Cope's rule and the adaptive landscape of dinosaur body size evolution

by ROGER B. J. BENSON[1*], GENE HUNT[2], MATTHEW T. CARRANO[2], NICOLÁS CAMPIONE[3]

[1]Department of Earth Sciences, University of Oxford, South Parks Road, Oxford OX2 3AN, United Kingdom; e-mail: roger.benson@ earth.ox.ac.uk

[2]Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, P.O. Box 37012, MRC 121, Washington, DC, USA; emails: HuntE@si.edu, CARRANOM@si.edu

[3]Palaeobiology Programme, Department of Earth Sciences, Uppsala University, Villavägen 16, 752 36 Uppsala, Sweden; email: nicolas.campione@geo.uu.se

*corresponding author

Abstract: The largest known dinosaurs weighed at least 20 million times as much as the smallest, indicating exceptional phenotypic divergence. Previous studies have focused on extreme giant sizes, tests of Cope's rule, and miniaturization on the line leading to birds. We use non-uniform macroevolutionary models based on Ornstein-Uhlenbeck and trend processes to unify these observations, asking: what patterns of evolutionary rates, directionality and constraint explain the diversification of dinosaur body mass?

We find that dinosaur evolution is constrained by attraction to discrete body size optima that undergo rare, but abrupt, evolutionary shifts. This model explains both the rarity of multi-lineage directional trends, and the occurrence of abrupt directional excursions during the origins of groups such as tiny pygostylian birds and giant sauropods. Most expansion of trait space results from rare, constraint-breaking innovations in just a small number of lineages. These lineages shifted rapidly into novel regions of trait space, occasionally to small sizes, but most often to large or giant sizes. As with Cenozoic mammals, intermediate body sizes were typically attained only transiently by lineages on a trajectory from small to large size. This demonstrates that bimodality in the macroevolutionary adaptive landscape for land vertebrates has existed for more than 200 million years.

Key words: dinosaur, body size, Cope's rule, adaptive landscape, Ornstein-Uhlenbeck models, trend models, phylogenetic Bayesian information criterion.

CRETACEOUS dinosaurs spanned more than six orders of magnitude in body size, from an estimated 15 g in some birds to at least 40 tonnes (Bakker 1971; Anderson *et al*. 1985; Bates *et al*. 2015) and perhaps as much as 90 tonnes (Colbert 1962; Mazzetta *et al*. 2004; Benson *et al*. 2014*a*; Lacovara *et al*. 2014; Carballido *et al*. 2017) in sauropods. This represents extraordinary phenotypic divergence from a Triassic ancestor living 140 million years earlier and weighing 10–30 kg (Sereno 1997; Benson *et al*. 2014*a*). Body size influences many aspects of animal biology including physiology, ecology, and life history energetics (e.g. Brown 1995), so exceptional variation in body size

signifies considerable variation in biological processes. *Mellisuga helenae* (bee hummingbird), the smallest living dinosaur, weighs 2 g (del Hoyo *et al*. 1999) and extends the range of body sizes achieved by dinosaurs to 20 million-fold, underscoring the evolutionary versatility of dinosaurs, and of the vertebrate bauplan in general.

Significant research has focused on estimating the extremely large body masses attained by some dinosaurs (Colbert 1962; Bakker 1971; Anderson *et al*. 1985; Burness *et al*. 2001; Mazetta *et al*. 2004; Carpenter 2006; Campione & Evans 2012; Benson *et al*. 2014*a*; Woodruff & Foster 2014; Bates *et al*. 2015) and on framing hypotheses of the physiological, environmental, ecological, and life history factors that made large sizes possible (Alexander 1989; Janis & Carrano 1992; Burness *et al*. 2001; Sander & Clauss 2008; O'Connor 2009; Sander *et al*. 2010; Werner & Griebeler 2011; Sookias *et al*. 2012; Erickson 2014). Because some taxa attained giant sizes, and because much smaller sizes appeared on the evolutionary line leading to birds, quantitative research into dinosaur body size evolution has generally been divided between studies of avian miniaturization (Turner *et al*. 2007; Novas *et al*. 2012; Lee *et al*. 2014; Puttick *et al*. 2014) and studies examining multi-lineage directional trends, especially trends of body size increase ('Cope's rule'; Hone *et al*. 2005; Carrano 2006; Zanno & Makovicky 2013; De Souza & Santucci 2014).

However, this division is artificial. In fact, these topics represent facets of a single, broader goal to characterize patterns of body size evolution in dinosaurs and their underlying macroevolutionary adaptive landscape. Characterization of these patterns for dinosaurs has lagged significantly

behind that for land mammals (e.g. Alroy 1999; Smith *et al*. 2012; Saarinen *et al*. 2014; Baker *et al*. 2015), which also evolved to a large range of body sizes. Comparison of these patterns provides a test of the hypothesis that a distinct dinosaurian life history (Janis & Carrano 1992; Varicchio 2011) resulted in a unique adaptive landscape that drove idiosyncratic macroevolutionary patterns during the Mesozoic (Codron *et al*. 2012; O'Gorman & Hone 2013).

Characterizing patterns of dinosaur body size evolution is also relevant to longstanding questions about how evolution has generated the phenomenal disparity of organismal phenotypes observed both today and in the geological past (Foote 1997*a*). For example, palaeontological time series indicate that early rapid increases in disparity are common among major animal groups (e.g. Hughes *et al*. 2013), but it is not clear whether these patterns primarily result from high early rates of evolution (an 'early burst' model; Harmon *et al*. 2010) or from the existence of constraints on the range of phenotypes attainable by a clade, such that trait space becomes rapidly saturated (e.g. Slater 2013; Oyston *et al*. 2015). Dinosaurs have well-resolved phylogenies compared to many fossil groups, and so provide a model system for addressing this question using phylogenetic comparative methods.

Here, we use non-uniform, model-based approaches to quantify patterns of dinosaur body size evolution, asking what macroevolutionary processes drove the diversification of dinosaur body sizes during the Mesozoic? These models are non-uniform because they allow multiple macroevolutionary regimes to exist on a phylogeny, each with its own set of model parameters. We specifically compare models based on Ornstein-

Uhlenbeck (OU) dynamics (Hansen 1997; Butler & King 2004; Beaulieu *et al*. 2012) to those based on directional trend-like dynamics (Pagel 2002; Hunt 2008; Hunt & Carrano 2010). These models imply different interpretations of long-term shifts in the body size distribution of species. Trend models encompass a style of macroevolution that is unbounded and based on directional evolution, whereas OU models describe constrained phenotypic divergence within adaptive zones (Hansen 2013). Trend models ascribe long-term directionality to a pervasive tendency for trait values to increase (or decrease) over time, and are supported when set of independently evolving lineages show changes in trait values that go preferentially in one direction over another. This is consistent with many explanations of Cope's rule, which focus on the broad advantages of ever larger body size (e.g., Brown & Maurer 1986; Van Valkenburgh *et al* 2004; Kingsolver & Pfennig 2004). In contrast, multi-peak OU models construe divergence as resulting from relatively few discrete shifts to new adaptive zones, with constrained evolutionary change occurring within these zones. This is consistent with more nuanced attempts to explain the frequent evolution of large size in land vertebrates (Stanley 1973; Hansen 1997; Alroy 1999). Within this framework, directionality occurs when a shift to a new adaptive zone occurs, and is limited to just a few instances or branches in the clade, rather than reflecting a general evolutionary tendency across multiple lineages.

Abbreviations.

Measurements. FAP, minimum anteroposterior diameter of femoral shaft; FC, minimum circumference around femoral shaft; FL, proximodistal length of

femur; FML, minimum mediolateral diameter of femoral shaft; TAP, minimum anteroposterior diameter of tibia shaft; HAP, minimum anteroposterior diameter of humeral shaft; HC, minimum circumference around humeral shaft; HL, proximodistal length of humerus; HML, minimum mediolateral diameter of humeral shaft; RAP, minimum anteroposterior diameter of radial shaft; RC, minimum circumference around radial shaft; RL, proximodistal length of radius; RML, minimum mediolateral diameter of radial shaft; TC, minimum circumference around tibia shaft; TL, proximodistal length of tibia; TML, minimum mediolateral diameter of tibia shaft;

Models of trait evolution. BM, Brownian motion model; OU, Ornstein-Uhlenbeck model; OU1, single-peak Ornstein-Uhlenbeck model (single, fixed θ); OUM, multi-peak Ornstein-Uhlenbeck model (multiple regimes within individual θ values) with fixed values of α and σ; OUMV, multi-peak Ornstein-Uhlenbeck model with fixed value of α and σ as a free parameter; OUMA, multi-peak Ornstein-Uhlenbeck model with fixed values of σ and α as a free parameter; OUMVA, multi-peak Ornstein-Uhlenbeck model with both σ and α as free parameters.

Model parameters. α, constraint or attraction parameter of OU model; θ, trait mean, or 'optimum' of OU model; λ, Pagel's λ, phylogenetic signal parameter; σ, Brownian variance or rate parameter of a BM or OU model.

Optimality criteria. AIC, Akaike's information criterion; AICc, Akaike's information criterion for finite sample sizes; BIC, Bayesian information criterion; pBIC, phylogenetic Bayesian information criterion.

METHODS

Phylogeny.

All our analyses were conducted using an updated version of the composite phylogeny of Benson *et al*. (2014*a*; see Appendix S1). Preliminary analyses indicated highly complex patterns of body size evolution in Late Cretaceous theropods. This is consistent with the high frequency of large changes in trait value on single-branches documented among Late Cretaceous theropods in our previous study (Benson *et al*. 2014*a*). Because of this complexity, when post-Aptian theropods (i.e. those occurring from the late Early Cretaceous onwards) were included in our analyses, meaningful regime shifts could not be recognized. In principle, frequent, large changes in trait values could result from either fast background rates and low constraint (under Brownian motion) or from frequent shifts between short-lived, constrained regimes (under multi-peak Ornstein-Uhlenbeck models; explained below. It should be difficult to distinguish between the two models when regime shifts are frequent, and the equilibrial, constrained phase become difficult to recognise. Therefore we limited analyses of Theropoda to taxa occurring only up to the end of the Aptian. This amounts to analyzing a large time slice, that extends from the Triassic up to the Aptian. Time-slicing of fossil phylogenies is appropriate if patterns of evolution change through time and younger patterns have the

potential to over-write the signatures of older patterns. This has been proposed for cladistic biogeographic methods (Hunn & Upchurch 2001; Upchurch & Hunn 2002) and approaches to estimating diversification rate shifts using tree symmetry (Tarver & Donoghue 2011), and is also appropriate here because it is impossible for the evolutionary patterns of lineages occurring in a later interval to actually change those occurring in an earlier interval. Nevertheless, complexity is evident in patterns of post-Aptian theropod evolution (Benson *et al*. 2014*a*), and we take account of this in our interpretations (see *Discussion*).

To reduce the size of the dataset and make our analyses computationally tractable, we subsetted our data as follows: (1) Triassic–Jurassic Dinosauria (intended only to establish the existence of a plesiomorphic dinosaur body size regime); (2) Triassic–Cretaceous Sauropodomorpha; (3) Triassic–Cretaceous Ornithischia; (4) Triassic–Aptian Theropoda. Subsetting and other phylogenetic functions were performed using functions from the package ape version 4.1 (Paradis *et al*. 2004) and phytools version 0.6-00 (Revell 2012) in R version 3.3.3 (R Core Team 2017).

Our trees contain polytomies that represent areas of continuing uncertainty in dinosaur phylogeny. To accommodate this uncertainty, analyses were conducted multiple times across a set of 20 phylogenies in which these polytomies were resolved at random. Furthermore, two alternative topologies were used for early sauropodomorphs, those of Yates (2007; Apaldetti *et al*. 2013) and Upchurch *et al*. (2007), which vary in the number of taxa included in a monophyletic ‘prosauropod’ clade (more taxa were found as pectinate outgroups to Sauropoda by Yates [2007]). A second

source of uncertainty concerns the ages of terminal taxa in our tree, which are frequently known only within bounds of several million years. Accordingly, we drew ages for each phylogeny from uniform distributions between the maximum and minimum possible ages for each taxon using a custom script.

Time-scaling the phylogeny.

In contrast to the ages of terminal taxa in our phylogeny, which are constrained by the stratigraphy of their occurrences, we can only reconstruct the ages of the nodes in our phylogeny based on indirect information. Various methods have been proposed to assign node ages (= divergence times) to trees of fossil taxa (Bapst 2012, 2013, 2014*a*; Lloyd *et al*. 2016). Bapst (2014*b*) used simulations to determine how well these methods performed compared to the true tree when estimating rates and modes of univariate trait evolution, and recommended using the *cal3* probabilistic method instead of the minimum branch length (mbl) method. Considering only the simulation scenario most similar to our data (Bapst 2014b, fig. 6F: fossils occur as terminal taxa with random times of observation between their apparent first and last appearance dates), the median AICc weight for Brownian motion (compared to OU; models described below) when Brownian motion (BM) was in fact the true generating model was approximately 0.3 when node ages were estimated using mbl, compared to approximately 0.5 when *cal3* was used. Therefore, *cal3* is marginally less biased towards supporting OU than is mbl. Note, however, that these simulations represent essentially a best-case scenario for *cal3*, because phylogenies were simulated under a time-homogeneous birth-death-sampling model (as assumed by *cal3*), whereas it

is likely that real sampling rates, speciation, and extinction rates are highly heterogeneous (e.g. Bapst & Hopkins 2017).

Following the above consideration, we are uncertain as to the best approach to estimating divergence times in dinosaur phylogeny. Here, and previously, we calibrated our tree to stratigraphy using the "mbl" (minimum branch length) method (e.g. Laurin 2004; Bapst 2012), setting a minimum branch duration of 1 Ma (“mbl1”), which results in post-Palaeozoic divergence times for Dinosauria (Benson *et al*. 2014*a*). We also perform our initial analyses (i.e. those using SURFACE and comparing to trend-based models) using two other methods (*cal3* and the extended Hedman method of Lloyd *et al*. 2016). All three methods were applied using the R packages paleotree version 2.9 (Bapst 2012; *cal3*, mbl) and a custom script provided by G. T Lloyd (http://www.graemetlloyd.com/pubdata/functions_7.r).

The *cal3* method uses a birth-death-sampling model (similar to the fossilized birth-death process of Heath *et al*. 2014) to estimate node ages. However, this is only possible when sampling rates, speciation rates, and extinction rates can be estimated *a priori* (Bapst 2013, 2014*b*). We believe this to be difficult for dinosaurs, in which vanishingly few genera or species have occurrences in multiple time intervals (i.e. most dinosaurs are singleton occurrences without meaningful range data).

In spite of this difficulty, Lloyd *et al*. (2016) recently used *cal3* to estimate node ages for a large phylogeny of dinosaurs. Lloyd *et al*. (2016) obtained their sampling, speciation and extinction rates (following Foote 1997*b*) from the *apparent* range-frequency distribution of dinosaur taxa as represented in the Paleobiology Database (paleobiodb.org/). It is likely these

'range' data used by Lloyd *et al*. (2016) at least indirectly reflect variation in the intervals of stratigraphic uncertainty in the placement of specimens, or the occurrence of wastebasket taxa or species misidentification. We do not advocate using this as a substitute for quality-controlled range data on fossil taxon ranges. Nevertheless, the resulting parameter estimates are qualitatively reasonable and we used them to test the sensitivity of our results to choice of time-scaling method, by calibrating a set of dinosaur trees to stratigraphy using *cal3* (extinction and speciation rate = 0.935; sampling rate = 0.018; D. Bapst, pers. comm. 17 April 2017).

Lloyd *et al*. (2016) also presented a modified probabilistic method based on Hedman (2010), which uses the ages of a sequence of outgroups to a node to estimate the age of that node. Lloyd *et al*. (2016) compared the performance of this method to that of *cal3*, demonstrating that it could yield similar estimates of divergence times that are also similar to those obtained by our previous work and here using mbl1 (Benson *et al*. 2014*a*; e.g. an Early Triassic age for root of Dinosauria).

The simulations of Bapst (2014*b*) suggest that all methods for node age estimation in fossil trees result in a bias towards finding support for the Ornstein-Uhlenbeck (OU) model of evolution relative to Brownian motion (Bapst 2014b), and in weak overestimation of rates of evolution. Furthermore, both mbl and *cal3* performed poorly compared to the true tree (median AICc weight for OU of approximately 0.8 when BM is the true model). This occurs because mis-estimation of the phylogeny (including its branch lengths) leads to inflated support for OU (Bapst 2014*b*; Bapst & Hopkins 2017), and no method for estimating the node ages of fossil trees proposed so far performs

perfectly. Nevertheless, depending on the specific analyses that are carried out, this bias might be considered to be small: an AICc weight of 0.3 for BM (described above) amounts to an AICc score for OU that is only 1.7 points better than that for BM. This is a small difference (Burnham & Anderson 2004). Furthermore, BM is a nested case of OU (the constraint parameter α = 0 in BM, but is free to vary in OU), and OU with a near-zero α parameter can be essentially identical to BM (described below), irrespective of its level of support from AICc weights. Therefore, the distinction between the values of α estimated from phylogenies with node ages calibrated using different methods might be a more important than their AICc weights. Because we find extremely strong support for the OU model in most clades, and because we find evidence of generally high values of α, we do not consider that choice of time-scaling method has been influential specifically on our finding of support for OU over BM.

Body mass estimates.

We used the non-phylogenetic versions of scaling equations provided by Campione & Evans (2012) and Campione *et al*. (2014) to estimate dinosaur body masses. These equations estimate tetrapod body mass using the minimum shaft circumferences of the humerus and femur (in quadrupeds) or that of the femur only (in bipeds):

[1] $mass_{quadruped} = (10^{(2.749*\log_{10}(FC + HC) - 1.104)}) / 1000$

[2] $mass_{biped} = (10^{(2.749*\log_{10}(FC * 2^{0.5}) - 1.104)}) / 1000$.

The decision to use non-phylogenetic equations resulted from comparison of non-phylogenetic (i.e. ordinary least squares; OLS) regression models to phylogenetic generalised least squares regression models (Garland & Ives 2001; implemented using the R packages ape version 4.1 and nlme version 3.1-131; Paradis *et al*. 2004; Pinheiro *et al*. 2013; using the tree of Campione & Evans [2012]). Ordinary least squares regression provides a substantially better explanation of the quadrupedal extant tetrapod data than does phylogenetic regression ($AICc_{OLS}$= -268; $AICc_{phylogenetic}$ = -232; AICc is Akaike's information criterion for finite sample sizes; Sugiura 1978; Burnham & Anderson 2004). This is consistent with the lack of support for differing relationships of body mass with stylopodial shaft circumferences among different clades and among taxa with different stances (Campione & Evans 2012). It indicates either that (1) stylopod shaft circumferences and tetrapod body mass are related to each other via a strong functional linkage that is constrained by the physical laws of the universe, with coefficients that do not vary substantially across the phylogeny (Motani & Schmitz 2011; Campione & Evans 2012), or (2) that the relationship between these variables evolves in a non-Brownian fashion. Examination of the residuals of these relationships supports the former hypothesis (physical constraint: the residuals are homoskedastic and phylogenetically normally-distributed, therefore providing no evidence of non-Brownian dynamics).

The ability of stylopodial circumferences to predict live body mass in tetrapods was initially documented by Anderson *et al*. (1985; primarily in mammals) and Campbell & Marcus (1992; in birds). Using a large dataset of

extant reptiles and mammals, Campione & Evans (2012) and Campione *et al*. (2014) showed that the combined humeral and femoral circumference is a robust proxy for estimating body mass that is largely independent of phylogenetic history, gait, and limb posture in non-avian tetrapods. Extant birds have a different scaling relationship between femoral shaft circumference and body mass than do other bipedal tetrapods, possibly due to their subhorizontal femoral orientation (Campione *et al*. 2014). However, the body proportions of non-avialan dinosaurs and most other Mesozoic stem-group birds indicate that they did not possess the apomorphic femoral orientation of extant birds (e.g. Carrano 1998, 2001; Campione *et al*. 2014), suggesting that the non-avian bipedal tetrapod scaling relationship of Campione *et al*. (2014) is appropriate for estimating the body masses of bipedal non-avian dinosaurs (Campione *et al*. 2014).

Previously, we were able to estimate the masses of either 441 or 426 dinosaur specimens (depending on whether questionably facultative quadrupeds were treated as bipeds [requiring less data for their mass estimation] or quadrupeds [requiring more data for mass estimation]). Of these, 310 were included in our phylogeny (Benson *et al*. 2014*a*). In the present study, we extended our dataset of mass estimates to 584 dinosaur specimens by estimating unknown femoral and humeral minimum shaft circumferences using other limb bone measurements. This was done using AICc-based comparisons to find the best generalized least squares regression model for each combination of variables from among the following options (Appendix S1): (1) varying the strength of phylogenetic signal using Pagel's lambda (Pagel 1999); (2) estimating a non-zero intercept or setting

the intercept to zero; and where relevant (3) including stance (quadrupedal or bipedal) or clade assignment (e.g. titanosaur | non-titanosaur; hadrosauroid | non-hadrosauroid; stegosaur | ankylosaur) as a covariate or interaction term.

Sets of model comparisons were conducted across all bipedal dinosaurs, across quadrupedal dinosaurs, and within groups of quadrupedal dinosaurs. Where multiple models of approximately equal goodness were available to predict unknown stylopod minimum shaft circumferences in a single dinosaur specimen, we used the model with the smallest estimated prediction error. The full set of estimates and their prediction errors are provided in Dataset S1.

It is important to account for errors in tip values when evaluating OU models, because failure to account for the error in estimated trait values can lead to spurious favoring of OU over BM-like models (Silvestro *et al*. 2015; Cooper *et al*. 2016). However, the calculation of prediction errors of some of our mass estimates was complicated by the fact that the limb shaft circumference measurements upon which they were based were also estimates, with associated prediction errors of their own.

We calculated the total error of each mass estimate using simulations that accounted for error propagation through multiple rounds of regression. The standard error of masses estimated directly from femoral and humeral circumferences has two contributing sources: (i) error in estimating that species mean femoral and humeral dimensions, and (ii) the error of the regressions used to estimate individual body masses. For nearly all dinosaurs, this species mean is estimated from a single individual animal. We assume within species that the limb dimensions are normally distributed with

a coefficient of variation equal to 5, which translates to a standard deviation on a $\log_{10}$ scale of 0.0217. This is a reasonable magnitude of variation for size-related traits in vertebrates (Yablokov 1974; see also Hunt & Carrano 2010, p. 256). Simulations combine this error in limb dimensions with the error in estimating the regression of body mass on limb shaft circumferences. When femoral/humeral circumference measurements were estimated from other variables via regression, the error was propagated through this regression, and then through the regression of the imputed limb data to produce mass estimates. In this way, the mass estimates are analyzed with uncertainties that reflect how they were calculated. These standard errors are incorporated into the likelihood functions in the standard way by increasing the expected variances of the tips by an amount equal to their squared standard error (O'Meara *et al*. 2006; Ives *et al*. 2007). R functions used to compute these standard errors are provided on Dryad (Benson *et al*. 2017).

Of our 584 mass estimates, 526 were from adult individuals, representing a total of 393 taxa included in our phylogeny. Only mass estimates of adult individuals were used in our analyses. Skeletal maturity was assessed from published histological studies (e.g., Lee & Werning 2008; Benton *et al*. 2010; Erickson *et al*. 2006, 2009a,b, 2010; Osi *et al*. 2012; Werning 2012) and qualitative indicators such as the fusion of neurocentral and neurocranial sutures.

Ornstein-Uhlenbeck (OU) models.

Our macroevolutionary analyses make use of Ornstein-Uhlenbeck (OU), or 'Hansen' models (Hansen 1997; Butler & King 2004; Beaulieu *et al*. 2012),

which include the parameters: Z_0, the estimated trait value at the root of the tree; β or σ^2, the Brownian variance, which describes the rate at which trait variance is expected to accumulate along phylogenetic lineages in Brownian motion models (Felsenstein 1985), and is a measure of stochastic spread of trait values over time (Hansen 1997; Beaulieu *et al*. 2012; Hunt 2012); θ, a macroevolutionary trait 'optimum'; and α, the strength of attraction to θ. Under OU, the expected change in a trait X within an infinitesimal time interval between t and t + dt is dX(t), and:

[3] $dX(t) = \alpha[\theta - X(t)]dt + \sigma dB(t)$.

Where X(t) is the value of X at time t, and the term dB(t) is a random variable with a mean of zero and variance of $\sigma^2 dt$ (Butler & King 2004; Beaulieu *et al*. 2012). This formulation includes a term describing trait attraction towards θ, which is the product of α and the difference between X(t) and θ:

[4] $\alpha[\theta - X(t)]dt$.

It also includes an independent term describing stochastic evolution in the form of Brownian motion (Felsenstein 1985):

[5] $\sigma dB(t)$.

When α = 0, term [4] becomes zero, resulting in Brownian motion (term [5]), as a special case of OU (Fig. 1A–B; e.g. Butler & King 2004; Slater 2013).

Interpretation of OU models can be complicated (Cooper *et al*. 2016) because OU models can also simulate the behavior of several other commonly used macroevolutionary models (Fig. 1):

(1) A 'trend' model in which Brownian motion is modified by the addition of a variable μ, describing the expected amount of directional change in trait values through time (Pagel 2002; Hunt & Carrano 2010). Trend-like behaviour is described by OU models when α becomes small and θ is outside the range of observed trait values, in which case α*θ approximates μ (Fig. 1E; Hansen 1997). Notably, trend models (and other models in which θ is different to Z_0) cannot be identified without the inclusion of fossil data (e.g. Slater *et al*. 2012).

(2) A stasis or 'white noise' model, in which trait values are drawn from a normal distribution with mean θ and a stable variance, independent of the phylogeny (Sheets & Mitchell 2001; Hunt 2006). OU models converge to stasis-like behavior through time when α is high, in which case instantaneous trait values [X(t)] are approximately equal to θ with a variance equal to $\sigma^2/2\alpha$. OU models describe stasis-like behavior from t = 0 when α is high and $\theta \approx Z_0$ (Fig. 1C; Hansen 1997).

The parameters of OU models therefore provide information on the mode of evolution. The phylogenetic half-life, $t_{0.5} = \ln(2) / \alpha$, is particularly useful value in this context as it describes the time taken for θ to become more influential than Z_0 in determining trait values within a regime (Hansen 1997; see also Slater 2015).

A key difference between BM and OU is that, under OU, non-zero values of α act to constrain trait values around θ, thereby limiting the

accumulation of trait variance through time (e.g. Hansen 1997; Butler & King 2004; Slater 2013). The expected variance in trait values among descendants (living at time = t) of a single common ancestor (that lived at time = 0) is $(\sigma^2/2\alpha)*(1 - \exp(-2\alpha t))$ (Hansen 1997), which asymptotes at $\sigma^2/2\alpha$ when t exceeds several phylogenetic half lives. This contrasts with the linear increase in variance with time under Brownian motion according to $\sigma^2 t$ (e.g. Felsenstein 1985; Hunt 2012; and compare Fig. 1A with Fig. 1C).

Characterising the macroevolutionary landscape of dinosaur body size—SURFACE algorithm.

We used a two-step approach to characterize dinosaur body size evolution. First, we used the R package SURFACE version 0.4-1 (Ingram & Mahler 2013). SURFACE implements an approach that locates a set of macroevolutionary regimes characterized by OU models with distinct trait optima (θ) on a phylogeny. To reduce computational demands, SURFACE assumes conserved, single values of α and σ^2 across the entire phylogeny. The locations of regime shifts are estimated using stepwise AICc (AICc = Akaike's information criterion for finite sample sizes; Akaike 1974; Sugiura 1978; Burnham & Anderson 2004), without prior specification of how the regimes should be distributed on the phylogeny (Ingram & Mahler 2013). SURFACE initially undertakes a forward phase, first fitting a two-regime model by identifying the best node at which to specify a regime shift using AICc. It then holds the position of that regime shift and iteratively searches for further shifts until no improvement in AICc can be attained. The algorithm then undertakes a backwards phase in which phylogenetic regimes are merged

together if there is a resulting improvement in AICc score, allowing the detection of evolutionary convergence (Ingram & Mahler 2013).

Ho & Ané (2014) demonstrated the existence of a 'large p small n' problem for fitting multi-regime OU models to comparative data. Because the number of possible shift configurations increases dramatically as shifts are added, and because AIC, AICc, and BIC (Bayesian information criterion) do not address the issue of false positives due to multiple comparisons, the SURFACE algorithm is liberal, and tends to support overly complex models (Ho & Ané 2014; Khabbazian *et al*. 2016; Davis & Betancur-R 2017). This has been forcefully demonstrated for ultrametric trees comprising only extant taxa (Ho & Ané 2014). Nevertheless, adding fossil taxa (i.e. analysing non-ultrametric trees, as done here), improves identifiability of the parameters of OU models (Slater 2013; Ho & Ané 2014), and it is possible that it also facilitates accurate regime shifts determinations. To address the problems with existing optimality criteria such as AICc, Khabbazian *et al*. (2016) proposed a new information criterion, "pBIC" (phylogenetic Bayesian information criterion). pBIC makes use of the effective sample size of taxa providing information about the trait optimum at each node on a phylogeny (Ané 2008), which is often considerably smaller than the number of taxa descended from that node (e.g. Ho & Ané 2014). The pBIC is conservative, with low rates of false positive identification of OU model regime shifts. However, until now it has not been implemented for non-ultrametric trees. We implemented a set of functions that calculate pBIC for SURFACE model fits, and conduct stepwise, SURFACE-like searches using pBIC instead of AICc. These were used to test whether our SURFACE fits were over-parameterized

by (1) Conducting fully-conservative stepwise-pBIC searches; (2) Conducting liberal forward-phase stepwise AICc searches and conservative backward-phase pBIC searches; and (3) By calculating pBIC for the shift configurations returned by (fully-liberal) forward- and backward-phase stepwise AICc searches. Ideally, pBIC would be used for all model comparisons throughout this paper (and others). However, because it is not implemented for most model implementations, we predominantly use AICc for model comparisons, using pBIC only to ensure that AICc does not unduly favour complex model fits for our data. Our pBIC functions are available on Dryad (Benson *et al*. 2017).

Characterising the macroevolutionary landscape of dinosaur body size—OUwie algorithm.

Having identified candidate macroevolutionary regimes using SURFACE, we estimated the full set of parameters (Z_0, θ, α, σ^2) of those regimes using maximum-likelihood in OUwie version 1.50 (Beaulieu *et al*. 2012). OUwie employs a model-fitting algorithm that potentially allows all key parameters to vary freely, including α and σ^2 (Beaulieu *et al*. 2012; models described in Table 1), and further differs from SUFACE in fixing the locations of regime shifts on the tree a priori. Our SURFACE results suggest that, in general, body size evolution occurs in a step-wise fashion, characterized by substantial values of α, and attraction to a set of distinct optima (θ) in trait space. However, if α actually varied among regimes, then we might incorrectly fit regime shifts to Brownian-like or trend-like portions of our phylogeny when holding α constant across regimes, as done by SURFACE. Furthermore,

SURFACE does not currently allow the inclusion of estimated measurement errors, whereas OUwie does allow these to be taken into account. For these two reasons, we used OUwie to estimate the full set of parameters for each regime, and to compare the fits of models in which different sets of parameters were allowed to vary freely, using AICc weights. This allowed us, for example, to determine whether allowing a distinct root node trait value (Z_0) improved upon a model in which this was set equal to the trait optimum (θ) for the regime present at the root of the tree, and whether allowing θ and α both to vary among regimes resulted in a better model than one in which only θ was allowed to vary.

Estimating the full set of parameters independently among regimes is computationally intensive (Beaulieu *et al*. 2012). Indeed, our analyses frequently recovered nonsensical parameter estimates for the most complex models (especially for OUMA and OUMVA; Table 1). For this reason the fits of complex models allowing most or all parameters to vary among regimes sometimes had to be discarded, and this was done using the following criteria: (1) Model fits that returned any highly precise (s.e. = 0) or imprecise (s.e. > 2) parameter estimates, except when implying trend-like dynamics with unrealized values of θ and low values of α (as described above; this occurs, e.g. on the single ornithischian lineage leading to the small-bodied ankylosaur *Struthiosaurus*). (2) Model fits that returned highly erroneous estimates of the ancestral body mass ($Z_0 < 0$ [= 1 kg] or $Z_0 > 2$ [= 100 kg]), as indicated by comparisons with other analyses (for example, analyses of Triassic–Jurassic Dinosauria were considered to provide robust estimates of the ancestral body mass of Ornithischia that should be replicated by analyses of Ornithischia). (3)

Entire sets of model fits for phylogenies for which all of the complex OU models (OUMV, OUMA, OUMVA) returned nonsensical parameter estimates by the preceding two criteria. Therefore, we simplified the models prior to analysis by analysing subtrees that contained relatively fewer shifts, broadly comprising Sauropodomorpha, Thyreophora, Marginocephalia, Iguanodontia and pre-Albian Coelurosauria. We also deleted taxa that were characterized by single, terminal-branch regimes not present at any internal branches of the phylogeny. All subtrees analysed included several early dinosaur taxa that provide information on the ancestral body size for Dinosauria: the early saurischians *Pampadromeus*, *Saturnalia*, *Chromogisaurus*, *Staurikosaurus*, *Eoraptor*, *Tawa*, and the early ornithischian *Pisanosaurus*.

Elaborations of the trend model.

The model of BM with a trend, as described above, posits that evolutionary dynamics hold uniformly over time and across branches of the tree (e.g. Hunt & Carrano 2010). However, it is possible that allowing for shifts in the trend parameter, μ, may provide a better account of the macroevolutionary processes operating in a clade, especially given the complexity of body size dynamics previously documented in dinosaurs (Carrano 2006; Benson *et al*. 2014*a*). We explored two kinds of elaborations of the uniform trend model: allowing for temporal shifts in μ (*time shift* models) and allowing for shifts in μ at specific nodes on the tree (*node shift* models) (Fig. 2). With temporal shifts, trend dynamics are uniform across all branches at any given instant of time, and, when a shift to a new value of μ occurs, it applies to all lineages alive at that time (Fig. 2A). Such a model might offer improvement if, for example,

body size increases are concentrated early or late in the history of a clade. Trends with dynamics that shift at nodes (Fig. 1E–F; Fig. 2B) allow for heterogeneity in body size evolution across subclades, describing a situation in which some clades evolve towards larger or smaller sizes through time whereas others do not.

The trend-based models were fit via maximum likelihood. Under a uniform trend model, tip values from a tree have a joint multivariate normal distribution (Hansen & Martins 1996). The multivariate vector of means, **m**, is equal to $Z_0 + \mu\mathbf{t}$, where, **t** is a vector of time spans between each terminal taxon and the root of the tree (e.g., t_i is the difference in age between the root of the tree and the i^{th} tip) and Z_0 is the trait value at the root. The covariance matrix among tips is the same as that for Brownian motion, $\sigma^2\mathbf{V}$, where **V** is the matrix that represents shared branch lengths among the tips (Martins & Hansen 1997). Thus, to compute the likelihood for a particular combination of parameters values (Z_0, μ, σ^2), one calculates **m** and **V** from the parameters, and then evaluates the density function of the multivariate normal distribution with inputs **m** and **V**.

The likelihood calculations are only slightly altered when μ varies over time or across branches. Now, the vector of means sums over multiple different μ along the path from the root to a particular tip such that the mean of the i^{th} terminal taxon is

$$[6]\ m_i = Z_0 + \mathbf{\Sigma}_j\, \mu_j t_j,$$

In which j indexes the (possibly) different trend regimes on the path from the root to the i^{th} tip, with μ_j as the trend parameter and t_j as the duration of time in each regime. The covariance computation is unchanged from uniform BM.

We found that trend models with temporal shifts often have multiple, local optima in the likelihood surface corresponding to different combinations of timings for the shifts in the μ parameter. These optima complicate hill-climbing searches, so we used a grid search to explore the space of temporal shift points while using hill-climbing algorithms to find maximum likelihood estimates for the remaining parameters.

The positions of branch shifts were identified using a stepwise procedure similar to forward phase of SURFACE. Starting with a uniform trend, a shift in the μ parameter was tested for each internal node of the tree and the node leading to the highest increase in AICc was retained. Next, a second shift point was searched for by testing each remaining node, and so on until no more complex model improves AICc. To simplify the parameter search, it was assumed that the shift occurred at the base of the branch leading to the node being tested.

We constrained the Brownian variance (σ^2) to be shared across all regimes in all models. When each regime was allowed to have its own μ and σ^2 parameters, the method would sometimes return shifts at nodes with near zero values for the Brownian variance when two sister taxa happened to have very similar body sizes. Constraining the models to a single σ^2 parameter shared across all trend regimes prevents such unrealistic parameter values.

Functions to fit these trend models were written in custom R code and rely heavily on functions from the ape (Paradis *et al*. 2004) and phytools (Revell 2012) packages. This code is provided on Dryad (Benson *et al*. 2017).

RESULTS

Body mass estimates—predicting unknown stylopodial shaft circumferences. Femoral and humeral minimum shaft circumferences (FC and HC) can be estimated as the perimeters of ovals with diameters equal to the measured minimum anteroposterior and mediolateral bone shaft diameters (equation [3] in Appendix S1). The relationships between these 'oval circumference estimates' and measured femoral and humeral shaft circumferences is described best by non-phylogenetic regression models (Table S1 and equations [4]–[5] in Appendix S1), which explain more than 99% of the variance in measured minimum shaft circumferences. These regression models can therefore be safely applied as a 'correction factor', allowing us to reliably estimate unknown stylopodial minimum shaft circumferences. However, it is only possible to use this approach when both shaft diameters are known, motivating the search for other relationships to predict FC and HC. These relationships are described below, and were used to generate the full set of mass estimates presented in Dataset S1. More detailed explanations of the model fits are provided in Appendix S1.

Bipedal dinosaurs. In bipedal dinosaurs, including all theropods, many early sauropodomorphs, and many ornithischians, FC can be predicted using other measurements of the femur and using measurements of the tibia,

demonstrating that bipedal taxa had relatively conservative hindlimb proportions (Table S2; equations [6]–[14] in Appendix S1). These measurements are, in order of predictive strength (as indicated by AICc-weights): femoral shaft minimum mediolateral diameter (FML), tibia minimum shaft circumference (TC), femoral shaft minimum anteroposterior diameter (FAP), and femoral length (FL). AICc weights also indicate that non-phylogenetic models predict FC better than phylogenetic models. Nevertheless, the relationships of FC with FL and with TL include clade assignment to Ornithischia, Sauropoda or Theropoda as a categorical covariate. This suggests that differences in the robustness (length:circumference ratio) of the femoral shaft evolved rapidly during basal divergences among major dinosaur clades, and were subsequently conserved, rather than evolving in a Brownian motion-like way across the phylogeny. The coefficients of the 'clade' covariate (equations [9]-[11] in Appendix S1) indicate that ornithischians and sauropodomorphs have proportionally longer tibiae compared to FC than do theropods. Furthermore, relative to shaft circumference, ornithischians have proportionally the shortest femora, and theropods have proportionally the longest femora. Insufficient femoral circumference measurements are available to determine whether, as in extant birds, Mesozoic birds (Avialae) had an abbreviated femur compared to other theropods, but this is unlikely in most cases as only a few Mesozoic taxa within Ornithuromorpha deviated from the typical hindlimb segment proportions of other small-bodied theropods (Benson & Choiniere 2013).

Quadrupedal dinosaurs. Dinosaurs evolved quadrupedal gaits from bipedal ancestors independently within Sauropodomorpha (Cooper 1984;

Yates *et al*. 2010; Bonaparte & Pumares 1995; Bonnan 2003; Yates & Kitching 2003; Yates 2007) and the ornithischian clades Thyreophora, Ceratopsia (Chinnery 2004; Chinnery & Horner 2007; Zhao *et al*. 2013) and Iguanodontia (Galton 1970, 1974; Norman 1980, 1986; Maidment *et al*. 2012) (Sereno 1999). Among transitional taxa within Sauropodomorpha, Thyreophora and Ceratopsia, and among iguanodontians, there is some ambiguity as to which are facultative rather than obligate quadrupeds. Furthermore, it is not clear what equation should be used to estimate body mass in facultative quadrupeds/bipeds. We previously estimated the masses of such taxa using the equation for quadrupeds (Campione & Evans 2012) and the equation for bipeds (Campione *et al*. 2014) and found that it made little difference in large-scale analytical studies (Benson *et al*. 2014*a*). Furthermore, Campione *et al*. (2014) also showed that when estimating the masses of facultative bipedal animals (such as macropods and primates), the quadrupedal equation still performed better than the bipedal correction, supporting use of the quadrupedal equation in those instances. Therefore, in the present work we estimate them all using the quadrupedal equation.

Substantial variation in interlimb and intra-hindlimb proportions has previously been noted among groups of quadrupedal dinosaurs (Maidment *et al*. 2012). It is not surprising, therefore, that our analyses find intermediate levels of phylogenetic signal, or the importance of clade assignment as a covariate term, when analysing all quadrupedal dinosaur together. This was found for the relationships between stylopodial shaft circumferences and other limb measurements, including those comparing forelimb measurements with hindlimb measurements (Tables S3, S4 in Appendix S1). Nevertheless, there

are some exceptions. Femoral circumferences of quadrupedal dinosaurs can be predicted using non-phylogenetic relationships with either femoral minimum mediolateral shaft diameter (FML) or femoral length (FL), and models for this relationship that include phylogenetic signal or clade assignment as a covariate term have negligible AICc weights (indicating that clade membership is not important for this relationship; Table S3; equations [15]-[16] in Appendix S1). The same is true when humeral minimum mediolateral (HML) or anteroposterior (HAP) shaft diameters are used to predict humeral minimum shaft circumferences (HC) (Table S4; equations [17]-[18] in Appendix S1). These relationships indicate that the proportions of both the femur and humerus are relatively conserved among quadrupedal dinosaurs, even though the proportional lengths of limb segments vary within and among limbs.

Quadrupedal sauropodomorphs. Many limb measurements provide poor predictions of stylopod minimum shaft circumferences in quadrupedal sauropodomorphs (Tables S5, S6). Nevertheless, FML and HC have strong relationships with FC (Table S5; equations [19]-[20] in Appendix S1), and HML and FC have strong relationships with HC (Table S6; equations [21]-[22] in Appendix S1). The best models of these relationships, according to AICc weights, lack phylogenetic signal. This indicates that the relative circumferences of the femoral and humeral shafts (FC and HC), and the mediolateral diameters of the shafts of these bones (FML, HML) do not change substantially across the phylogeny of quadrupedal sauropodomorphs.

Thyreophoran ornithischians. Many limb measurements provide poor predictions of stylopod minimum shaft circumferences in thyreophoran

ornithischians, regardless of whether phylogenetic or non-phylogenetic regression models are used (Tables S7, S8). Nevertheless, FML and HC have strong relationships with FC (Table S7; equations [23]-[24] in Appendix S1), and FC has a strong relationship with HC (Table S8; equation [25] in Appendix S1). The best models for regression of FC on HC and HC on FC are non-phylogenetic models. Models of the relationship between HC and FC including phylogenetic signal, or assignment to Stegosauria or Ankylosauria as a covariate or interaction term have negligible AICc weights (Tables S7, S8). However, The relationship between FML and FC has strong phylogenetic signal ($\lambda = 1.00$ or 0.98; Table S7), indicating that the eccentricity of the femur varies not only between Stegosauria and Ankylosauria, but also among more closely related taxa within those clades.

Ceratopsian ornithischians. Measurements of both the femur and tibia do a poor job of predicting FC in quadrupedal ceratopsians (Table S9), and so were not used. However, HC and HL have strong, non-phylogenetic relationships with FC (Table S9; equations [26]-[27] in Appendix S1). Humeral circumference (HC) is predicted well by non-phylogenetic relationships with HL and FC (Table S10; equations [28]-[29] in Appendix S1). Overall, these relationships indicate a conserved numerical relationship between FC, HC and HL in quadrupedal ceratopsians.

Iguanodontian ornithischians. Femoral measurements provide good estimates of FC in quadrupedal iguanodontians, and humeral measurements provide good estimates of HC. By contrast measurements of other bones provide poor estimates (Tables S11, S12; equations [30]-[35] in Appendix S1).

The best model of the relationship between FC and FML is non-phylogenetic (AICc-weight = 0.66), although the best models including a 'clade' covariate specifying assignment to Hadrosauroidea or to non-hadrosauroid iguanodontians have non-negligible AICc weights (= 0.14; Table S11). The best model of the relationship between FC and FL includes this 'clade' term (AICc-weight = 0.26; Table S11), although models lacking this term have similar AICc weights (best AICc-weight = 0.20). This indicates that hadrosauroids have proportionally longer femora than do non-hadrosauroid iguanodontians (equation [30] in Appendix S1).

The best models of the relationships between HC and other humeral measurements are non-phylogenetic regression models. However, a model including a 'clade' term, specifying assignment to Hadrosauroidea or to non-hadrosauroid iguanodontians has a non-negligible AICc-weight (= 0.10) for the relationship between HC and HL (best, non-phylogenetic model: AICc-weight = 0.54), and models including strong phylogenetic signal ($\lambda = 1.00$) have comparable AICc weights to non-phylogenetic models of the relationship between HC and HAP. It is of note that igunodontians are the only clade of possibly quadrupedal dinosaurs that do not exhibit a conserved relationship between the femoral and humeral shaft circumferences. This occurs because hadrosauroids have proportionally gracile humeri compared to iguanodontians.

Characterising the macroevolutionary landscape of dinosaur body size evolution.

Locating macroevolutionary regimes using stepwise AIC. Comparison of AICc scores for multi-regime OU models fit using SURFACE to those of our trend-based models demonstrates overwhelming support for OU on phylogenies calibrated using mbl1 (Fig. 3), *cal3* and the Hedman method (Figs S1, S2). In fact, trend-based models have comparable AICc weights to single-regime Brownian motion models for Theropoda and Dinosauria on many of our trees (Fig. 3; Dataset S2). In contrast, for Ornithischia, and especially Sauropodomorpha, node shift trend-based models have better (i.e. lower) AICc weights than either Brownian motion or temporal shift models for almost all phylogenies, especially when calibrated using mbl1 (Fig. 3; Dataset S2). Nevertheless, even for these clades, trend-based models perform very poorly compared to OU-based models (Figs 3, S1, S2).

Stepwise AICc searches from SURFACE (Ingram & Mahler 2013) for multi-regime OU models recovered different distributions of macroevolutionary regimes across alternative phylogenies used in our analyses (Appendix S2). However, much of this variation was trivial. The only substantive element of variation concerns bimodal distributions of AICc values obtained across phylogenies for Ornithischia, Triassic–Aptian Theropoda, and Triassic–Jurassic Dinosauria (shown for mbl trees in Fig. 4, and for *cal3* and Hedman trees in Appendix S2). These formed two clusters: one with low AICc values and high α, and another with high AICc values and low α (Fig. 4). The populations of high-AICc/small α model fits are characterised by the occurrence of fewer macroevolutionary regimes shifts than are present in the population of better models (e.g. Fig. 5; full results are presented in Appendix S2). The most extreme of these models are very weakly constrained and so

approximate Brownian motion (α <0.005 [small]; phylogenetic half life > 60 Ma; θ = Z_0), with a small number of regime shifts capturing ephemeral, high-magnitude trend-like dynamics within some groups (α = small; θ is outside the range of observed trait values) (Fig. 2B).

The recovery of two sets of model fits for some groups illustrates the difficulty of fitting complex phylogenetic models to phenotypic datasets. It is not possible to manually explore the full range of candidate models that could be fit to each phylogeny. Nevertheless, because we calibrated the same topologies using each method of divergence time estimation (mbl, *cal3*, Hedman), and because the divergence time methods yielded different regime configurations, we are able to show using OUwie that SURFACE recovered suboptimal models fits for at least some of our phylogenies when it returned results in the high-AICc/low-α class (Appendix S3). Focusing on our mbl1 trees: low-α solutions were initially fit to tree topologies 4, 15 and 18 for Ornithischia. However, a high-α solution was found for all these topologies when timescaled using the Hedman algorithm (henceforth: the Hedman regimes), and both the AICc and pBIC scores for the Hedman regimes mapped to the mbl timescaled phylogeny were better than for the mbl regimes mapped to the same topology using any of the timescaling methods (Appendix S3). The same result was also found for almost all the tree topologies that had initially recovered low-α solutions from stepwise AICc searches on mbl1 trees for Dinosauria (2, 3, 10, 16, 18, 20) and Theropoda (1, 4, 9, 12, 16, 17, 19; and cal3 regimes provided the best fit to topology 7). Only in a small number of cases did the low-α solution perform better than any other found by our searches either by a substantial (topology 5 for Dinosauria)

or negligible margin (topology 6 for Dinosauria; topology 15 for Theropoda). Difficulty finding the best fits might be expected for complex datasets: stepwise optimisation methods are not guaranteed to find the best solution to complex model fits, and some statisticians have cautioned against their use (e.g. Burnham *et al*. 2011). Out result in general suggest that the high-AICc/low-α class of SURFACE model fits are likely to be suboptimal and are not discussed further here.

The low-AICc/high-α class of SURFACE model fits includes relatively large numbers of regime shifts for most clades (5-7 for Sauropodomorpha; 6-10 for Theropoda; 15–18 for Ornithischia; 11–15 for Dinosauria). However, stepwise AICc is known to provide support for overly complex models compared to the conservative pBIC criterion (Ho & Ané 2014; Khabbazian *et al*. 2016; Davis & Betancur-R 2017). We therefore also implemented modified versions of the SURFACE algorithm using pBIC in both the forward and backward phases, and using AICc in the forward phase followed by pBIC in the backward phase (mixed algorithm). Comparable results were found for Sauropodomorpha by all search methods (Appendix S4; small differences noted below). For the other clades (Dinosauria, Ornithischia, Theropoda), the mixed algorithm finds identical models to our stepwise AICc searches for almost 100% of cases (Appendix S4), whereas full stepwise pBIC searches find considerably simpler models (Appendix S4) than the low-AICc/high-α class of stepwise AICc model fits. Nevertheless, in all of these cases, the pBIC values of the simpler models are worse than those of the complex models, indicating a failure of the algorithm to sufficiently explore parameter space. These are therefore not discussed further. Visualisation of the search

paths (Appendix S4) shows that the stepwise pBIC searches are unable to cross a central region of the model space characterised by intermediate complexity and suboptimal pBIC, but highly optimal AICc. Stepwise AICc searches can cross this central valley and discover more complex models with highly optimal pBIC scores. In summary, both the AICc and pBIC optimality criteria provide support the selection of complex multi-regime models, although hill-climbing searches using pBIC do not generally discover these models (Appendix S4).

Our SURFACE results across multiple phylogenies for each dinosaur group show congruent distributions of macroevolutionary regimes with only minor variations (Appendix S2). The descriptions given here are based on our mbl1 trees. However, model fits to our other trees are shown in Appendix S2, and show few differences. All model fits for Triassic–Jurassic Dinosauria specify a small-bodied ancestral regime, with estimates of θ (= Z_0 for Dinosauria) ranging from 14–24 kg (Fig. 6). This 'basal dinosaur' regime was inherited by the earliest theropods (e.g. *Coelophysis*, *Staurikosaurus*, *Tawa*), sauropodomorphs (*Chromogisaurus*, *Efraasia*, *Pampadromaeus*, *Saturnalia*) and ornithischians (*Pisanosaurus*, *Scutellosaurus*, *Stormbergia*). This is congruent with estimates of the primitive dinosaurian body mass presented in our previous work (16–24 kg for Dinosauria and 27 kg for Theropoda; Benson *et al*. 2014*a*, table 2 [Z_0 values]), and is approximately one order of magnitude smaller than the estimated mass of the theropod ancestor presented by Lee *et al*. (2014; 175 kg).

Sauropodomorph evolution is characterized by a Triassic regime shift to larger body masses in the grade that includes *Plateosauravus* or *Ruehleia*

and taxa that are more closely related to sauropods, such as *Massospondylus* and *Lufengosaurus* (Fig. 7; 'prosauropods' herein; θ = 1,100 kg to 1,900 kg; Appendix S2; phenograms shown in Appendix S5), and subsequently to even larger masses in the clade comprising *Isanosaurus*, *Vulcanodon* or *Tazoudasaurus*, and all more derived taxa, including *Vulcanodon* (Fig. 7; 'sauropods' herein; θ = 15,000 kg to 17,000 kg). All members of the 'prosauropod' regime became extinct during the Early Jurassic in a size-selective extinction that was only survived by giant sauropodomorphs of the sauropod regime. There is some variance within the 'prosauropod' regime due to the occurrence of smaller body sizes in taxa such as *Anchisaurus* and *Sarahsaurus*, which are assigned a separate macroevolutionary regime in some iterations of our SURFACE analyses. Furthermore, the 'sauropod' regime shift could be located at a slightly more or less inclusive node than that defined by *Isanosaurus*, and includes terminal single-branch regimes that model the occurrence of body size reduction in the island dwarf sauropod *Europasaurus* (1,000 kg), and body size increase in the gigantic taxon *Argentinosaurus* (95,000 kg), and sometimes in *Ruyangosaurus* (54,000 kg) (Bonaparte & Coria 1993; Sander *et al*. 2006; Lü *et al*. 2009). Unlike for the other clades, stepwise pBIC searches for Sauropodomorpha generally find slightly better models (according to pBIC) than stepwise AICc searches. These models include slightly fewer regimes by generally including *Argentinosaurus* (and *Ruyangosaurus*) in the same regime as other sauropods, and *Anchisaurus* and *Sarahsaurus* with other Late Triassic/Early Jurassic non-sauropodan sauropodomorphs.

Ornithischian body size evolution (Fig. 8) is characterized by a Triassic shift to small body sizes within Heterodontosauridae (θ = 0.7 to 1.6 kg), Middle–Late Jurassic shifts to larger body sizes in thyreophorans (Stegosauria θ = 3,100 to 11,000 kg; Ankylosauria θ = 1,000 to 1,200 kg) and iguanodontians (which convergently share the regime seen in ankylosaurs in many solutions), and to smaller body sizes in early ceratopsians such as *Psittacosaurus* (θ = 6.5 to 8.4 kg) (phenograms shown in Appendix S5). The Cretaceous saw further shifts to larger body sizes in the ceratopsian clade that includes leptoceratopsids and ceratopsids (θ = 190 to 350 kg; and θ = 4,500 to 9,900 kg in Ceratopsidae), and in euhadrosaurian iguanodontians (which convergently share the regime seen in Ceratopsidae in many solutions).

Against the background patterns of ornithischian body size evolution (Fig. 8), several single-lineage regimes shifts are seen consistently on the single branches leading to exceptionally large- or small-bodied taxa (Appendices S2, S5): the small Triassic taxa *Lesothosaurus* (6.3 kg) and *Abrictosaurus* (1.4 kg), the large-bodied non-hadrosauroid iguanodontians *Iguanodon* (8,300 kg) and *Muttaburrasaurus* (5,200 kg), the rhabdodontid iguanodontian and possible island dwarf *Mochlodon* (41 kg), the gigantic hadrosaurine iguanodontian *Shantungosaurus* (17,400 kg), the large-bodied thescelosaurid, *Thescelosaurus neglectus* (340 kg), the small-bodied non-ceratopsid ceratopsians *Graciliceratops* (6.7 kg) and *Bagaceratops* (5.7 kg), the giant pachycephalosaur *Pachycephalosaurus* (370 kg), the possible island dwarf ankylosaur *Struthiosaurus* (130 kg), and the largest Early Jurassic ornithischian: the thyreophoran *Scelidosaurus* (650 kg) (Appendix S5; see

Benton *et al*. 2010; Osi *et al*. 2012; for discussions of island dwarfism in dinosaurs).

The overall pattern in Triassic–Aptian theropod body size evolution (Figs 9–10) starts with a shift from the 'basal dinosaur' regime to a regime characterized by larger body size in the clade comprising *Liliensternus* plus the speciose, long-lived clades Tetanurae and Ceratosauria (θ = 780 to 960 kg or 130 to 150 kg; explained below), which originated in the Late Triassic. This was followed by a series of shifts towards smaller body sizes on the line leading to birds, within Coelurosauria.

The earliest-diverging coelurosaurs, including early tyrannosauroids and taxa such as *Zuolong*, have smaller body masses than those seen in allosauroids, megalosauroids, and many ceratosaurs. In many of our phylogenies, this is identified as a shift to a smaller body size regime at the base of Coelurosauria (θ = 120 to 150 kg) from a primitive large body size regime (θ = 780 to 960 kg) (Fig. 9). However, in other phylogenies the basal coelurosaurian regime is inherited unmodified from the Late Triassic regime (θ = 130 to 150 kg), and the larger body sizes seen in some ceratosaurs, and non-coelurosaurian tetanurans (allosauroids and megalosauroids) represent three separate evolutionary shifts to a shared large body size regime (θ = 1100 to 1200 kg) in those clades (Fig. 10).

Subsequently, a shift to a smaller body size regime shared with early dinosaurs occurred in the clade comprising ornithomimosaurs and all more derived coelurosaurs (θ = 11 to 14 kg). A further shift to a smaller body size regime occurred in Paraves (θ = 1.0 to 1.2 kg), consistent with the evolutionary miniaturization event proposed by Turner *et al*. (2007) (Figs 9–

10). This ~ 1 kg regime was inherited by long-tailed, early birds (Avialae) such as *Archaeopteryx* (Appendix S2). A second shift to a smaller body size regime occurred in pygostylian birds (θ = 0.093 to 0.110 kg), and a further shift to an even smaller size regime in the enantiornithine clade including *Iberomesornis* and more derived taxa is found on some phylogenies (Appendices S2, S5; pygostylian θ = 0.170 to 0.180; enantiornithine θ = 0.053 to 0.055).

An evolutionary shift to a larger body size regime is consistently identified in the herbivorous therizinosaurs (θ =120 to 150 kg), on the terminal branches leading to the giant Triassic taxon *Herrerasaurus* (270 kg), in the large-bodied dromaeosaurids *Utahraptor* (250 kg) and *Tianyuraptor* (20 kg), and sometimes in the large ornithuromorph bird *Yanornis* (1.4 kg) (Figs 9–10; Appendices S2, S5). In model fits in which tetanurans have a primitively large body size regime, terminal-branch regimes are also required to describe the origins of small-body size in the ceratosaur *Limusaurus* (20 kg) and the basal tetanuran *Chuandongocoelurus* (18 kg) (Appendix S2).

Parameterisations of macroevolutionary regimes. Fully parameterized multi-regime OU models fit using OUwie (Beaulieu *et al*. 2012), and taking into account the errors associated with our body mass estimates, were generally much better than single-regime Brownian motion models, single-regime OU models, or our trend-based models. This is indicated, for example, by differences in AICc between multi-regime OU models and trend-based models (Fig. 11). Of these comparisons, BM1 was only supported for one tree in Thyreophora (tree 11), and one tree in Theropoda (tree 11), and trend-based models were only supported for tree 9 in Ornithopoda and tree 10 in Theropoda.

The solution for theropod tree 10 includes a small, but positive trend of body mass increase in non-pygostylian coelurosaurs (μ = 0.019 per Myr), with a transition to weakly decreasing body size in pygostylians (μ = -0.022 per Myr) (Fig. 12A). Among non-pygostylians, the genus *Microraptor* exhibits a strong trend towards miniaturization (μ = -0.15 per Myr). Among pygostylians, the Songlinornithidae, a clade of ornithuromorphs that includes *Yanornis* and *Yixianornis* in our trees, exhibit a strong trend of body size increase (μ = 0.098 per Myr). The solution for ornithopod tree 9 is typical of the well-supported trend models for ornithopods (Fig. 12B). The base of the clade experiences BM-like evolution with a trend parameter very close to zero. There is a shift to a moderately increasing trend (μ = 0.15 per Myr) within Thescelosauridae, giving rise to relatively large body sizes in this clade by the end of the Cretaceous, and a parallel shift to a similar increasing trend (μ = 0.10 per Myr) near the base of Iguanodontia.

The three main macroevolutionary regimes identified by SURFACE for sauropodomorphs ('basal dinosaur' | 'prosauropod' | 'sauropod') are best characterized by complex OU-based models such as OUMV (α > 0; θ and σ vary among regimes), OUMA (θ and α vary among regimes) and OUMVA (θ, σ and α vary among regimes), with weakly non-negligible fits for variable rate Brownian motion models (BMS; α = 0 and σ varies among regimes) on two trees (tree 8: AICc-weight = 0.09 for BMS compared to 0.79 for OUMV; tree 15: AICc-weight = 0.11 compared to 0.65 for OUMV).

OUMA (θ and α vary among regimes; Table 1) and OUMVA (θ, σ and α vary among regimes; Table 1) models frequently returned nonsensical parameter estimates (defined above) and could not be compared to other

models on all trees. In such cases, OUMV (θ and σ vary among regimes; Table 1) was the most complex model that could be tested. Across the results for all trees, a clear pattern is evident in which OUMV, OUMA or OUMVA models of sauropodomorph body size evolution were overwhelmingly supported over simpler models by AICc weights (Fig. 11A; Dataset S3). OUM, OU1, BM1 and BMS models (defined in Table 1) are never present among the set of non-negligible models (i.e. those with AICc weights at least 10% those of the best model; Dataset S3). Estimates of α for regimes other than the primitive dinosaur regime ranged from 0.008 to 0.085, indicating phylogenetic half-lives ($t_{0.5}$) ranging from 86.6 Ma to 8.25 Ma, which is shorter than the 140 Ma duration of Sauropoda, which exhibit the longest-lived macroevolutionary regime within Sauropodomorpha. This indicates the predominance of constrained evolution, in which trait optima (θ) are more influential than variance (σ) in determining individual body masses. Consistent with this inference, substantial body size disparity failed to accrue though the entire evolutionary history of sauropods, and sauropod body size disparity was remarkably consistent throughout this time (Fig. 13A)

Our analyses of the ornithischian subclades Thyreophora, Marginocephalia and Ornithopoda provide further support for OU models on most trees (Dataset S3). For thyreophorans (Ankylosauria + Stegosauria), OU-based models received overwhelming support from AICc-weights for almost all trees (BM1 was best supported on tree 11; Fig. 11B), including OUM, OUMV, and OUMA models with values of α ranging from 0.032 ($t_{0.5}$ = 22 Ma) to 0.089 ($t_{0.5}$ = 7.8 Ma) for regimes other than the primitive dinosaur regime. For marginocephalians (Ceratopsia + Pachycephalosauria), OUM

models with values of α ranging from 0.012 ($t_{0.5}$ = 58 Ma) and 0.172 ($t_{0.5}$ = 4.0 Ma) received overwhelmingly the best AICc-weights. Although OU-based models also generally provided the best explanation of ornithopod body size, many of these had values of α that were approximately equal to zero, and BMS (variable-rate Brownian motion) models received the best AICc-weights for trees 8 and 20, and non-negligible AICc-weights for tree 2 (Dataset S3).

It was difficult to fit models to theropod body size evolution using the entire tree of Theropoda, and many preliminary searches yielded nonsensical parameter values. However, models were fit successfully for six trees, using the portion of the tree that includes ornithomimosaurs and all more derived coelurosaurs, excluding *Utahraptor* and therizinosaurs (when these were placed in the same regime as *Utahraptor* by SURFACE, as for most trees), which were identified by SURFACE as having distinct, large-bodied tip regimes. Model comparisons indicate strong support for complex OU-based models (OUMA and OUMVA) across at least this portion of the theropod tree. Values of α range from 0.014 ($t_{0.5}$ = 50 Ma) to 1.165 ($t_{0.5}$ = 0.6 Ma) (Dataset S3).

DISCUSSION

Phylogenetic studies of trait evolution in extinct groups have largely focused on two questions. (1) Are patterns of trait evolution consistent with niche-filling models of adaptive radiation (e.g. Slater 2013, 2015; Benson *et al*. 2014*a*; Close *et al*. 2015; Hopkins & Smith 2015; Stubbs & Benton 2016; Cantalapiedra *et al*. 2017)? And (2) do phylogenetic lineages of animals,

especially of tetrapods, collectively show a directional tendency to increase in body mass through evolutionary time (i.e. Cope's rule: Stanley 1973; McShea 1994; Van Valkenburgh *et al*. 2004; Hone & Benton 2005; Carrano 2006; Hunt & Roy 2006; Hunt & Carrano 2010; Benson *et al*. 2012; Sookias *et al*. 2012; Zanno & Makovicky 2013; Benson *et al*. 2014*b*; Slater *et al.* 2017)? Our results provide a single framework under which both questions can be addressed.

In this discussion, we initially focus on previous hypotheses of trend-like dynamics, including Cope's Rule and evolutionary patterns along the avian stem lineage. We then consider the broader patterns of dinosaurian body mass evolution and their implications for adaptive landscapes, and the roles of niche-filling models and constraint in generating observed patterns of phenotypic disparity.

Cope's Rule.

Given the astounding sizes reached by some Mesozoic dinosaurs, it is not surprising that there has been extensive interest in testing Cope's Rule (or Depéret's Rule; sensu Simpson 1953) in the group (Sereno 1997; Hone & Benton 2005; Hone *et al*. 2005; Carrano 2006; Butler & Goswami 2008; Hone *et al*. 2008; Hunt & Carrano 2010; Sookias *et al*. 2012a; Zanno & Makovicky 2013). Edward D. Cope informally discussed a general accumulation of sizes over evolutionary time (Cope, 1887), and Cope's Rule has commonly been defined as an active directional trend towards larger body sizes (e.g. Depéret 1907; McShea 1994; Carrano 2006). 'Active' trends involve similar directional changes occurring simultaneously among sets of independent lineages.

Under this paradigm, the observation of a large sample of lineages, most of which show change in the same direction, provides statistical evidence of selection towards larger size through time (McShea 1994). The outcome of this process therefore involves increases in both the mimimum and maximum body sizes of descendants (McShea 1994). 'Passive' expansion instead involves an increase in the total range of trait values seen among the descendants of a small-bodied ancestor due to non-directional, diffusive changes in body size. This model, which resembles Brownian motion also involves increased maximum body sizes. However, it differs from an 'active' trend model because the minimum body sizes of descendants are decreased or unaltered (McShea 1994). Alroy (2000) has noted that a wider range of dynamics are possible, a point with which we agree (and see Carrano 2006).

Previous studies on non-avian dinosaurs have recovered support for overall increases in mean body size through time (Sereno 1997; Hone *et al*. 2005), and detailed study has indicated that this generally resulted from passive expansion (Carrano 2006; Sookias *et al*. 2012). Evidence of active trends, embodied by 'trend-based' models here, has been scarce, although it has been demonstrated among ornithischians (Hunt & Carrano 2010; and ornithopods [herein]; and also in pterodactyloid pterosaurs; Benson *et al*. 2014*b*).

Notably, few of our analyses find support for unconstrained, 'active' trend models in dinosaur body mass evolution, which classically have been associated with 'Cope's rule'. Furthermore, where it occurs, this support is equivocal (Fig. 11). Nevertheless, we also reject strictly uniform passive expansion (i.e. Brownian motion dynamics). Instead, our results provide

strong support for multi-peak Ornstein-Uhlenbeck models, which describe the exploration of a macroevolutionary landscape by phylogenetically defined regimes that undergo constrained evolution around distinct trait optima (e.g. Hansen 1997, 2013; Butler & King 2004; Beaulieau *et al*. 2012; Ingram & Mahler 2013; Mahler & Ingram 2014; Lapiedra *et al*. 2013).

Miniaturisation on the avian stem lineage.

Modern and Mesozoic birds (Avialae) include small-bodied taxa often weighing only tens of grams. Small body size is likely associated with the origin of bird flight, as the size of aerodynamic surfaces scale with the square of linear dimensions, whereas body mass scales with their cube. This makes it easier for smaller animals to fly. Therefore, patterns of body size miniaturisation on the avian stem lineage have received significant research attention (Turner *et al*. 2007; Novas *et al*. 2012; Benson *et al*. 2014*a*; Lee *et al*. 2014; Puttick *et al*. 2014). Previous studies have documented the origins of body masses around 1 kg in early paravians such as *Archaeopteryx* (Turner *et al*. 2007), and also present in several other dinosaur lineages (e.g. Butler *et al*. 2009; Novas *et al*. 2015). In fact, 1 kg is large for extant bird species, which have a modal body mass around 100 g (Brown 1995). Such small body sizes are absent among adult dinosaurs (including *Archaeopteryx*), other than in Pygostylia (and possibly also Alvarezsauroidea; Table 2). Therefore, because of a focus on *Archaeopteryx*, evolutionary patterns associated with the attainment of uniquely bird-like small body masses in Mesozoic dinosaurs have not been addressed until now.

We find that avian miniaturization resulted from a serial pattern of shifts to macroevolutionary regimes with smaller body size optima (Figs 9–10). This model describes a stepwise, rather than gradual, pattern of body size decreases along the avian stem lineage (also documented by Novas *et al*. 2012). Regime shifts are concentrated during two intervals: (1) the Early Jurassic/Middle Jurassic boundary interval, which saw decreases in body size from the tetanuran or coelurosaurian ancestor, culminating in a paravian ancestor weighing an estimated 1 kg (e.g. Turner *et al*. 2007), and (2) the latest Jurassic or Early Cretaceous, giving rise to considerably smaller body sizes in Pygostylia. Most Early Cretaceous pygostylians have body masses in the range of 13 g (*Iberomesornis*) to 307 g (*Pengornis*) (Figs 9–10), similar to those of many extant birds, but smaller than all other dinosaurs.

Lee *et al*. (2014) reported sustained, gradual evolutionary body size reduction along the avian stem lineage, from an ancestral theropod dinosaur weighing 175 kg to masses of 1 kg in Avialae. This contrasts strongly with our findings, which indicate a stepwise pattern and involve at least some body size increases during early theropod evolution, from an ancestor weighing 10-30 kg (Figs 9–10).

In fact, the higher estimated body mass obtained by Lee *et al*. (2014) for the theropod ancestor is an artefact of incomplete and biased sampling of early dinosaur taxa in their analysis. As evidence of this: we can replicate their result by imposing their taxon sample on our dataset. Lee *et al*. (2014) selected only two Triassic dinosaurs for inclusion in their analysis, both of which were theropods and have anomalously high body masses compared to other early dinosaurs (*Herrerasaurus*, 274 kg, *Liliensternus*, 84 kg; Table 3;

and see Appendices S2, S5). Most other Triassic theropods have masses of 30 kg or less (i.e. *Coelophysis*, *Staurikosaurus*, *Tawa*; excepting *Gojirasaurus*; Table 3), and similar small body masses are seen among the earliest sauropodomorphs and ornithischians (Table 3: Carnian). These small-bodied early dinosaurs are informative outgroups to Theropoda, but were omitted from the analysis of Lee *et al*. (2014).

Maximum-likelihood estimation (Felsenstein 1973; Schluter *et al*. 1997; Paradis *et al*. 2004) using our dataset and a pruned tree of theropod dinosaurs that excludes the smaller-bodied Triassic dinosaur taxa (mimicking the analysis of Lee *et al*. 2014), finds the mass of the ancestral theropod to be 253 kg (95% confidence interval: 45.9–1390 kg). This is similar to the value obtained by Lee *et al*. (2014), but differs from the 10–30 kg estimate obtained by our full analysis, and clearly demonstrates the importance of representative taxon sampling in comparative analyses of trait evolution. Our finding that the ancestral theropod was much smaller that proposed by Lee *et al*. (2014) clearly refutes the hypothesis of sustained, directional body size reduction from this ancestral theropod dinosaur along the lineage leading to birds.

Paravian miniaturization event. Miniaturization in early paravians was first proposed by Turner *et al*. (2007; see also Novas *et al*. 2012; Benson *et al*. 2014*a*; Puttick *et al*. 2014). This miniaturization resulted from several downwards regime shifts from ancestral tetanurans or coelurosaurs, which were characterized by larger body sizes (Figs 9–10). The body masses of tetanuran and coelurosaurian ancestors are uncertain: our multi-peak OU models suggest two alternative possible optima for the primitive tetanuran macroevolutionary regime. The first possibility suggests a relatively larger

optimum ~1,000 kg, with a shift to a smaller optimum within Coelurosauria. This is congruent with classical notions of theropod body size evolution (Fig. 9; Novas *et al*. 2012). Alternatively, a smaller optimum of 150 kg is possible among early tetanurans, and was maintained in coelurosaurs. Under this second possibility, the larger body sizes observed in non-coelurosaurian tetanurans such as allosauroids and megalosauroids represent independent instances of clade-specific body size increases (Fig. 10).

Miniaturization to sizes of ~1 kg is not unique to paravians, or even to theropods. Similar body masses evolved under similar evolutionary dynamics in heterodontosaurid ornithischians (Butler *et al*. 2010), the ceratopsian *Graciliceratops* (Maryanska & Osmólska 1975), mononychine theropods (which have small body sizes but occur after the Aptian and so could not be analysed in our study), and also the early possible tetanuran *Chilesaurus* (~ 1 kg; immature individual; Novas *et al*. 2015) (Table 2). *Chilesaurus* was not included in our analyses because adult mass estimates are not available. However, although another small-bodied basal tetanuran, *Chuandongocoelurus* (18 kg) is currently accommodated by our models as the product of a lineage-specific miniaturization event (Appendices S2, S5). Nevertheless, it is still possible that continued sampling of the Early–Middle Jurassic theropod fossil record will indicate much smaller body size for the earliest tetanurans, more congruent with the masses of these apparently exceptional, small-bodied taxa. This is especially probable given the known biases against preservation of small-bodied species in at least parts of the dinosaurian fossil record (Brown *et al*. 2013).

Pygostylian miniaturization event. The second set of either one or two regime shifts on the line leading to birds occurred during the latest Jurassic or Early Cretaceous. This culminated in body sizes predominantly in the range of 15 to 300 g in Pygostylia (Figs 9–10; Appendixes S2, S5). This event has significance as the evolutionary origin of the modal body size observed among extant birds (100 kg; Brown 1995), and is unique among dinosaurs (i.e. such small body sizes only occur within Pygostylia; Fig. 14F). Furthermore, small body size in Pygostylia co-occurs alongside several other features of the skeleton and integument that improve flight ability (O'Connor *et al*. 2011), and with relaxed constraints on hindlimb morphology (Benson & Choiniere 2013). Pygostylian miniaturization, and not paravian miniaturization to 1 kg, might therefore be classed among the key features that facilitated evolutionary advancements in powered flight on the avian stem lineage.

Apparently in contradiction of our results, Puttick *et al*. (2014) demonstrated that Avialae shared a macroevolutionary regime with Paraves. If this were correct, then we would expect Mesozoic birds to occupy the same range of body sizes as their close dinosaurian relatives. This is not the case. Instead, pygostylian body masses are distinct from those of other paravians (430 g [*Microraptor*] – 600 kg [*Austroraptor*]; Tables 2, 4). Puttick *et al*. (2014) rejected the possibility of a distinct regime within Avialae based on tests of a set of a priori hypotheses that did not include the possibility of a distinct pygostylian regime, as detected here by our exploratory SURFACE searches. We recommend as good practice the examination of datasets using exploratory methods during a hypothesis-formulation phase in future studies.

Explaining dinosaur body size diversification.

Dinosaur body size evolution is best explained by multi-regime OU models; models applying a single regime to the entire phylogeny are poorly supported. Our results therefore suggest that dinosaur body mass evolution can be described by a set of constrained models within phylogenetically determined regimes, each characterized by attraction to a distinct macroevolutionary body size optimum.

A recent paper cautioned users against uncritical acceptance of results that favour OU models (Cooper *et al*. 2016). When they examined the performance of statistical tests of OU versus BM, these authors reported elevated rejection rates for BM in favor of OU, especially when error in tip values was not accounted for. Given that we accounted for error in body mass estimates in a realistic manner, we do not expect that this bias applies to our analyses. Nevertheless, Cooper *et al*. (2016) also noted that, in some cases, OU models were statistically favored by model comparison, but the resulting estimates of the α parameter implied evolutionary dynamics that were only trivially different from BM. This observation is consistent with the finding of Ho & Ané (2014), that low values the constraint parameter α (i.e. values that imply a long phylogenetic half-life relative to the total tree height, and therefore Brownian motion-like dynamics) were likely to be over-estimated, or falsely found to be different from zero. However, examination of the α parameter estimates (and the corresponding phylogenetic half-lives) of our model fits indicated dynamics substantially different from BM, a conclusion that is supported by the very large AICc differences we almost always found between OU and BM or trend models.

Shifts in the optima of our OU models cause transient, strongly directional evolution bridging order-of-magnitude body size transitions (i.e. large changes in θ). Because the statistical power to detect regime shifts depends more strongly on the size of change in θ between regimes than on counts of taxa analysed (Ho & Ané 2014), it is possible that regime shifts with smaller changes of θ exist, but were not detected. Therefore, our results provide a summary of only the larger events in the evolution of dinosaur body size. The distribution of dinosaur body sizes can be described as the distribution of originations of descendent regimes from ancestor regimes through time, the direction of body size change involved in regime originations, and the pattern of regime extinctions through time. We examined these patterns to explain body size diversification in Dinosauria. In doing so, we made a distinction between regime shifts that led to clades containing multiple taxa, which often represent long-lived clades (Fig. 14A,B: large symbols), and 'singleton' regimes containing only single taxa, which may represent transient instances of gigantism or dwarfism as an evolutionary response to local environmental conditions (Fig. 14A,B: small symbols).

Originations giving rise to larger-bodied descendant regimes are more frequent (11 occurrences) than those giving rise to smaller-bodied descendant regimes (six occurrences) (Fig. 14A: large symbols representing originations of multi-taxon regimes). This is common in ornithischians, and especially in sauropodomorphs (for which transitions to small-bodied multi-taxon regimes are absent). By contrast, increases are approximately as common as decreases in pre-Albian theropods, a pattern that is entirely due to the

occurrence of successive, stepwise decreases on the lineage leading to birds (Figs 9–10).

Constraint-breaking pulses of body size evolution generated most dinosaurian body size disparity. Ornstein-Uhlenbeck models describe constrained patterns of evolution, in which the total disparity obtainable within a regime is limited. For dinosaur body mass, the expected asymptotic trait variances ($\sigma^2/2\alpha$; Hansen 1997) of individual regimes are typically much less than one order of magnitude (Dataset S3: taking maximum values in $\log_{10}$ kg of 0.579 [Sauropoda], 0.132 [Thyreophora], 0.406 [Marginocephalia], 0.188 [Ornithopoda], and 0.278 [Theropoda: Coelurosauria]; all results excluding the primitive dinosaur regimes). There are a few exceptions to this, among 'prosauropods' (non-sauropodan sauropdomorphs) for some trees, and for the primitive dinosaur regime on trees of Sauropodomorpha and Theropoda (however, the regimes characterizing these relatively short-lived grades may not have persisted for sufficiently long to constrain model fits accurate).

The expected asymptotic trait variance is an estimate of the disparity that could have accumulated under a single macroevolutionary regime, and our confidence in these estimates if boosted by the observation that they are similar to the observed clade disparities (e.g. Fig. 13A for sauropods). Expected asymptotic trait variances in our models are small compared to differences between trait optima (θ) for successive regimes (Fig. 14A–B: typically 1–2 orders of magnitude [1–2 $\log_{10}$ kg]). This pattern may largely explain the goodness of fit of multi-regime OU models, and indicates that most of the variance of dinosaur body sizes that evolved during the Mesozoic resulted from stepwise shifts in the 'optimal' body sizes of groups of lineages,

and not from the gradual expansion of lineages within regimes to wider trait spaces. In other words: large, constraint-breaking body size changes occur on a minority of branches corresponding to the transitions between macroevolutionary regimes. These branches are associated with high magnitude shifts in trait values, as documented in previous work (Benson *et al*. 2014*a*: as high rates of body size change at nodes). And this overall pattern is also consistent with that inferred from patterns of body size change at different taxonomic ranks among Phanerozoic marine invertebrates (Heim *et al*. 2015).

Transitions between macroevolutionary regimes caused rapid, directional trait evolution, under which most dinosaurian body size disparity was accumulated. For example, sauropodomorphs explored roughly three orders of magnitude of body size variation due to a minimum of two regime shifts that occurred during less than 40 million years of the Triassic (Fig. 7). These gave rise to high body size disparity among Triassic and Early Jurassic sauropodomorphs, which was comparable to that of all dinosaurs during those intervals (Fig. 13A). During their subsequent evolution, which was dominated by occupancy of a single regime, sauropodomorphs explored little more than a single order of magnitude in body size (Figs 7,13A). In an even more extreme example, tetanuran theropods explored more than three orders of magnitude of body size variation though a minimum of three regime shifts lasting less than 10 million years during the Middle Jurassic (Figs 9–10). This culminated in apparently the greatest body size disparity of (non-pygostylian) theropod evolutionary history during the Middle Jurassic (Fig. 13B). Also during the Early–Middle Jurassic, expansions over two orders of magnitude

resulted from three regime shifts among ornithischian dinosaurs (Fig. 8; in Thyreophora, Marginocephalia and Iguanodontia). The hypothesis that dinosaur body size disparity resulted from gradual dispersal of lineages through trait space is easy to reject: even variable-rate Brownian motion models provide substantially worse fit to the data than do multi-peak OU models.

The adaptive landscape and suboptimal or unrealised regions of dinosaurian body size space. The pulsed pattern of body size evolution documented here is consistent with the hypothesis of quantum evolution, or the existence of a heterogeneous macroevolutionary adaptive landscape (Simpson 1953; Hansen 2013). Under this paradigm, transitions between macroevolutionary adaptive zones or peaks involve rapid, directional evolutionary treks across apparently unstable regions of trait space. Given that the correspondence between body size and ecological niches is coarse, it is clear that the peaks represented by our macroevolutionary regimes are broad entities. We therefore view them as comparable to the 'maximum adaptive zones' of Stanley (1973) rather than as fitness functions that are exactly shared among species (see also Hansen 1997, 2013 for a discussion of the difficulties scaling microevolutionary models to macroevolutionary observations). Furthermore, the apparent timescales of transitions between regimes span several million years, a long duration compared to microevolutionary timescales that are measured in generations. In general, macroevolutionary changes occur at slower rates than predicted from simple extrapolation of microevolutionary changes to longer timescales (e.g. Uyeda *et al*. 2011). Hansen (2013) has described a model in which microevolutionary

changes result from rapid attraction to adaptive peaks, whereas macroevolutionary changes reflect shifts in the positions of those peaks in the adaptive landscape due to changes in the abiotic and biotic environments. In this context, our study provides information on the topology of the macroevolutionary adaptive landscape for dinosaurs.

There may be ultimate limits to the maximum and minimum body sizes attainable by terrestrial vertebrates. Dinosaurs seem to have approached both during their evolutionary history, attaining maximum body masses as small as 2 g in birds, and as large as 40 tonnes (Mazetta *et al*. 2004; Lacovara *et al* 2014; Bates *et al*. 2015) and perhaps 90 tonnes (*Argentinosaurus*, herein) in giant sauropods (Table 4). The regimes relevant to these extremely large and extremely small body sizes are both unique, having each evolved only once among Mesozoic dinosaurs (Fig. 14D,F). This contrasts with the more frequent attainment of less extreme body size regimes in other groups of dinosaurs (Fig. 14C–F). One possible explanation is that extreme small and large body sizes both require fundamental structural reorganization. If so, then terrestrial tetrapods significantly outside of the body size range seen in dinosaurs will rarely, or never, evolve.

An apparently underexploited region of trait space between 30 and 300 kg (1.5–2.5 $\log_{10}$ kg; grey rectangles in Figs 14–15) was represented by the trait optimum of only one regime, meaning that it was attained only once during dinosaur evolution, in early tetanuran or coelurosaurian theropods (Figs 9–10). In fact, many coelurosaur species of this size are known, so the central portion of dinosaurian body size space is densely populated among theropods (Figs 14–15). Sauropodomorpha and Ornithischia do contain taxa

within this size range. However these taxa belong either to lineages making rapid transitions from smaller- to larger-body sized regimes, or are at the margins of the body size distributions for their clades. This is most evident in the body mass frequency distributions for Jurassic–Early Cretaceous ornithischians (Fig. 15E,H,L) and Triassic sauropodomorphs (Fig. 15A). It is less evident during other times because large-bodied ornithischians are not known from the Triassic (Fig. 15B) and small-bodied sauropodomorphs are absent by the Middle Jurassic (Fig. 15G). It is likely that the proportion of dinosaur species at smaller body sizes has been underestimated due to taphonomic biases that limit the preservation potential of individuals weighing less than 60 kg in some deposits (Brown *et al*. 2013). Nevertheless, our data confirm the presence of a bimodal or trimodal body mass distribution in dinosaurs: fitting a set of mixture models (McLachlan & Basford 1988) using the R package mclust (Fraley *et al*. 2012) finds support for models positing that the distribution of log-transformed dinosaur body sizes is composed of a mixture of either two or three normally distributed sub-distributions. These models receive similar levels of support from Bayesian information criteria (BIC), and are substantially better (ΔBIC > 100) than a model with a single, normal distribution for dinosaur body size. A bimodal body size distribution was also proposed by some previous studies (Codron *et al*. 2012; O'Gorman and Hone 2012), and this is explained here by the bimodal distribution of body size optima for macroevolutionary regimes (Fig. 14).

Previous studies (Codron *et al*. 2012; O'Gorman & Hone 2012) invoked the non-uniformitarian life history mode of dinosaurs (e.g. Janis & Carrano 1992; Varicchio 2011) to explain bimodality of dinosaur body sizes, arguing

that oviparity and high fecundity resulted in an abundance of medium-sized individuals representing juveniles of large-bodied dinosaur species in Mesozoic ecosystems, reducing the fitness of medium-sized adults. This hypothesis is somewhat consistent with our models, although the theropod clade Coelurosauria included high species diversity at intermediate body masses (Figs 9–10, 14E, 15Q: Late Cretaceous Theropoda), which is not consistent with ecological exclusion. Furthermore, a bimodal pattern of body size evolution is also seen in Cenozoic mammals, in which lineages are either constrained by attraction to small, plesiomorphic body masses or otherwise attracted to large, derived body masses, with an apparent lack of optima at intermediate sizes (0.4 kg–8 kg; Alroy 1999). The different numerical values of dinosaurian and mammalian body size optima suggest that these patterns are analogous, rather than strictly equivalent, and may have different causes. Nevertheless, this presence of bimodality in mammalian body size evolution suggests that distinct life history factors need not be invoked to explain the same pattern in dinosaurs. Instead, bimodal body size dynamics have characterized terrestrial vertebrates since early in the Mesozoic. The microevolutionary explanation of this pattern is currently unknown. Furthermore, some of the best-sampled dinosaur assemblages are biased against preservation of smaller taxa (Brown *et al*. 2013) and detailed appraisals of the effect of this bias on body size distributions and macroevolutionary patterns, such as those obtained here, have yet to be carried out.

Niche-filling (early burst) or constrained models of trait evolution?

Rapid increases in disparity are observed early in the fossil records of many

animal groups (e.g. Foote 1997*a*; Hughes *et al*. 2013). A similar pattern occurs for the body mass disparity of non-pygostylian dinosaurs, and in the major subclades Sauropodomorpha (which peaked in the Triassic), Ornithischia (which peaked in the Late Jurassic) and non-pygostylian Theropoda (which peaked in the Middle Jurassic) (Fig. 13). A key question that can be addressed by phylogenetic studies of trait evolution such as ours is whether these patterns result from high, but decelerating, early rates of evolution (the 'early burst' model; Harmon *et al*. 2010) or from constraints on the range of phenotypes attainable by a clade. Under the constraint hypothesis, trait space becomes rapidly saturated (e.g. Slater 2013; Oyston *et al*. 2015) even if evolutionary rates do not decelerate. The early burst model was specifically formulated to test Simpson's niche-filling hypothesis of adaptive radiation (Simpson 1954; Harmon *et al*. 2008). This hypothesis suggests a link between rates of evolution and the availability of ecological niches, which could be made available by key innovations that provide access to new adaptive zones, or by other factors such as mass extinctions, or geographic dispersal (reviewed by Etienne & Haegeman 2012).

A 'constrained' pattern, as represented by Ornstein-Uhlenbeck (OU) models, is evident from our results, and is consistent with the widespread existence of constraints on body size evolution within animal groups (Heim *et al*. 2015: marine animals). Nevertheless, pulses of body size evolution that cause increases in phenotypic disparity are associated with macroevolutionary regime shifts, and these are not homogeneously distributed through time. In fact, sauropodomorph and ornithischian regime shifts are concentrated in the first half of the evolutionary history of dinosaurs

(Fig. 14B), in the Triassic (sauropodomorphs) and Triassic–Jurassic (ornithischians; especially the Early/Middle Jurassic following a major extinction event at the Triassic/Jurassic boundary; e.g. Raup & Sepkoski 1982; Brusatte *et al*. 2008). Therefore, the temporal distribution of macroevolutionary pulses seen in non-theropod dinosaurs is arguably consistent with a broad scale niche-filling model of dinosaurian body size evolution (i.e. adaptive radiation). Under this hypothesis, opportunities (i.e. access to unfilled niches) resulted from fundamental key innovations of the dinosaur body plan in the Triassic, and from extinctions among of potential competitors around the Triassic/Jurassic boundary.

In contrast, theropod dinosaurs underwent a large number of high magnitude body size changes in the Cretaceous (Benson *et al*. 2014*a*), to the extent that post-Aptian macroevolutionary regimes of theropods could not be characterized in the present study, and will likely not be characterized until we obtain a significantly greater sample of their fossil diversity. In illustration of the intensity with which Cretaceous theropods explored body size space, just one clade, pygostylian birds made their first appearance in the Early Cretaceous (Zhang *et al*. 2000, 2008; Wang *et al*. 2015), and explored a body size range of approximately 4.5 orders of magnitude by the end of the Cretaceous (Fig. 14,15N,R; Tables 2, 5), over a time interval lasting only one-third of the total evolutionary history of Dinosauria (Fig. 13C). Cretaceous pygostylians range from 15 g in taxa such as *Qiliania* and *Iberomesornis* to 190 kg in *Gargantuavis* (Buffetaut & Le Loeuff 1998). This range is only slightly less than the range of body masses achieved by non-avian dinosaurs through the entire Mesozoic (maximum 5.5 orders of magnitude: from 434 g in

Microraptor, the smallest certainly non-juvenile non-avian dinosaur up to a maximal estimate of 95,000 kg in *Argentinosaurus*; smaller masses are recorded in some alvarezsauroids: *Parvicursor* [153 g] and *Ceratonykus* [262 g], but we cannot confirm that these specimens represent adults; Table 3). Given the expansive pattern of body size evolution seen in Cretaceous pygostylians, and the high frequency of large evolutionary body size changes among other Cretaceous theropods (Benson *et al*. 2014*a*), the pattern of body size evolution seen in theropod dinosaurs is not consistent with any conception of the 'early burst' model. Nevertheless, it could be reconciled with niche-filling as a mechanism of evolutionary diversification if some groups are able to maintain high levels of biological versatility (Vermeij 1973; = 'evolvability') that allow the continued discovery of new niches, whilst others do not (Benson *et al*. 2014*a*).

CONCLUSIONS

(1) We infer scaling relationships among measurements of dinosaur limb bones to estimate the adult body masses of 526 dinosaur species.

(2) Scaling relationships among the hindlimb elements of bipedal dinosaurs, and among the forelimb and hindlimb elements of quadrupedal dinosaurs are typically conserved within groups such as Ornithischia, Sauropodomorpha and Theropoda (bipedal) and Sauropoda, Thyreophora and Ceratopsidae (quadrupedal), as they do not show strong phylogenetic signal. This suggests that locomotor design is functionally constrained within clades with distinct

body plans, but that it varies among those clades, resulting from shifts in large limb proportions at just a few nodes on the dinosaurian tree.

(3) Trend-based models, which describe strict conceptions of Cope's rule, provide poor explanations of dinosaurian body mass evolution in almost all cases. We therefore reject the widespread presence of multi-lineage directional trends of body size evolution in dinosaurs.

(4) Instead, dinosaurian body mass can be explained by multi-regime Ornstein-Uhlenbeck models, which describe constrained evolution around macroevolutionary 'optima' or adaptive zones (Stanley 1972; Hansen 2013). Rare shifts in these optima cause lineages to undergo rapid excursions through trait space characterized as 'quantum evolution' by Simpson (1954), and under which most of the disparity of dinosaurian body masses accumulated.

(5) Our analyses reject two recent hypotheses regarding miniaturization on the bird stem lineage. (i) Theropod dinosaurs did not undergo sustained evolutionary body size reductions from the Triassic to the Late Jurassic. Support for this hypothesis resulted from overestimation of the ancestral theropod body mass. (ii) Pygostylian birds, principally in the range of 15–300 g do not share a body size optimum with other paravians (minimum body masses ~ 400 g; maximum: 600 kg; Tables 3, 5).

(6) Dinosaur body mass optima are unevenly distributed among clades and through time. Most sauropodomorph regime shifts occurred in the Triassic, most ornithischian regime shifts occurred in the Triassic–Jurassic, and theropods continued to undergo regime shifts for their entire evolutionary history. This pattern might be explained by the distribution of opportunity

resulting from key innovations during the early evolutionary history of Dinosauria, extinctions of potentially competing taxa at the Triassic/Jurassic boundary, and potentially from greater evolutionary versatility in theropods than in other dinosaurs.

(7) Body mass optima are unevenly distributed in trait space. Shifts to larger body size are more frequent than to smaller size in ornithischians and sauropodomorphs, and the distribution of body size optima is bimodal, with a sparsely populated intermediate region of trait space around 30–300 kg. This explains previous observations of bimodality in dinosaur body mass distributions. However, as bimodality is also present in mammalian body size evolution, the occurrence of bimodal dinosaur body mass evolution may not require an explanation that invokes specific attributes of dinosaur life history.

(8) Dinosaur body size disparity can be explained predominantly by constrained (Ornstein-Uhlenbeck) models of evolution. This suggests that the early attainment of maximal disparity seen in many groups of animals may result from constraints on the range of phenotypes achievable by a clade.

Acknowledgements. We thank Jeremy Beaulieu for advice on the implementation of OUwie, and Graham Slater and Graeme Lloyd for discussion. Cecile Ané gave invaluable guidance on implementation of pBIC for SURFACE. David Bapst, Marcello Ruta and Stephen Brusatte provided critical comments in review, which improved the manuscript. Parts of this work were funded by the European Union's Horizon 2020 research and innovation programme 2014–2018 under grant agreement 677774 (ERC Starting Grant:

TEMPO) to RBJB. Dinosaur silhouettes used with thanks to the artist, Scott Hartman (non-avian dinosaurs) and Nobu Tamura (bird silhouette).

SUPPORTING INFORMATION

Additional Supporting Information can be found in the online version of this article. Data for this study are available in the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.1t3r4. [please note that the data for this paper are not yet published and this temporary link should not be shared without the express permission of the author].

Appendix S1. Methods and results for dinosaur body mass estimates, including Tables S1–S12.

Appendix S2. Summary figures from SURFACE analyses for Triassic–Jurassic Dinosauria, Sauropodomorpha, Ornithischia and Triassic–Aptian Theropoda, ordered by AICc or pBIC score.

Appendix S3. Diagrams comparing the pBIC scores of SURFACE regime configurations from analysis of timescaled phylogenies mapped to the same phylogenies timescaled using other methods.

Appendix S4. Diagrams illustrating trajectories of stepwise-AICc and stepwise-pBIC searches following the SUFACE algorithm and resulting regime configurations.

Appendix S5. Phenograms illustrating regime configurations mapped to phylogenies evolving in body size space through time.

Dataset S1. Measurements, metadata and body mass estimates for Dinosauria.

Dataset S2. AICc scores for trend-based model fits compared to Brownian motion.

Dataset S3. AICc scores and model fits for OU and other models implemented in Ouwie.

Dataset S4. Phylogenetic trees used in our analyses, in Newick format.

Scripts S1. Scripts used in R version 3.3.3 to conduct analyses in the current paper including estimation of standard errors of variables estimated through regression, fitting models of trends that can shift over time or across nodes, and versions of surfaceForward and surfaceBackward (Ingram & Mahler 2013) that use pBIC (Khabbazian et al. 2016), rather than AICc.

REFERENCES

AKAIKE, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.

ALEXANDER, R. M. 1998. All-time giants: the largest animals and their problems. *Palaeontology,* **41**,1231–1245.

ALROY, J. A. 1999. Cope's rule and the dynamics of body mass evolution in North American fossil mammals. *Science*, **280**, 731–734.

ANDERSON, J. F., HALL-MARTIN, A. and 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.

ANÉ, C. 2008. Analysis of comparative data with hierarchical autocorrelation. *Annals of Applied Systematics*, **2**, 1078–1102.

APALDETTI, C., POL, D. and YATES, A. 2013. The postcranial anatomy of *Coloradisaurus brevis* (Dinosauria: Sauropodomorpha) from the Late Triassic of Argentina and its phylogenetic implications. *Palaeontology*, **56**, 277–301.

BAKER, J., MEADE, A., PAGEL, M. and VENDITTI, C. 2015. Adaptive evolution toward larger size in mammals. *Proceedings of the National Academy of Sciences of the USA*, **112**, 5093–5098.

BAKKER, R. T. 1971. Ecology of the brontosaurs. *Nature*, **229**, 172–174.

BAPST, D. W. 2012. paleotree: an R package for paleontological and phylogenetic analyses of evolution. *Methods in Ecology and Evolution*, **3**, 803–807.

— 2013. A stochastic rate-calibrated method for time-scaling phylogenies of fossil taxa. *Methods in Ecology and Evolution*, **4**, 724–733.

— 2014*a*. Preparing paleontological datasets for phylogenetic comparative methods. 515–544. *In* GARAMSZEGI, L. Z. (ed.). *Modern Phylogenetic*

Comparative Methods and Their Application in Evolutionary Biology. Springer: Heidelberg.

— 2014*b*. Assessing the effect of time-scaling methods on phylogeny-based analyses in the fossil record. *Paleobiology*, **40**, 331–351.

— and HOPKINS S. 2017. Comparing *cal3* with other a posteriori time-scaling approaches in a case study with the pterocephaliid trilobites. *Paleobiology* **43**, 49–67.

BATES, K. T., FALKINGHAM, P. L., MACAULAY, S,. BRASSEY, C. and MAIDMENT, S. C. R. 2015. Downsizing a giant: re-evaluating *Dreadnoughtus* body mass. *Biology Letters*, **11**, 20150215.

BEAULIEAU, J. M., JHUWUENG, D. –C., BOETTIGER, C. and O'MEARA, BC. 2012. Modelling stablizing selection: expanding the Ornstein–Uhlenbeck model of adaptive evolution. *Evolution*, **66**, 2369–2383.

BENSON, R. B. J. and CHOINIERE, J. N. 2013. Rates of dinosaur limb evolution provide evidence for exceptional radiation in Mesozoic birds. *Proceedings of the Royal Society B*, **280**, 20131780.

—, BUTLER, R. J., CARRANO, M. T. and O'CONNOR, P. M. 2012. Air-filled postcranial bones in theropod dinosaurs: physiological implications and the 'reptile'-bird transition. *Biological Reviews*, **87**, 168–193.

—, CAMPIONE, N. E., CARRANO, M. T., MANNION, P. D., SULLIVAN, C., UPCHURCH, P. and EVANS, D. C. 2014*a*. Rates of dinosaur body mass evolution indicate 170 million years of sustained ecological innovation on the avian stem lineage. *PLOS Biology*, **12**(5), 1001853.

—, FRIGOT, R. A., GOSWAMI, A., ANDRES, B. and BUTLER, R. J. 2014*b*. Competition and constraint drove Cope's rule in the evolution of giant flying reptiles. *Nature Communications*, **5**, 3567.

—, HUNT, G., CARRANO, M. T. and CAMPIONE, N. 2017. Data from: Cope's rule and the adaptive landscape of dinosaur body size evolution. *Dryad Digital Repository*. doi.org/10.5061/dryad.1t3r4.

BENTON, M. J., CSIKI, Z., GRIGORESCU, D., REDELSTORFF, R., SANDER, P. M., STEIN, K. and WEISHAMPEL, D. B. 2010. Dinosaurs and the island rule: The dwarfed dinosaurs from Haţeg Island. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **293**, 438–454.

BONAPARTE, J. F. and CORIA, R. A. 1993. Un Nuevo y gigantesco sauropodo titanosaurio de la Formacian Rio Limay (Albiano-Ceonmaniano) de la Provincia del Neuquen, Argentina. *Ameghiniana*, **30**, 271–282.

— and PUMARES, J. A. 1995. Notas sobre el primer craneo de *Riojasaurus incertus* (Dinosauria, Prosauropoda, Melanorosauridae) del Triasico Superios de La Rioja, Argentina. *Ameghiniana*, **32**, 341–349.

BONNAN, M. F. 2003. The evolution of manus shape in sauropod dinosaurs: implications for functional morphology, forelimb orientation, and phylogeny. *Journal of Vertebrate Paleontology*, **23**, 595–613.

BROWN, J. H. 1995 *Macroecology*. University of Chicago Press, Chicago, 269 pp.

— and MAURER, B. A. 1986. Body size, ecological dominance, and Cope's Rule. *Nature*, **324**, 248–250.

BROWN, C. M., EVANS, D. C., CAMPIONE, N. E., O'BRIEN, L. J. and EBERTH, D. A. 2013. Evidence for taphonomic size bias in the Dinosaur Park Formation (Campanian, Alberta), a model Mesozoic terrestrial alluvial-paralic system. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **372**, 108–122.

BRUSATTE, S. L., BENTON, M. J., RUTA, M. and LLOYD, G. T. 2008. Superiority, competition, and opportunism in the evolutionary radiation of dinosaurs. *Science*, **321**, 1485–1487.

BUFFETAUT, E., LE LOEUFF, J., MECHIN, P. and MECHIN-SALESSY, A. 1995. A large French Cretaceous bird. *Nature*, **377**, 110.

BURNESS, G. P., DIAMOND, J. and FLANNERY, T. 2001. Dinosaurs, dragons, and dwarfs: the evolution of maximal body size. *Proceedings of the National Academy of Sciences of the USA*, **25**, 14518–14523.

BURNHAM, K. P. and ANDERSON, D. R. 2004 *Model selection and multimodel inference*. Springer, New York, 488 pp.

—, — and HUYVAERT, K. P. 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioural Ecology and Sociobiology*, **65**, 23–35.

BUTLER, M. A. and KING, A. A. 2004. Phylogenetic comparative analysis: a modelling approach for adaptive radiation. *American Naturalist*, **164**, 683–695.

BUTLER, R. J., GALTON, P. M., PORRO, L. B., CHIAPPE, L. M., HENDERSON, D. M. and 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.

CAMPBELL, K. E. and 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.

CAMPIONE, N. E. and EVANS, D. C. 2012. A universal scaling relationship between body mass and proximal limb bone dimensions in quadrupedal terrestrial tetrapods. *BMC Biology*, **10**, 1–21.

—, EVANS, D. C., BROWN, C. M. and CARRANO, M. T. 2014. Body mass estimation in non-avian bipeds using a theoretical conversion to quadrupedal stylopodial proportions. *Methods in Ecology and Evolution*, **5**, 913–923.

CANTALAPIEDRA J. L., PRADO, J. L., FERNÁNDEZ M. H. and ALBERDI, M. T. 2017. Decoupled ecomorphological evolution and diversification in Neogene–Quaternary horses. *Science*, **355**, 627–230.

CARBALLIDO, J. L., POL, D., OTERO, A., CERDA, I. A., SALGADO, L., GARRIDO, A. C., RAMEZANI, J., CÚENO, N. R. and KRAUSSE, J. M. A new giant titanosaur sheds light on body mass evolution among sauropod dinosaurs. *Proceedings of the Royal Society B*, **284**, 20171219.

CARPENTER, K. 2006. Biggest of the big: a critical re-evaluation of the mega-sauropod *Amphicoelias fragillimus*. 131–138, *In* FOSTER, J. R. and LUCAS, S. G., (eds). *Paleontology and Geology of the Upper Jurassic Morrison Formation*. *New Mexico Museum of Natural History and Science Bulletin* 36.

CARRANO, M. T. 1998 Locomotion in non-avian dinosaurs: integrating data from hindlimb kinematics, in vivo strains, and bone morphology. *Paleobiology*, **34**, 450–469.

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

— 2006. Body-size evolution in the Dinosauria. 225–268. *In* CARRANO, M. T., GAUDIN, T. J., BLOB, R. W. and WIBLE, J. R. (eds). *Amniote Paleobiology*. University of Chicago Press, Chicago.

— 2007. The appendicular skeleton of *Majungasaurus crenatissimus* (Theropoda: Abelisauridae) from the Late Cretaceous of Madagascar. *Journal of Vertebrate Paleontology*, **27**(suppl. 2), 163–179.

CHINNERY, B. 2004. Morphometric analysis of evolutionary trends in the ceratopsian postcranial skeleton. *Journal of Vertebrate Paleontology*, **24**, 591–609.

— and HORNER J. R. 2007. A new neoceratopsian dinosaur linking North American and Asian taxa. *Journal of Vertebrate Paleontology*, **27**, 625–641.

CLOSE, R. A., FRIEDMAN M., LLOYD, G. T. and BENSON, R. B. J. 2015. Evidence for a Mid-Jurassic Adaptive Radiation in Mammals. *Current Biology*, **25**, 2137–2142.

CODRON, D., CARBONE, C., MÜLLER, D. W. H. and CLAUSS, M. 2012. 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.

COOPER, M. R. A reassessment of *Vulcanodon karibaensis* Raath (Dinosauria: Saurischia) and the origin of Sauropoda. *Palaeontologia Africana*, **25**, 203–231.

COOPER, N., THOMAS, G. H., VENDITTI, C., MEADE, A. and FRECKLETON, R. P. 2016. A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. *Biological Journal of the Linnean Society*, **118**, 64–77.

DAVIS, A. M. and BETANCUR-R, R. 2017. Widespread ecomorphological convergence in multiple fish families spanning the marine-freshwater interface. *Proceedings of the Royal Society B*, **284**, 20170565.

DE SOUZA, L. M. and SANTUCCI, R. M. 2014. Body size in Titanosauriformes (Sauropoda, Macronaria). *Journal of Evolutionary Biology*, **27**, 2001–2012.

DEL HOYO, J., ELLIOTT, A. and SARGATAL, J. 1999. *Handbook of the Birds of the World*, volume 5: Barn Owls to Hummingbirds. Lynx Edicions, Barcelona, 759 pp.

ERICKSON, G. M. 2014. On dinosaur growth. *Annual Review of Earth and Planetary Sciences*, **42**, 675–697.

—, CURRIE, P. J., INOUYE, B. D. and WINN, A. A. 2006. Tyrannosaur life tables: The first look at non-avian dinosaur population biology. *Science*, **313**, 213–217.

—, RAUHUT, O. W. M., ZHOU, Z. –H., TURNER, A. H., INOUYE, B. D., HU, D. –Y, NORELL, M. A. 2009*a* Was dinosaurian physiology inherited by birds? Reconciling slow growth in *Archaeopteryx*. *PLOS ONE,* **4**, e7390.

—, MAKOVICKY, P. J., INOUYE, B. D., ZHOU, C. –F. and GAO, K. 2009*b*. Life table for *Psittacosaurus lujiatunensis*: The first glimpse into ornithischian dinosaur population biology. *Anatomical Record*, **292**, 1514–1521.

—, CURRIE, P. J., INOUYE, B. D., and WINN, A. A. 2010. A revised Life table and survivorship curve for *Albertosaurus sarcophagus* based on the Dry Island mass death assemblage. *Canadian Journal of Earth Sciences*, **47**, 1269–1275.

ETIENNE, R. S. and HAEGEMANN, B. 2012. A conceptual and statistical framework for adaptive radiations with a key role for diversity dependence. *American Naturalist*, **180**, E75–E89.

FELSENSTEIN, J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. *American Journal of Human Genetics*, **25**, 471–492.

— 1985. Phylogenies and the comparative method. *American Naturalist*, **125**, 1–15.

FOOTE, M. 1997*a*. The evolution of morphological diversity. *Annual Review of Ecology and Systematics*, **28**, 129–152.

— 1997*b*. Estimating taxonomic durations and preservation probability. *Paleobiology*, **23**, 278–300.

FRALEY, C., RAFTERY, A. E., MURPHY, T. B. and SCRUCCA, L. 2012. mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation Technical Report No. 597, Department of Statistics, University of Washington.

GALTON, P. M. 1970. The posture of hadrosaurian dinosaurs. *Journal of Paleontology*, **44**, 464–473.

— 1974. Notes on *Thescelosaurus*, a conservative ornithischian dinosaur from the Upper Cretaceous of North America, with comments on ornithopod classification. *Journal of Paleontology*, **48**, 1048–1067.

GARLAND, T. JR. and IVES, A. R. 2001. Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. *American Naturalist*, **155**, 346–364.

GRILLO, O. and DELCOURT, R. 2017. Allometry and body length of abelisauroid theropods: *Pycnonemosaurus nevesi* is the new king. *Cretaceous Research*, **69**, 71–89.

HANSEN, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution*, **51**, 1341–1351.

— 2013. Adaptive landscapes and macroevolutionary dynamics. 205–226. *In* SVENSSON E. and CALSBEEK R. (eds). *The adaptive landscape in evolutionary biology*. Oxford University Press, Oxford.

— and MARTINS EP. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution*, **50**, 1404–1417.

HARMON, L. J., LOSOS, J. B., DAVIES, T. J., GILLESPIE, R. G., GITTLEMAN, J. L., JENNINGS, W. B., KOZAK, K. H., MCPEEK, M. A. MORENO-ROARK, F., NEAR, T. J., PURVIS, A., RICKLEFS, R. E., SCHLUTER, D., SCHULTE, J. A. II, SEEHAUSEN, O., SIDLAUSKAS, B. L., TORRES-CARAVAJAL, O., WEIR, J. T. and MOOERS, A. O. 2010.

Early bursts of body size and shape evolution are rare in comparative data. *Evolution*, **64**, 2385–2396.

HEATH, T. A., HUELSENBECK, J. P. and STADLER, T. 2014. The fossilized birth-death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences of the USA*. **111**, E2957–E2966.

HEDMAN, M. 2010. Constraints on clade ages from fossil outgroups. *Paleobiology*, **36**, 16–31.

HEIM, N. A., KNOPE, M. L., SCHAAL, E. K., WANG, S. C. and PAYNE, J. L. 2015. Cope's rule in the evolution of marine animals. *Science*, **347**, 867–870.

HO, L. S. T. and ANÉ, C. 2014. Intrinsic inference difficulties for trait evolution with Ornstein-Uhlenbeck models. *Methods in Ecology and Evolution*, **5**, 1133–1146.

HONE, D. W. E. and BENTON, M. J. 2005. The evolution of large size: how does Cope's Rule work? *Trends in Ecology and Evolution*, **20**, 4–6.

— KEESEY T. M., PISANI D. and PURVIS A. 2005. Macroevolutionary trends in the Dinosauria: Cope's rule. *Journal of Evolutionary Biology*, **18**, 587–595.

HOPKINS M. J. and SMITH A. B. 2015. Dynamic evolutionary change in post-Paleozoic echinoids and the importance of scale when interpreting changes in rates of evolution. *Proceedings of the National Academy of Sciences of the U.S.A.*, **112**, 3758–3763.

HUGHES, M., GERBER, S. and WILLS, M. A. 2013. Clades reach highest morphological disparity early in their evolution. *Proceedings of the National Academy of Sciences of the USA*, **110**, 13875–13879.

HUNN, C. A. and UPCHURCH, P. 2001. The importance of time/space in diagnosing the causality of phylogenetic events: towards a 'chronobiogeographical' paradigm? *Systematic Biology*, **50**, 1–17.

HUNT, G. 2006. Fitting and comparing models of phyletic evolution:random walks and beyond. *Paleobiology*, **32**, 578–601.

— 2008. Evolutionary patterns within fossil lineages: model-based ASSESSMENT of modes, rates, punctuations and process. 117–131. *In* BAMBACH, R. K. and KELLEY, P. H. (eds). *From Evolution to Geobiology: Research Questions Driving Paleontology at the Start of a New Century*. The Paleontological Society.

— 2012. Measuring rates of phenotypic evolution and the inseparability of tempo and mode. *Paleobiology*, **38**, 351–373.

— and CARRANO, M. T. 2010. Models and methods for analyzing phenotypic evolution in lineages and clades. *Special Papers of the Paleontological Society*, **16**, 245–269.

— and ROY, K. 2006. Climate change, body size evolution, and Cope's Rule in deep-sea ostracodes. *Proceedings of the National Academy of Sciences of the USA*, **103**, 1347–1352.

INGRAM, T. and MAHLER, D. L. 2013. SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. *Methods in Ecology and Evolution*, **4**, 416–425.

IVES, A. R., MIDFORD, P. E. and GARLAND, T. JR. 2007. Within-species variation and measurement error in phylogenetic comparative methods. *Systematic Biology*, **56**, 252–270.

JANIS, C. M. and CARRANO, M. T. 1992. Scaling of reproductive turnover in archosaurs and mammals: why are large terrestrial animals so rare? *Acta Zoological Fennica*, **28**, 201–216.

KELLNER, A. W. A. and CAMPOS, D. A. 2002. On a theropod dinosaur (Abelisauria) from the continental Cretaceous of Brazil. *Arquivos do Museu Nacional Rio de Janeiro*, **60**, 163–170.

KHABBAZIAN M., KRIEBEL R., ROHE K. and ANÉ C. 2016. Fast and accurate detection of evolutionary shifts in Ornstein-Uhlenbeck models. *Methods in Ecology and Evolution*, **7**, 811–824.

KINGSOLVER, J. G. and PFENNIG, D. W. 2004. Individual-level selection as a cause of Cope's rule of phyletic size increase. *Evolution*, **58**, 1608–1612.

LACOVARA, K. J., IBIRICU, L. M., LAMANNA, M. C., POOLE, J. C., SCHROETER, E. R., ULLMANN, P. V., VOEGELE, K. K., BOLES, Z. M., EGERTON, V. M., HARRIS, J. D., MARTÍNEZ, R. D. and NOVAS, F. E. 2014. A gigantic, exceptionally complete titanosaurian sauropod dinosaur from southern Patagonia, Argentina. *Scientific Reports*, **4**, 6196.

LAPIEDRA, O., SOL, D., CARRANZA, S. and BEAULIEU, J. M. 2013. Behavioural changes and the adaptive diversification of pigeons and doves. *Proceedings of the Royal Society B*, **280**, 20122893.

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

LEE, A. H. and WERNING, S. 2008. Sexual maturity in growing dinosaurs does not fit reptilian growth models. *Proceedings of the National Academy of Sciences of the USA*, **105**, 582–587.

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

LLOYD, G. T., BAPST, D. W., FRIEDMAN, M. and DAVIS, K. E. 2016. Probabilistic divergence time estimation without branch lengths: dating the origins of dinosaurs, avian flight and crown birds. *Biology Letters*, **12**, 20160609.

LÜ, J. –C., XU, L., JIA, S. –H. and JI, Q. 2009. A new gigantic sauropod dinosaur from the Cretaceous of Ruyang, Henan, China. *Geological Bulletin of China*, **28**, 1–10.

MAHLER, D. L. and INGRAM, T. 2014. Phylogenetic comparative methods for studying clade-wide convergence. 425–450. *In* GARAMSZEGI, L. Z. (ed.). *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology*. Springer-Verlag, Berlin.

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

MARTINS, E. P. and HANSEN, T. F. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information

into the analysis of interspecific data. *American Naturalist*, **149**, 646–667.

MARYANSKA, T. and OSMÓLSKA, H. 1975. Protoceratopsidae (Dinosauria) of Asia. *Palaeontologia Polonica*, **33**,133–181.

MATLEY, C. A. 1924. Note on an armoured dinosaur from the Lameta beds of Jubbulpore. *Records of the Geological Survey of India*, **55**, 105–109.

MAZZETTA, G. V., CHRISTIANSEN, P. and FARIÑA, R. A. 2004. Giants and bizarres: body size of some southern South American Cretaceous dinosaurs. *Historical Biology*, **2004**, 1–13.

MCSHEA, D. W. 1994. Mechanisms of large-scale evolutionary trends. *Evolution*, **48**, 1747–1763.

MCLACHLAN, G. J. and BASFORD, K. E. 1988. *Mixture models: Inference and applications to clustering*. Marcel Dekker, New York, 253 pp.

MOTANI, R. and SCHMITZ, L. 2011. Phylogenetic versus functional signals in the evolution of form-function relatonships in terrestrial vision. *Evolution*, **65**, 2245–2257.

NORMAN, D. B. 1980. On the ornithischian dinosaur Iguanodon bernissatensis from the Lower Cretaceous of Bernissart (Belgium). *Mémoires de l'Institut Royal des Sciences Naturelles de Belgique*, **178**, 1–103.

— 1986. On the anatomy of *Iguanodon atherfieldensis* (Ornithischia: Ornithopoda). *Bulletin de l'Institut Royal des Sciences Naturelle de Belgique: Sciences de la Terre*, **56**, 281–372.

NOVAS, F. E., EZCURRA, M. D., AGNOLÍN, F. L., POL, D. and ORTÍZ, R. 2012. New Patagonian Cretaceous theropod sheds light about the early

radiation of Coelurosauria. *Revista del Museo Argentino de Ciencias Naturales, nueva serie*, **14**, 57–81.

— SALGADO, L., SUÁREZ, M., AGNOLÍN, F. L., EZCURRA, M. N. D., CHIMENTO, N. S. R., DE LA CRUZ, R., ISASI, M. P., VARGAS, A. O. and RUBILAR-ROGERS, D. 2015. An enigmatic plant-eating theropod from the Late Jurassic period of Chile. *Nature*, **522**, 331–334.

O'CONNOR, P. M. 2009 Evolution of archosaurian body plans: skeletal adaptations of an air-sac-based breathing apparatus in birds and other archosaurs. *Journal of Experimental Zoology*, **311A**, 629–646.

O'CONNOR, J., CHIAPPE, L. M. and BELL, A. 2011. *Pre-modern birds: avian divergences in the Mesozoic*. 39–116. *In* DYKE, G. and KAISER, G. (eds). *Living Dinosaurs: the Evolutionary History of Modern Birds*. Wiley-Blackwell, Oxford.

O'GORMAN, E. J. and HONE, D. W. E. 2012. Body size distribution of the dinosaurs. *PLOS ONE*, **7**(12), e51925.

O'MEARA, B. C., ANÉ, C., SANDERSON, M. J. and WAINWRIGHT, P. C. 2006. Testing for different rates of continuous trait evolution using likelihood. *Evolution*, **60**, 922–933.

ŐSI, A., PRONDVAI, E., BUTLER, R. and 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**(9), e44318.

OYSTON, J. W., HUGHES, M., WAGNER, P. J., GERBER, S. and WILLS, M. A. 2015. What limits the morphological disparity of clades? *Interface Focus* **5**, 20150042.

PAGEL, M. 1999. Inferring the historical patterns of biological evolution. *Nature*, **401**, 877–884.

— 2002. Modelling the evolution of continuously varying characters on phylogenetic trees: the case of hominid cranial capacity. 269–286. *In* MACLEOD, N. and FOREY, P. L. (eds). *Morphology, Shape and Phylogeny*. Taylor and Francis, London.

PARADIS, E., CLAUDE, J. and STRIMMER, K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.

PINHEIRO, J., BATES, D., DEBROY, S., SARKAR, D. and R CORE TEAM. 2014. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-117, <URL: http://CRAN.R-project.org/package=nlme>.

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

R CORE TEAM. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

RAUP, D. M. and SEPKOSKI, J. J. 1976. Mass extinctions in the marine fossil record. *Science*, **215**, 1501–1503.

REVELL, L. J. 2012. phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, **3**, 217–223.

SAARINEN, J. J., BOYER, A. G., BROWN, J. H., COSTA, D. P., ERNEST, S. K. M., EVANS, A. R., FORTELIUS, M., GITTLEMAN, J. L., HAMILTON, M. J., HARDING, L. E., LINTULAAKSO, K., LYONS, S. K., OKIE, J. G., SIBLY, R. M., STEPHENS, P. R. THEODOR, J. UHEN, M. D. and

SMITH, F. A. 2014. Patterns of maximum body size evolution in Cenozoic land mammals: eco-evolutionary processes and abiotic forcing. *Proceedings of the Royal Society B*, **281**, 20132049.

SANDER, P. M. and CLAUSS, M. 2008. Perspective: sauropod gigantism. *Science*, **322**, 200–201.

—, MATEUS, O., LAVEN, T. and KNÖTSCHKE, N. 2006. Bone histology indicates insular dwarfism in a new Late Jurassic sauropod dinosaur. *Nature*, **441**, 739–741.

—, CHRISTIAN, A., CLAUSS, M., FECHNER, R., GEE, C. T.,GRIEBLER, E. -M., GUNGA, H. -C., HUMMEL, J., MALLISON, H., PERRY, S. F., PREUSCHOFT, H., RAUHUT, O. W. M., REMES, K., TÜTKEN, T., WINGS, O. and WITZEL, U. 2010. Biology of the sauropod dinosaurs: the evolution of gigantism. *Biological Reviews*, **86**, 117–155.

SCHLUTER, D., PRICE, T., MOOERS, A. O. and LUDWIG, D. 1997. Likelihood of ancestor states in adaptive radiation. *Evolution*, **51**, 1699–1711.

SERENO, P. C. 1997. The origin and evolution of dinosaurs. *Annual Reviews of Earth and Planetary Sciences*, **25**, 435–489.

— 1999. The evolution of dinosaurs. *Science*, **284**, 2137–2147.

SHEETS, H. D. and MITCHELL, C. E. 2001. Why the null matters: statistical tests, random walks and evolution. *Genetica*, **112-113**, 105–125.

SILVESTRO, D., ANTONELLI, A., SALAMIN, N. and QUENTAL, T. B. 2015. The role of competition in evolutionary replacements of North American canids. *Proceedings of the National Academy of Sciences of the USA*, **112**, 8684–8689.

SIMPSON. G. G. 1953. The Major Features of Evolution. Columbia University Press, New York, 434 pp.

SLATER. G. J. 2013. Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the Cretaceous-Palaeogene boundary. *Methods in Ecology and Evolution*, **4**, 734–744.

— 2015. Iterative adaptive radiations of fossil canids show no evidence for diversity-dependent trait evolution. Proceedings of the National Academy of Sciences of the U.S.A., **112**, 4897–4902.

—, HARMON. L. J. and ALFARO. M. E. 2012. Integrating fossils wit molecular phylogenies improves inference of trait evolution. *Evolution*, **66**, 3931–3944.

—, GOLDBOGEN, J. A. and PYENSON, N. D. 2017. Independent evolution of baleen whale gigantism linked to Plio-Pleistocene ocean dynamics. *Proceedings of the Royal Society B*, **284**, 20170546.

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. and UHEN. M. D. 2012. The evolution of maximum body size of terrestrial mammals. *Science*, **330**, 1216–1219.

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

STANLEY, S. M. 1973. An explanation for Cope's rule. *Evolution*, **27**, 1–26.

STUBBS, T. L. and BENTON, M. J. 2016. Ecomorphological diversifications of Mesozoic marine reptiles, the roles of ecological opportunity and extinction. *Paleobiology*, **42**, 547–573.

SUGIURA, N. 1978. Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics — Theory and Methods*, **7**, 13–26.

TARVER, J. E. and DONOGHUE, P. C. J. 2011. The trouble with topology: phylogenies without fossils provide a revisionist perspective of evolutionary history in topological analyses of diversity. *Systematic Biology*, **60**, 700–712.

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

UPCHURCH, P., BARRETT, P. M. and GALTON, P. M. 2007. A phylogenetic analysis of basal sauropodomorph relationships: Implications for the origin of sauropod dinosaurs. *Special Papers in Palaeontology*, **77**, 57–90.

—, HUNN C. A. 2002. “Time”: the neglected dimension in cladistic biogeography. *Geobios*, **35**, 277–286.

UYEDA, J. C., HANSEN, T. F., ARNOLD, S. J. and PIENAAR, J. 2011. The million-year wait for macroevolutionary bursts. *Proceedings of the National Academy of Sciences of the USA*, **108**, 15908–15913.

VAN VALKENBURGH, B., WANG, X. -M. and DAMUTH, J. 2004. Cope’s rule, hypercarnivory, and extinction in North American canids. *Science*, **306**, 101–104.

VARICCHIO, D. J. 2011. A distinct dinosaur life history? *Historical Biology*, **23**, 91–107.

VERMEIJ, G. J. 1973. Biological versatility and Earth history. *Proceedings of the National Academy of Sciences of the* USA, **70**, 1936–1938.

WANG, M., ZHENG, X., O'CONNOR, J. K., LLOYD, G. T., WANG, X., WANG, Y., ZHANG, X. and ZHOU, Z. 2015. The oldest record of Ornithuromorpha from the Early Cretaceous of China. *Nature Communications*, **6**, 6987.

WERNER, J. and GRIEBELER, E. M. 2011. Reproductive biology and its impact on body size: comparative analysis of mammalian, avian and dinosaurian reproduction. *PLOS ONE*, **6**(12), e28442.

WERNING, S. 2012. The ontogenetic osteohistology of *Tenontosaurus tilletti*. *PLOS One*, **7**, e33539.

WOODRUFF, D. C. and FOSTER, J. R. 2014. The fragile legacy of *Amphicoelias fragillimus* (Dinosauria: Sauropoda; Morrison Formation – latest Jurassic). *Volumina Jurassica*, **12**, 211–220.

YATES, A. M. 2007 The first complete skull of the Triassic dinosaur *Melanorosaurus* Haughton (Sauropodomorpha: Anchisauria). *Special Papers in Palaeontology*, **77**, 9–55.

— and KITCHING, J. W. 2003. The earliest known sauropod dinosaur and the first steps towards sauropod locomotion. *Proceedings of the Royal Society B*, **270**, 1753–1758.

—, BONNAN, M. F., NEVELING, J., CHINSAMY, A. and BLACKBEARD, M. G. 2010. A new transitional sauropodomorph dinosaur from the Early

Jurassic of South Africa and the evolution of sauropod feeding and quadrupedalism. *Proceedings of the Royal Society B*, **277**, 787–794.

ZANNO, L. E. and MAKOVICKY, P. J. 2013. No evidence for directional evolution of body mass in herbivorous theropod dinosaurs. *Proceedings of the Royal Society B*, **280**, 20122526.

ZHANG, F. and ZHOU, Z. 2000. A primitive enantiornithine bird and the origin of feathers. *Science*, **290**, 1955–1959.

ZHANG, F., ZHOU, Z. and BENTON, M. J. 2008. A primitive confuciusornithid bird from China and its implications for early avian flight. *Science in China, series D — Earth Sciences*, **51**, 625–639.

ZHAO, Q., BENTON, M. J., SULLIVAN, C., SANDER, P. M. and XU, X. 2013. Histology and postural change during the growth of the ceratopsian dinosaur *Psittacosaurus lujiatunensis*. *Nature Communications*, **4**, 2079.

FIGURE CAPTIONS

FIG. 1. Simulated behavior of macroevolutionary models under various parameter values. A, B, Brownian motion (BM) under high (A) and low (B) values of the Brownian variance parameter showing diffusive behavior and unbiased directionality of trait evolution along lineages; C, D, Ornstein-Uhlenbeck (OU) models showing bounded or constrained evolution under (C) a single regime in which the trait optimum (θ) equals the starting values (Z_0), and (D) in which a second regime originates from the first and has a trait optimum that is greater than the starting value; E, F, trend models originating at the base of the phylogeny (E) and descending from an ancestral OU

regime (F). For this figure, a trend model with μ = 0.7 was approximated using an OU model with low α (= 0.005) and distant, unrealized θ (= 150).

FIG. 2. Simulated behavior of new trend-based models under various parameter values. A, node-shift model, showing a change in the trend parameter μ at a node on the phylogeny. B, time-shift model, showing a change in the trend parameter μ at a time during the evolutionary history of a group. Z0 is the trait value at the root node and σ is the Brownian variance. For this figure, a trend model with μ = 0.7 was approximated using an OU model with low α and distant, unrealized θ.

FIG. 3. Comparisons of AICc scores for SURFACE (multi-regime Ornstein-Uhlenbeck), best trend-based, and Brownian motion (BM1) models. A, Triassic–Jurassic Dinosauria; B, Sauropodomorpha; C, Ornithischia; and Triassic–Aptian Theropoda. Models were fit across 20 phylogenies scaled to time using the mbl1 algorithm, and results for each phylogeny are connected by lines. Results based on other timescaling algorithms were essentially identical. AICc scores for all trend-based models and BM1 models are given in Dataset S2. AICc scores and visualizations of SURFACE models are given in Appendix S2.

FIG. 4. Plots of log10(α) on AICc across 20 phylogenies for SURFACE model fits showing bimodal distribution of model outcomes for most groups. A, Triassic–Jurassic Dinosauria; B, Sauropodomorpha; C, Ornithischia; and Triassic–Aptian Theropoda. Models were fit across 20 phylogenies scaled to

time using the mbl1 algorithm.,Tree numbers correspond to those used throughout this paper, e.g. Appendix S2 and elsewhere.

FIG. 5. Comparison of low-AICc/high-α class and high-AICc/low-α class SURFACE model fits for Triassic–Jurassic Dinosauria. A, low-AICc/high-α class of model based on tree 9; B, high-AICc/low-α class of model based on tree 10. The suboptimal model fit is characterized by a low value of α (= 0.005 in (A) compared to 0.074 in (B)) and few distinct regimes. Red and blue lineages exhibit trend-like attraction to unrealized low (blue) or high (red) trait optima.

FIG. 6. SURFACE stepwise-AICc model for phylogeny 12 of Triassic–Jurassic Dinosauria. Results for other phylogenies show little variation from this (except that described above; Figs 4–5), and are presented in Appendix S2.

FIG. 7. SURFACE model fit and regime evolution through time for Sauropodomorpha. A, SURFACE stepwise-AICc model for phylogeny 17 of Sauropodomorpha. Results for other phylogenies show little variation from this and are presented in Appendix S2. B, Evolution of body size regimes in Sauropodomorpha simplified from (A) by collapsing each phylogenetically-independent multi-taxon regime to a single branch.

FIG. 8. SURFACE model fit and regime evolution through time for Ornithischia. A, SURFACE stepwise-AICc model for phylogeny 5 of Ornithischia. Results for other phylogenies show little variation from this and

are presented in Appendix S2. B, Evolution of body size regimes in Ornithischia, simplified from (A) by collapsing each phylogenetically-independent multi-taxon regime to a single branch.

FIG. 9. SURFACE model fit and regime evolution through time for Theropoda, showing a regime configuration consistent with a large ancestral body size for Tetanurae. A, SURFACE stepwise-AICc model for phylogeny 6 of Theropoda. B, Evolution of body size regimes in Theropoda simplified from (A) by collapsing each phylogenetically-independent multi-taxon regime to a single branch.

FIG. 10. SURFACE model fit and regime evolution through time for Theropoda, showing a regime configuration consistent with a smaller ancestral body size for Tetanurae, with multiple independent origins of larger body size (within Ceratosauria, Megalosauroidea, and Allosauroidea). A, SURFACE stepwise-AICc model for phylogeny 3 of Theropoda. B, Evolution of body size regimes in Theropoda simplified from (A) by collapsing each phylogenetically-independent multi-taxon regime to a single branch.

FIG. 11. Comparisons of AICc scores for best OU-based models fit using OUwie (Table 1; Dataset S3), best trend-based (Dataset S2), and Brownian motion (BM1) models. A, Sauropodomorpha; B, Triassic–Aptian Theropoda; C, thyreophoran ornithischians; D, marginocephalian ornithischians; and E, ornithopod ornithischians. Taxonomic content of ornithischian subtrees corresponds to Appendix S5. Models were fit across 20 phylogenies for

Sauropodomorpha, and to 16 phylogenies representing the low-AICc/high-α cluster for Ornithischia (Fig. 4C). Two phylogenies for Sauropodomorpha (trees 2 and 8) returned nonsensical parameter estimates (defined in text) and were discarded. Results for each phylogeny are connected by lines. Trees supporting Brownian motion or trend-based models are indicated and explained in the text.

FIG. 12. Visualisation of best trend model results for trees on which trend models had the best AICc weights (Fig. 11). A, Trend-based model for tree 10 of Theropoda. B, Trend-based model for tree 9 of ornithopod ornithischians. Although the entire trees of Theropoda and Ornithischia are shown, grey-shaded internodes and terminals were not included in this analysis, which focused on comparison with OUwie (taxonomic inclusion described in text). 'Songlingorn.' Is an abbreviation of Songlongornithidae, including *Yanornis* and *Yixianornis* in our tree.

FIG. 13. Disparity of non-pygostylian Dinosauria (shaded polygon) and dinosaurian subclades (symbols and dashed lines) through the Mesozoic. A, Sauropodomorpha (non-sauropodan sauropodomorphs) and Sauropoda; B, Ornithischia; C, Theropoda (non-pygostylian theropods) and Pygostylia. Analysis uses the following time bins: Middle Triassic, Late Triassic, Early Jurassic, Middle Jurassic, Late Jurassic, Early Cretaceous, Cenomanian–Santonian, Campanian, and Maastrichtian. 'Disparity' is the standard deviation of $\log_{10}$ body mass for each clade, and error bars are standard interquartile ranges of this value from 1000 bootstrapping replicates. These

plots were constructed using the full set of N = 526 specimens for which adult body masses were available.

FIG. 14. Distribution of shifts in optimal body size (θ) between regimes (A–B) and body mass frequency distributions for dinosaurs (unfilled bars = all dinosaurs; filled bars = dinosaur subgroups) throughout the Mesozoic (C–F). A–B regime shifts plotted against θ of the ancestral regime (A) and (B) time. Large circles represent clade-level regime shifts, and small circles represent 'singleton' regime shifts. The vertical dashed line in A indicates θ for the ancestral dinosaurian body size regime and the grey box indicates a sparsely populated region of trait space at intermediate body masses. Dashed vertical lines in B indicate period and epoch boundaries. Vertical arrows extending up and down from the x axis indicate the initiation of 'trend-like' dynamics on single branches that give rise to exceptionally large-bodied (up) or small-bodied (down) taxa. These regimes are characterized by high-magnitude, unrealised values of theta. Body size frequency distributions for C, Sauropodomorpha; D, Ornithischia; E, non-pygostylian Theropoda; and F, Pygostylia. Grey boxes indicates a sparsely-populated region of trait space at intermediate body masses. Dashed vertical lines indicate the approximate range (1 kg – 10,000 kg) of body masses classes seen among non-pygostylian, non-sauropod dinosaurs (with a few exceptions among small ornithischians, small non-pygostylian theropods, and large ornithischians). The smallest size classes are entirely occupied by pygostylian birds, and the largest size class is entirely occupied by sauropod dinosaurs. Panels C–F were constructed using the full set of N = 526 specimens for which adult body

masses were available. “Sauropodom.” Is an abbreviation of Sauropodomorpha.

FIG. 15. Histograms showing (log_{10}) body mass distributions for clades of dinosaurs among Mesozoic intervals (coloured bars): A–C, Triassic; D–F, Early Jurassic, G–J, Middle–Late Jurassic; K–N, Early Cretaceous; and O–P, Late Cretaceous. The unfilled bars indicate the body mass distributions for all dinosaurs during each interval. The grey rectangle indicates the underpopulated ‘intermediate’ range of dinosaur body masses. Dashed lines bracket the approximate minimum (1 kg) and maximum (10,000 kg) body masses for most dinosaurs (especially Theropoda and Ornithischia). These plots were constructed using the full set of N = 526 specimens for which adult body masses were available. “Sauropodom.” Is an abbreviation of Sauropodomorpha.

TABLES

Model	Description	Regimes	Parameters varying among regimes
BM1	Brownian motion	Single regime	NA
BMS	Brownian motion	Multi-regime	σ^2
OU1	Ornstein-Uhlenbeck	Single regime	(Z_0)
OUM	Ornstein-Uhlenbeck	Multi-regime	θ, (Z_0)
OUMV	Ornstein-Uhlenbeck	Multi-regime	θ, σ^2 (Z_0)
OUMA	Ornstein-Uhlenbeck	Multi-regime	θ, α (Z_0)
OUMVA	Ornstein-Uhlenbeck	Multi-regime	θ, σ^2, α (Z_0)

Table 1. List of Ornstein-Uhlenbeck (OU)-based models and Brownian-motion (BM) models compared in the current work. Z_0 is listed in parentheses because each model was tested both allowing Z_0 to be distinct from θ and constraining Z_0 to equal θ for the regime present at the root node of the phylogeny.

	Group	Age	Mass estimate (kg)
Theropoda: Pygostylia			
Iberomesornis romerali	Enantiornithes	Early Cretaceous	0.013 (0.0053–0.033)
Paraprotopteryx gracilis	Enantiornithes	Early Cretaceous	0.016 (0.0064–0.040)
Qiliania graffini	Enantiornithes	Early Cretaceous	0.017 (0.0076–0.040)
Huoshanornis huji	Enantiornithes	Early Cretaceous	0.022 (0.0090–0.055)
Cathayornis yandica	Enantiornithes	Early Cretaceous	0.03 (0.013–0.067)
Sinornis santensis	Enantiornithes	Early Cretaceous	0.035 (0.015–0.080)
Alexornis antecedens	Enantiornithes	Early Cretaceous	0.037 (0.018–0.079)
Concornis lacustris	Enantiornithes	Early Cretaceous	0.039 (0.016–0.098)
Theropoda:Maniraptora			
*Parvicursor remotus**	Alvarezsauroidea	Late Cretaceous	0.15 (0.073–0.32)
*Ceratonykus oculatus**	Alvarezsauroidea	Late Cretaceous	0.26 (0.11–0.65)
Microraptor zhaoianus	Dromaeosauridae	Early Cretaceous	0.43 (0.20–0.92)
Epidexipteryx hui	Scansoriopterygidae	Early Cretaceous	0.47 (0.21–1.1)
Rahonavis ostromi	Dromaeosauridae	Late Cretaceous	0.58 (0.31–1.1)
Anchiornis huxleyi	Troodontidae	Late Jurassic	0.58 (0.23–1.4)
Xixianykus zhangi	Alvarezsauroidea	Late Cretaceous	0.74 (0.29–1.9)
Mahakala omnogovae	Dromaeosauridae	Late Cretaceous	0.76 (0.41–1.4)
Theropoda			
Procompsognathus triassicus	Theropoda	Triassic	1.1 (0.61–2.1)
Sinosauropteryx prima	Coelurosauria	Early Cretaceous	1.6 (0.65–4.1)
Compsognathus longipes	Coelurosauria	Late Jurassic	2.6 (1.0–6.5)
Velocisaurus unicus	Ceratosauria	Late Cretaceous	3.2 (1.3–7.8)
Tawa hallae	Theropoda	Triassic	3.2 (1.4–7.3)
*Hexing qingyi**	Ornithomimosauria	Early Cretaceous	4.2 (1.5–12)
Segisaurus halli	Theropoda	Early Jurassic	5 (1.7–14)
Eodromaeus murphi	Theropoda	Triassic	7.1 (2.5–20)
Sauropodomorpha			
Pampadromaeus barberenai	Sauropodomorpha	Triassic	8.5 (4.6–16)
Saturnalia tupiniquim	Sauropodomorpha	Triassic	11 (5.7–20)

Chromogisaurus novasi	Sauropodomorpha	Triassic	13 (7.0–24)
Plateosaurus engelhardti	Sauropodomorpha	Triassic	920 (490–1700)
Ornithischia			
Fruitadens haagarorum	Heterodontosauridae	Late Jurassic	0.73 (0.39-1.4)
Tianyulong confuciusi	Heterodontosauridae	Late Jurassic	0.89 (0.42–1.9)
Abrictosaurus consors	Heterodontosauridae	Early Jurassic	1.4 (0.68–3.0)
Psittacosaurus sinensis	Ceratopsia	Early Cretaceous	4.6 (1.8–11)
Yueosaurus tiantaiensis	Thescelosauridae	Early Cretaceous	4.6 (2.2–9.6)
Gongbusaurus wucaiwanensis	Ornithischia	Late Jurassic	4.9 (2.6–9.2)
Heterodontosaurus tucki	Heterodontosauridae	Early Jurassic	5.2 (1.8–15)
Xiaosaurus dashanpensis	Ornithischia	Late Jurassic	5.3 (2.3–12)

Table 2. Body mass estimates for the smallest dinosaurs. Body mass estimates and their ±95% confidence intervals (in brackets) are given for major groups within Dinosauria, demonstrating the uniquely small body masses of Pygostylian birds. A complete set of mass estimates and their standard errors is given in Dataset S1.

	Group	Age	Mass estimate (kg)
Saurischia: ?Theropoda			
Eoraptor lunensis	Theropoda	Triassic: Carnian	17 (9.3–32)
Herrerasaurus ischigualastensis	Theropoda	Triassic: Carnian	270 (150–510)
Theropoda			
Staurikosaurus pricei	Theropoda	Triassic: Carnian	23 (12–43)
Coelophysis bauri	Theropoda	Triassic: Norian	9.8 (5.3–18)
Gojirasaurus quayi	Theropoda	Triassic: Norian	190 (45–820)
Guaibasaurus candelariensis	Theropoda	Triassic: Norian	33 (16–72)
Procompsognathus triassicus	Theropoda	Triassic: Norian	1.1 (0.61–2.1)
Tawa hallae	Theropoda	Triassic: Norian	3.2 (1.4–7.3)
Liliensternus liliensterni	Theropoda	Triassic: Rhaetian	84 (45–160)
Saurischia: ?Sauropodomorpha			
Eodromaeus murphi	Theropoda	Triassic: Carnian	7.1 (2.5–20)
Sauropodomorpha			
Chromogisaurus novasi	Sauropodomorpha	Triassic: Carnian	13 (7.0–24)
Pampadromaeus barberenai	Sauropodomorpha	Triassic: Carnian	8.5 (4.6–16)
Panphagia protos	Sauropodomorpha	Triassic: Carnian	12 (3.2–49)
Saturnalia tupiniquim	Sauropodomorpha	Triassic: Carnian	11 (5.7–20)
Ornitshichia			
Pisanosaurus mertii	Ornithischia	Triassic: Carnian	35 (17–76)
Eocursor parvus	Ornithischia	Triassic: Norian/Rhaetian	4.2 (2.0–8.8)

Table 3. Body mass estimates for Triassic dinosaurs, especially for the earliest Triassic dinosaurs (Carnian stage). Body mass estimates and their ±95% confidence intervals (in brackets) are given for major groups within Dinosauria, demonstrating the autapomorphically large body mass of

Herrerasarus compared to other Carnian dinosaurs (see also Benson *et al*. 2014*a*). A complete set of mass estimates and their standard errors is given in Dataset S1.

	Group	Age	Mass estimate (kg)
Theropoda: Maniraptora			
Deinocheirus mirificus	Ornithomimosauria	Late Cretaceous	7300 (3900-14000)
Segnosaurus galbinensis	Therizinosauria	Late Cretaceous	4600 (2000–11000)
Suzhousaurus megatherioides	Therizinosauria	Early Cretaceous	3100 (1500–6700)
Gigantoraptor erlianensis	Oviraptorosauria	Late Cretaceous	2000 (1100-3700)
Nothronychus graffami	Therizinosauria	Late Cretaceous	1200 (640–2200)
Austroraptor cabazai	Dromaeosauridae	Late Cretaceous	600 (260–1400)
Beishanlong grandis	Ornithomimosauria	Late Cretaceous	580 (250–1300)
Nothronychus mckinleyi	Therizinosauria	Late Cretaceous	570 (130–2500)
Theropoda			
Tyrannosaurus rex	Tyrannosauroidea	Late Cretaceous	7700 (4100-14000)
Giganotosaurus carolinii[1]	Allosauroidea	Late Cretaceous	6100 (3300-11000)
Tyrannotitan chubutensis	Allosauroidea	Early Cretaceous	5400 (2900-10000)
Saurophaganax maximus	Allosauroidea	Late Jurassic	3800 (2000-7000)
Chilantaisaurus tashuikouensis	Allosauroidea	Late Cretaceous	3600 (1900-6700)
Lametasaurus indicus[2]	Ceratosauria	Late Cretaceous	3600 (1400–9200)
Acrocanthosaurus atokensis	Allosauroidea	Early Cretaceous	3500 (1900-6400)
Carcharodontosaurus saharicus	Allosauroidea	Late Cretaceous	3300 (1400–7600)
Sauropoda			
Argentinosaurus huinculensis[3]	Titanosauria	Late Cretaceous	95000 (32000–280000)
Dreadnoughtus schrani	Titanosauria	Late Cretaceous	59000 (29000–120000)
Brachiosaurus altithorax	Macronaria	Late Jurassic	58000 (23000–150000)
Ruyangosaurus giganteus	Titanosauriformes	Late Cretaceous	54000 (19000–160000)
Turiasaurus riodevensis	Turiasauria	Late Jurassic	51000 (25000–100000)
Paralititan stromeri	Titanosauria	Late Cretaceous	51000 (17000–150000)
Losillasaurus giganteus	Turiasauria	Late Jurassic/ Early Cretaceous	47000 (17000–130000)
Argyrosaurus superbus	Titanosauria	Late Cretaceous	42000 (14000–120000)
Ornithischia			
Shantungosaurus giganteus	Hadrosauroidea	Late Cretaceous	17000 (8600–35000)

Iguanodon seelyi	Iguanodontia	Early Cretaceous	15000 (7600–31000)
Triceratops horridus	Ceratopsia	Late Cretaceous	13000 (6500–27000)
Pentaceratops sternbergii	Ceratopsia	Late Cretaceous	11000 (5300–22000)
Torosaurus latus	Ceratopsia	Late Cretaceous	9700 (1400–66000)
Magnapaulia laticaudus	Hadrosauroidea	Late Cretaceous	8600 (2000–38000)
Iguanodon bernissartensis	Iguanodontia	Early Cretaceous	8300 (4100-17000)
Edmontosaurus regalis	Hadrosauroidea	Late Cretaceous	7600 (3700–15000)

Table 4. Body mass estimates for the largest dinosaurs. Body mass estimates and their ±95% confidence intervals (in brackets) are given for major groups within Dinosauria, demonstrating the uniquely large body masses of sauropod dinosaurs. A complete set of mass estimates and their standard errors is given in Dataset S1. Notes: 1, The body mass of *Giganotosarus* is given here, but the phenotypically similar taxon *Mapusaurus* also attained large body size. 2, *Lametasaurus indicus* has a highly robust tibia (Matley 1924; Carrano 2007) that provides our highest estimated mass among ceratosaurian theropods. However, *Pycnonemosaurus*, which was not included in our dataset, has comparable measurements of the tibia shaft and is likely of comparable body mass (Kellner and Campos 2002; Grillo and Delcourt 2017). 3, Our estimated body mass for *Argentinosaurus* is based on a referred femur housed at Museo de la Plata, La Plata, Argentina. A robust body mass estimate around 70 tonnes was reported using multiple methods for a nearly-complete skeleton of a related taxon, *Patagotitan*, shortly before publication of the current work (Carballido *et al*. 2017).

	Group	Age	Mass estimate (kg)
Pygostylia			
Gargantuavis philoinos	Avialae indet.	Late Cretaceous	190 (100–350)
Hesperornis crassipes	Ornithuromorpha	Late Cretaceous	25 (11–53)
Baptornis varneri	Ornithuromorpha	Late Cretaceous	11 (5.0–23)
Didactylornis jii (= *Sapeornis*)	Avialae	Early Cretaceous	2.9 (1.1–7.2)
Hollanda luceria	Ornithuromorpha	Late Cretaceous	2.8 (1.3–6.0)
Patagopteryx deferrariisi	Ornithuromorpha	Late Cretaceous	2.6 (1.2–5.5)
Jeholornis prima	Avialae	Early Cretaceous	2.4 (0.95–6.1)
Vorona berivotrensis	Ornithuromorpha	Late Cretaceous	1.8 (0.85–3.9)
Sauropoda			
Magyarosaurus dacus	Titanosauria	Late Cretaceous	750 (370–1500)
Europasaurus holgeri	Titanosauriformes	Late Jurassic	1050 (510–2100)
Lirainosaurus astibae	Titanosauria	Late Cretaceous	1800 (740–4400)
Dystrophaeus viaemalae	Diplodocoidea	Late Jurassic	3100 (1100–8400)
Gondwanatitan faustoi	Titanosauria	Late Cretaceous	3600 (1300–9700)

Table 5. Body mass estimates for the largest pygostylians and smallest sauropods. Body mass estimates and their ±95% confidence intervals (in brackets) are given. A complete set of mass estimates and their standard errors is given in Dataset S1. Body masses for the smallest pygostylians and largest sauropods are shown in Tables 3 and 4.

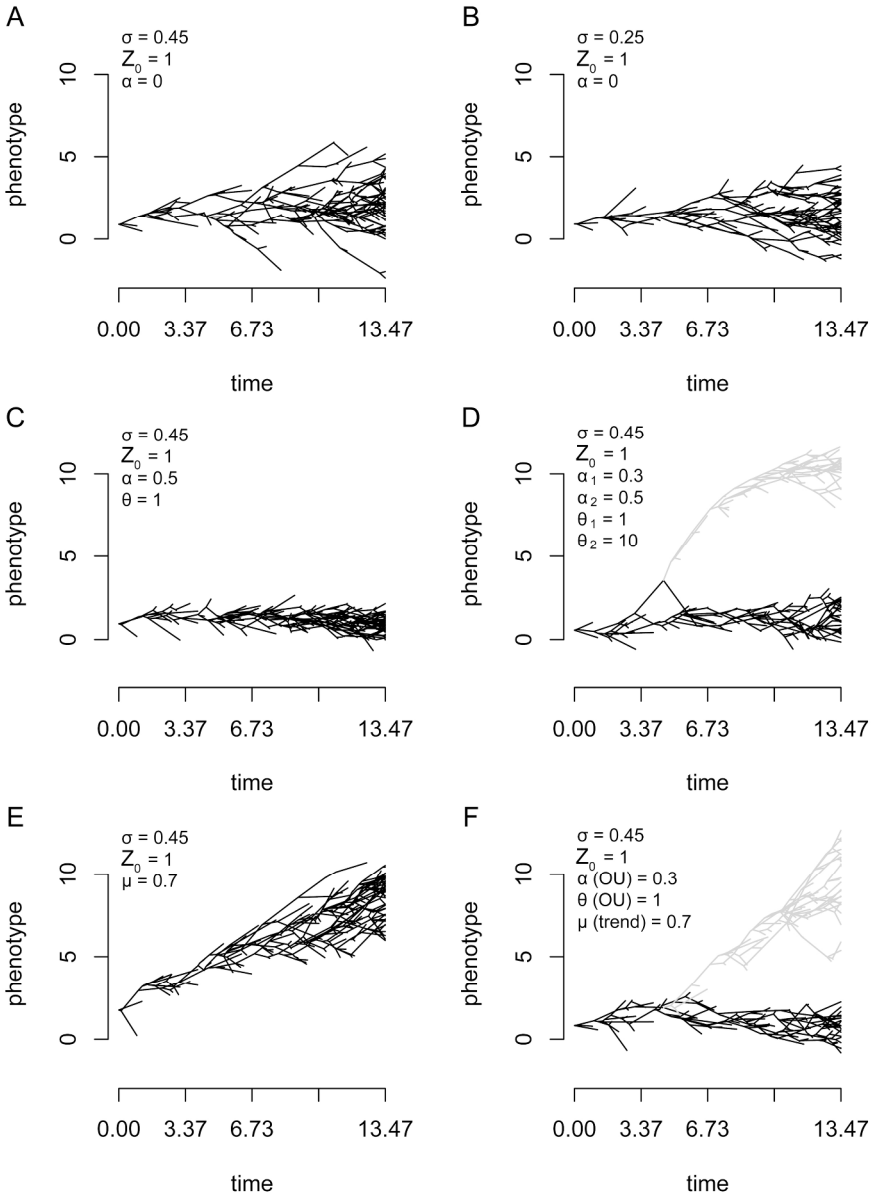

FIG. 1. Simulated behavior of macroevolutionary models under various parameter values. A, B, Brownian motion (BM) under high (A) and low (B) values of the Brownian variance parameter showing diffusive behavior and unbiased directionality of trait evolution along lineages; C, D, Ornstein-Uhlenbeck (OU) models showing bounded or constrained evolution under (C) a single regime in which the trait optimum (θ) equals the starting values (Z0), and (D) in which a second regime originates from the first and has a trait optimum that is greater than the starting value; E, F, trend models originating at the base of the phylogeny (E) and descending from an ancestral OU regime (F). For this figure, a trend model with μ = 0.7 was approximated using an OU model with low α (= 0.005) and distant, unrealized θ (= 150).

218x287mm (300 x 300 DPI)

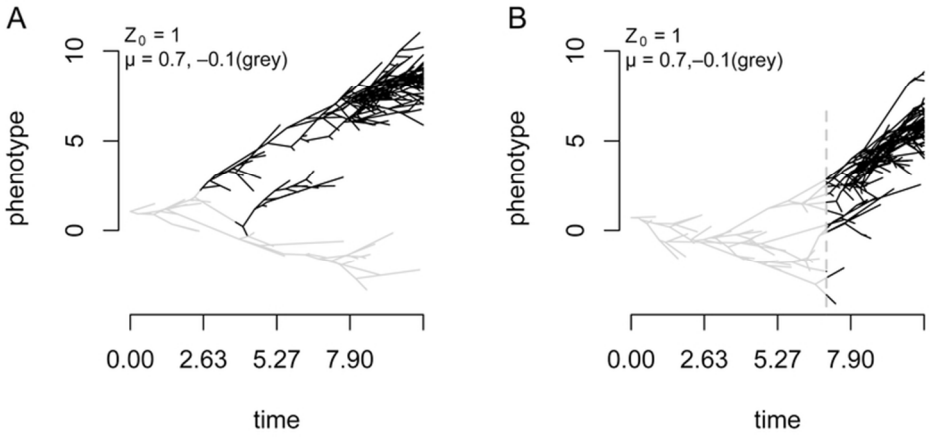

FIG. 2. Simulated behavior of new trend-based models under various parameter values. A, node-shift model, showing a change in the trend parameter μ at a node on the phylogeny. B, time-shift model, showing a change in the trend parameter μ at a time during the evolutionary history of a group. Z0 is the trait value at the root node and σ is the Brownian variance. For this figure, a trend model with μ = 0.7 was approximated using an OU model with low α and distant, unrealized θ.

72x31mm (300 x 300 DPI)

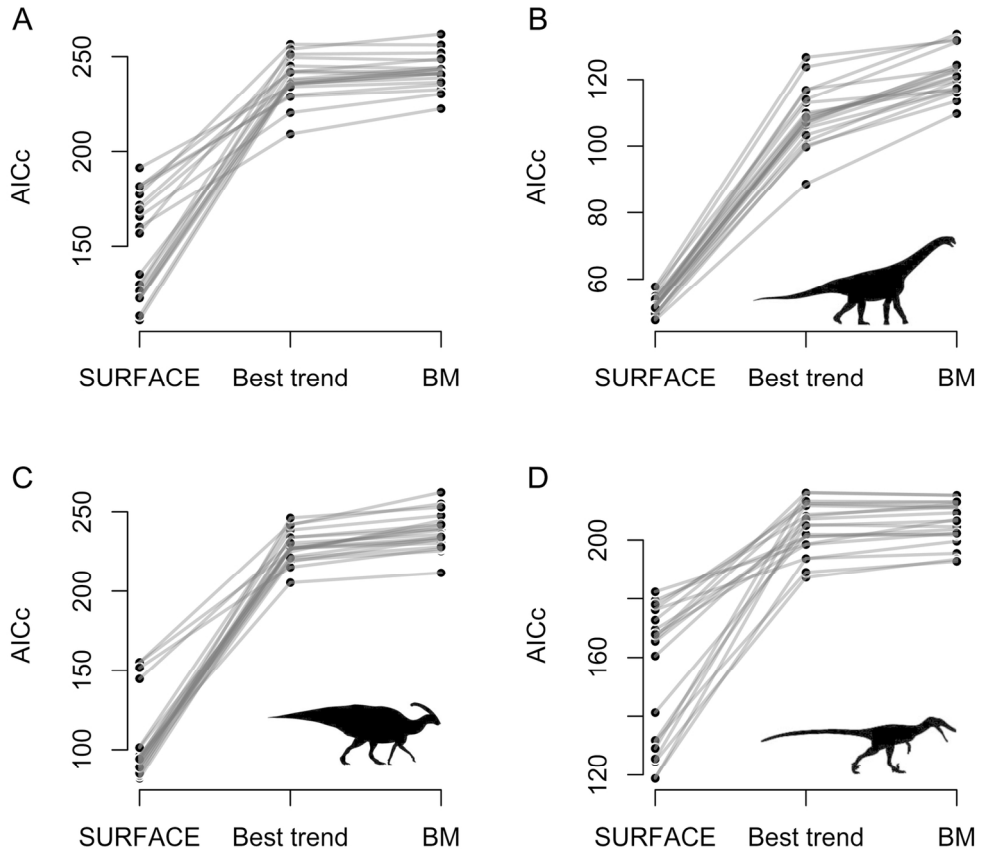

FIG. 3. Comparisons of AICc scores for SURFACE (multi-regime Ornstein-Uhlenbeck), best trend-based, and Brownian motion (BM1) models. A, Triassic–Jurassic Dinosauria; B, Sauropodomorpha; C, Ornithischia; and Triassic–Aptian Theropoda. Models were fit across 20 phylogenies scaled to time using the mbl1 algorithm, and results for each phylogeny are connected by lines. Results based on other timescaling algorithms were essentially identical. AICc scores for all trend-based models and BM1 models are given in Dataset S2. AICc scores and visualizations of SURFACE models are given in Appendix S2.

143x123mm (300 x 300 DPI)

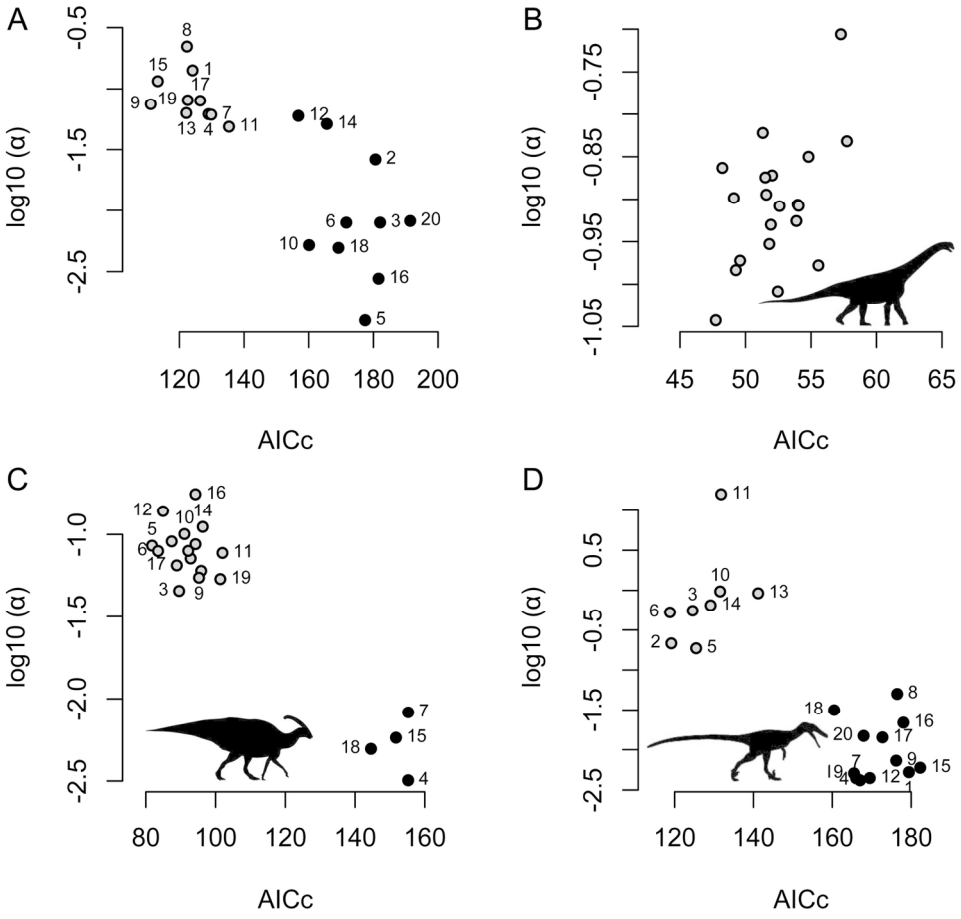

FIG. 4. Plots of log10(α) on AICc across 20 phylogenies for SURFACE model fits showing bimodal distribution of model outcomes for most groups. A, Triassic–Jurassic Dinosauria; B, Sauropodomorpha; C, Ornithischia; and Triassic–Aptian Theropoda. Models were fit across 20 phylogenies scaled to time using the mbl1 algorithm.,Tree numbers correspond to those used throughout this paper, e.g. Appendix S2 and elsewhere.

152x139mm (300 x 300 DPI)

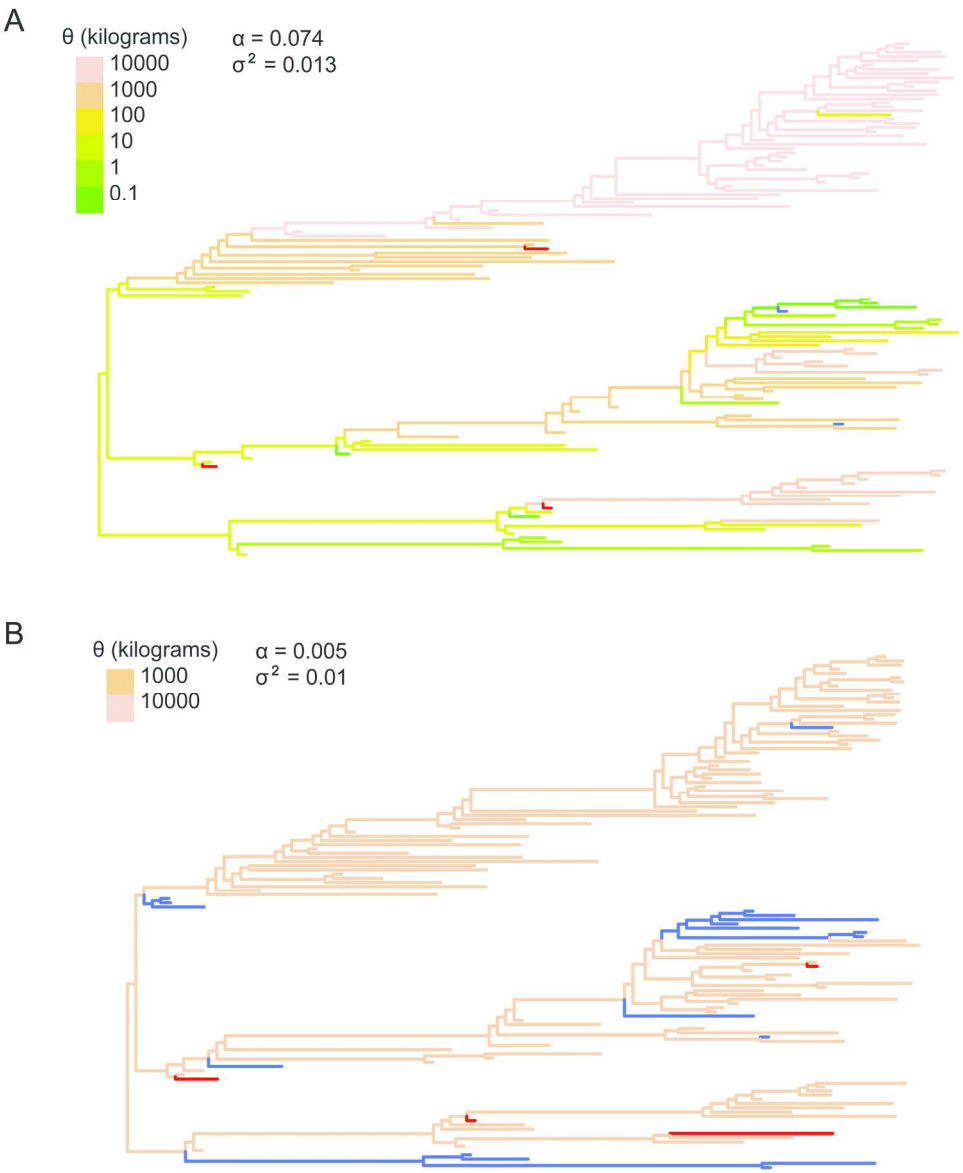

FIG. 5. Comparison of low-AICc/high-α class and high-AICc/low-α class SURFACE model fits for Triassic–Jurassic Dinosauria. A, low-AICc/high-α class of model based on tree 9; B, high-AICc/low-α class of model based on tree 10. The suboptimal model fit is characterized by a low value of α (= 0.005 in (A) compared to 0.074 in (B)) and few distinct regimes. Red and blue lineages exhibit trend-like attraction to unrealized low (blue) or high (red) trait optima.

196x231mm (300 x 300 DPI)

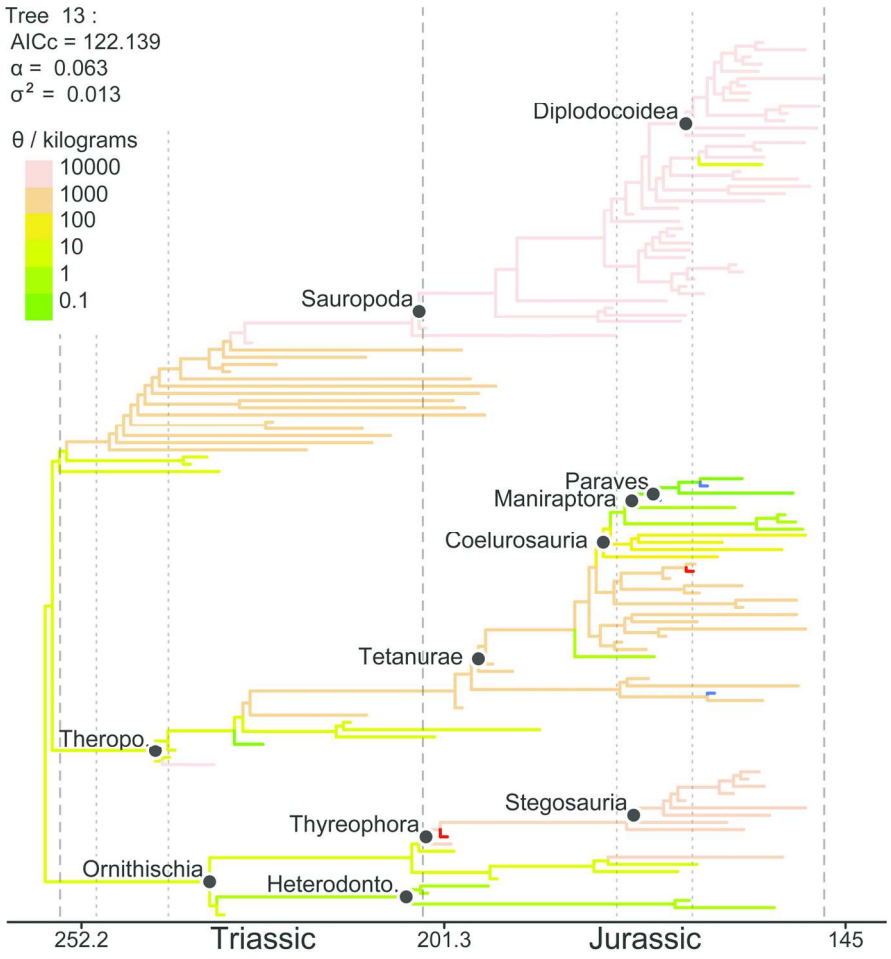

FIG. 6. SURFACE stepwise-AICc model for phylogeny 12 of Triassic–Jurassic Dinosauria. Results for other phylogenies show little variation from this (except that described above; Figs 4–5), and are presented in Appendix S2.

159x153mm (300 x 300 DPI)

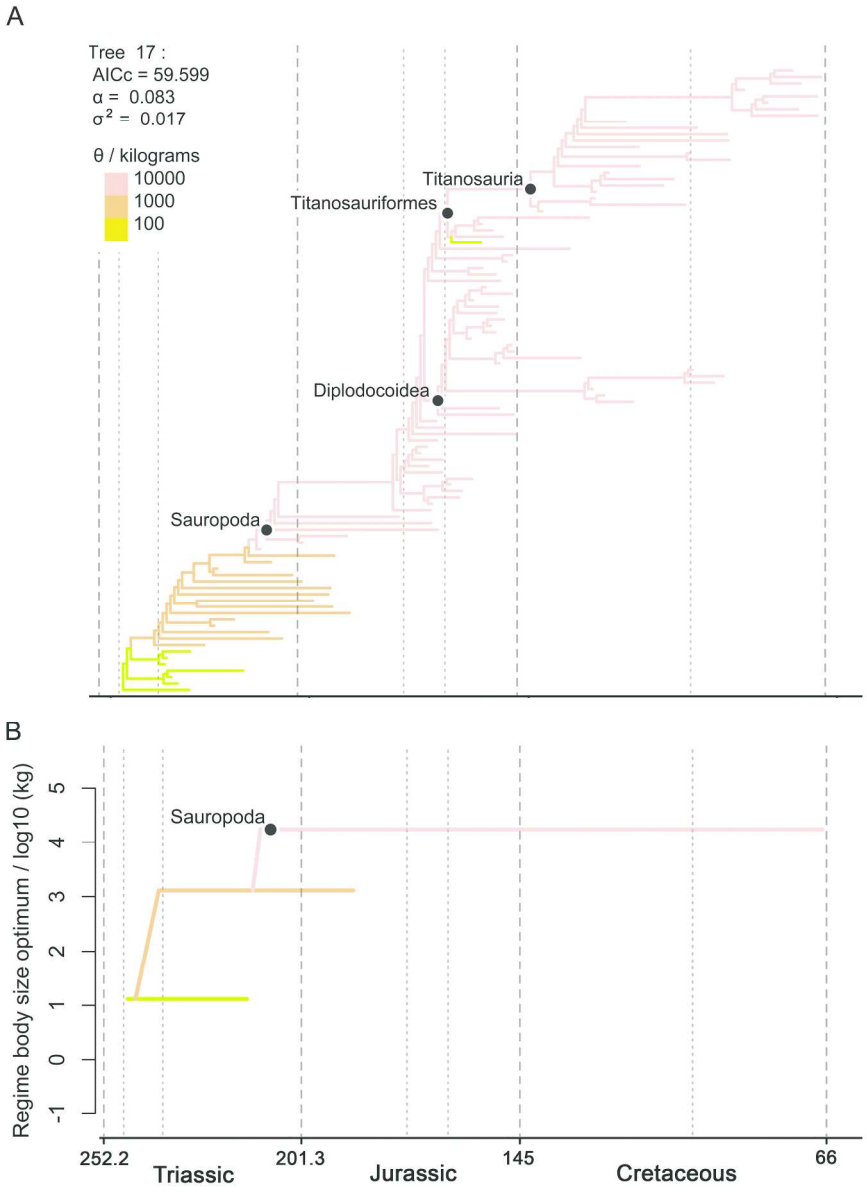

FIG. 7. SURFACE model fit and regime evolution through time for Sauropodomorpha. A, SURFACE stepwise-AICc model for phylogeny 17 of Sauropodomorpha. Results for other phylogenies show little variation from this and are presented in Appendix S2. B, Evolution of body size regimes in Sauropodomorpha simplified from (A) by collapsing each phylogenetically-independent multi-taxon regime to a single branch.

226x307mm (300 x 300 DPI)

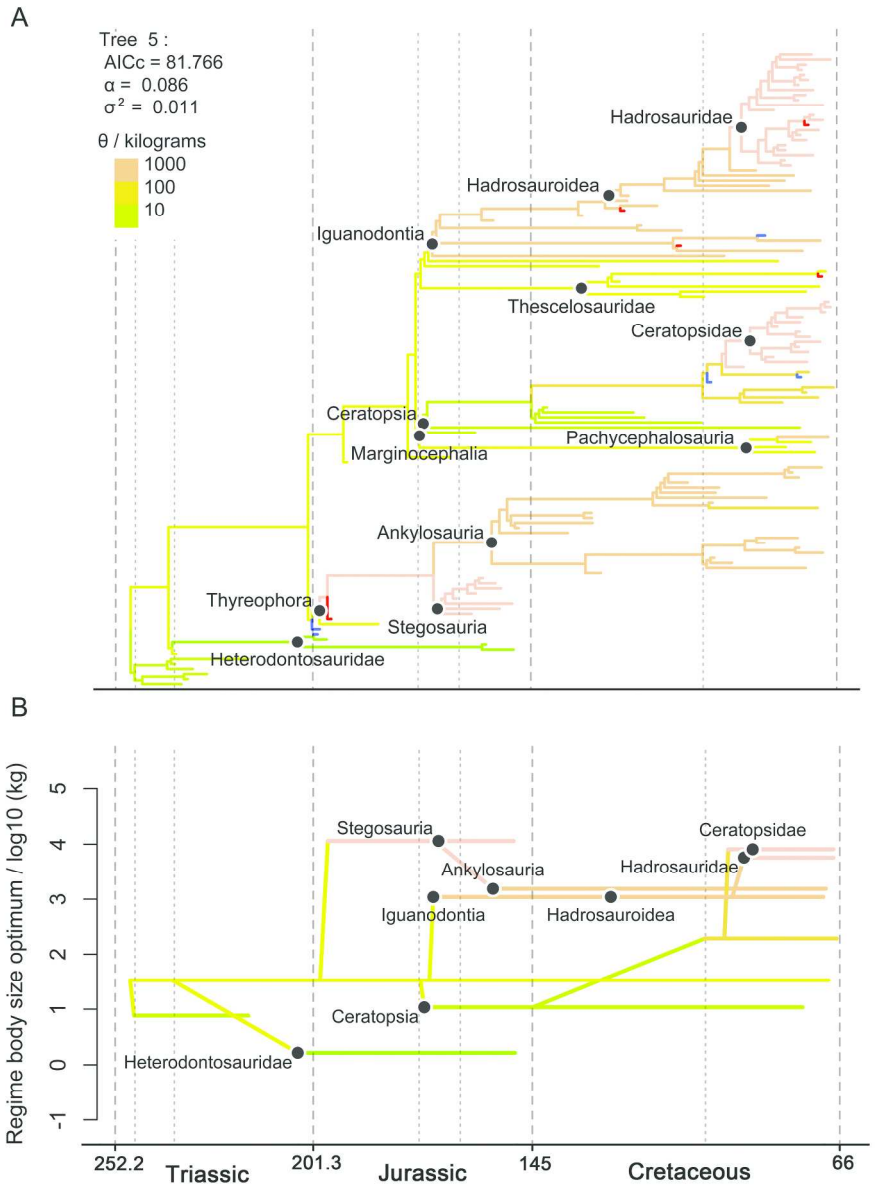

FIG. 8. SURFACE model fit and regime evolution through time for Ornithischia. A, SURFACE stepwise-AICc model for phylogeny 5 of Ornithischia. Results for other phylogenies show little variation from this and are presented in Appendix S2. B, Evolution of body size regimes in Ornithischia, simplified from (A) by collapsing each phylogenetically-independent multi-taxon regime to a single branch.

222x297mm (300 x 300 DPI)

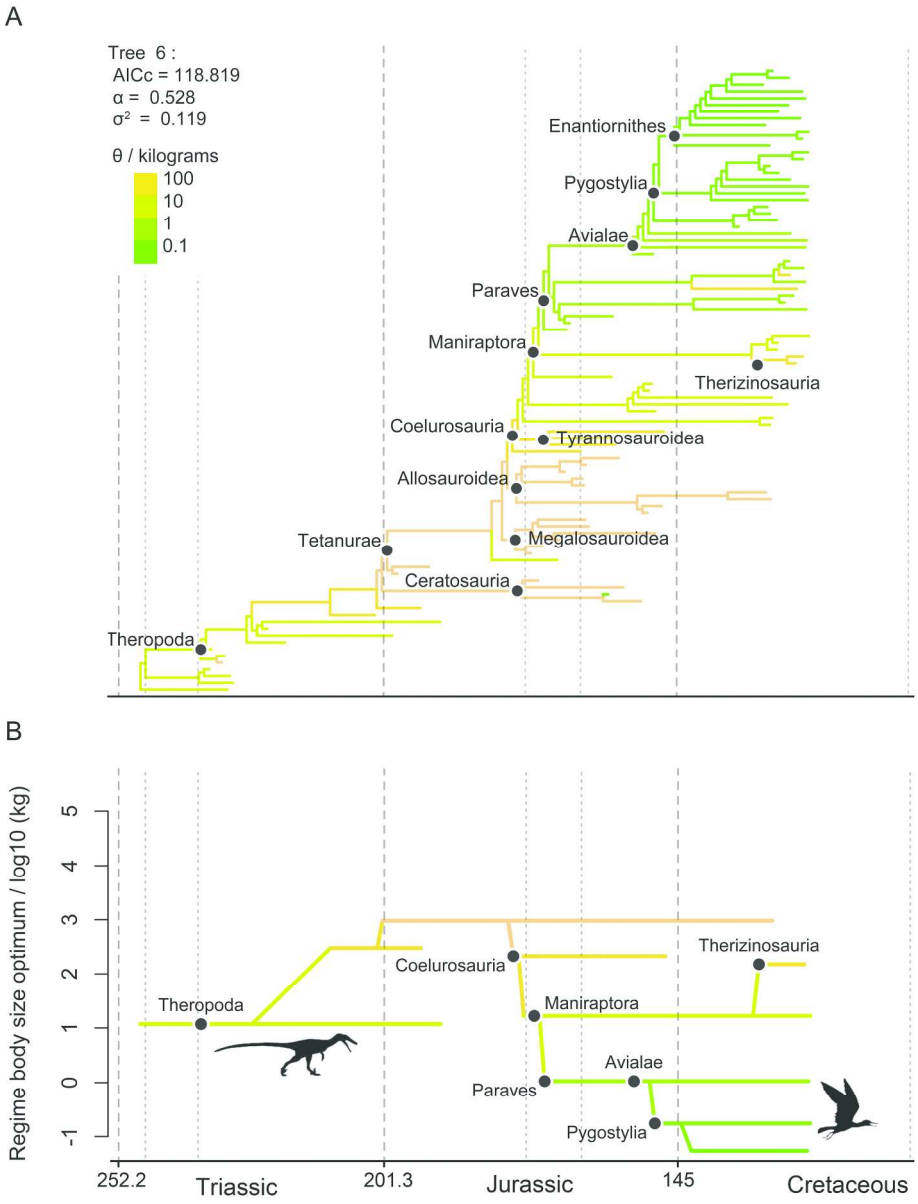

FIG. 9. SURFACE model fit and regime evolution through time for Theropoda, showing a regime configuration consistent with a large ancestral body size for Tetanurae. A, SURFACE stepwise-AICc model for phylogeny 6 of Theropoda. B, Evolution of body size regimes in Theropoda simplified from (A) by collapsing each phylogenetically-independent multi-taxon regime to a single branch.

213x275mm (300 x 300 DPI)

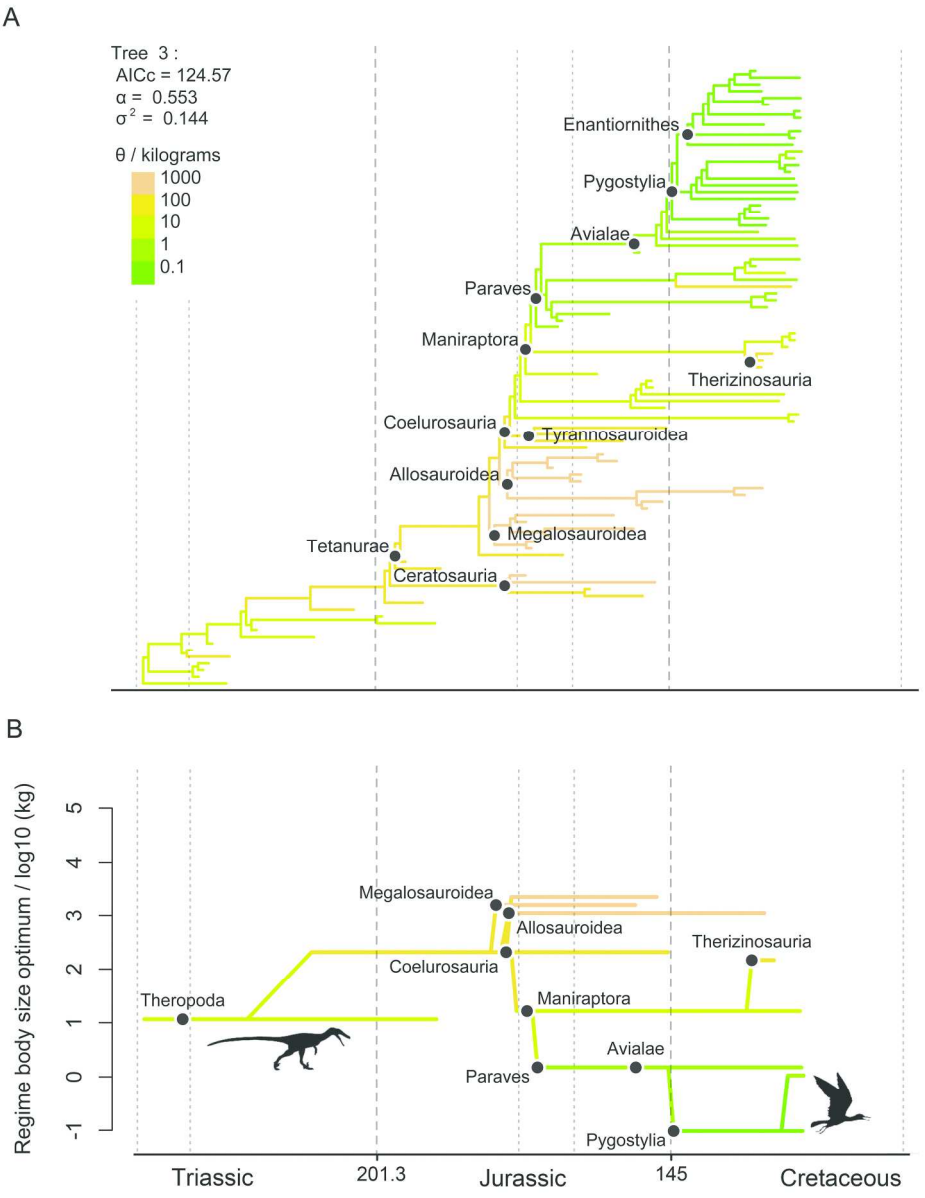

FIG. 10. SURFACE model fit and regime evolution through time for Theropoda, showing a regime configuration consistent with a smaller ancestral body size for Tetanurae, with multiple independent origins of larger body size (within Ceratosauria, Megalosauroidea, and Allosauroidea). A, SURFACE stepwise-AICc model for phylogeny 3 of Theropoda. B, Evolution of body size regimes in Theropoda simplified from (A) by collapsing each phylogenetically-independent multi-taxon regime to a single branch.

213x275mm (300 x 300 DPI)

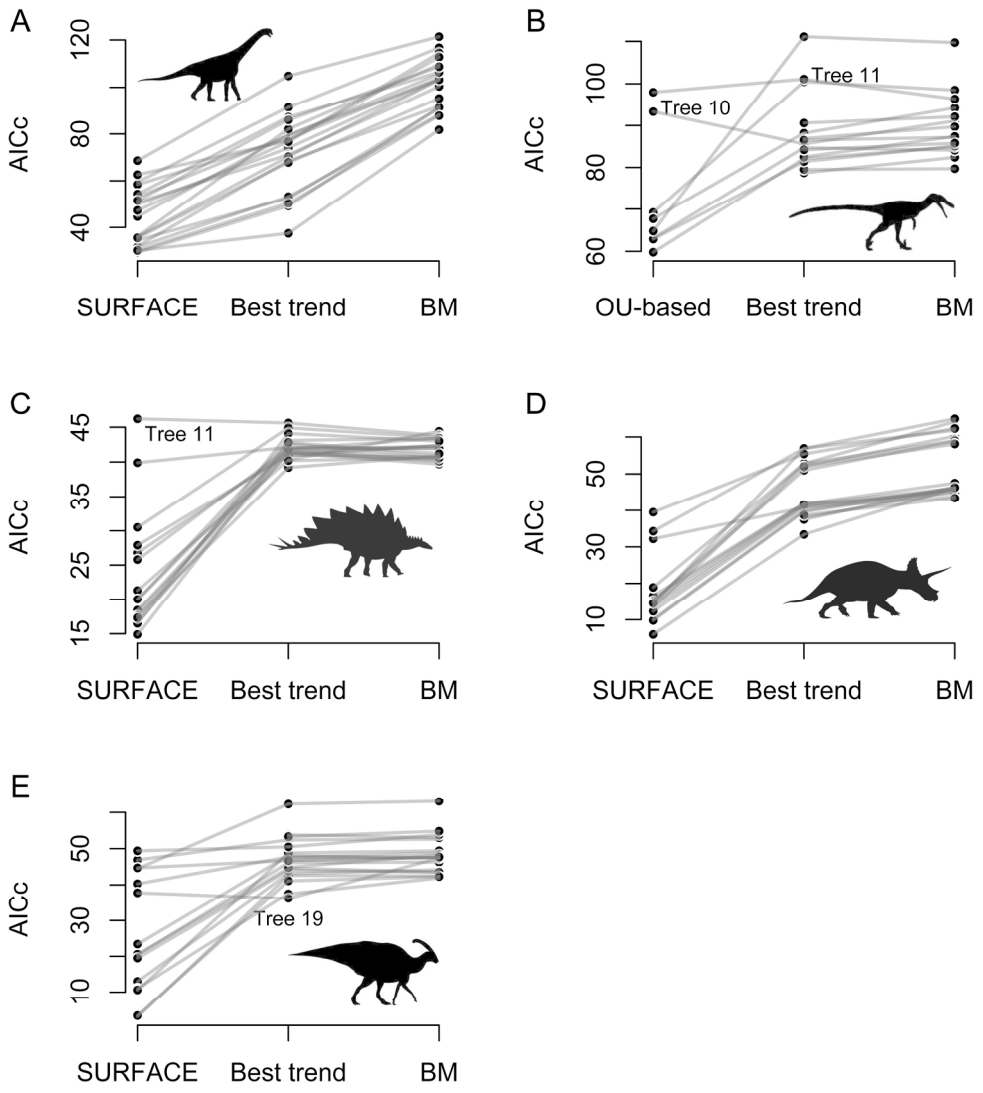

FIG. 11. Comparisons of AICc scores for best OU-based models fit using OUwie (Table 1; Dataset S3), best trend-based (Dataset S2), and Brownian motion (BM1) models. A, Sauropodomorpha; B, Triassic–Aptian Theropoda; C, thyreophoran ornithischians; D, marginocephalian ornithischians; and E, ornithopod ornithischians. Taxonomic content of ornithischian subtrees corresponds to Appendix S5. Models were fit across 20 phylogenies for Sauropodomorpha, and to 16 phylogenies representing the low-AICc/high-α cluster for Ornithischia (Fig. 4C). Two phylogenies for Sauropodomorpha (trees 2 and 8) returned nonsensical parameter estimates (defined in text) and were discarded. Results for each phylogeny are connected by lines. Trees supporting Brownian motion or trend-based models are indicated and explained in the text.

181x198mm (300 x 300 DPI)

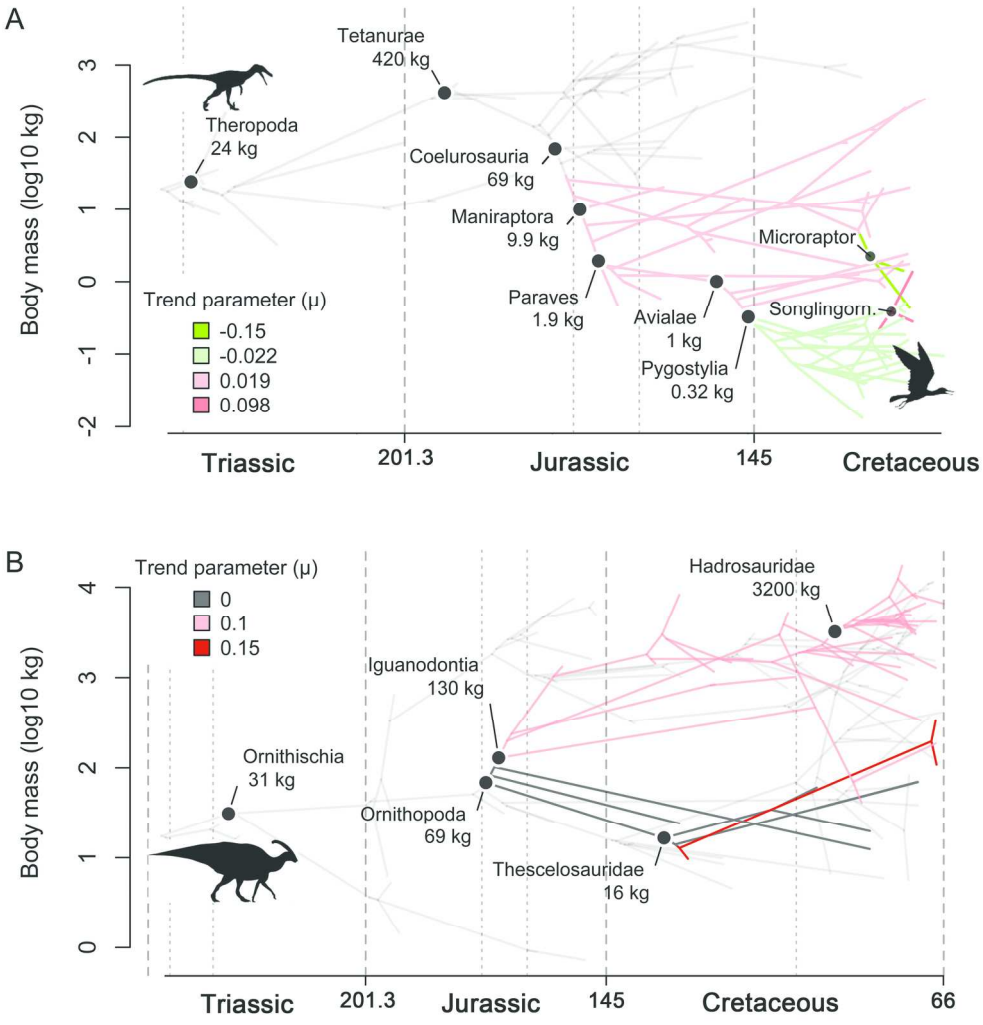

FIG. 12. Visualisation of best trend model results for trees on which trend models had the best AICc weights (Fig. 11). A, Trend-based model for tree 10 of Theropoda. B, Trend-based model for tree 9 of ornithopod ornithischians. Although the entire trees of Theropoda and Ornithischia are shown, grey-shaded internodes and terminals were not included in this analysis, which focused on comparison with OUwie (taxonomic inclusion described in text). 'Songlingorn.' Is an abbreviation of Songlongornithidae, including Yanornis and Yixianornis in our tree.

171x176mm (300 x 300 DPI)

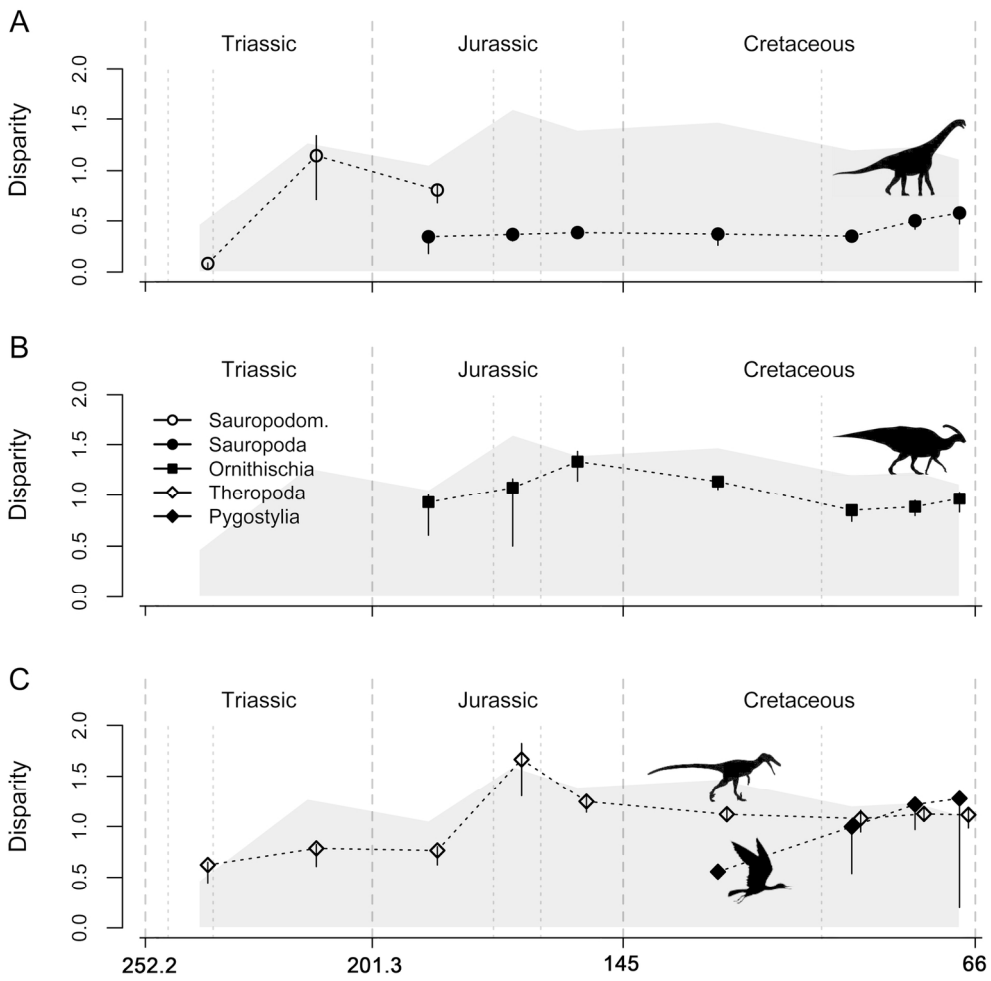

FIG. 13. Disparity of non-pygostylian Dinosauria (shaded polygon) and dinosaurian subclades (symbols and dashed lines) through the Mesozoic. A, Sauropodomorpha (non-sauropodan sauropodomorphs) and Sauropoda; B, Ornithischia; C, Theropoda (non-pygostylian theropods) and Pygostylia. Analysis uses the following time bins: Middle Triassic, Late Triassic, Early Jurassic, Middle Jurassic, Late Jurassic, Early Cretaceous, Cenomanian–Santonian, Campanian, and Maastrichtian. 'Disparity' is the standard deviation of log10 body mass for each clade, and error bars are standard interquartile ranges of this value from 1000 bootstrapping replicates. These plots were constructed using the full set of N = 526 specimens for which adult body masses were available.

163x160mm (300 x 300 DPI)

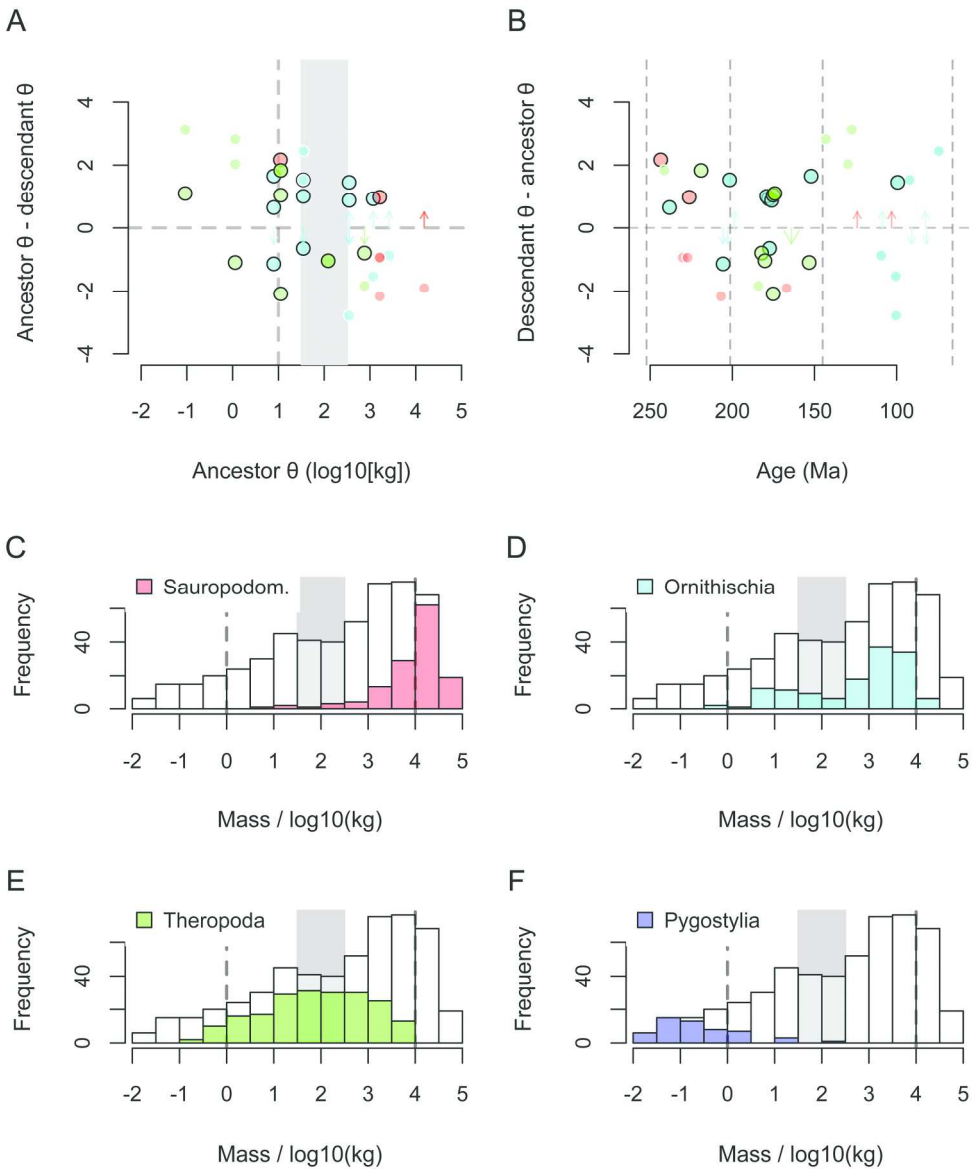

FIG. 14. Distribution of shifts in optimal body size (θ) between regimes (A–B) and body mass frequency distributions for dinosaurs (unfilled bars = all dinosaurs; filled bars = dinosaur subgroups) throughout the Mesozoic (C–F). A–B regime shifts plotted against θ of the ancestral regime (A) and (B) time. Large circles represent clade-level regime shifts, and small circles represent 'singleton' regime shifts. The vertical dashed line in A indicates θ for the ancestral dinosaurian body size regime and the grey box indicates a sparsely populated region of trait space at intermediate body masses. Dashed vertical lines in B indicate period and epoch boundaries. Vertical arrows extending up and down from the x axis indicate the initiation of 'trend-like' dynamics on single branches that give rise to exceptionally large-bodied (up) or small-bodied (down) taxa. These regimes are characterized by high-magnitude, unrealised values of theta. Body size frequency distributions for C, Sauropodomorpha; D, Ornithischia; E, non-pygostylian Theropoda; and F, Pygostylia. Grey boxes indicates a sparsely-populated region of trait space at intermediate body masses. Dashed vertical lines indicate the approximate range (1 kg – 10,000 kg) of body masses classes seen among non-pygostylian, non-sauropod dinosaurs (with a few exceptions among small ornithischians, small non-pygostylian theropods, and large ornithischians). The smallest size classes are entirely occupied by

pygostylian birds, and the largest size class is entirely occupied by sauropod dinosaurs. Panels C–F were constructed using the full set of N = 526 specimens for which adult body masses were available. "Sauropodom." Is an abbreviation of Sauropodomorpha.

196x232mm (300 x 300 DPI)

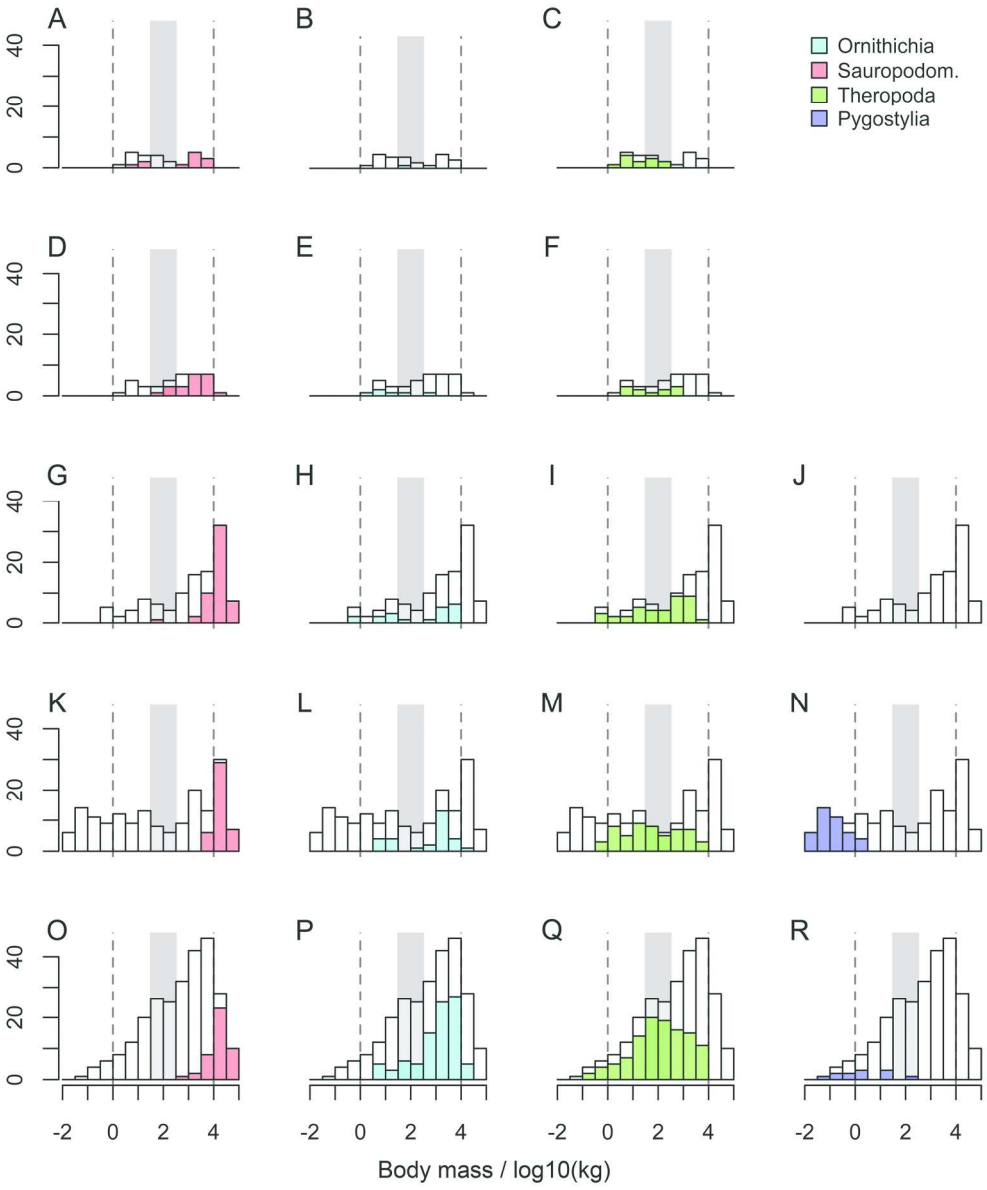

FIG. 15. Histograms showing (log10) body mass distributions for clades of dinosaurs among Mesozoic intervals (coloured bars): A–C, Triassic; D–F, Early Jurassic, G–J, Middle–Late Jurassic; K–N, Early Cretaceous; and O–P, Late Cretaceous. The unfilled bars indicate the body mass distributions for all dinosaurs during each interval. The grey rectangle indicates the underpopulated 'intermediate' range of dinosaur body masses. Dashed lines bracket the approximate minimum (1 kg) and maximum (10,000 kg) body masses for most dinosaurs (especially Theropoda and Ornithischia). These plots were constructed using the full set of N = 526 specimens for which adult body masses were available. "Sauropodom." Is an abbreviation of Sauropodomorpha.

197x236mm (300 x 300 DPI)