**SUPPLEMENTARY MATERIAL**

**Bite Force Data**

Bite force data were collected primarily through the literature, augmented with novel estimates (Table S1). We took the maximum bite force for each taxon as the taxon-representative bite force (Table S1). These are typically bite forces taken at the posterior-most position along the tooth row – i.e. molars in mammals – but are often at more anterior positions, particularly in smaller animals such as lizards, bats and finches, in which standardized biting may be difficult owing to the relative sizes of the bite force transducers – i.e. small animals are incapable of biting comparatively large force plates at their posterior-most biting positions. More importantly, some taxa may have behavioural or morphological constraints for them to bite at anterior positions. Mechanically, anterior biting positions are lower in magnitude compared to posterior biting positions, so there are concerns that differences in biting positions may bias regression coefficients (higher intercepts for posterior bite forces and slopes likely to differ). In order to take into account such potential biases, each bite force record was categorized into one of two biting position categories (Bite Point): anterior and posterior. Mid-jaw bite points were categorized as “anterior” since anterior biting positions are often defined differently across studies – for instance, the biting point for which the tooth comes into contact with food/force transducers first (e.g., caniniform teeth instead of the more anterior incisiform teeth).

Similarly, bite force records were categorized into bite force acquisition type (Bite Type): *in vivo* or estimates. *In vivo* bite forces are recordings from live individuals available from literature (Table S1). On the other hand, estimated bite forces are any values that were obtained from methods other than live recordings (i.e. indirectly): simple 2D static lever models; multibody dynamics analyses (MDA); finite elements analyses (FEA); tooth fractures; and tooth indentation simulations. These methods were treated as a single category since some methods are only associated with an insufficient number of bite force records to warrant separate categories in a regression framework.

Novel bite forces were calculated for felids and archosaurs using two-dimensional lever models (see relevant sections below). Literature-based bite force estimates were standardized as follows: 1) Dry skull estimates in carnivores using muscle stress value *σ* = 300 kN [1] were readjusted to reflect *σ* = 370 kN [after 2]. Dry skull estimates are known to underestimate *in vivo* bite force measurements as well as bite force estimates derived from muscle architecture data [3] and while correction factors do exist [3], they are limited to canine bite force – i.e. non-maximal bite force – and thus we do not perform any corrections other than the aforementioned adjustment to *σ*. 2) Bite force indexes in anthropoids [4] were converted to absolute bite forces using *σ* = 350 kN [after 5], and doubled to reflect bilateral bites. Adjusting primate bite forces [4-6] reflecting *σ* = 370 kN [after 2] made no qualitative difference to the results, and we used bite force estimates as published for these taxa [except for 5]. 3) Unilateral bite force estimates [7-10] were doubled to reflect bilateral bites, but not unilateral *in vivo* bite force measurements, which were used unadjusted, as they can often and regularly attain maximal bite force [11].

We retained only reliable measures of bite force by excluding records using the following criteria: 1) method used to obtain bite force is unclear – e.g. from old literature sources; 2) bite force value is a clear outlier – e.g. spurious recording/estimate (see below for some examples in dinosaurs); 3) bite force value clearly does not represent maximum muscle-driven bite force – e.g. bite force of juvenile specimens or recorded from non-maximal biting behaviour or indirect traces such as bite marks and indentations (see below); and 4) the taxon is not included in our phylogeny (see below). Out of the 732 records collected, our final dataset comprises bite force spanning 434 species (Table S1; Data S1-S2).

**Estimating Bite Force in Extant and Extinct Felids**

Bite forces were estimated in 51 species of extant and extinct felids using the “dry skull” method [3], in which cross-sectional areas of adductor muscles (*A*Musc) were estimated from dimensions of the cranial adductor fossae from photographs of dry skull specimens [1-3, 12, 13]. Felid bite forces build on the dataset of [13] with additional fossil taxa including seven species of sabre-toothed cats, two species of basal felids (*Proailurus lemanensis* and *Hyperailurictis validus*), the Eurasian cave lion *Panthera spelaea*, the fossil jaguar *P. augusta* and an unnamed fossil felid (F:AM 62192) most likely either a stem feline (basal to crown Felinae) [14] or a stem non-*Panthera* feline.

**Estimating bite force in extant and extinct archosaurs**

We estimated bite forces in five species of extant birds (*Branta canadensis*, *Buteo buteo*, *Larus fuscus*, *Gallus gallus*, and *Struthio camelus*) and 17 species of extinct archosaurs including 13 species of dinosaurs (Table S1).

Bite forces were estimated in extant and extinct archosaurs using simple two-dimensional lever models. Levers were measured from landmarks placed on lateral view images of the skull and mandible following [15]. Muscle levers (*m*) were taken for three functional muscle groups [after 16]: the temporal group (mTemp, musculus adductor mandibulae externus [mAME] + m. pseudotemporalis superficialis [mPsTS]); the quadratic group (mQuad, m. pseudotemporalis profundus [mPsTP] + m. adductor mandibulae posterior [mAMP]); and the pterygoid group (mPt, m. pterygoideus [mPt]) (muscle homology and nomenclature following [17]). Bite levers (*b*) were taken at three biting positions along the tooth row, the anterior-most, midpoint, and posterior-most positions, with the posterior-most position serving as the bite point for the maximum bite force.

Static bite force *F*Bite was calculated as follows:

*F*Bite = 2*σ*Σ(*A*i*m*i) / *b*, (1)

where *σ* is the specific muscle tension (300 kPa [18]), *A*i is the *A*Musc for the ith muscle group, *m*i is the lever for the ith muscle group, and *b* is the lever for the bite point. *F*Bite is doubled to reflect a bilateral bite force. As with the muscle lever *m*, *A*Musc were grouped into the three functional units described above.

For the extant archosaurs, *A*Musc were obtained through dissection, either as the physiological cross-sectional area (*A*Phys) estimated from muscle architecture measurements following [19], or the gross cross-sectional area (*A*Gross), as simply the cross-section taken perpendicular to the long axis of the muscle body [16]. *A*Gross was measured in one specimen each of *S. camelus*, *B. buteo*, and *B. canadensis*, and two specimens each of *L. fuscus*, and *G. gallus*. The muscle body was sectioned roughly perpendicular to the major axis of the muscle body, and its *A*Gross was digitally measured using Image J [20]. The mean value of the left and right sides was taken as the final *A*Gross value. Muscle measurements for *A*Phys calculations were taken for two specimens of *B. buteo* and one specimen each of *L. fuscus* and *S. camelus*. Muscles were weighed prior to sectioning. Muscles were carefully sectioned under a microscope using a sharp scalpel. Incisions were made parallel to the length of the muscle fibers as much as possible. Fiber lengths (*L*) and pennation angles (*θ*) were measured using ImageJ. For all specimens, fiber lengths were taken at multiple locations on one or more sections through each muscle, the mean of which was taken as *L*. *A*Phys for each muscle was then calculated as:

*A*Phys = (*M*cos*θ*) / (*ρL*), (2)

following [21], where *M* is the wet weight of the muscle body (g), *θ* is the mean pennation angle, *ρ* is the specific density (1.056×10-3 g/mm3 [22]), and *L* (mm) is mean fiber length. In the case of parallel fibers *θ* is 0° and thus cos*θ* is 1.

For extinct archosaurs, cross-sectional areas of the jaw adductor muscles were estimated as *A*Gross using a variant of the dry skull method [3] whereby cranio-mandibular dimensions (namely the areas of the supratemporal, subtemporal, and mandibular fenestrae) were used to bound the *A*Gross of individual jaw adductor muscles. We measured *A*Gross of the mAME, mPsTS, mPsTP+mAMP and mPt on photographs and diagrams of reconstructed skulls taken at various angles of view. These are conceptually similar to previously published methods to estimate *A*Gross in extinct dinosaurs [7, 8, 16]. We further applied muscle pennation angle *θ*=45° for the mTemp group, *θ*=0° for the mQuad group, and *θ*=30° for the mPt group, based on average pennation angles in our extant archosaur samples. We applied the effects of pennation on to *A*Gross through division of *A*Gross by sin*θ*. This approximates *A*Phys in fossil archosaurs.

**Validating bite force estimates**

We validated our bite force estimates based on *A*Musc measurements using a 20.3 kg specimen of *Alligator mississippiensis* (KPM-NFR 000016). Bite force calculated using the method outlined above, was 1.1 kN. Using ontogenetic bite force and body size data in *A. mississippiensis* [23, 24] (data kindly provided by G.M. Erickson to MS), we can predict the bite force corresponding to an individual of 20.3 kg from an ordinary least squares (OLS) regression model fitted on log10*F*Bite against log10*M*Body. The predicted bite force for 20.3 kg is 1.2 kN (95% prediction intervals: 0.65-2.2 kN), nearly identical to our estimated bite force.

Further, our estimated bite forces for other taxa are also comparable – i.e. in the same order of magnitude – to previous estimates [25-28] (Table S1). For instance, our estimate for a 0.69 kg common buzzard *Buteo buteo* (8.9 N) is comparable to published bite force estimates of similarly sized falconiforms, *Falco peregrinus* and *F. mexicanus* (9.2-13 and 13.8 N respectively) [25]. We also estimate a bite force of 33 N for the ostrich *Struthio camelus* based on our *A*Musc measurements, while that calculated from *A*Musc data found in the literature is 52.7 N [26].

Similarly, for fossil taxa, we estimate bite force for *Allosaurus fragilis* at 9.4 kN (published estimate at 8.7 kN [27]), for *Majungasaurus crenatissimus* at 7.8 kN (published estimate for *Carnotaurus sastrei*, a closely related taxon, at 5.3 kN [7]), and for *Tyrannosaurus rex* at 45-49 kN (published estimate at 57 kN [27]). Slight deviations may owe to individual variation (different specimens used), differences in muscle reconstructions (multi-body dynamic modelling employed by [27] is a dynamic approach that can accommodate detailed muscle parameterization), and measurement error, but the nature of the discrepancies are not easily determined. Nonetheless, our bite force estimates fall within expected variance of *in vivo* measurements given the empirical relationship between log10*F*Bite and log10*M*Body.

**Excluded bite force data**

Some bite force data were excluded from the analysis, and we comment here on some of the notable fossil examples.

***Tyrannosaurus rex.*** *Tyrannosaurus rex* has receiveddisproportionate interest with respect to its palaeobiology and biomechanics, including bite force [27-32]. In particular [27-29, 31] have presented estimates of maximum bite force in *T. rex*: 6.41-13.4 kN [29], 183-235 kN [31], 4.0kN and 57 kN (for juvenile and adult respectively)[27], and a range of values, 8.5-35 kN, spanning multiple specimens [28]. Bite force estimates of Gignac and Erickson [28] were not included in our dataset owing to the fact that the publication appeared after the cutoff for our data collection period. Estimates from Erickson et al. [29] are based on puncture experiments using modern bovine bone and predicting forces necessary to puncture cortical bones to the depth recorded on a *Triceratops* pelvis. Thus, it is unlikely to represent maximum bite force but some normal feeding bites, and we have not included these values in our final dataset. On the other hand, bite forces from Meers [31] are an order of magnitude higher than those of [27, 28] as well as those estimated here (Table S1). Meers [31] extrapolated bite force from a bivariate relationship between bite force and body mass in extant animals (to our knowledge, the largest such dataset published at that point in time) using a relatively conservative body mass estimate for *T. rex*. However, some univariate *in vivo* bite force values (specifically those of *Crocuta crocuta* [33] and *Alligator mississippiensis* [23, 24]) were doubled to reflect bivariate bites [31]. Doubling is necessary when estimating static bite force from dry skulls or muscle architecture data to incorporate the effects of bilateral muscle contractile forces into the bite force lever model. However, *in vivo* unilateral bite force recordings often are not simply half the values of bilateral recordings, but ~75% to 100% [11, 34]. This most likely owes to the muscles on the balancing side (contralateral; on the opposite side of a unilateral bite point) contributing along with the muscles on the working side (ipsilateral; on the same side of the bite point) to generate unilateral bite force, albeit not necessarily maximally – e.g. muscle activity on the ipsilateral side is 30% higher than on the contralateral side in humans (and indeed bilateral bite force is also about 30% higher than unilateral bite forces) [34]. Involuntary muscle stimulation in opossums demonstrate that at least in some species under certain conditions (i.e., involuntary stimulation while under anesthesia) maximum unilateral bite force can reach forces attained in maximum bilateral bites [11]. Taken together, we can safely conclude that unilateral bite forces should be taken at face value as a near approximation of maximum bite force in the absence of a known correction factor (e.g. 1.35-1.45 [34]), and most definitely should not be doubled. Thus, we have selected the 57 kN bite force estimate of Bates and Falkingham [27] for our taxon-representative bite force value for *T. rex*.

***Allosaurus fragilis.*** Perhaps less popular as the focus of biomechanical studies compared to *T. rex*, *Allosaurus fragilis* nonetheless has been included in many biomechanical studies of biting performance [15, 16, 27, 35-38], but bite force estimates are comparatively rarer [16, 27] (this study; Table S1). While our bite force estimates (9.4 kN) are comparable to that of Bates and Falkingham [27] (8.7 kN), those of Rayfield et al. [16] vary considerably depending on the method used. For instance, a purely muscle-driven bite force at the posterior-most biting position is estimated at 3.6 kN, whereas the maximum bilateral force that can be applied at maxillary teeth 3, 4, and 5 before the skull yields is estimated at 55 kN using Finite Elements (FE) modelling [16]. Accounting for muscle forces, then the skull is estimated to yield at 19 kN [16]. These latter estimates for the yielding force of the skull are, while biomechanically and biologically interesting, probably not appropriate as proxies of bite force. It is unlikely that muscle-generated maximum bite force will reach magnitudes at which the skull breaks – the more realistic interpretation would be that the skull is “over-engineered” [16].

***Deinonychus antirrhopus.*** Owing to the fragmentary nature of its fossil skull materials, *Deinonychus antirrhopus* has received much less attention with respect to biomechanical performance of biting. Nonetheless, we are aware of two studies that include *Deinonychus* [38, 39], with Gignac et al. [39] presenting bite force estimates. As Therrien et al. [38] presented relative mandibular force profiles, which should not be confused with bite force for similar reasons to FE estimates [16], we will focus our discussion on the bite force estimates of Gignac et al. [39]. Gignac et al. [39] used the same approach as in Erickson et al. [29] and predicted the force necessary to have produced bite marks at the depths as preserved on *Tentosaurus* bones (namely the radius). These indentations have been attributed to *Deinonychus* based on comparisons between the casts of the indentations and tooth morphology of contemporary theropod taxa [39]. The force necessary to produce the fossil bite mark was 3.1 kN, but up to 4.1 kN after accounting for increased loading rates and soft tissue resistance, with an extrapolated maximum (posterior-most) bite force of 6.3 and 8.2 kN, unadjusted and adjusted respectively. The only comparable bite force available in similarly-sized dromaeosaurs is our estimate for *Dromaeosaurus albertensis* (Table S1) at 0.89 kN, an order of magnitude lower than that of Gignac et al. [39].

*Dromaeosaurus* has a skull length of ~230 mm and a skull width of ~103 mm. Dinosaurian taxa in our dataset with approximately similarly-sized skulls include *Herrerasaurus* (*L*Sk, 310 mm; *W*Sk, ~108 mm; *F*Bite, 1.9 kN), *Erlikosaurus* (*L*Sk, 265 mm; *W*Sk, 100 mm; *F*Bite, 0.22-0.27 kN), and *Struthio* (*L*Sk, ~200 mm; *W*Sk, ~94 mm; *F*Bite, 0.01-0.05 kN). Comparably-sized mammalian taxa include *Canis latrans* (*L*Sk, ~190 mm; *W*Sk, ~100 mm; *F*Bite, 0.68 kN), *Canis lupus* (New Guinea singing dog and dingo; *L*Sk, ~180 mm; *W*Sk, ~100 mm; *F*Bite, 0.6 kN), and *Cuon apinus* (*L*Sk, ~180 mm; *W*Sk, ~108 mm; *F*Bite, 0.67 kN). Allowing for short-skulled mammalian taxa with similar skull widths will increase our comparative sample to include 6 species of extant and extinct felids spanning bite forces between 0.5 and 0.97 kN: *Leopardus pardalis* (0.52 kN); *Pardofelis temminckii* (0.57 kN); *Lynx pardinus* (0.63 kN); *Lynx lynx* (0.67 kN); *Metailurus parvulus* (0.8 kN); and *Proailurus lemanensis* (0.97 kN). Only *Herrerasaurus* is estimated to have bite force above 1 kN, and it is larger than *Dromaeosaurus*.

Given that: 1) our approach results in bite force estimates for dinosaurs that are comparable to previous estimates [27, 28]; 2) that they also correspond well with extant taxa of comparable skull sizes; and 3) that the fossil bite marks attributed to *Deinonychus* are matched to the premaxillary teeth of the producer [39] and it is inconceivable that an anterior-most bite force of a long-snouted theropod would be higher than the posterior-most bite force in similarly-sized taxa, we conclude that the bite force of *Deinonychus* as presented by Gignac et al. [39] is an overestimate. Contrasted with the bite marks of *T. rex* [29, 40], which we interpret as being the result of non-maximal feeding bites, it is conceivable that the bite marks of *Deinonychus* [39] were produced as a result of more aggressive kinetic biting behaviors involving movement of the whole head and neck [41, 42] rather than by a static muscle-driven bite. Therefore, bite force for *Deinonychus* is not included in our analysis.

**Body Mass Data**

Body mass data were collected primarily through the literature, prioritizing values associated with bite force records – especially for individual records in which bite force and body mass were collected for the same individual (Table S1). For bite force records in which associated body mass data were not available, taxon-representative body mass data were taken from widely used resources including [43] for birds and [44] for mammals (see Table S1 for additional sources). For extinct species, we either relied on published estimates (e.g. [45]) or predicted from cranio-dental measurements and published regression equations [46] (Table S1). For species in which there are multiple body mass data available, we took the mean value as the taxon-representative body mass.

The choice of proxy for size can have profound impact on model outcomes and interpretations of trait evolution. For instance, Law et al. [47] found that the evolution of body mass and body length are decoupled in musteloid mammals, while Kilbourne and Hoffman [48] found that scaling of limb inertial properties in mammals differ when limb length or body mass are used as a proxy for size. In the context of our study, since bite force is to a great extent determined by the size of the head, some measure of head size (skull length, skull width, jaw length) would be ideal if the objective of our study were to account for geometric scaling.

However, we chose body mass as a proxy of size over head size, for several reasons: 1) body mass is most readily available in the literature compared to other candidate proxies such as head length or head width; 2) body mass is not directly redundant with bite force – i.e., head length is redundant with biting levers while head width is redundant with muscle diameter; and 3) body mass is tightly associated with ecology. The last reason is perhaps the most important. As we are interested in detecting instances of exceptional rates of adaptations as a response to ecological selection, the relative bite force for an animal of a certain body size that fills a particular size-associated ecological niche should be what natural selection would be working on. For instance, we would expect the herbivorous dinosaurs *Stegosaurus* and *Plateosaurus*, to have occupied some specific ecological niche associated with their large body sizes, not with their disproportionally smaller head sizes.

**Phylogeny**

We used the Time Tree of Life (TTOL) [49] as the backbone phylogeny, with fossil tips inserted at the appropriate phylogenetic and temporal positions according to recent or widely accepted phylogenetic placements as follows. The extinct Miocene thylacine, *Nimbacinus dicksoni* (15.97-11.608 Mya), was inserted to the TTOL as sister to the recently extinct thylacine, *Thylacinus cynocephalus* [50], with their divergence based on the former’s first appearance date (FAD). The Cretaceous marsupial *Didelphodon vorax* (69-66 Mya) was inserted along the stem of crown marsupials [51], using the FAD of the family, Stagodontidae, 105.3 Mya to date its divergence. The divergence between *Euparkeria* (247.2-242 Mya) and other archosaurs [52] was determined using the FAD for the family, Euparkeriidae, at 251.2 Mya. The origin date for the phytosaur *Parasuchus* (237-221.5 Mya) was determined from the FAD of the clade Phytosauria (Carnian: 237 Mya) and inserted in the relevant phylogenetic position following [52]. The divergence between the clade comprising *Ornithosuchus* (237-208.5 Mya) and *Riojasuchus tenuisceps* (228-208.5 Mya) and other archosaurs (phylogenetic positions according to [52]) was determined from the FAD for the clade Ornithosuchidae (Carnian: 237 Mya). Fossil dates are from the Paleobiology Database (accessed 9 Feb 2017).

We replaced the tip branch of *Homo sapiens* in the TTOL with the dated phylogeny of fossil hominins taken from [53].

For Dinosauria, we took the phylogeny and dates from [54]. For Felidae, we combined two published trees [14, 55], slightly modifying interrelationships between smilodontin sabre-toothed cats recovered by [14] to better reflect current consensus [56]. Divergence dates were determined from their first appearance dates (FAD) with zero length branches scaled using the “equal” method [57], and tips extended to their last appearance dates (LAD). Branches were scaled using the paleotree R package [58]. Scaling branches using the total FAD-LAD ranges ensures the maximum amount of time available for evolution to occur, forming the most conservative basis for rate shift detection – i.e. given the longest possible tree length, rate shifts must be truly exceptional in order to be recognized. Conversely, this means that any younger FAD and/or older LAD will shorten the branches resulting in a more generous basis for rate shift detection – i.e., potentially more rate shifts may be identified than presented here. The dinosaurian and felid trees were grafted onto the appropriate positions in the TTOL. Felid taxa in the TTOL were pruned save one, along the branch of which the scaled felid tree including the fossil taxa as describe above was attached at 34 Mya (after the Eocene/Oligocene boundary, the possible appearance of earliest felids [56]). The clade Aves was taken from the TTOL and grafted onto the Dinosauria phylogeny, along the branch leading to the Late Cretaceous bird *Vegavis* (66 Mya) at 118.46 Mya (or 52.46 Myr deep from the *Vegavis* tip age). *Vegavis* was subsequently dropped from the tree as it is not in the final bite force data. All avian taxa in the TTOL were pruned save for one, along the branch of which the scaled dinosaur tree including all extant birds as described above was attached at the basal divergence date between Aves and Crocodilia as in the original TTOL (238.34 Mya). The combined tree was pruned to reflect the number of species in the bite force data set (*N* = 434).

**Model grouping structure**

Each taxon was assigned to one of five groups: mammals excluding bats (hereafter “Mammals”), bats, finches (Fringillidae, Estrildidae, and Darwin’s finches); non-finch dinosaurs (including other birds, hereafter “Dinosaurs”); and non-dinosaurian diapsid reptiles (hereafter “Reptiles”) (Fig. 1; Table S1). We chose these five groups because they act as good descriptors of the distribution of data (Fig. S3) – previous studies show that finches and bats have steeper slopes [59, 60] compared to other vertebrates, which generally scale close to the theoretical isometric slope of 0.67 [61] – as well as conforming to widely recognized taxonomic groups (mammals and reptiles were divided into separate groups to avoid polyphyletic grouping). Birds and dinosaurs were grouped together, as fossil evidence points to a blurred distinction in their physiology and biology [62-65], while dinosaurs are very different from other reptiles [66].

**Potential confounding factor: biting position**

To test whether different biting positions affect the scaling relationship between bite force and body mass, a variable-rates regression model of the form log10*F*Bite ~ BitePoint + log10*M*Body + log10*M*Body:BitePoint + log10*M*Body:Bats + log10*M*Body:Finches, where BitePoint is a categorical variable on biting position (anterior or posterior), Bats and Finches are categorical variables representing group memberships respectively, was fitted. Interaction terms between log10*M*Body and categorical variables are the separate slope terms. Intercept differences for Bats and Finches are not modeled (see text). Since both Bats and Finches comprise entirely of anterior bite forces, while having significantly different slopes from the other groups (Mammals, Dinosaurs, and Reptiles) we mainly tested the effects of bite point on the entirety of the non-chiropteran, non-finch amniotes by estimating separate slopes for anterior and posterior biting positions. We did not further test differences within the three groups – e.g. Dinosaur:BitePoint – owing to small sample sizes in certain categories. There is no significant difference in slopes between the anterior and posterior biting positions (*p*MCMC = 0.473), but there is a significant difference in intercepts between the two biting position categories (*p*MCMC < 0.05). See Fig S6.

**Potential confounding factor: bite force type**

To test whether different bite force types (*in vivo* measurements or indirect estimates) affect the scaling relationship between bite force and body mass, a variable-rates regression model of the form log10*F*Bite ~ BiteType + log10*M*Body + log10*M*Body:BiteType + log10*M*Body:Bats + log10*M*Body:Finches, where BiteType is a categorical variable of bite force type (*in vivo* or estimate), Bats and Finches are categorical variables representing group memberships respectively, was fitted. Interaction terms between log10*M*Body and categorical variables are the separate slope terms. Intercept differences for Bats and Finches are not modeled (see text). Since both Bats and Finches comprise entirely of *in vivo* bite forces, while having significantly different slopes from the other groups (Mammals, Dinosaurs, and Reptiles) we mainly tested the effects of bite force type on the entirety of the non-chiropteran, non-finch amniotes by estimating separate slopes for *in vivo* bite force measurements and indirect bite force estimates. We did not further test differences within the three groups – e.g. Dinosaur:BiteType – owing to small sample sizes in certain categories. There is no significant difference in slopes (*p*MCMC = 0.42), nor is there a significant difference in intercepts between the two bite force categories (*p*MCMC = 0.358). See Fig S7.

**Priors in BayesTraits**

Default priors were used for all VR models fitted using BayesTraits:

* Regression coefficients (β): uniform distribution (-100, 100)
* Rate scalars (branch and node): scaled gamma distribution (α=1.10; β=1.00)

Unlike some other methods to detect rate heterogeneity in the evolution of a continuously varying trait – e.g., auteur [67], BAMM [68] – the VR model as implemented in BayesTraits does not require a prior on the number of rate shifts. Rather, each branch on the phylogenetic tree has equal probability of being scaled by a rate scalar (*r*) such that the amount of trait change along each branch conforms to Brownian Motion. The rate scalars themselves are drawn from a prior distribution (a gamma distribution [α=1.10; β=1.00] rescaled so that the median is one). Thus, for ***every branch that is scaled***, a parameter *r* is estimated through MCMC.

**Detecting rate shifts in bite force evolution**

We tested for instances of exceptional bursts of evolutionary change in bite force based on rate shifts along branches on the phylogeny [69]. The VR regression model as implemented in BayesTraits works to modify the branch lengths to “detect heterogeneity in rates of phylogenetically structured residual errors” [69]. That is, once the appropriate level of variance in the response variable – e.g., bite force – is explained by some predictor variable(s) – e.g., body mass – outlying deviations from the regression line will be explained as rate shifts (Fig. S1). Under Brownian motion, bite force – after accounting for body mass and other confounding variables – evolves at a rate proportional to time (and an estimated background variance) across the phylogeny. For any evolutionary change along a given branch that is greater/less than the expected amount of change for the duration of time to occur (given body mass), the corresponding branch must be stretched/compressed in length in proportion to the observed amount of phenotypic change – corresponding to a rate increase/decrease. In a regression context, modifying the branch lengths corresponds to modifying the variance-covariance matrix of the trait values according to the variance-covariance matrix of the scaled phylogenetic tree [70]. This is conceptually identical to simpler homogenous scaling parameters such as λ, δ, κ [71] or α of the OU model [72] – only that branch scaling can be applied to any branch or clade anywhere on the tree in the VR model rather than to the entire tree.

The magnitude of branch stretching/compressing is the rate scalar (*r*). It follows that, phenotypic changes owing to adaptations (potentially as a response to strong selective pressure) would be proportional to *r*. Thus, we define exceptional change following the criteria of [69]: 1) certainty of rate shifts, the branch in question must be scaled in >95% of the posterior sample of scaled trees; and 2) magnitude of rate shifts, the *r* in question must be greater than two. This is akin to how positive genetic selection is detected using the famous dN/dS ratio [69]. Rate heterogeneity that do not fulfil these two criteria were not considered as instances of exceptional rates of adaptations. We determined whether rate-shifts constituted exceptional rates of adaptations if they satisfied the criteria set out above in all of three independent replicate Markov Chain Monte Carlo (MCMC) chains.

In order to determine if rate-heterogeneity was statistically significant, we fitted an *equal-rate* *(ER) model* (or Brownian motion) as a simpler alternative to each of our VR models. Model selection was performed using the Bayes Factor (BF): BF is defined as twice the difference in log marginal likelihood (*m*) between the complex model (model1) and the simple model (model0) – i.e., BF = 2 × (*m*1 – *m*0). Marginal likelihood is essentially the likelihood of the model after accounting for the number of parameters. We computed BF using *m* from our 5-Group VR model and the simple alternative 5-Group ER model, and selected the VR model over the ER model when BF value was greater than 2 [73]. For completeness, Raftery’s[73] interpretations of BF are: < 0, negative support; 0-2, barely worth mentioning; 2-5, positive; 5-10, strong; > 10, very strong.

We ran our MCMC chains for 109 iterations, with a burn-in period of 108 iterations, sampling every 105 iterations, resulting in a posterior sample of 900 modified VR trees and model estimates, for each regression model. We used stepping stone sampling to compute marginal likelihoods from which BF were calculated. Post-processing of the BayesTraits outputs were conducted using an online post-processor (available at www.evolution.reading.ac.uk/VarRatesWebPP), as well as in R [74].

If exceptional rates of adaptations are detected in bite force evolution, then we would expect there to be two types of rate shifts on the phylogenetic tree (Fig. S1). The first type of rate shifts, branch-wise rate shifts (branch shifts), are cases in which significant increases in rates with respect to background rate are detected along individual branches. These may be along terminal branches or internal branches often subtending monophyletic clades. Terminal branch shifts represent independent acceleration of evolution in individual species since their most recent divergence from close relatives. On the other hand, internal branch shifts represent cases in which evolution sped up in an ancestral lineage, with traits rapidly moving away from the ancestral trait value (Fig 1). Subsequent evolution in descendant branches may revert back to the ancestral rate. In a regression context, such internal branch shifts will be the equivalent of an intercept offset (a grade shift) in the focal clade compared to the rest of the sample (Fig. S1). With respect to bite force evolution, we would expect to find internal branch shifts where an ancestral lineage rapidly increased/reduced bite force to evolve into a drastically different functional niche, with the descendants remaining in that niche and evolving at a constant rate. For instance, sabre-toothed cats have long been known to have had substantially lower bite forces compared to extant carnivores of similar sizes and if such a reduction in bite force occurred rapidly, then we would expect a branch shift along the branch subtending the clade Machairodontinae (although we do not find this pattern).

The second type of rate shift, clade-wise rate shifts (clade shifts), occur across all branches within a clade and represent cases in which rapid divergences in trait values have occurred (Fig. S1). In a regression context clade shifts will be associated with an increase in variance about the best-fit regression line in the focal clade compared to the rest of the sample (Fig. S1). With respect to bite force evolution, clade shifts will be expected in clades that have undergone rapid ecological expansion, often exploiting ecological opportunities available following the colonisation of a new geographical region (e.g., Darwin’s finches on the Galapagos Islands) or even after a mass extinction event. The expectation here is that exceptional rates of adaptations manifest in lineages of a clade in a way that bite force values rapidly diverge and continuously evolve in response to strong and diverse functional demands.

**The VR model in relation to other models of evolution**

There is increasing evidence indicating that departure from simple uniform evolutionary processes like Brownian Motion (BM) is common and widespread across various phenotypic traits. Perhaps the most commonly used alternative model to simple BM is the Ornstein-Uhlenbeck (OU) model. The OU model (like the VR model) is an extension of the BM model with two additional parameters: θ, the evolutionary optimum; and α, the strength of the pull towards θ. The evolutionary interpretation of the OU model is that of stabilizing selection working on the trait of interest towards θ or an adaptive peak. Multi-peak OU models allow for the estimation of multiple values of θ and α applied to various parts of the phylogenetic tree.

Practically, the OU model can (though not necessarily) be fitted as a BM model on a scaled phylogenetic tree in which the branches have been modified according to αi. This is conceptually similar (if not identical) to how the VR model is fitted – only that the VR model scales the branches/nodes of the phylogeny separately and not according to any *a priori* evolutionary process. Thus, the VR model is more flexible in simultaneously placing branch and node scalars, both stretching and compressing individual branches and across entire clades. This flexibility allows the VR model to fit rates that can explain patterns arising from an OU process of trait evolution. A branch scalar (*r* >1) applied to a branch subtending a clade of interest (branch- or mean-shift) is akin to a shift to a new adaptive regime, while a clade-wide rate scalar (*r* <<1; extreme rate decrease) applied to the branches across that clade corresponds to stabilizing selection (or even stasis) preventing trait values within the clade deviating away from the adaptive peak (the clade-mean trait value). However, while the OU model explicitly models departures from BM with stabilising selection as the only evolutionary process, the VR model can flexibly model evolutionary rate heterogeneity (increases as well as decreases), along individual branches and across entire clades (and a combination of each). Thus, the VR model not only is capable of modelling processes similar to stabilising selection, it is not restricted to it as the OU model is.

**Rates of phenotypic evolution as proxies for adaptations**

Macroevolutionary rates have long been equated with adaptations and the strengths of selection [69, 75-79]. That is, given the expected amount of trait change owing to the passage of time and evolutionary history (e.g., effects of shared ancestry and background evolutionary rate), any amount of phenotypic change that exceeds (or under achieves) that expectation must be because of extrinsic factors – namely selection. Faster rates of evolution would signify strong selection for trait change while slower rates would indicate strong selection for no change (stabilizing selection).

Additionally, in this context, phenotypic evolutionary changes following Brownian Motion can also be interpreted as adaptations, in the sense that small incremental changes in phenotype can be the result of adaptations in response to natural selection. Strictly speaking the Brownian motion model of trait evolution does not include selection as a process of evolution, implying that changes entirely owe to drift. In the context of phenotypic macroevolution, we can hypothesize that while drift may be responsible for the underlying mechanism generating genotypic and thus phenotypic variation, such phenotypic changes owing to drift must be suitable adaptations given the environments – otherwise they would be deleterious mutations – especially for phenotypes to persist from generation to generation over millions of years. Thus, ultimately, selection is a driving force underlying phenotypic macro-evolution even under constant-rate processes.

One way to think about this is in the context of a gradually shifting adaptive landscape [80]. That is, on a local and short time scale, selection may work as a stabilizing force towards some evolutionary optimum, but that such optima change over time. Such fluctuating optima [80] may result in a process not dissimilar to Brownian motion on a macro-evolutionary scale, as evolutionary optima and adaptive peaks are not stable nor consistent across such longer time scales. The adaptive landscape itself changes through time and phenotypic evolution is simply following suit. This was illustrated rather nicely by Simpson [75] (p. 204; Fig 33c), in what he called a “shifting adaptive zone”.

Incidentally, we refer to “exceptional rates of adaptations” as exceptionally outstanding instances of rate shifts (r > 2) in line with how positive genetic selection is detected using the dN/dS ratio [69, 81, 82]. Exceptional rates of adaptations thus signify an exceptionally strong influence of some factor – that is external to evolutionary history (inherited background rate) and passage of time (branch length) – that must drive the excess amount of trait change. Again, in essence, this is how dN/dS is interpreted.

**Impact of Sampling on Rate Shift Detection**

While our bite force data set presented here is by far the largest compiled for amniotes to date, with only 434 species sampled, it is hardly representative of the total diversity of amniotes, with over 25,000 extant species and over 300 Myr of evolutionary history. It is entirely possible (and within expectations) that we are missing rate shifts along unsampled lineages (especially long branches) in which bite force data do not currently exist.

However, more importantly, we would not expect rate shifts that we do detect, given our sample, to disappear in a larger sample, should bite force data become available. That is, adding more bite force data in perching birds to break up the long branch subtending the Passeroidea clade is not expected to eliminate the rate shift observed along that lineage (though the exact phylogenetic distribution of rate shifts along the series of branches may differ somewhat). This owes to the fact that rate shifts represent changes along branches that exceed the magnitude of change expected given the amount of time elapsed along those branches – the total sum of changes along paths of relevant branches would not change whether more lineages are sampled along those paths.

Conversely, if we artificially reduced the number of sampled taxa in clades where we find clade-shifts, we will undoubtedly reduce the power to detect rate-shifts in the corresponding clade of the reduced sample.

**Assessment of incorrectly identifying rate shifts through simulations**

Here we assess whether the VR model incorrectly detects rate shifts in cases where there are no rate shifts in trait data – e.g., that exceptional rate shifts in Darwin’s finches and hominins are merely because of their recent and rapid diversifications (short branches) and not real rate shifts. Using the amniote phylogeny, we simulated 100 traits under Brownian Motion using the fastBM function in the phytools R package [83] with the variance and root value set to the estimates taken from our 5-Group+BitePoint VR model. This is akin to generating residuals under BM. We then adjusted the simulated traits using the regression coefficient estimates obtained from the VR model to simulate the dependent variable given the actual body mass and modelled relationship between bite force and body mass [69]. For each simulated dependent trait we fitted equal-rate (ER) and variable-rates (VR) regression models using BayesTraits V3 with group-wise body mass and bite point as the independent variables. We compared log marginal likelihoods (*m*) (computed via stepping stone sampling) using Bayes Factors (BF = 2 × [*m*VR - *m*ER]) to determine whether incorporating rate heterogeneity improved model fit over a constant rate model.

We find that in all 100 simulated traits, the VR model has substantially lower *m* compared to the ER model (BF ranging from -16 to -31) meaning that the ER model is a far better fit to data generated under BM than the VR model. Thus, the VR model as implemented in BayesTraits does not incorrectly detect rate shifts where there are none, and that rate shifts that we do detect in our 5-Group+BitePoint VR regression model are real signals contained in the data. This demonstrates that topology and branch lengths – e.g., large number of tips with short branches (densely sampled clades) – do not bias results in favour of more rate shifts when rate shifts are actually not present.

**Bite force – brain size relationship in hominins**

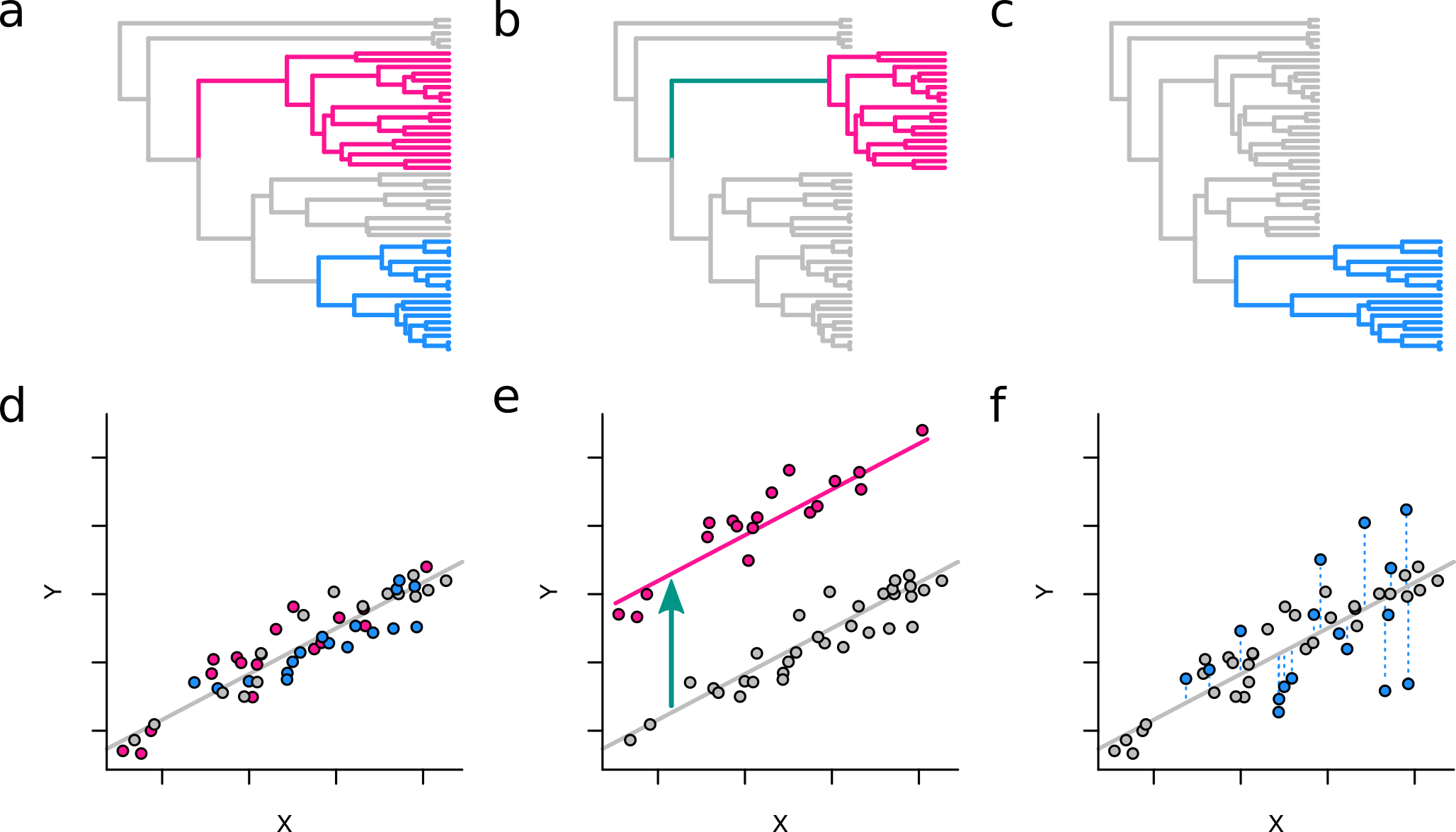
To investigate whether bite force decreased with increasing brain size in hominins, we collected brain size data (endocranial volume [ECV]) [84-86] for hominid taxa (*Pongo*, *Gorilla*, *Pan* and hominins) for which we have bite force and body mass data. For taxa where multiple ECV records were available, we used the taxon mean value. We then fitted a phylogenetic generalized least squares regression (pGLS) using the R package ‘caper’ [87] with log10*F*Bite as the response variable and log10ECV and log10*M*Body as the predictor variables. We assessed significance of regression coefficients using one-tailed tests as we have prior expectations of scaling directionality (positive for *M*Body and negative for ECV).

Results show that as expected, bite force has a positive relationship with body size (slope = 1.35, *p*one-tailed = 0.0459), and a negative relationship with ECV (slope = -1.49, *p*one-tailed = 0.0346). This indicates that after accounting for scaling effects owing to body size, bite force does indeed scale negatively with brain enlargement in hominids (Fig. S9).

**Residuals from a non-phylogenetic regression are misleading**

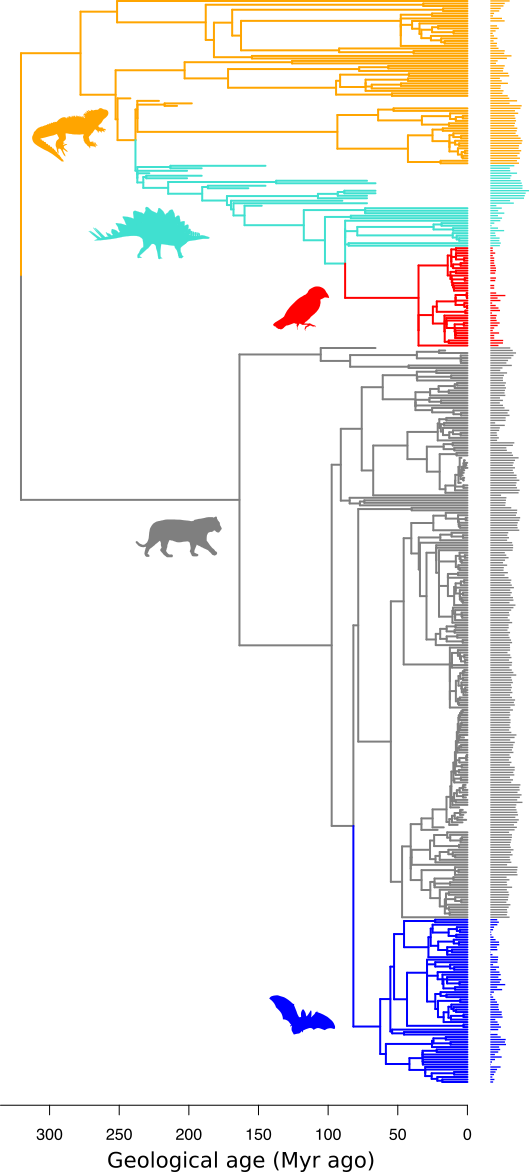
That bite force scales with body mass has been long known [1-3, 13, 23, 24, 31, 51, 88, 89], and deviations from the scaling relationship – i.e. regression residuals – have frequently been used to make inferences about functional adaptations [1, 2, 13, 51, 90, 91]. However non-phylogenetic regression residuals can be misleading. For instance, residuals from a non-phylogenetic regression (5-Group multiple regression model) identify 27 outlying taxa (based on 95% prediction intervals; Table S6), which would normally be considered as taxa exhibiting exceptional levels of adaptations in bite force compared to the majority of the taxa in the dataset. While most of these outlying taxa are also identified as exhibiting exceptional change in our variable-rates model, the magnitudes of associated rate scalars do not correspond with those of the non-phylogenetic residuals (Table S6) – taxa with the highest residuals do not necessarily have the highest rates of evolution. Further, such non-phylogenetic outlier tests result in some cases that are positively misleading – i.e. the two turtles *Chelus fimbriata* and *Pelomedusa subrufa*, and the lizard *Gallotia galloti*, have very large non-phylogenetic residuals but their evolutionary rates are indistinguishable from background evolution. Even more strikingly, non-phylogenetic outlier tests fail to identify the majority of the taxa for which elevated evolutionary rates were reconstructed – e.g. the majority of finches, hominins and maniraptoran dinosaurs (Table S6). Most importantly, non-phylogenetic approaches can only test for deviances from trend lines at the tips (species data), while phylogenetic approaches – because they inherently incorporate evolutionary history – can shed light on deep-time, broad-scale adaptive evolutionary dynamics. Accordingly, we identify multiple cases of clade-wise adaptations that would otherwise not be detected using non-phylogenetic approaches. Thus, while the non-phylogenetic approach may identify *T. rex* as being an outlier (though marginally here, Table S6), phylogenetic analyses reveal that *T. rex* has not undergone exceptional adaptation.

Non-phylogenetic regression residuals can be misleading owing to the fact that the effects of shared ancestry and evolutionary history (passage of time and inherited rate of evolution) are not appropriately taken into account. A residual of high absolute magnitude may not be exceptional if evolutionary time and rate are considered – i.e. the passage of time and the evolutionary rate may be sufficient to explain such a large change in bite force. Thus, it is crucial that an evolutionary phylogenetic approach be taken when identifying instances of exceptional rates of adaptations in biomechanical traits.

****

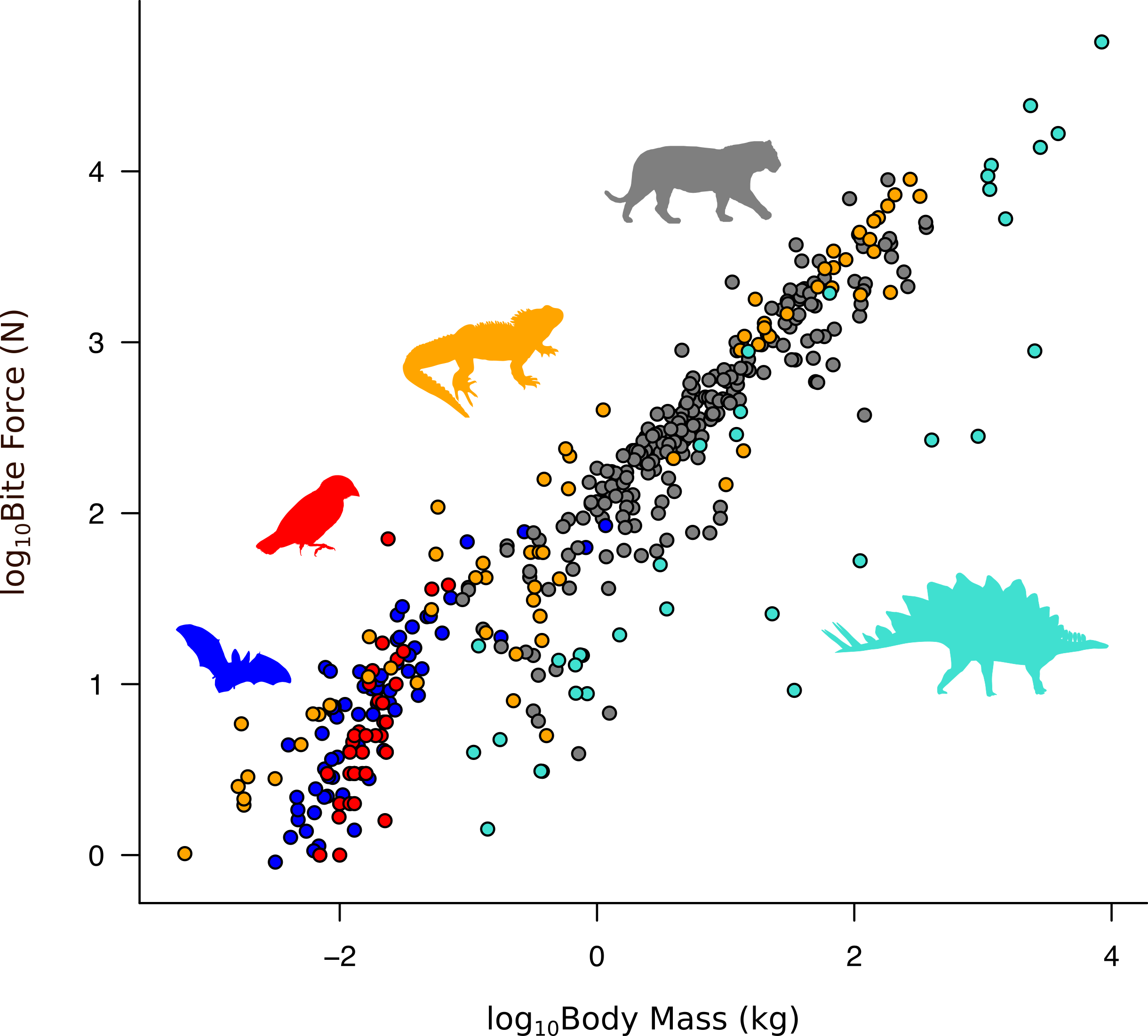
**Figure S1. Theoretical expectations for branch-wise and clade-wise rate shifts in a variable-rate model**

Three hypothetical scenarios of trait evolution are shown as modified branch lengths accompanying a variable-rate model (a-c). Branch lengths are in units of time when traits evolve proportional to time (i.e., Brownian Motion; a) but are modified accordingly when there is a single-branch rate shift (branch-shift; b) or clade-wide rate shifts (clade-shift; c). The branch-shift example here (b) is one occurring along an internal branch (green) subtending the focal clade (pink) and is represented by stretching the corresponding branch in the time tree (a) by a scaling factor (rate scalar, *r*) so that the modified branch length is proportional to the amount of change in trait value. Similarly, a clade-shift (c) is depicted as scaled branches along all branches in the focal clade (blue). In a regression context, evolutionary rate is in Y given X (i.e., regression residuals) and would be randomly distributed about the fitted regression line when rates are constant (d). When a branch-shift has occurred at the base of the focal clade (pink), then corresponding Y values will be offset from the global regression line proportional to the magnitude of the rate-shift (e). Conversely, a clade-shift in the focal clade (blue) will be represented as increased variance in corresponding Y values about the global regression line (f).

****

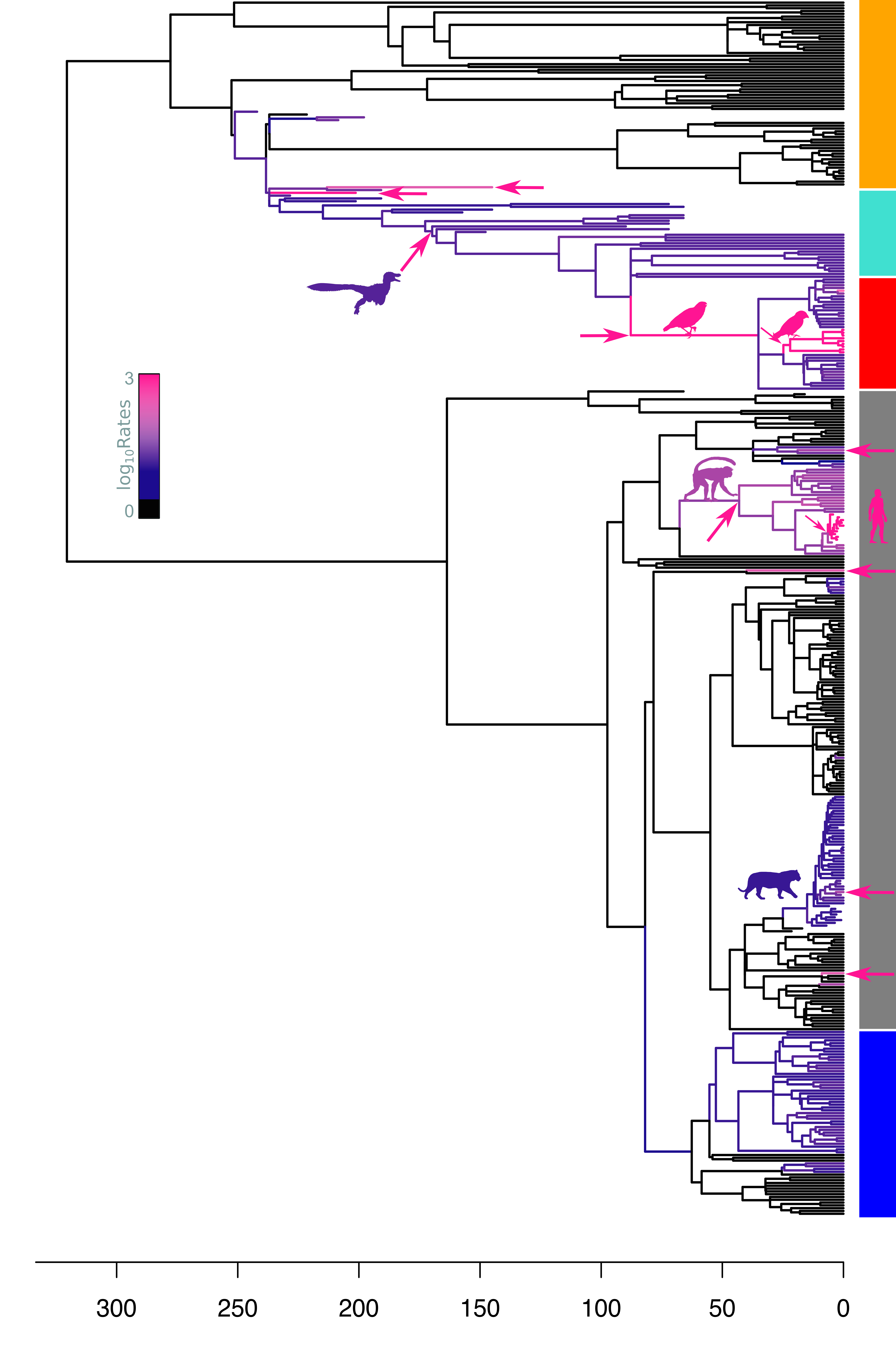
**Figure S2. Distribution of bite force across the phylogeny of amniotes**

Log10 bite force values for each taxon are displayed along the tips of the amniote phylogenetic tree as horizontal bar plots. Branches are coloured according to the five groups of interest: bats, blue; mammals excluding bats, grey; finches, red; dinosaurs excluding finches, turquoise; and reptiles excluding dinosaurs, orange.

****

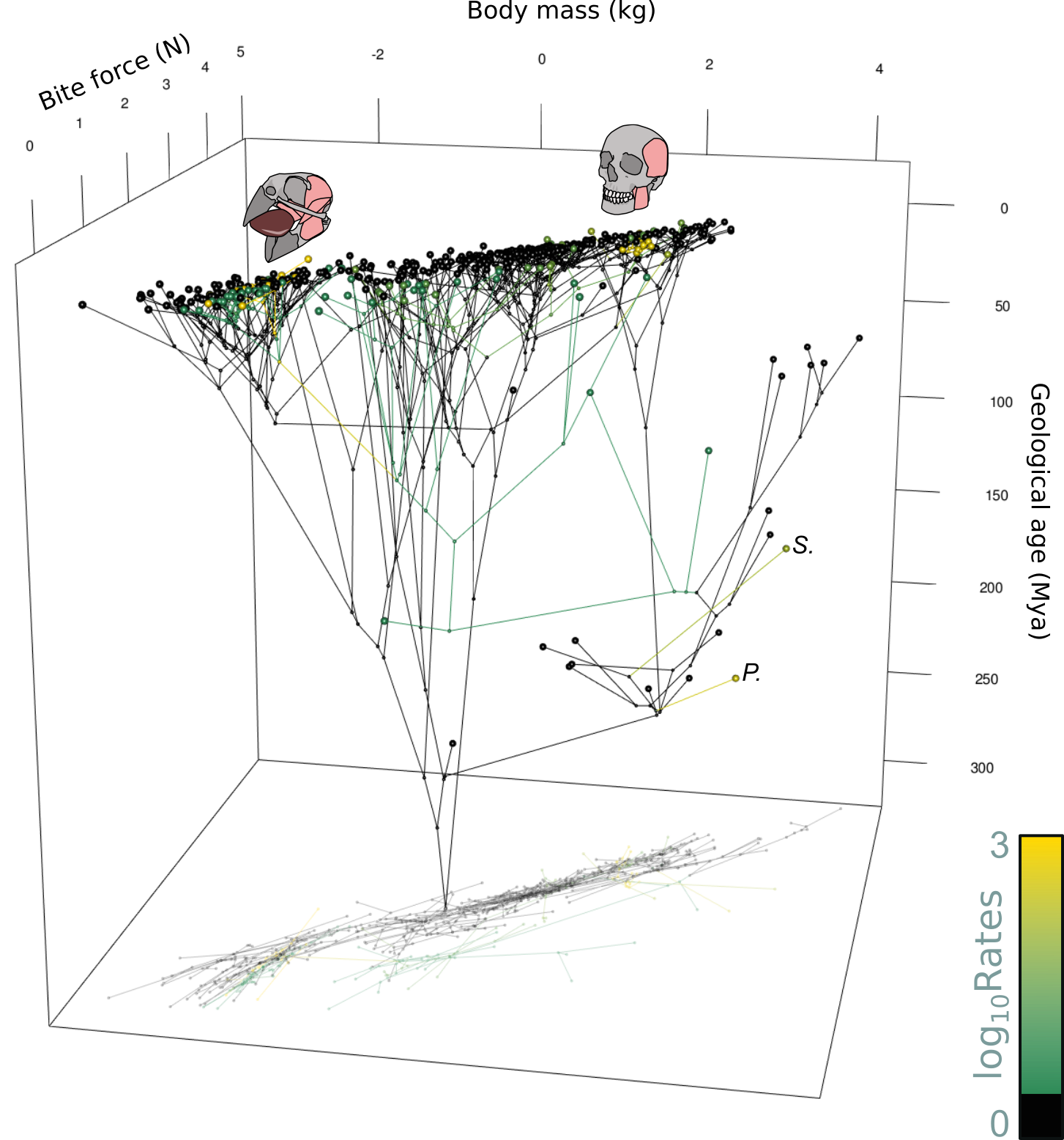
**Figure S3. Relationship between bite force and body mass in amniotes**

The bivariate relationship between log10 bite force and log10 body mass are shown, coloured according to the five groups as in Fig. 1. Overall, there is a clear relationship between bite force and body mass, but with some interesting patterns. Bats (blue) and finches (red) appear to scale differently to the rest of the taxa and we expect a model with separate slopes would explain the data better than one with a single slope. Bite force in many dinosaur species (turquoise) appear to be substantially lower than expected given their body size. These include extant birds but also large-bodied taxa such as *Erlikosaurus*, *Plateosaurus* and *Stegosaurus*. On the other hand, carnivorous taxa such as *Herrerasaurus*, *Dromaeosaurus*, ceratosaurs, allosaurs and tyrannosaurs have bite forces in line with what one would expect from their body sizes. We expect a variable-rate model combined with a separate-slope model (5-Group VR model) to explain the high variability in dinosaurian bite force better than an equal-rate model.



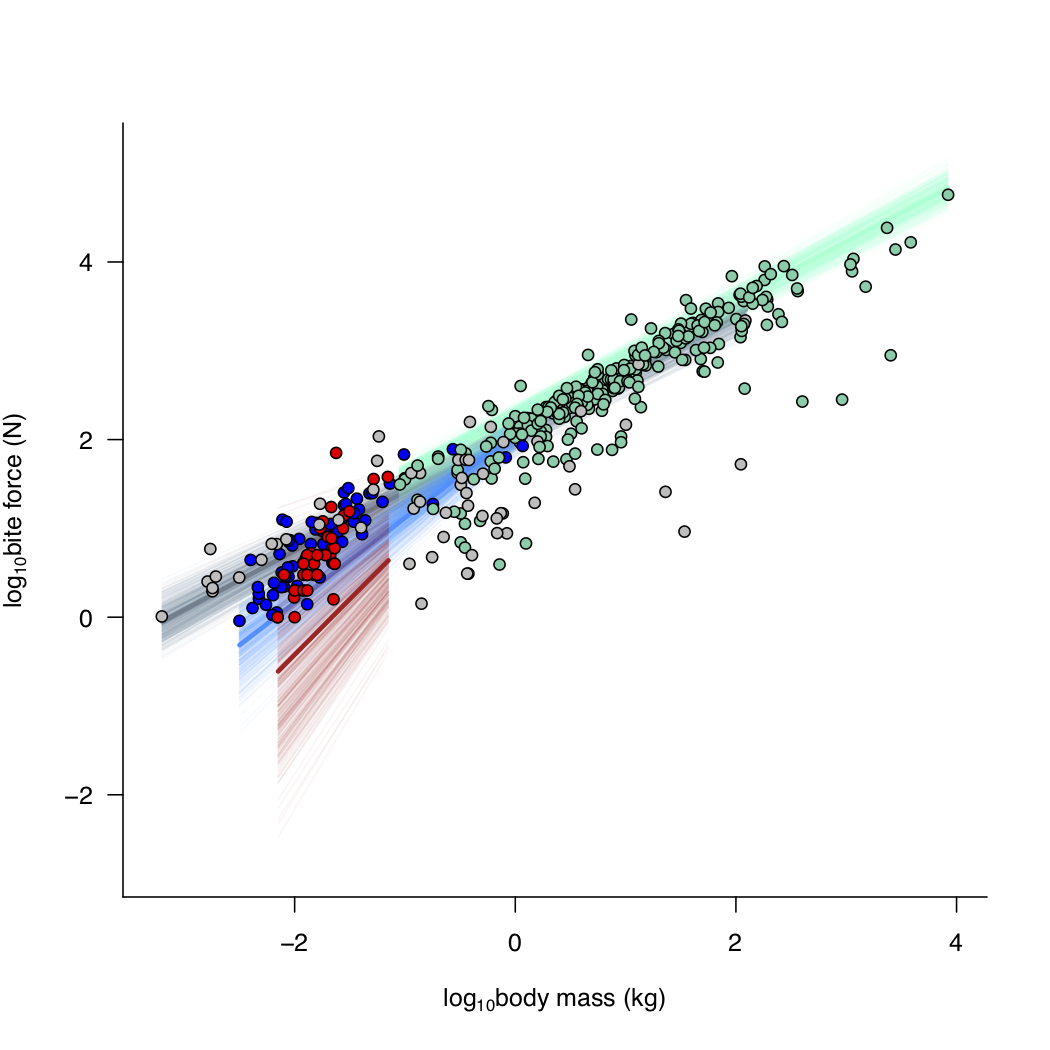
**Figure S4. Rate heterogeneity in the evolution of bite force**

Elevated rates of bite force evolution (in >50% of the posterior sample) are shown as a colour gradient (blue violet to hot pink) on corresponding branches of the phylogenetic tree used in this study, while branches indistinguishable from the background rate are shown in black. Branches and clades in which exceptional rates of adaptations are detected (Fig. 1) are indicated by pink arrows; horizontal arrows indicate branch-shifts while diagonal arrows indicate clade-shifts. Silhouettes show clades of interest, namely those undergoing exceptional rates of adaptations (Fig. 1), but also Felidae. Felids exhibit elevated rates in over 50% of the posterior sample across all branches (clade-wide rate shift) but not in over 95% thus not fulfilling one of the two criteria to be identified as instances of exceptional rates of adaptations. Similarly, chiropteran subclades and all dinosaurian branches also show high proportions of elevated rates in the posterior sample, but not reaching the 95% threshold (except maniraptorans).

****

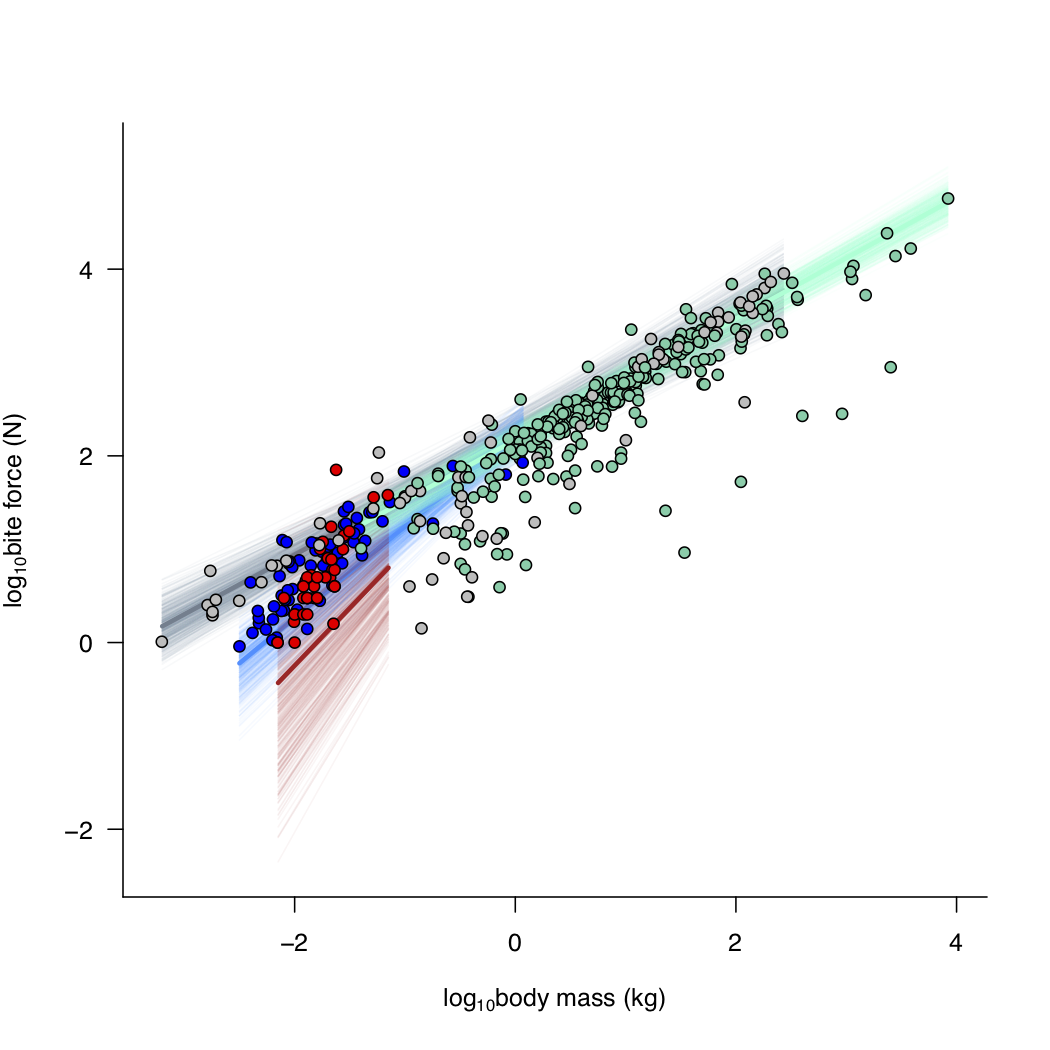
**Figure S5. Relationship between bite force and body size across phylogeny and through time**

The evolution of bite force across phylogeny and through time with respect to body size shows multiple bursts of accelerated changes (non-black branches). Evolutionary rates are as in Figs 1 and 2 and are shown along a color gradient between green (low) and gold (exceptional). Branches with rates indistinguishable from the background rate are in black. Skull drawings with reconstructed jaw muscles represent some clades of interest: Darwin’s finches (*Geospiza magnirostris*) and the homin lineage (*Homo sapiens*). *Plateosaurus* (*P.*) and *Stegosaurus* (*S.*) are associated with one of the highest rates across the tree, mostly owing to their disproportionately small heads given their large body sizes (see text for discussion). A two-dimensional projection (conventional bivariate phylomorphospace [92]) is cast as a shadow.



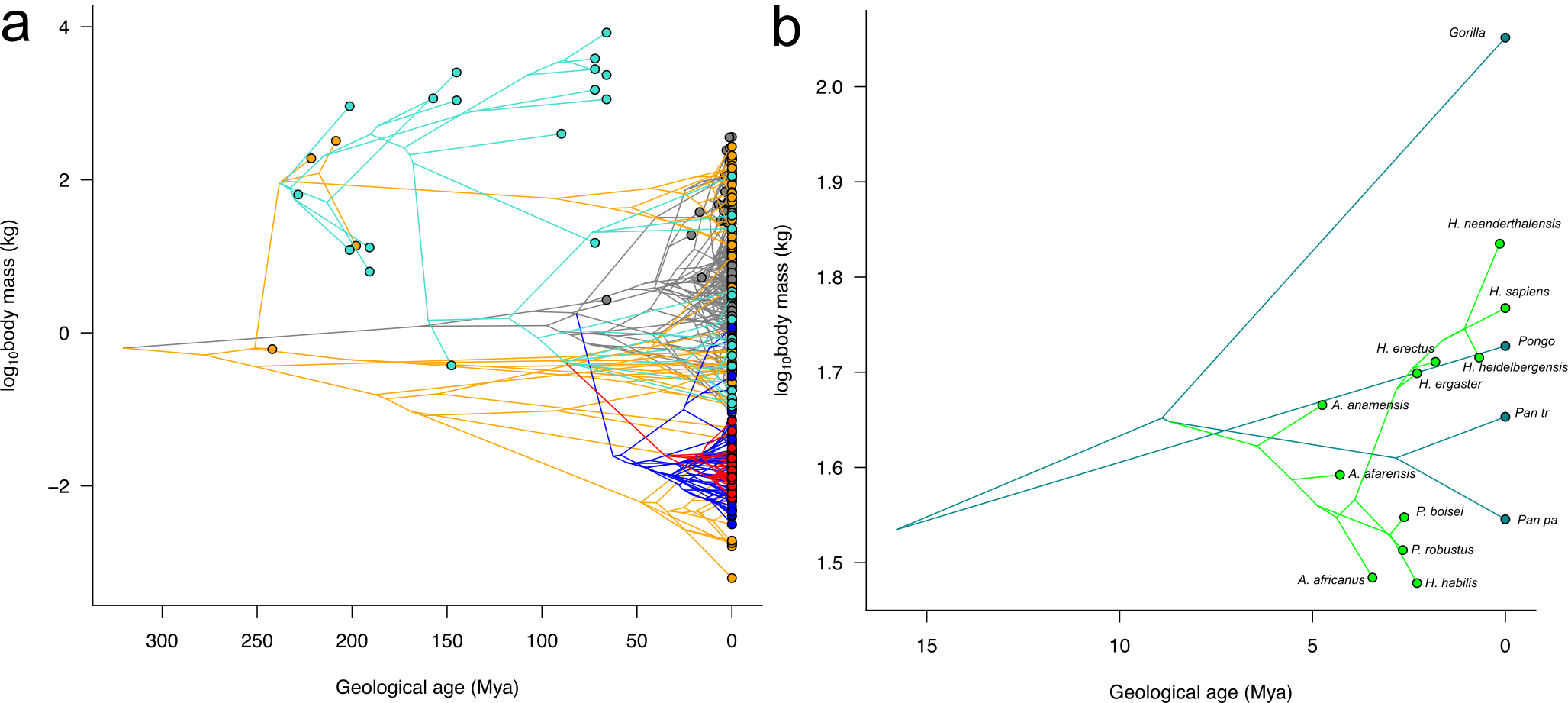
**Figure S6. Difference in slopes between biting positions**

To test whether different biting positions affect the scaling relationship between bite force and body mass, a variable-rates regression model of the form log10*F*Bite ~ BitePoint + log10*M*Body + log10*M*Body:BitePoint + log10*M*Body:Bats + log10*M*Body:Finches, where BitePoint is a categorical variable on biting position (anterior or posterior), Bats and Finches are categorical variables representing group memberships respectively, was fitted. Interaction terms between log10*M*Body and categorical variables are the separate slope terms. Intercept differences for Bats and Finches are not modeled (see text). Since both Bats and Finches comprise entirely of anterior bite forces, while having significantly different slopes from the other groups (Mammals, Dinosaurs, and Reptiles) we mainly tested the effects of bite point on the entirety of the non-chiropteran, non-finch amniotes by estimating separate slopes for anterior and posterior biting positions. We did not further test differences within the three groups – e.g. Dinosaur:BitePoint – owing to small sample sizes in certain categories. There is no significant difference in slopes between the anterior and posterior biting positions (*p*MCMC = 0.473), but there is a significant difference in intercepts between the two biting position categories (*p*MCMC < 0.05). Colors denote different groups and biting positions: red, finches; blue, bats; grey, anterior biting position; and aquamarine, posterior biting position.



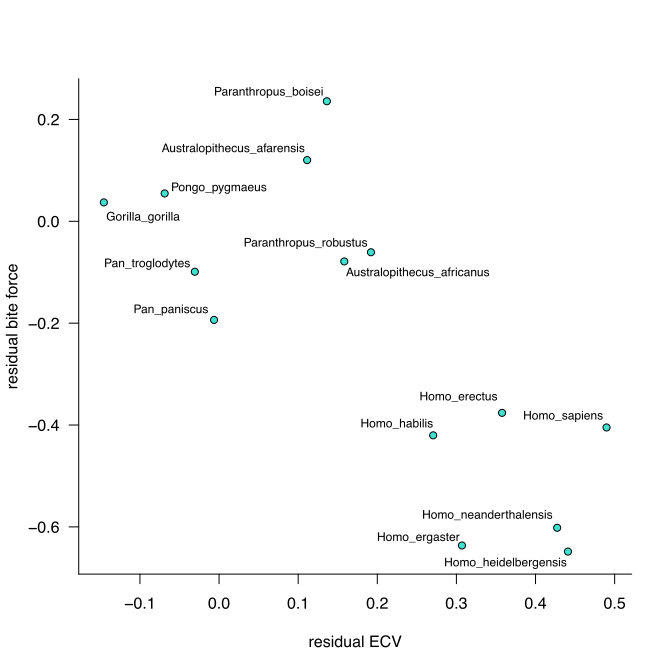
**Figure S7. Difference in slopes between bite force types**

To test whether different bite force types (*in vivo* measurements or indirect estimates) affect the scaling relationship between bite force and body mass, a variable-rates regression model of the form log10*F*Bite ~ BiteType + log10*M*Body + log10*M*Body:BiteType + log10*M*Body:Bats + log10*M*Body:Finches, where BiteType is a categorical variable on bite force type (*in vivo* or estimate), Bats and Finches are categorical variables representing group memberships respectively, was fitted. Interaction terms between log10*M*Body and categorical variables are the separate slope terms. Intercept differences for Bats and Finches are not modeled (see text). Since both Bats and Finches comprise entirely of *in vivo* bite forces, while having significantly different slopes from the other groups (Mammals, Dinosaurs, and Reptiles) we mainly tested the effects of bite force type on the entirety of the non-chiropteran, non-finch amniotes by estimating separate slopes for *in vivo* bite force measurements and indirect bite force estimates. We did not further test differences within the three groups – e.g. Dinosaur:BiteType – owing to small sample sizes in certain categories. There is no significant difference in slopes (*p*MCMC = 0.42), nor is there a significant difference in intercepts between the two bite force categories (*p*MCMC = 0.358). Colors denote different groups and bite force types: red, finches; blue, bats; grey, *in vivo* bite force; and aquamarine, estimated bite force.



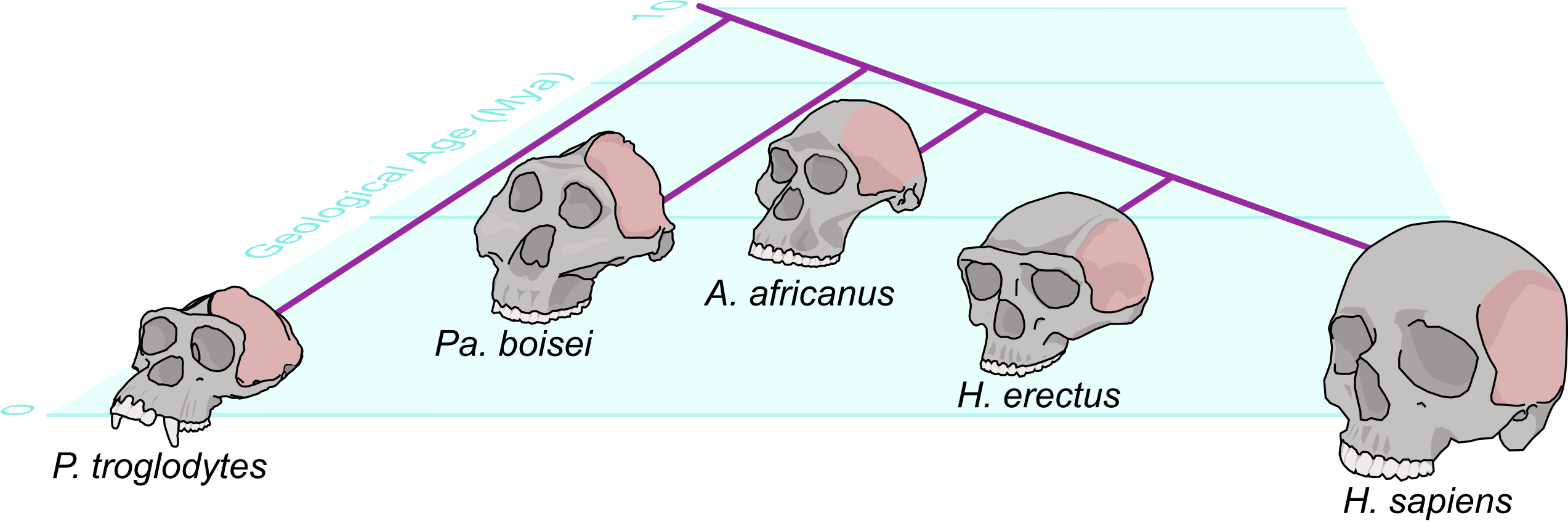
**Figure S8. Ancestral reconstruction of body size across phylogeny and through time**

Evolution of body size while accounting for variable rates is shown to compare with the evolution of bite force (Fig 2). (a) Overall, bite force evolution (Fig 2) exhibits similar patterns to body size evolution, demonstrating the effect of scale on bite force evolution. (b) However, evolution of body size in hominins (bright green) is inconsistent with evolution of bite force (Fig 2c inset), where body size in *Homo* follows a positive directional evolutionary trend while bite force shows the opposite.

****

**Figure S9. The relationship between bite force and brain size in hominids after accounting for body size**

The relationship between bite force and ECV using the residuals of a phylogenetic multiple regression model between log10*F*Bite against log10ECV and log10*M*Body. Results show that as expected, bite force has a positive relationship with body size (slope = 1.35, *p*one-tailed = 0.0459), and a negative relationship with ECV (slope = -1.49, *p*one-tailed = 0.0346). This indicates that after accounting for scaling effects owing to body size, bite force does indeed scale negatively with brain enlargement in hominids.

****

**Figure S10. Reduction of the jaw closing muscles in hominins**

The cranial attachments for the temporal muscles migrate ventrally through hominin evolution. The chimpanzee and the early human *Paranthropus*, show strong sagittal crests along the dorsal midline of the skull, which would have served as anchoring sites for large, well-developed temporal muscles. *Australopithecus* exhibits an intermediate stage with a weaker sagittal crest but muscle attachments extending to the skull midline postero-dorsally. The temporal muscles in the *Homo* species – e.g. *H. erectus* and *H. sapiens* – do not extend more dorsally than the lateral surfaces of the cranium and are far from joining at the cranial midline. Reduction in bite force – as a necessary consequence of smaller temporal muscles – correspond strongly with brain enlargement (see Fig. S9), indicating that a trade-off between bite force and brain size is highly likely in human evolution.

**TABLES**

**Table S2. Statistical summaries of VR and ER regression models**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Single-slope VR model** | |  |  | **Single-slope ER model** | |  |
| *Lh* | 107.551 |  |  | *Lh* | -208.530 |  |
| *R*2 | 0.788 |  |  | *R*2 | 0.438 |  |
| **Coefficient** | **Estimate** | ***p*MCMC** |  | **Coefficient** | **Estimate** | ***p*MCMC** |
| Intercept | 2.171 | 0.000\* |  | Intercept | 2.183 | 0.000\* |
| *M*Body | 0.674 | 0.000\* |  | *M*Body | 0.722 | 0.000\* |
| **5-Group VR model** | |  |  | **5-Group ER Model** | |  |
| *Lh* | 127.790 |  |  | *Lh* | -173.351 |  |
| *R*2 | 0.795 |  |  | *R*2 | 0.518 |  |
| **Coefficient** | **Estimate** | ***p*MCMC** |  | **Coefficient** | **Estimate** | ***p*MCMC** |
| Intercept | 2.195 | 0.000\* |  | Intercept | 2.218 | 0.000\* |
| *M*Body | 0.635 | 0.000\* |  | *M*Body | 0.621 | 0.000\* |
| *M*Body\_Bats | 0.330 | 0.000\* |  | *M*Body\_Bats | 0.342 | 0.008\* |
| *M*Body\_Dinosaurs | 0.087 | 0.170 |  | *M*Body\_Dinosaurs | 0.002 | 0.468 |
| *M*Body\_Finches | 0.618 | 0.021\* |  | *M*Body\_Finches | 1.488 | 0.000\* |
| *M*Body\_Reptiles | 0.059 | 0.144 |  | *M*Body\_Reptiles | 0.067 | 0.277 |
| **5-Group+BitePoint VR model** | | |  | **5-Group+BitePoint ER model** | | |
| *Lh* | 127.381 |  |  | *Lh* | -172.590 |  |
| *R*2 | 0.809 |  |  | *R*2 | 0.519 |  |
| **Coefficient** | **Estimate** | ***p*MCMC** |  | **Coefficient** | **Estimate** | ***p*MCMC** |
| Intercept | 2.087 | 0.000\* |  | Intercept | 2.078 | 0.000\* |
| BitePoint | 0.173 | 0.000\* |  | BitePoint | 0.170 | 0.106 |
| *M*Body | 0.650 | 0.000\* |  | *M*Body | 0.628 | 0.000\* |
| *M*Body\_Bats | 0.305 | 0.001\* |  | *M*Body\_Bats | 0.321 | 0.018\* |
| *M*Body\_Dinosaurs | 0.064 | 0.241 |  | *M*Body\_Dinosaurs | -0.006 | 0.484 |
| *M*Body\_Finches | 0.613 | 0.018\* |  | *M*Body\_Finches | 1.455 | 0.000\* |
| *M*Body\_Reptiles | 0.035 | 0.248 |  | *M*Body\_Reptiles | 0.050 | 0.317 |
| **5-Group+BiteType VR model** | | |  | **5-Group+BiteType ER model** | | |
| *Lh* | 125.242 |  |  | *Lh* | -173.472 |  |
| *R*2 | 0.794 |  |  | *R*2 | 0.517 |  |
| **Coefficient** | **Estimate** | ***p*MCMC** |  | **Coefficient** | **Estimate** | ***p*MCMC** |
| Intercept | 2.205 | 0.000\* |  | Intercept | 2.264 | 0.000\* |
| BiteType | -0.050 | 0.394 |  | BiteType | -0.061 | 0.348 |
| *M*Body | 0.637 | 0.000\* |  | *M*Body | 0.620 | 0.000\* |
| *M*Body\_Bats | 0.331 | 0.000\* |  | *M*Body\_Bats | 0.355 | 0.013\* |
| *M*Body\_Dinosaurs | 0.097 | 0.151 |  | *M*Body\_Dinosaurs | 0.011 | 0.442 |
| *M*Body\_Finches | 0.591 | 0.021\* |  | *M*Body\_Finches | 1.489 | 0.000\* |
| *M*Body\_Reptiles | 0.056 | 0.136 |  | *M*Body\_Reptiles | 0.066 | 0.276 |

\*Significant: *p*MCMC < 0.05

**Table S3. Pair-wise differences in slopes amongst the five groups (*p*MCMC)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Mammals | | Bats | Dinosaurs | Finches | Reptiles |
| Mammal | | - | 0.001\* | 0.241 | 0.018 | 0.248 |
| Bats | |  | - | 0.027\* | 0.192 | 0.004\* |
| Dinosaurs | |  |  | - | 0.036\* | 0.352 |
| Finches | |  |  |  | - | 0.027\* |
| Reptiles | |  |  |  |  | - |

\*Significant: *p*MCMC < 0.05

**Table S4. Differences in group-wise slopes from a theoretical scaling factor of 0.67**

|  |  |  |
| --- | --- | --- |
| Group | Slope | *p*MCMC0.67 |
| Finches | 1.260 | 0.021\* |
| Bats | 0.954 | 0.001\* |
| Mammals | 0.650 | 0.183 |
| Dinosaurs | 0.714 | 0.298 |
| Reptiles | 0.685 | 0.353 |

\*Significant: *p*MCMC < 0.05

**Movie S1. Rate of bite force evolution across phylogeny and through time with respect to body mass**

The evolution of bite force in relation to body mass is shown as a phylomorphospace [92] through time (“chronophylomorphospace” [14]) with branches colored according to exceptional rate increases.

**Movie S2. Group-wise bite force evolution across phylogeny and through time with respect to body mass**

The evolution of bite force in relation to body mass is shown as a phylomorphospace [92] through time (“chronophylomorphospace” [14]) with branches colored according to the five groups.

**SUPPLEMENTARY REFERENCES**

[1] Wroe, S., McHenry, C. & Thomason, J. 2005 Bite club: comparative bite force in big biting mammals and the prediction of predatory behaviour in fossil taxa. *Proceedings of the Royal Society B-Biological Sciences* **272**, 619-625.

[2] Christiansen, P. & Wroe, S. 2007 Bite forces and evolutionary adaptations to feeding ecology in carnivores. *Ecology* **88**, 347-358.

[3] Thomason, J.J. 1991 Cranial Strength in Relation to Estimated Biting Forces in Some Mammals. *Canadian Journal of Zoology-Revue Canadienne De Zoologie* **69**, 2326-2333.

[4] Demes, B. & Creel, N. 1988 Bite force, diet, and cranial morphology of fossil hominids. *Journal of Human Evolution* **17**, 657-670. (doi:http://dx.doi.org/10.1016/0047-2484(88)90023-1).

[5] Eng, C.M., Lieberman, D.E., Zink, K.D. & Peters, M.A. 2013 Bite force and occlusal stress production in hominin evolution. *American Journal of Physical Anthropology* **151**, 544-557. (doi:10.1002/ajpa.22296).

[6] Perry, J.M.G., Hartstone-Rose, A. & Logan, R.L. 2011 *The Jaw Adductor Resultant and Estimated Bite Force in Primates*11 p.

[7] Mazzetta, G.V., Cisilino, A.P., Blanco, R.E. & Calvo, N. 2009 Cranial Mechanics and Functional Interpretation of the Horned Carnivorous Dinosaur Carnotaurus Sastrei. *Journal of Vertebrate Paleontology* **29**, 822--830.

[8] Reichel, M. 2010 A model for the bite mechanics in the herbivorous dinosaur Stegosaurus (Ornithischia, Stegosauridae). *Swiss Journal of Geosciences* **103**, 235-240. (doi:10.1007/s00015-010-0025-1).

[9] Lautenschlager, S. 2013 Cranial myology and bite force performance of Erlikosaurus andrewsi: a novel approach for digital muscle reconstructions. *Journal of Anatomy* **222**, 260-272. (doi:10.1111/joa.12000).

[10] Lautenschlager, S., Brassey, C.A., Button, D.J. & Barrett, P.M. 2016 Decoupled form and function in disparate herbivorous dinosaur clades. *Sci Rep-Uk* **6**, 26495. (doi:10.1038/srep26495

https://www.nature.com/articles/srep26495#supplementary-information).

[11] Thomason, J.J., Russell, A.P. & Morgeli, M. 1990 Forces of biting, body size, and masticatory muscle tension in the opossum Didelphis virginiana. *Canadian Journal of Zoology-Revue Canadienne De Zoologie* **68**, 318-324.

[12] Christiansen, P. & Adolfssen, J.S. 2005 Bite forces, canine strength and skull allometry in carnivores (Mammalia, Carnivora). *Journal of Zoology* **266**, 133-151.

[13] Sakamoto, M., Lloyd, G.T. & Benton, M.J. 2010 Phylogenetically structured variance in felid bite force: the role of phylogeny in the evolution of biting performance. *Journal of Evolutionary Biology* **23**, 463-478. (doi:10.1111/j.1420-9101.2009.01922.x).

[14] Sakamoto, M. & Ruta, M. 2012 Convergence and divergence in the evolution of cat skulls: Temporal and spatial patterns of morphological diversity. *Plos One* **7**, e39752.

[15] Sakamoto, M. 2010 Jaw biomechanics and the evolution of biting performance in theropod dinosaurs. *Proceedings of the Royal Society B: Biological Sciences* **277**, 3327-3333.

[16] Rayfield, E.J., Norman, D.B., Horner, C.C., Horner, J.R., Smith, P.M., Thomason, J.J. & Upchurch, P. 2001 Cranial design and function in a large theropod dinosaur. *Nature* **409**, 1033-1037.

[17] Holliday, C.M. & Witmer, L.M. 2007 Archosaur adductor chamber evolution: Integration of musculoskeletal and topological criteria in jaw muscle homology. *Journal of Morphology* **268**, 457-484.

[18] Sinclair, A.G. & Alexander, R.M. 1987 Estimates of Forces Exerted by the Jaw Muscles of Some Reptiles. *Journal of Zoology* **213**, 107-115.

[19] Biewener, A.A. & Full, R.J. 1992 Force platform and kinematic analysis. In *Biomechanics - Structures and Systems* (ed. A.A. Biewener), pp. 45-73, Oxford University Press.

[20] Rasband, W.S. 2012 ImageJ. (1.43U ed. Bethesda, Maryland, U. S. National Institute of Health.

[21] Sacks, R.D. & Roy, R.R. 1982 Architecture of the Hindlimb Muscles of Cats - Functional-Significance. *Journal of Morphology* **173**, 185-195.

[22] Burkholder, T.J., Fingado, B., Baron, S. & Lieber, R.L. 1994 Relationship between Muscle-Fiber Types and Sizes and Muscle Architectural Properties in the Mouse Hindlimb. *Journal of Morphology* **221**, 177-190.

[23] Erickson, G.M., Lappin, A.K. & Vliet, K.A. 2003 The ontogeny of bite-force performance in American alligator (Alligator mississippiensis). *Journal of Zoology* **260**, 317-327.

[24] Erickson, G.M., Lappin, A.K., Parker, T. & Vliet, K.A. 2004 Comparison of bite-force performance between long-term captive and wild American alligators (Alligator mississippiensis). *Journal of Zoology* **262**, 21-28.

[25] Sustaita, D. & Hertel, F. 2010 *In vivo* bite and grip forces, morphology and prey-killing behavior of North American accipiters (Accipitridae) and falcons (Falconidae). *The Journal of Experimental Biology* **213**, 2617-2628.

[26] Gussekloo, S.W.S. & Bout, R.G. 2005 Cranial kinesis in palaeognathous birds. *Journal of Experimental Biology* **208**, 3409-3419.

[27] Bates, K.T. & Falkingham, P.L. 2012 Estimating maximum bite performance in *Tyrannosaurus rex* using multi-body dynamics. *Biology Letters* **8**, 660-664.

[28] Gignac, P.M. & Erickson, G.M. 2017 The Biomechanics Behind Extreme Osteophagy in Tyrannosaurus rex. *Sci Rep-Uk* **7**, 2012. (doi:10.1038/s41598-017-02161-w).

[29] Erickson, G.M., VanKirk, S.D., Su, J.T., Levenston, M.E., Caler, W.E. & Carter, D.R. 1996 Bite-force estimation for Tyrannosaurus rex from tooth-marked bones. *Nature* **382**, 706-708.

[30] Molnar, R.E. 1998 Mechanical factors in the design of the skull of Tyrannosaurus rex (Osborn, 1905). *Gaia* **15**, 193-218.

[31] Meers, M.B. 2002 Maximum bite force and prey size of Tyrannosaurus rex and their relationships to the inference of feeding behaviour. *Historical Biology* **16**, 1-12.

[32] Rayfield, E.J. 2004 Cranial mechanics and feeding in Tyrannosaurus rex. *Proceedings of The Royal Society of London Series B-Biological Sciences* **271**, 1451-1459.

[33] Binder, W.J. & Van Valkenburgh, B. 2000 Development of bite strength and feeding behaviour in juvenile spotted hyenas (Crocuta crocuta). *Journal of Zoology* **252**, 273-283.

[34] van der Bilt, A., Tekamp, A., van der Glas, H. & Abbink, J. 2008 Bite force and electromyograpy during maximum unilateral and bilateral clenching. *European Journal of Oral Sciences* **116**, 217-222.

[35] Bakker, R.T. 1998 Brontosaur killers: Late Jurassic allosaurids as sabre-tooth cat analogues. *Gaia* **15**, 145-158.

[36] Henderson, D.M. 2002 The eyes have it: The sizes, shapes, and orientations of theropod orbits as indicators of skull strength and bite force. *Journal of Vertebrate Paleontology* **22**, 766-778.

[37] Rayfield, E.J. 2005 Aspects of comparative cranial mechanics in the theropod dinosaurs Coelophysis, Allosaurus and Tyrannosaurus. *Zoological Journal of The Linnean Society* **144**, 309-316.

[38] Therrien, F., Henderson, D.M. & Ruff, C.B. 2005 Bite me: biomechanical models of theropod mandibles and implications for feeding behaviour. In *The Carnivorous Dinosaurs* (ed. K. Carpenter), pp. 179-237, Indiana University Press.

[39] Gignac, P.M., Makovicky, P.J., Erickson, G.M. & Walsh, R.P. 2010 A description of Deinonychus antirrhopus bite marks and estimates of bite force using tooth indentation simulations. *Journal of Vertebrate Paleontology* **30**, 1169-1177. (doi:Doi 10.1080/02724634.2010.483535

Pii 924297860).

[40] Erickson, G.M. & Olson, K.H. 1996 Bite marks attributable to Tyrannosaurus rex: Preliminary description and implications. *Journal of Vertebrate Paleontology* **16**, 175-178.

[41] D'Amore, D.C. & Blumenschine, R.J. 2009 Komodo monitor (Varanus komodoensis) feeding behavior and dental function reflected through tooth marks on bone surfaces, and the application to ziphodont paleobiology. *Paleobiology* **35**, 525-552. (doi:10.1666/0094-8373-35.4.525).

[42] Snively, E., Cotton, J.R., Ridgely, R. & Witmer, L.M. 2013 Multibody dynamics model of head and neck function in Allosaurus (Dinosauria, Theropoda). *Palaeontologia Electronica* **16**.

[43] Dunning, J.B. 2007 CRC Handbook of Avian Body Masses.

[44] Jones, K.E., Bielby, J., Cardillo, M., Fritz, S.A., O'Dell, J., Orme, C.D.L., Safi, K., Sechrest, W., Boakes, E.H., Carbone, C., et al. 2009 PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology* **90**, 2648-2648. (doi:10.1890/08-1494.1).

[45] Henderson, D.M. & Snively, E. 2004 Tyrannosaurus en pointe: allometry minimized rotational inertia of large carnivorous dinosaurs. *Proceedings Of The Royal Society Of London Series B-Biological Sciences* **271**, S57-S60.

[46] Van Valkenburgh, B. 1990 Skeletal and dental predictors of body mass in carnivores. In *Body Size in Mammalian Paleobiology: Estimation and Biological Implications* (eds. J. Damuth & B.J. MacFadden), pp. 181-205, Cambridge University Press.

[47] Law, C.J., Slater, G.J. & Mehta, R.S. 2018 Lineage Diversity and Size Disparity in Musteloidea: Testing Patterns of Adaptive Radiation Using Molecular and Fossil-Based Methods. *Systematic Biology* **67**, 127-144. (doi:10.1093/sysbio/syx047).

[48] Kilbourne, B.M. & Hoffman, L.C. 2015 Energetic benefits and adaptations in mammalian limbs: Scale effects and selective pressures. *Evolution* **69**, 1546-1559. (doi:10.1111/evo.12675).

[49] Hedges, S.B., Marin, J., Suleski, M., Paymer, M. & Kumar, S. 2015 Tree of Life Reveals Clock-Like Speciation and Diversification. *Molecular Biology and Evolution* **32**, 835-845. (doi:10.1093/molbev/msv037).

[50] Muirhead, J. & Archer, M. 1990 Nimbacinus dicksoni, a plesiomorphic thylacine (Marsupialia: Thylacinidae) from Tertiary deposits of Queensland and the Northern Territory. *Memoirs of the Queensland Museum* **28**, 203-221.

[51] Wilson, G.P., Ekdale, E.G., Hoganson, J.W., Calede, J.J. & Vander Linden, A. 2016 A large carnivorous mammal from the Late Cretaceous and the North American origin of marsupials. *Nat Commun* **7**, 13734. (doi:10.1038/ncomms13734

http://www.nature.com/articles/ncomms13734#supplementary-information).

[52] Nesbitt, S.J. 2011 The early evolution of archosaurs: Relationships and the origin of major clades. *Bulletin of the American Museum of Natural History*, 1-288.

[53] Dembo, M., Matzke, N.J., Mooers, A.Ø. & Collard, M. 2015 Bayesian analysis of a morphological supermatrix sheds light on controversial fossil hominin relationships. *Proceedings of the Royal Society B: Biological Sciences* **282**.

[54] Benson, R.B., Campione, N.E., Carrano, M.T., Mannion, P.D., Sullivan, C., Upchurch, P. & Evans, D.C. 2014 Rates of dinosaur body mass evolution indicate 170 million years of sustained ecological innovation on the avian stem lineage. *PLoS Biol* **12**, e1001853. (doi:10.1371/journal.pbio.1001853).

[55] Johnson, W.E., Eizirik, E., Pecon-Slattery, J., Murphy, W.J., Antunes, A., Teeling, E. & O'Brien, S.J. 2006 The Late Miocene radiation of modern Felidae: A genetic assessment. *Science* **311**, 73-77.

[56] Werdelin, L., Yamaguchi, N., Johnson, W.E. & O'Brien, S.J. 2010 Phylogeny and evolution of cats (Felidae). In *Biology and conservation of wild felids* (eds. D.W. Macdonald & A.J. Loveridge), pp. 59-82. Oxford, Oxford University Press.

[57] Brusatte, S.L. & Sereno, P.C. 2008 Phylogeny of Allosauroidea (Dinosauria : Theropoda): Comparative analysis and resolution. *Journal Of Systematic Palaeontology* **6**, 155--182.

[58] Bapst, D.W. 2012 paleotree: an R package for paleontological and phylogenetic analyses of evolution. *Methods in Ecology and Evolution* **3**, 803-807. (doi:10.1111/j.2041-210X.2012.00223.x).

[59] Aguirre, L.F., Herrel, A., van Damme, R. & Matthysen, E. 2002 Ecomorphological analysis of trophic niche partitioning in a tropical savannah bat community. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **269**, 1271-1278.

[60] van der Meij, M.A.A. & Bout, R.G. 2004 Scaling of jaw muscle size and maximal bite force in finches. *Journal Of Experimental Biology* **207**, 2745-2753.

[61] Schmidt-Nielsen, K. 1984 *Scaling: Why is animal size so important?* Cambridge, Cambridge University Press.

[62] Brusatte, S.L., Lloyd, G.T., Wang, S.C. & Norell, M.A. 2014 Gradual assembly of avian body plan culminated in rapid rates of evolution across the dinosaur-bird transition. *Current Biology* **24**, 2386-2392. (doi:10.1016/j.cub.2014.08.034).

[63] Lee, M.S.Y., Cau, A., Naish, D. & Dyke, G.J. 2014 Sustained miniaturization and anatomical innovation in the dinosaurian ancestors of birds. *Science* **345**, 562-566. (doi:10.1126/science.1252243).

[64] Puttick, M.N., Thomas, G.H. & Benton, M.J. 2014 High rates of evolution preceded the origin of birds. *Evolution* **68**, 1497-1510. (doi:10.1111/evo.12363).

[65] Brusatte, Stephen L., O’Connor, Jingmai K. & Jarvis, Erich D. 2015 The Origin and Diversification of Birds. *Current Biology* **25**, R888-R898. (doi:https://doi.org/10.1016/j.cub.2015.08.003).

[66] Brusatte, S.L., Nesbitt, S.J., Irmis, R.B., Butler, R.J., Benton, M.J. & Norell, M.A. 2010 The origin and early radiation of dinosaurs. *Earth-Science Reviews* **101**, 68-100. (doi:10.1016/j.earscirev.2010.04.001).

[67] Eastman, J.M., Alfaro, M.E., Joyce, P., Hipp, A.L. & Harmon, L.J. 2011 A Novel Comparative Method for Identifying Shifts in the Rate of Character Evolution on Trees. *Evolution* **65**, 3578-3589. (doi:10.1111/j.1558-5646.2011.01401.x).

[68] Rabosky, D.L. 2014 Automatic Detection of Key Innovations, Rate Shifts, and Diversity-Dependence on Phylogenetic Trees. *Plos One* **9**, e89543. (doi:10.1371/journal.pone.0089543).

[69] Baker, J., Meade, A., Pagel, M. & Venditti, C. 2016 Positive phenotypic selection inferred from phylogenies. *Biological Journal of the Linnean Society* **118**, 95-115. (doi:10.1111/bij.12649).

[70] Venditti, C., Meade, A. & Pagel, M. 2011 Multiple routes to mammalian diversity. *Nature* **479**, 393-396. (doi:10.1038/nature10516).

[71] Pagel, M. 1999 Inferring the historical patterns of biological evolution. *Nature* **401**, 877-884.

[72] Butler, M.A. & King, A.A. 2004 Phylogenetic comparative analysis: A modeling approach for adaptive evolution. *American Naturalist* **164**, 683-695. (doi:10.1086/426002).

[73] Raftery, A.E. 1996 Hypothesis testing and model selection. In *Markov Chain Monte Carlo in Practice* (eds. W.R. Gilks, S. Richardson & D.J. Spiegelhalter), pp. 163-187. London, Great Britain, Chapman & Hall.

[74] R Core Development Team. 2011 R: A language and environment for statistical computing. (2.10.1 ed. Vienna, Austria, R Foundation for Statistical Computing.

[75] Simpson, G.G. 1944 *Tempo and mode in evolution*. New York, Columbia University Press; 237 p.

[76] Dawkins, R. & Krebs, J.R. 1979 Arms races between and within species. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **205**, 489-511. (doi:10.1098/rspb.1979.0081).

[77] Pagel, M. 1994 The adaptationist wager. In *Phylogenetics and Ecology* (eds. P. Eggleton & R.I. Vane-Wright), pp. 29-51. London, Academic Press.

[78] Benton, M.J. 2009 The Red Queen and the Court Jester: Species Diversity and the Role of Biotic and Abiotic Factors Through Time. *Science* **323**, 728-732. (doi:10.1126/science.1157719).

[79] Baker, J., Meade, A., Pagel, M. & Venditti, C. 2015 Adaptive evolution toward larger size in mammals. *Proceedings of the National Academy of Sciences* **112**, 5093-5098. (doi:10.1073/pnas.1419823112).

[80] Voje, K.L., Starrfelt, J. & Liow, L.H. 2018 Model Adequacy and Microevolutionary Explanations for Stasis in the Fossil Record. *The American Naturalist* **191**, 509-523. (doi:10.1086/696265).

[81] Yang, Z. 2002 Inference of selection from multiple species alignments. *Current Opinion in Genetics & Development* **12**, 688-694. (doi:http://dx.doi.org/10.1016/S0959-437X(02)00348-9).

[82] Yang, Z. 2006 *Computational Molecular Evolution*. Oxford, Oxford University Press; 376 p.

[83] Revell, L.J. 2012 phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**, 217-223. (doi:10.1111/j.2041-210X.2011.00169.x).

[84] DeSilva, J.M. & Lesnik, J.J. 2008 Brain size at birth throughout human evolution: A new method for estimating neonatal brain size in hominins. *Journal of Human Evolution* **55**, 1064-1074. (doi:http://dx.doi.org/10.1016/j.jhevol.2008.07.008).

[85] Isler, K., Christopher Kirk, E., Miller, J.M.A., Albrecht, G.A., Gelvin, B.R. & Martin, R.D. 2008 Endocranial volumes of primate species: scaling analyses using a comprehensive and reliable data set. *Journal of Human Evolution* **55**, 967-978. (doi:http://dx.doi.org/10.1016/j.jhevol.2008.08.004).

[86] Rightmire, G.P. 2004 Brain size and encephalization in early to Mid-Pleistocene Homo. *American Journal of Physical Anthropology* **124**, 109-123. (doi:10.1002/ajpa.10346).

[87] Orme, D., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Isaac, N. & Pearse, W. 2013 caper: Comparative Analyses of Phylogenetics and Evolution in R. R package version 0.5.2. (R package version 0.5.2 ed.

[88] Huber, D.R., Eason, T.G., Hueter, R.E. & Motta, P.J. 2005 Analysis of the bite force and mechanical design of the feeding mechanism of the durophagous horn shark Heterodontus francisci. *Journal of Experimental Biology* **208**, 3553-3571.

[89] Erickson, G.M., Gignac, P.M., Steppan, S.J., Lappin, A.K., Vliet, K.A., Brueggen, J.D., Inouye, B.D., Kledzik, D. & Webb, G.J.W. 2012 Insights into the Ecology and Evolutionary Success of Crocodilians Revealed through Bite-Force and Tooth-Pressure Experimentation. *Plos One* **7**, e31781. (doi:10.1371/journal.pone.0031781).

[90] Herrel, A., Podos, J., Huber, S.K. & Hendry, A.P. 2005 Bite performance and morphology in a population of Darwin's finches: implications for the evolution of beak shape. *Functional Ecology* **19**, 43-48.

[91] Herrel, A., Podos, J., Huber, S.K. & Hendry, A.P. 2005 Evolution of bite force in Darwin's finches: a key role for head width. *Journal of Evolutionary Biology* **18**, 669-675.

[92] Sidlauskas, B. 2008 Continuous and arrested morphological diversification in sister clades of characiform fishes: a phylomorphospace approach. *Evolution* **62**, 3135--3156.