. Indiana University Press, Bloomington, ID.

KUBO, T. & KUBO, M. O. (2016). Nonplantigrade foot posture: a constraint on dinosaur body size. *PLoS One* **11**, e0145716.

LACOVARA, K. J., LAMANNA, M. C., IBIRICU, L. M., POOLE, J. C., SCHROETER, E. R., ULLMANN, P. V., VOEGELE, K. K., BOLES, Z. M., CARTER, A. M., FOWLER, E. K., EGERTON, V. M., MOYER, A. E., COUGHENOUR, C. L., SCHEIN, J. P., HARRIS, J. D., et al. (2014). A gigantic, exceptionally complete titanosaurian sauropod dinosaur from southern Patagonia, Argentina. *Scientific Reports* **4**, 6196.

LARRAMENDI, A. & PALOMBO, M. R. (2015). Body size, biology and encephalization quotient of *Palaeoloxodon* ex gr. *P. falconeri* from Spinagallo cave (Hyblean plateau, Sicily). *Hystrix-Italian Journal of Mammalogy* **26**, 102–109.

LAURIN, M. (2004). The evolution of body size, Cope's rule and the origin of amniotes. *Systematic Biology* **53**, 594–622.

LEE, M. S. Y., CAU, A., NAISH, D. & DYKE, G. J. (2014). Sustained miniaturization and anatomical innovation in the dinosaurian ancestors of birds. *Science* **345**, 562–566.

LONGRICH, N. (2010). The function of large eyes in *Protoceratops*: a nocturnal Ceratopsian. In *New Perspectives on Horned Dinosaurs: The Royal Tyrrell Museum Ceratopsian Symposium* (eds M. J. RYAN, B. J. CHINNERY-ALLGEIER and D. A. EBERTH), pp. 308–327. Indiana Univeristy Press, Bloomington.

MACAULAY, S., HUTCHINSON, J. R. & BATES, K. T. (2017). A quantitative evaluation of physical and digital approaches to centre of mass estimation. *Journal of Anatomy* **231**, 758–775.

MADSEN, J. H. JR. (1976). *Allosaurus fragilis*: a revised osteology. *Utah Geological Survey Bulletin* **109**, 1–163.

MAIDMENT, S. C. R. & BARRETT, P. M. (2012). Does morphological convergence imply functional similarity? A test using the evolution of quadrupedalism in ornithischian dinosaurs. *Proceedings of the Royal Society B: Biological Sciences* **279**, 3765–3771.

MAIDMENT, S. C. R., LINTON, D. H., UPCHURCH, P. & BARRETT, P. M. (2012). Limb-bone scaling indicates diverse stance and gait in quadrupedal ornithischian dinosaurs. *PLoS One* **7**, e36904.

MAIDMENT, S. C. R., HENDERSON, D. M. & BARRETT, P. M. (2014). What drove reversions to quadrupedality in ornithischian dinosaurs? Testing hypotheses using centre of mass modelling. *Naturwissenschaften*, **101**, 989–1001.

*MAKOVICKY, P. J., APESTEGUÍA, S. & AGNOLÍN, F. L. (2005). The earliest dromaeosaurid theropod from South America. *Nature* **437**, 1007–1011.

MALLISON, H. (2010). The digital *Plateosaurus* I: body mass, mass distribution, and posture assessed using CAD and CAE on a digitally mounted complete skeleton. *Palaeontologia Electronica* **13**, 26p.

MALLISON, H. (2011a). Defense capabilities of Kentrosaurus aethiopicus Hennig, 1915. *Palaeontologia Electronica* **14**, 1–25.

MALLISON, H. (2011b). Plateosaurus in 3D: how CAD models and kinetic-dynamic Modeling bring an extinct animal to life. In *Biology of the Sauropod Dinosaurs:*

Understanding the Life of Giants (eds N. KLEIN, K. REMES, C. T. GEE and P. M. SANDER), pp. 219–236. Indiana University Press, Bloomington.

MALLISON, H. (2011c). Rearing giants: kinetic-dynamic modeling of sauropod bipedal and tripodal poses. In *Biology of the Sauropod Dinosaurs: Understanding the Life of Giants* (eds N. KLEIN, K. REMES, C. T. GEE and P. M. SANDER), pp. 237–250. Indiana University Press, Bloomington.

MALLISON, H. & WINGS, O. (2014). Photogrammetry in paleontology: a practical guide. *Journal of Paleontological Techniques* 12, 1–31.

MANNION, P. D. & UPCHURCH, P. (2010). Completeness metrics and the quality of the sauropodomorph fossil record through geological and historical time. *Paleobiology* 36, 283–302.

DE MARGERIE, E., SANCHEZ, S., CUBO, J. & CASTANET, J. (2005). Torsional resistance as a principal component of the structural design of long bones: comparative multivariate evidence in birds. *The Anatomical Record Part A* 282A, 49–66.

MASSARE, J. A. (1988). Swimming capabilities of Mesozoic marine reptiles: implications for method of predation. *Paleobiology* 14, 187–205.

MAZZETTA, G. V., FARIÑA, R. A. & VIZCAÍNO, S. F. (1998). On the palaeobiology of the South American horned theropod *Carnotaurus sastrei* Bonaparte. *Gaia* 15, 185–192.

MAZZETTA, G. V., CHRISTIANSEN, P. & FARIÑA, R. A. (2004). Giants and bizarres: body size of some southern South American Cretaceous dinosaurs. *Historical Biology* 16, 71–83.

MCCLAIN, C. R. & BOYER, A. G. (2009). Biodiversity and body size are linked across metazoans. *Proceedings of the Royal Society B* 276, 2209–2215.

MCDONALD, A. T. (2012). The status of *Dollodon* and other basal iguanodonts (Dinosauria: Ornithischia) from the upper Wealden beds (Lower Cretaceous) of Europe. *Cretaceous Research* 33, 1–6.

MCINTOSH, J. S., BRETT-SURMAN, M. K. & FARLOW, J. O. (1997). Sauropods. In *The Complete Dinosaur* (eds J. O. FARLOW and M. K. BRETT-SURMAN), pp. 264–190. Indiana University Press, Indianapolis.

MILLIEN, V. (2008). The largest among the smallest: the body mass of the giant rodent *Josephoatigasia monesi*. *Proceedings of the Royal Society B* 275, 1953–1955.

MILLIEN, V. & BOVY, H. (2010). When teeth and bones disagree: body mass estimates in a giant extinct rodent. *Journal of Mammalogy* 91, 11–18.

MOTANI, R. (2001). Estimating body mass from silhouettes: testing the assumption of elliptical body cross-sections. *Paleobiology* 27, 735–750.

MYHRVOLD, N. P. (2013). Revisiting the estimation of dinosaur growth rates. *PLoS One* 8, e81917.

*NORELL, M. A. & MAKOVICKY, P. J. (1999). Important features of the dromaeosaur skeleton II: information from newly collected specimens of *Velociraptor mongoliensis*. *American Museum Novitates* 3282, 1–45.

*NOVAS, F. E. & PUERTA, P. F. (1997). New evidence cencerning avian origins from the late cretaceous of Patagonia. *Nature* 387, 390–392.

*NOVAS, F. E., POL, D., CANALE, J. I., PORFIRI, J. D. & CALVO, J. O. (2009). A bizarre Cretaceous theropod dinosaur from Patagonia and the evolution of Gondwanan dromaeosaurids. *Proceedings of the Royal Society B: Biological Sciences* 276, 1101–1107.

O'GORMAN, E. J. & HONE, D. W. E. (2012). Body size distribution of the dinosaurs. *PLoS One* 7, e51925.

ŐSI, A., PRONDVAI, E., BUTLER, R. & WEISHAMPEL, D. B. (2012). Phylogeny, histology and inferred body size evolution in a new rhabdodontid dinosaur from the Late Cretaceous of Hungary. *PLoS One* 7, e44318.

OTERO, A., CUFF, A. R., ALLEN, V., SUMNER-ROONEY, L., POL, D. & HUTCHINSON, J. R. (2019). Ontogenetic changes in the body plan of the sauropodomorph dinosaur *Mussaurus patagonicus* reveal shifts of locomotor stance during growth. *Scientific Reports* 9, 7614.

PADIAN, K. & HORNER, J. R. (2014). The species recognition hypothesis explains exaggerated structures in non-avialan dinosaurs better than sexual selection does. *Comptes Rendus Palevol* 13, 97–107.

PAUL, G. S. (1988). *Predatory Dinosaurs of the World: A Complete Illustrated Guide*. Simon & Schuster, New York.

PAUL, G. S. (1990). Massospondylus. In *Encyclopedia of Dinosaurs* (ed. P. DODSON). Beekman House, New York.

PAUL, G. S. (1997). Dinosaur models: the good, the bad, and using them to estimate the mass of dinosaurs. In *Dinofest International* (eds D. L. WOLBERG, E. STUMP and G. D. ROSENBERG), pp. 129–142. Phoenix: The Academy of Natural Sciences, Arizona State University.

PAUL, G. S. (1998). Terramegathermy and Cope's Rule in the lane of titans. *Modern Geology* 23, 179–218.

PAUL, G. S. (2008). A revised taxonomy of the iguanodont dinosaur genera and species. *Cretaceous Research* 29, 192–216.

PAUL, G. S. (2010). *The Princeton Field Guide to Dinosaurs*, 1st Edition. Princeton University Press, Princeton.

PAUL, G. (2019). Determining the largest known land animal: a critical comparison of differing methods for restoring the volume and mass of extinct animals. *Annals of Carnegie Museum* 85(4), 335–358.

*PERLE, A., NORELL, M. & CLARK, J. M. (1999) *A New Maniraptoran Theropod, Achillobator Giganticus (Dromaeosauridae), from the Upper Cretaceous of Burkhant*, pp. 1–104. Ulaanbaatar: National University of Mongolia.

PERSONS, W. S. IV & CURRIE, P. J. (2011). The tail of *Tyrannosaurus*: reassessing the size and locomotive importance of the m. caudofemoralis in non-avian theropods. *The Anatomical Record* 294, 119–131.

PETERS, R. H. (1983). *The Ecological Implications of Body Size*. Cambridge University Press, New York.

PONTZER, H., ALLEN, V. & HUTCHINSON, J. R. (2009). Biomechanics of running indicates endothermy in bipedal dinosaurs. *PLoS One* 4, 1–9.

PUTTICK, M. N., THOMAS, G. H. & BENTON, M. J. (2014). High rates of evolution preceded the origin of birds. *Evolution* 68, 1497–1510.

RICHARDS, H. L., WELLS, R. T., EVANS, A. R., FITZGERALD, E. M. G. & ADAMS, J. W. (2019). The extraordinary osteology and functional morphology of the limbs in Palorchestidae, a family of strange extinct marsupial giants. *PLoS One* 14, e0221824.

RINDERKNECHT, A. & BLANCO, R. E. (2008). The largest fossil rodent. *Proceedings of the Royal Society B: Biological Sciences* 275, 923–928.

ROMANO, M. & MANUCCI, F. (2019). Resizing *Lisowicia bojani*: volumetric body mass estimate and 3D reconstruction of the giant Late Triassic dicynodont. *Historical Biology* 1–6.

ROMANO, M. & RUBIDGE, B. (2019). First 3D reconstruction and volumetric body mass estimate of the tapinocephalid dinocephalian *Tapinocaninus pamelae* (Synapsida: Therapsida). *Historical Biology* 1–8.

ROMER, A. S. & PRICE, L. I. (1940). Review of the Pelycosauria. *Geological Society of America* Special Paper 28.

*ROYO-TORRES, R., COBOS, A. & ALCALÁ, L. (2006). A giant European dinosaur and a new sauropod clade. *Science* 314, 1925–1927.

SALGADO, L. & BONAPARTE, J. F. (1991). A new dicraeosaurid sauropod, *Amargasaurus cazaui* gen. et sp. nov., from the La Amarga Formation, Neocomian of Neuquén Province, Argentina. *Ameghiniana* 28, 333–346.

SAMPSON, S. D. (1997). Bizarre structures and dinosaur evolution. In *Dinofest International* (eds D. L. WOLBERG, E. STUMP and G. D. ROSENBERG), pp. 39–45. Phoenix: The Academy of Natural Sciences, Arizona State University.

SAMPSON, S. D. (2009). *Dinosaur Odyssey: Fossil Threads in the Web of Life*. University of California Press, Berkeley.

SANDER, P. M., CHRISTIAN, A., CLAUSS, M., FECHNER, R., GEE, C. T., GRIEBELER, E.-M., GUNGA, H.-C., HUMMEL, J., MALLISON, H., PERRY, S. F., PREUSCHOFT, H., RAUHUT, O. W. M., REMES, K., TÜTKEN, T., WINGS, O. & WITZEL, U. (2011). Biology of the sauropod dinosaurs: the evolution of gigantism. *Biological Reviews* 86, 117–155.

SEEBACHER, F. (2001). A new method to calculate allometric length-mass relationships of dinosaurs. *Journal of Vertebrate Paleontology* 21, 51–60.

SEEBACHER, F. (2003). Dinosaur body temperatures: the occurrence of endothermy and ectothermy. *Paleobiology* 29, 105–122.

SELLERS, W. I., HEPWORTH-BELL, J., FALKINGHAM, P. L., BATES, K. T., BRASSEY, C. A., EGERTON, V. M. & MANNING, P. L. (2012). Minimum convex hull mass estimations of complete mounted skeletons. *Biology Letters* 8, 842–845.

SELLERS, W. I., MARGETTS, L., CORIA, R. A. & MANNING, P. L. (2013). March of the titans: the locomotor capabilities of sauropod dinosaurs. *PLoS One* 8, e78733.

SELLERS, W. I., POND, S. B., BRASSEY, C. A., MANNING, P. L. & BATES, K. T. (2017). Investigating the running abilities of tyrannosaurus rex using stress-constrained multibody dynamic analysis. *PeerJ* 5, e3420.

SERENO, P. C., DUTHEIL, D. B., IAROCHENE, M., LARSSON, H. C. E., LYON, G. H., MAGWENE, P. M., SIDOR, C. A., VARRICCHIO, D. J. & WILSON, J. A. (1996). Predatory dinosaur from the Sahara and late cretaceous fanuan differentiation. *Science* 272, 986–991.

SERENO, P. C., MARTÍNEZ, R. N. & ALCOBER, O. A. (2012). Osteology of Eoraptor lunensis (Dinosauria, Sauropodomorpha). *Journal of Vertebrate Paleontology* 32, 83–179.

SERRANO, F. J., PALMQVIST, P. & SANZ, J. L. (2015). Multivariate analysis of neognath skeletal measurements: implications for body mass estimation in Mesozoic birds. *Zoological Journal of the Linnean Society* 173, 929–955.

SMITH, R. J. (1980). Rethinking allometry. *Journal of Theoretical Biology* 87, 97–111.

SMITH, R. J. (2002). Lead review: estimation of body mass in paleontology. *Journal of Human Evolution* 43, 271–287.

SMITH, F. A., BOYER, A. G., BROWN, J. H., COSTA, D. P., DAYAN, T., ERNEST, S. K. M., EVANS, A. R., FORTELIUS, M., GITTLEMAN, J. L., HAMILTON, M. J., HARDING, L. E., LINTULAAKSO, K., LYONS, S. K., MCCAIN, C., OKIE, J. G., SAARINEN, J. J., SIBLY, R. M., STEPHENS, P. R., THEODOR, J. & UHEN, M. D. (2010). The evolution of maximum body size of terrestrial mammals. *Science* 330, 1216–1219.

SNIVELY, E., O'BRIEN, H., HENDERSON, D. M., MALLISON, H., SURRING, L. A., BURNS, M. E., HOLTZ, T. R. JR., RUSSELL, A. P., WITMER, L. M., CURRIE, P. J., HARTMAN, S. A. & COTTON, J. R. (2019). Lower rotational inertia and larger leg muscles indicate more rapid turns in tyrannosaurids than in other large theropods. *PeerJ* 7, e6432.

SOOKIAS, R. B., BENSON, R. B. J. & BUTLER, R. J. (2012a). Biology, not environment, drives major patterns in maximum tetrapod body size through time. *Biology Letters* 8, 674–677.

SOOKIAS, R. B., BUTLER, R. J. & BENSON, R. B. J. (2012b). Rise of dinosaurs reveals major body-size transitions are driven by passive processes of trait evolution. *Proceedings of the Royal Society B: Biological Sciences* 279, 2180–2187.

STOINSKI, S., SUTHAU, T. & GUNGA, H.-C. (2011). Reconstructing Body volume and surface area of dinosaurs using laser scanning and photogrammetry. In *Biology of the Sauropod Dinosaurs: Understanding the Life of Giants* (eds N. KLEIN, K. REMES, C. T. GEE and P. M. SANDER), pp. 94–104. Indiana University Press, Bloomington.

THERRIEN, F. & HENDERSON, D. M. (2007). My theropod is bigger than yours…or not: estimating body size from skull length in theropods. *Journal of Vertebrate Paleontology* **27**, 108–115.

TSAI, H. P. & HOLLIDAY, C. M. (2015). Articular soft tissue anatomy of the archosaur hip joint: structural homology and functional implications. *Journal of Morphology* **276**, 601–630.

TSCHOPP, E., MATEUS, O. & BENSON, R. B. J. (2015). A specimen-level phylogenetic analysis and taxonomic revision of Diplodocidae (Dinosauria, Sauropoda). *PeerJ* **3**, e857.

TURNER, A. H., POL, D., CLARKE, J. A., ERICKSON, G. M. & NORELL, M. A. (2007). A basal dromaeosaurid and size evolution preceding avian flight. *Science* **317**, 1378–1381.

VANBUREN, C. S., CAMPIONE, N. E. & EVANS, D. C. (2015). Head size, weaponry, and cervical adaptation: testing craniocervical evolutionary hypotheses in Ceratopsia. *Evolution* **69**, 1728–1744.

VARRICCHIO, D. J., MARTIN, A. J. & KATSURA, Y. (2007). First trace and body fossil evidence of a burrowing, denning dinosaur. *Proceedings of the Royal Society B* **274**, 1361–1368.

VARRICCHIO, D. J., MOORE, J. R., ERICKSON, G. M., NORELL, M. A., JACKSON, F. D. & BORKOWSKI, J. J. (2008). Avian paternal care had dinosaur origin. *Science* **322**, 1826–1828.

VIZCAINO, S. F., BLANCO, R. E., BENDER, J. B. & MILNE, N. (2011). Proportions and function of the limbs of glyptodonts. *Lethaia* **44**, 93–101.

WANG, M., O'CONNOR, J. K., XU, X. & ZHOU, Z. (2019). A new Jurassic scansoriopterygid and the loss of membranous wings in theropod dinosaurs. *Nature* **569**, 256–259.

WARTON, D. I., DUURSMA, R. A., FALSTER, D. S. & TASKINEN, S. (2012). Smatr 3-an R package for estimation and inference about allometric lines. *Methods in Ecology and Evolution* **3**, 257–259.

WEDEL, M. J. (2003). Vertebral pneumaticity, air sacs, and the physiology of sauropod dinosaurs. *Paleobiology* **29**, 243–255.

WIEDEMANN, A., SUTHAU, T. & ALBERTZ, J. (1999). Photogrammetric survey of dinosaur skeletons. *Mitteilungen aus dem Museum für Naturkunde der Humboldt-Universität Berlin, Geowissenschaftliche Reihe* **2**, 113–119.

*XU, X. & NORELL, M. A. (2004). A new troodontid dinosaur from China with avian-like sleeping posture. *Nature* **431**, 838–841.

*XU, X., WANG, X.-L. & WU, X.-C. (1999). A dromaeosaurid dinosaur with a filamentous integument from the Yixian Formation of China. *Nature* **401**, 262–266.

*XU, X., ZHOU, Z. & WANG, X. (2000). The smallest known non-avian theropod dinosaur. *Nature* **408**, 705–708.

*XU, X., NORELL, M. A., WANG, X.-L., MAKOVICKY, P. J. & WU, X.-C. (2002). A basal troodontid from the Early Cretaceous of China. *Nature* **415**, 780–784.

XU, X., YOU, H., DU, K. & HAN, F. (2011). An *Archaeopteryx*-like theropod from China and the origin of Avialae. *Nature* **475**, 465–470.

*XU, X., WANG, K., ZHANG, K., MA, Q., XING, L., SULLIVAN, C., HU, D., CHENG, S. & WANG, S. (2012). A gigantic feathered dinosaur from the lower cretaceous of China. *Nature* **484**, 92–95.

ZANNO, L. E. & MAKOVICKY, P. J. (2011). Herbivorous ecomorphology and specialization patterns in theropod dinosaur evolution. *Proceedings of the National Academy of Sciences* **108**, 232–237.

ZHANG, F., ZHOU, Z., XU, X., WANG, X. & SULLIVAN, C. (2008). A bizarre Jurassic maniraptoran from China with elongate ribbon-like feathers. *Nature* **455**, 1105–1108.

*ZHOU, Z. & ZHANG, F. (2002). Largest bird from the Early Cretaceous and its implications for the earliest avian ecological diversification. *Naturwissenschaften* **89**, 34–38.

VIII. Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1. Metadata and mass estimates of volumetric-density (VD)-based reconstructions generated between 1905 and 2020 (see ReadMe file for full explanation).

Appendix S2. R Source code.