UNIVERSITÀ DEGLI STUDI DI NAPOLI

FEDERICO II

**DIPARTIMENTO DI SCIENZE DELLA TERRA,
DELL'AMBIENTE E DELLE RISORSE**

DOTTORATO IN

"SCIENZE DELLA TERRA, DELL'AMBIENTE E

DELLE RISORSE"

XXXIII CICLO

DIVERSIFICATION TRENDS INFERRED FROM THE FOSSIL RECORD

COORDINATORE

Ch.mo Prof. Maurizio Fedi

CANDIDATA

Dott.ssa Silvia Castiglione

TUTOR

Ch.mo Prof. Pasquale Raia

CO-TUTOR

Ch.mo Prof. Filippo Barattolo

Anno Accademico 2020-2021

SUMMARY

INTRODUCTION AND AIM OF THE PROJECT

Macroevolution focuses on patterns and processes occurring above the level of species and over geological timescales (Raia, 2016). Investigating diversification processes, both morphological and taxonomical, gives a chance to answer important questions in evolutionary biology. Why do some clades have more species than others? Why do some groups undergo striking adaptive radiations, but others persist for millions of years as living fossils? Why do some groups have much more ecological or morphological diversity than others? Does anything limit the number of species on earth and what? These complex questions share a common underlying feature: all, to some degree, concern rates of macroevolutionary change that occur across geological timescales (Rabosky & Slater, 2014).

The aim of my project was to produce a coherent array of new methods to investigate phenotypic and taxonomic diversification by using phylogenies including extinct species. I started by developing *RRphylo*, a new phylogenetic comparative method based on phylogenetic ridge regression, which works with a phylogenetic tree and phenotypic data (univariate or multivariate either) to estimate branch-wise rates of phenotypic evolution and ancestral characters simultaneously. The main innovations, which translate in advantages of *RRphylo* over existing methods, lies in the absence of any *a priori* hypothesis about the mode of phenotypic evolution and in its ability to deal with fossil phylogenies. Both these factors make *RRphylo* very suitable to study phenotypic evolution in its different facets.

I further implemented *RRphylo* to locate clade- or state-related shifts in absolute rates of phenotypic evolution, to integrate the effect of additional (to the phylogeny) predictors on rates estimation, to identify temporal trends in phenotypic mean and evolutionary rates occurring on the entire tree or pertaining individual clades, to identify instances of morphological convergence, to include ancestral character information derived from the fossil record, and to work with discrete variables.

All these tools are collected into the `RRphylo` R package, online from April 2018, and counting > 14000 downloads on CRAN to date. I have been handling the maintenance and updates of the `RRphylo` package for both the release (https://cran.r-project.org/web/packages/RRphylo/index.html) and development (https://github.com/pasraia/RRphylo) versions, and creating/updating explanatory vignettes to facilitate its usage.

COMPARATIVE METHODS AND TRAIT EVOLUTION

The comparison between two phenotypes across a group of species or between a phenotype and an environmental component has long been assessed by the comparative method (Felsenstein, 1985). Such approach looked for possible relation between variables by performing statistical analyses, like regression, correlation, contingency table methods, which typically assume the variables are independent draws derived from the same distribution. When dealing with species-related variables (e.g. phenotypes, physiological limits, ecological preferences), though, this assumption is obviously violated as species are related to each other in a hierarchically structured phylogeny. In fact, it is quite common to observe that close related species resemble each other more than species drawn at random from the tree, thus assuming their independence in statistical tests could lead to overstatement of significance (Felsenstein, 1985).

Phylogenetic comparative methods (PCMs) allow to quantitatively integrate phylogenetic information in evolutionary analyses involving species-related variables. The PCMs framework is based on the phylogenetic variance-covariance matrix (**C**), that is a quantitative representation of phylogenetic relations among species. In **C**, diagonal elements represent evolutionary variance of individual tips, corresponding to the time distance of the tip from the tree root (Fig. 1). Off-diagonal elements represent the covariance, in terms of shared evolutionary time, between tip pairs, corresponding to the time distance from the tree root to the most recent common ancestor (MRCA) of each pair (Fig. 1). The product of **C** times the rate of evolution along the tree branches (σ^2) outputs the evolutionary variance-covariance matrix ($\mathbf{V}=\mathbf{C}*\sigma^2$).

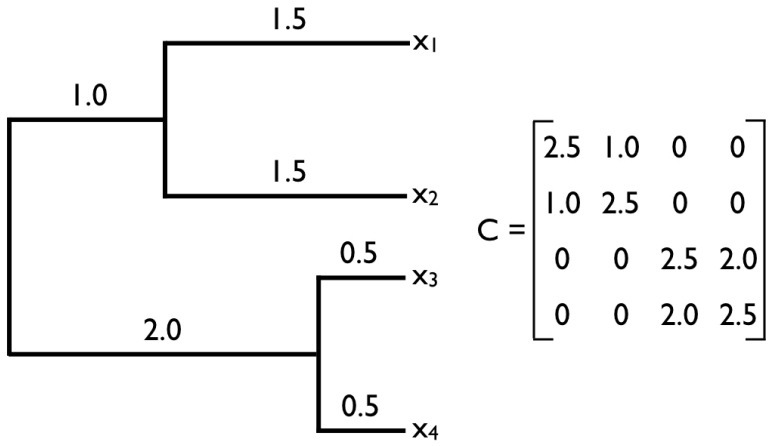

Figure 1. Example of phylogenetic tree and its related variance-covariance matrix. Image from Harmon 2019, available for reuse under Creative Commons Attribution 4.0 International license (https://lukejharmon.github.io/pcm/chapter3_bmintro/#section-3.4-brownian-motion-on-a-phylogenetic-tree).

Traditional PCMs are model-based methods as they consist in fitting different *a priori* evolutionary models to phylogenetic tree and species-related data (from now on I will talk about phenotypic traits, but this equally applies to other species-related variables). Each model is based on different modifications of **C** and has a specific number of parameters to be evaluated by means of maximum likelihood estimation.

Brownian motion (BM) is the basic model in testing likelihood because it approximates the evolution of traits under many different scenarios (Harmon, 2019), and also because it well represents the Darwinian concept of evolution, meant as a gradual increase of phenotypic differences resulting from cumulating genetic mutations in time (Elliot & Mooers, 2014; Pennell et al., 2014). The BM model assumes phenotypic evolution to be a function of time, with the mean of phenotypes remaining constant through it and the variance being directly proportional to σ^2 (Fig. 2), so that at any point in time (t) the distribution of phenotypes will be proportional to y_0+**V**, where y_0 is the phenotype at the tree root and **V** is the evolutionary variance-covariance matrix. At each instant in time the trait values can increase or decrease independently from their current state, describing a "random walk" around the starting value (y_0), but the net change in average phenotype

values is always zero. The only parameters to be estimated are thus y_0 and the $\boldsymbol{\sigma^2}$, which are used to evaluate the distribution of estimated characters at tips.

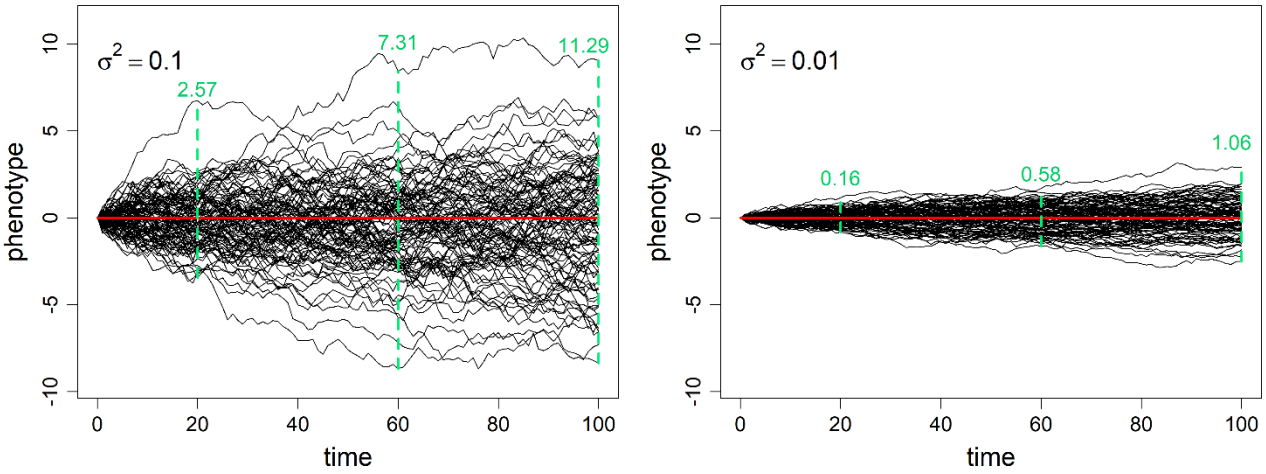

Figure 2. Example of Brownian motion. The plots show 100 Brownian motion simulations spanning the same time intervals (t = 100), with the same mean trait value (= 0; red solid line), and different rate parameters (0.1 on the left, 0.01 on the right). Green dashed lines and related numbers indicate the range of phenotypic values at t = 20, t = 60, and t = 100.

While BM uses the variance-covariance matrix just as it is, more complex models imply modifications of **C** altering the relationships between species. This is accomplished by tree stretching (O'Meara, 2012), which consists in algorithmically transforming a set of branch lengths thanks to new fitted parameters.

Pagel's delta (δ), lambda (λ), and kappa (κ) are the most common stretching frameworks (Pagel, 1999). The δ parameterization fits a model where the rates of evolution exponentially increase or decrease over time. It raises all the elements of **C** to a positive power δ. As the elements in **C** correspond to the heights of nodes in the tree, this transformation actually shifts the position of nodes with respect to the tree root. δ equal to one produce no alteration of the tree, thus corresponding to a time-constant BM model. For δ less than one, branch lengths are unevenly reduced with shallower branches being more altered than deeper ones (Fig. 3 upper left). This moves the nodes to the recent and delays the time of separation between species. Such model

represents a slowdown in the rate of evolution. For δ larger than one, shallower branches are stretched more than deeper ones, pulling the nodes back in time, thus resulting in earlier phenotypic divergence of species (Fig. 3 bottom left). This produces a model where the rate of evolution speeds up through time.

The λ parameter is often used to measure the phylogenetic signal, that is a measure of similarity among species' phenotypes due to their phylogenetic relationships. With the λ transformation the variance covariance matrix is modified by multiplying the off-diagonal elements by an estimated parameter (λ) bounded between zero and one. When λ = 1 the traits covariance remains unchanged, thus the model is equal to BM. As λ approaches zero, the covariance decreases, and species become less related than expected under BM (Fig. 3 upper right).

Pagel's κ is a punctuational model in which phenotypic variation is related to the number of speciation events. It modifies the phylogeny by raising the branch lengths to the estimated value of κ, ranging between zero and one. For κ = 1 the branch lengths do not change, while for κ = 0 all the lengths become one, thus eliminating the effect of time proportionality on phenotypes (Fig. 3 bottom right).

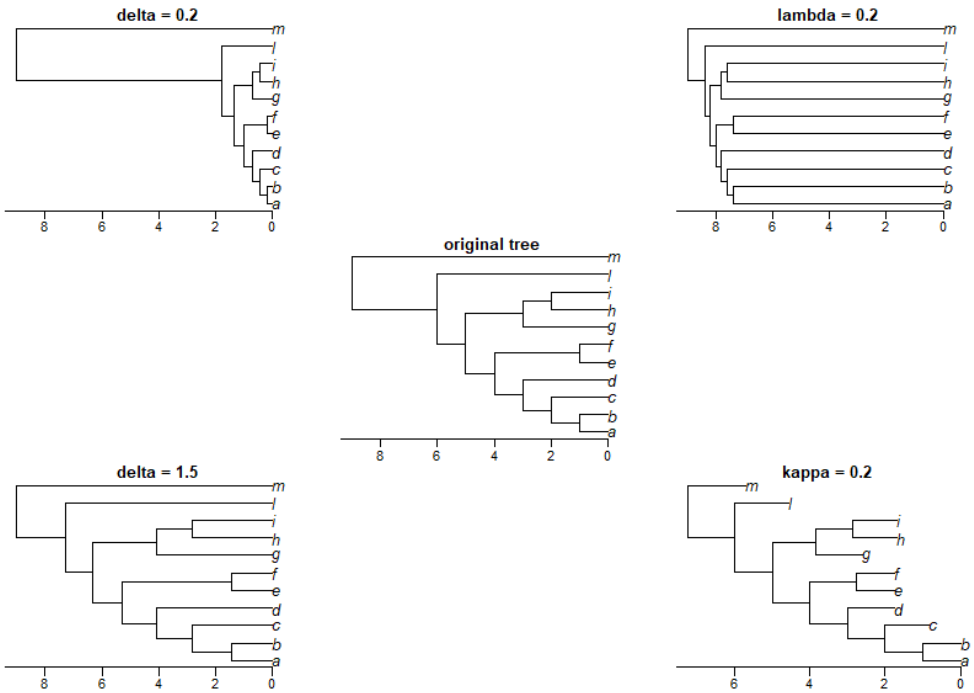

Figure 3. The original tree compared to tree stretching under different Pagel's transformations.

Another set of PCM frameworks implies a directional change of the average trait value in time. The Ornstein-Uhlenbeck model (Hansen, 1997; Butler & King, 2004) is one of the most commonly used BM-like models. It simulates the presence of one or more phenotypic optima (θ) attracting trait values with a strength α (Fig. 4c). The parameter θ is the long-term mean: once the mean trait reaches it, phenotypic evolution proceeds as a random walk around it. An OU model with α = 0 and θ equal to the starting value approximates a simple BM model (Fig. 4a). In case of θ different from the starting value, the model represents a BM with phenotypic trend through time (Cooper et al., 2016; Fig. 4b).

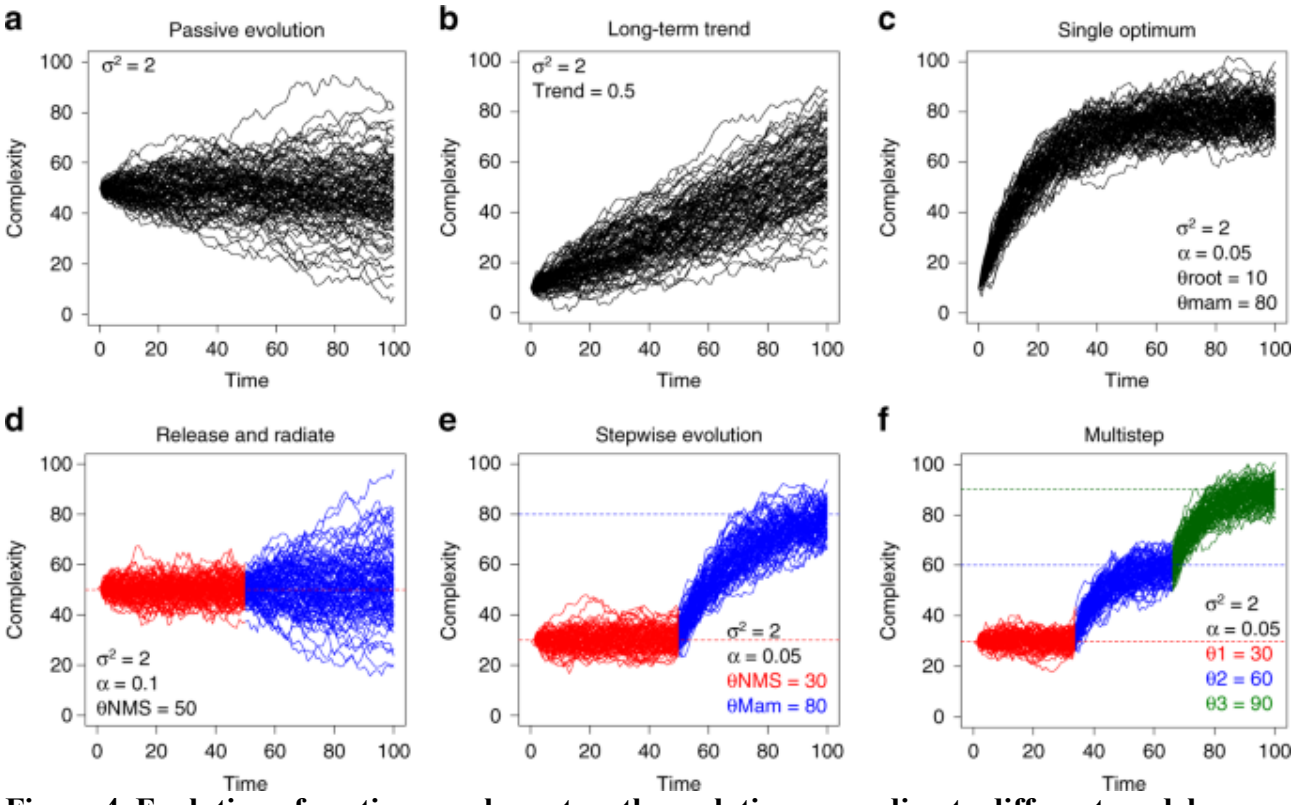

Figure 4. Evolution of continuous characters through time according to different models or combinations of them. a) Brownian Motion; b) Brownian Motion with trend; c) Ornstein-Uhlenbeck with single optimum; d) Release from constraint (OU followed by BM); e) Stepwise evolution toward higher optimum (two consecutive OU models); f) Multistep evolution with three optima shifts (three consecutive OU models). Image from Jones et al. 2019 (https://www.nature.com/articles/s41467-019-13026-3/figures/1).

The models described so far always include the rate of trait evolution as estimated parameter, assuming it to be constant along time and across the whole phylogeny. By combining these same models, it is possible to test for time- or clade-related changes of σ^2. Examples of time-related shifts in the rate of evolution are the "Release and radiate" model (Fig. 4d), mimicking a release from constraint on phenotypic evolution by arranging OU and BM, and stepwise models (Fig. 4e-f) which combine subsequent OU models to simulate a shift in phenotypic optimum. Similarly, the "Censored" and "Non-censored" approaches from O'Meara et al. (2006) test for clade-related changes in σ^2 by applying different BM rates to different parts of the tree.

Whether they estimate only one σ^2 or more, each of these frameworks is based on specific assumptions about phenotypic evolution and is built to test just that particular hypothesis. This kind of approaches is advantageous as it keeps low the number of estimated parameters, thus reducing the error around the likelihood estimates. In some cases, though, the oversimplification may produce low-likelihood models, which bad represent the real distribution of traits.

Recently developed methods overcome this problem by calculating distinct rates for each branch of the phylogeny and relaxing the assumption of any *a priori* evolutionary model. AUTEUR (Accommodating Uncertainty in Trait Evolution Using R; Eastman et al., 2011) is a Bayesian method using reversible-jumps Markov Chain Monte Carlo to model rate heterogeneity on phylogenies of living species. It allows exploring from the simplest BM model with a single rate value for the whole tree to a complex variable-rate model with different rates for each individual tree branch. The "models of trait macroevolution on trees" (MOTMOT; Thomas & Freckleton, 2012) is a package of tools apt to investigate patterns in the tempo of continuous trait evolution on phylogenetic trees by fitting a range of non-Brownian models of trait evolution. It compares single- and multiple-rates models by means of maximum likelihood estimation and Akaike Information Criterion (AIC) scores allowing to identify either single-branch, clade-wise or group-related shifts. BAMM (Bayesian Analysis of Macroevolutionary Mixtures; Rabosky, 2014) is a model-free

method built to deal with shifts in both phenotypic and taxonomic evolution in a phylogenetically explicit context. It uses reversible jump Markov Chain Monte Carlo to explore models with different numbers of evolutionary regimes (meant as transitions to new evolutionary processes defined by new speciation, extinction, and phenotypic evolutionary rates), finally returning the posterior probabilities of such models, the positions where regime shifts should occur, and marginal distributions of speciation and extinction rates for each branch of the tree.

Despite being valuable addition to the PCMs toolkit in providing means to get rid of model parameterizations, these methods have their own limitations, such as overparameterization, misleading results (BAMM, Meyer et al., 2018), time-consumption, and the biggest one: they only work with phylogenies of extant species, possibly preventing a real understanding of macroevolutionary patterns.

The *RRphylo* method is specifically meant to fill this gap. Starting from a phylogenetic tree and phenotypic data, the function performs phylogenetic ridge regression (Kratsch & McHardy, 2014) to estimate branch-wise rates of phenotypic evolution, and uses such rates together with observed phenotypic values to reconstruct ancestral characters at internal nodes. As it just uses ridge regression, *RRphylo* is free of any kind of superimposed evolutionary model and equally works with paleontological and neontological trees. Better still, including fossil information is welcome to increase the reliability of both evolutionary rates and ancestral character estimates.

Besides that, *RRphylo* has some other nice properties. Being a regression framework, it allows to easily account for the effect of a covariate (meant as variable affecting rates magnitude) or integrate additional predictors and fossil information at internal nodes in evolutionary rates estimation. Also, by providing branch-wise rates of evolution, it allows to scan the phylogeny for possible rate shifts occurring at clade level or related to a given state (i.e. discrete variable like diet, locomotory stance, sociality). Likewise, temporal trends in absolute evolutionary rates and phenotypic mean can be investigated, the latter taking advantage of the ancestral character estimates at internal nodes.

PAPERS COLLECTION

Castiglione et al. 2018 – Methods in Ecology and Evolution

In this paper, I presented the *RRphylo* method and its first application to identify clade-wise or state-related shifts in the rate of phenotypic evolution (*search.shift*). I applied *RRphylo* to simulated trees and data first, to assess its performance under different intensity and ubiquity of rate variation in phylogenetic trees. Eventually, I applied it to two different real case scenarios, the rate of evolution in insular mammals, which includes a rich record of recently extinct species, and is viewed as either exceptionally fast by some authors (Lister, 1989; Millien, 2011) or quite unexceptional by others (Raia & Meiri, 2011); and the evolution of body size in ornithodiran archosaurs, the clade that includes both the largest land vertebrates ever, sauropod dinosaurs (Sander et al., 2011), and the small, rapidly diversifying bird lineages (Lee et al., 2014a).

Materials and Methods

Phylogenetic Ridge Regression – RRphylo develops on phylogenetic ridge regression as described in Kratsch and McHardy (2014), and Gubry-Rangin et al. (2015). It applies penalized ridge regression to the tree and species data. The difference between the phenotype at each tip and the phenotype at the tree root is the sum of a vector of phenotypic transformations along the root to tip path, given by the equation $\Delta P = \beta_1 l_1 + \beta_2 l_2 + \ldots + \beta_n l_n$ where the β_{ith} and l_{ith} elements represent the regression coefficient and branch length, respectively, for each *ith* branch along the path. As regression slopes, the β coefficients represent the actual rate of phenotypic transformation along each branch. The matrix solution to find the vector of β coefficients for all the branches is given by the equation $\hat{\beta} = (\boldsymbol{L}^T\boldsymbol{L} + \lambda\boldsymbol{I})^{-1}\boldsymbol{L}^T y$ (James et al., 2013); where **L** is the matrix of tip to root distances of the tree (the branch lengths), having tips as rows. For each row of **L**, entries are zeroes for the branches outside the tip to root path, and actual branch lengths for those branches along the path. The vector $\hat{y}$ is the vector of phenotypes (tip values), $\hat{\beta}$ is the vector of regression coefficients

and λ is a penalization factor that avoids perfect predictions of $\hat{y}$, therefore allowing for the estimation of the vector of ancestral states (computed as $\hat{a} = \boldsymbol{L}'\beta$, where $\mathbf{L}'$ is the node to root path matrix, calculated in analogy to **L**, but with nodes as rows).

Kratsch and McHardy (2014) applied L2 (quadratic) penalization to estimate λ. I departed from this approach and applied a biologically oriented, conservative solution to the penalization problem, finding the maximum likelihood estimate of λ minimizing the rate (β coefficients) variation along the root to tip paths. Large absolute values of λ are consistent with very high phylogenetic signal (meaning that phylogenetically close species will tend to have very similar phenotypes). I empirically found that λ values from 0 to 1 are consistent with the Brownian motion model of evolution.

Searching for Rate Shifts – For a rate shift to be real, the β coefficients attached to the branches evolving under a distinctive rate regime must be either statistically larger or smaller than the coefficients calculated for the other branches of the tree. The basic machinery of *search.shift* (i.e. the function to assess hypotheses about rate shifts) is to compute the difference of mean rates between branches hypothesized to evolve under different rate regimes, and then to assess for the significance of such difference through randomizations. Since *RRphylo* assigns a specific rate to each branch of the tree, it is feasible to apply indifferently when the different rate regimes pertain to distinct clades, or to a number of unrelated species across the phylogeny. I refer to these two distinct situations as "clade"-level and "sparse" (phylogenetically) distributed conditions.

Eventually, both *RRphylo* and *search.shift* can further account for the effect of a covariate on the rate values by taking the residuals of the rate versus covariate regression (instead of the absolute rate values fitted by *RRphylo*) to contrast rates among different branches.

Testing Procedures – I devised a number of simulation experiments to assess the power of *RRphylo* and *search.shift* (under both "clade" and sparse" conditions), the incidence of Type I and

Type II errors, and the sensitivity of the method to varying tree size, rate change magnitude and commonness (i.e. of the proportion of species affected by the rate shift).

In addition to simulation experiments, I applied the method to search for rate shifts to real data. I focused on body size, which is the most obvious and readily available physical characteristic for fossil species, and the most common locus of investigations on possible instances of rate change and (the evolutionary consequences of) alleged evolutionary constraints (Alexander, 1998).

I started by assembling a phylogeny of ornithodirans taking the phylogeny and body mass data for dinosaurs from Benson et al. (2014), for early birds from Lee et al. (2014a, 2014b), and for pterosaurs from Villalobos et al. (2017). Data were supplemented with various sources. I used this phylogeny and data to test for rate shifts in body size evolution as propelled by the acquisition of flight, which impinges on the pterosaur and early bird and bird-like dinosaurs clades. I additionally tested for the existence of differences in the evolutionary rate regimes pertaining to different types of locomotion. Ornithodirans had four possible stances, bipedal (as in theropods, and early ceratopsians), quadrupedal (as in sauropods and large ornitischians), a combination of both (as in several hadrosaurs and prosauropods which were able to use both a two-limbs or four-limbs gaits; Maidment et al., 2012), and flight (in pterosaurs and birds). The body size rate regime linked to flight falls under the "clade" condition, whereas those involving transitions to bipedal, quadrupedal, mixed and flight locomotory types regard the "sparse" condition, since more than one stance occur within individual clades (e.g. ceratopsians, sauropodomorphs, theropods; Maidment et al., 2012).

In keeping with the reptile data, the second real case scenario regards the possible rate shift linked to flight in mammal body size (which is limited to bats). I finally tested the idea that insularity produces acceleration in the rate of body size evolution (e.g. Lister, 1989; Millien, 2006). This is a case of a phylogenetically "sparse" case, as insularity occurs ubiquitously across the mammal tree. I took the mammal data and insularity status from Raia et al. (2010a).

The rate values in *RRphylo* are regression coefficients between parent/descendant pairs. As such, the magnitude of the rates (i.e. coefficients) heavily depends on the trait values. With body

size data, this means rates between pairs of large species will be larger than rates between small species, even if they represent a smaller, proportional amount of phenotypic change. As explained above, the *RRphylo* method allows regressing the rates against the phenotype (or any other covariate upon indication), and using the absolute value of the residuals of such regression in the place of the original rate values. Whether to use the original rates or factoring out the effect of a covariate depends on the trait being considered, and the scientific question being answered. In both real case scenarios, I factored out body size from the rate calculation, after having verified that rates actually scale with body size. *RRphylo* allows both automatic detection of rate shifts and testing the specific hypotheses that the change in rate regards specific clades or unrelated species in the tree. For both ornithodirans and mammals, I ran the automatic detection first, and then tested the hypotheses about the effects of locomotor stance and insularity on the rate of body size evolution.

Results

As for simulation experiments, I found Type I error rate to be as low as 3% under the "clade" condition. The ability of *search.shift* to retrieve instances of rate change is close to 95% (on average). Under the "sparse" condition the incidence of Type II error is 3%. Sensitivity analysis revealed the probability to find the rate shift is significantly influenced by tree size and the magnitude of rate change, while the commonness of the altered rate regime across the tree is marginally significant.

Investigating the evolution of body size in ornithodirans, I found rates calculated for the bird (Avialae, exclusive of Confuciusornithidae) and pterosaur clades within ornithodirans significantly higher than for the rest of the tree ($p << 0.001$) also when tested alone ($p << 0.001$ for both clades). *search.shift* automatically recognized two additional clades showing significantly small rates (Fig. 5): Ornithopods and Sauropodomorpha (Fig. 5a).

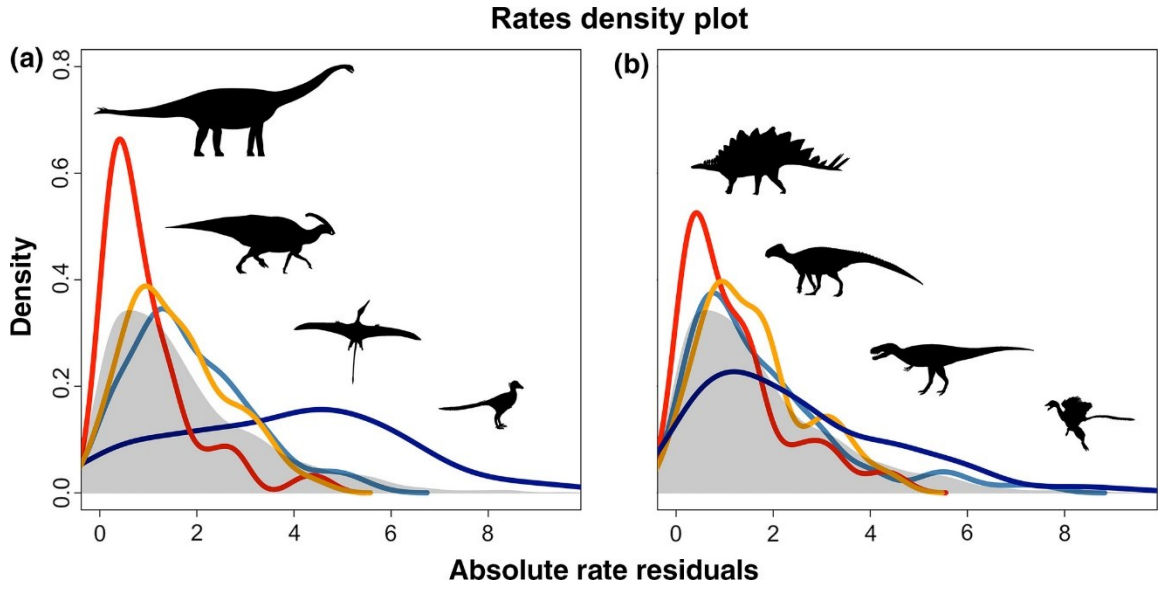

Figure 5. (a) Density plot of the absolute rate values versus body mass regression residuals, for the clades showing statistically distinctive rates. Pterosaurs (light blue) and early birds (deep blue) have significantly high rates. Ornithischians (orange) and Sauropoda (red) have significantly low rates. (b) Density plot of the absolute rate values versus body mass regression residuals per locomotory type in Ornithodirans. Flying (deep blue) species have significantly high rates. Quadrupedal species (red) have significantly low rates. The density plot of the entire Ornithodiran tree absolute rate residuals is shown in dark grey. Image from Castiglione et al. 2018.

In keeping with my hypothesis, I found statistically different rates to pertain to different stances in ornithodirans. Quadruped species show the smallest average rates, whereas flying species show significantly higher rates than any other stance type (Fig. 5b). By contrasting rates among species with different stances, I found quadruped species to show statistically smaller rates than bipedal ($p = 0.005$), and flying species ($p = 0.001$). No other pairwise difference was significant.

About mammals, the automatic detection procedure produced instances of significantly high rates for the clade "Afrotheria" (Fig. 6, node 7635), for dasyurids (Fig. 6, node 7866), leporids (Fig. 6, node 5714), anteaters and sloths (Fig. 6, node 7559), and sigmodontine rodents (Fig. 6, node 4818). Significantly small rates accrue to the white-toothed shrews (Fig. 6, node 7371), and to murine rodents of the *Rattus* division (Fig. 6, node 4416). The most interesting case regards bats.

While Yinpterochiroptera bats (exclusive of flying foxes) show significantly large rates, Vespertilionidae offer a case of small rates (Fig. 6, nodes 6709 and 7074 respectively). Yet, the bat clade tested as a whole shows significantly larger rates than the rest of the tree ($p < .0001$). Eventually, I found no statistically significant difference in rates of body size evolution between continental, and insular endemic mammals ($p = 0.149$).

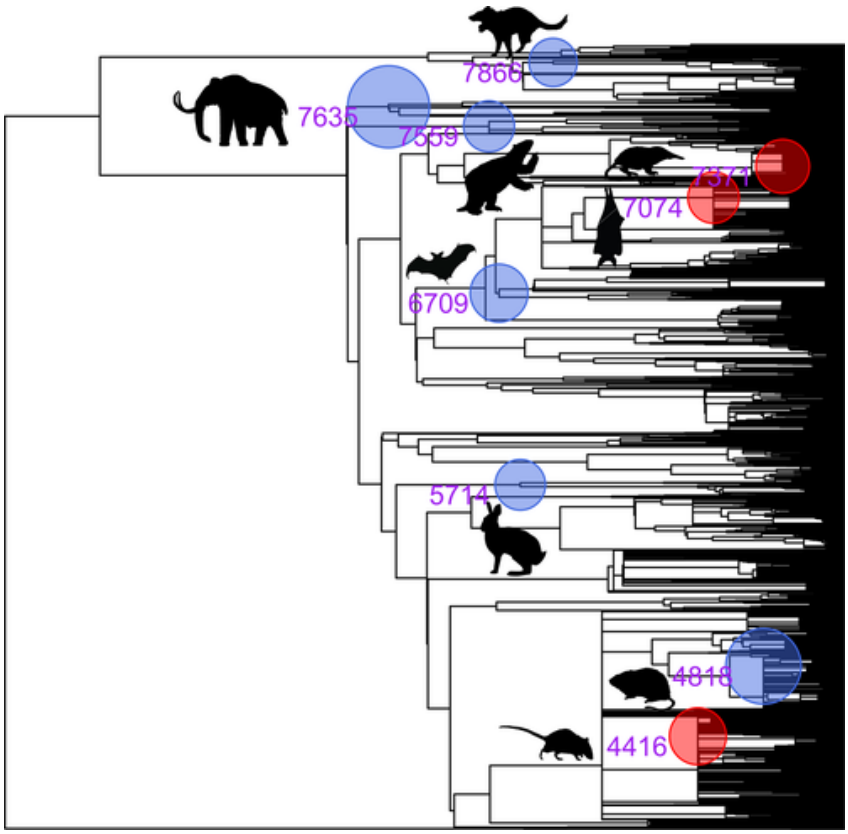

Figure 6. The mammal tree with clades having significantly low (red) or high (blue) rates indicated, taking the residuals of the absolute rate values versus body mass regression. Nodes: 4416 murinae rodents (Rattus division), 4818 sigmodontine rodents, 5714 leporids, 6709 Yinpterochiroptera, 7074 vesper bats, 7371 white-toothed shrews, 7559 anteaters and sloths, 7635 afrotheria, 7866 dasyurids. Image from Castiglione et al. 2018.

Discussion

I devised a new method, *RRphylo*, to compute branch-wise rates of phenotypic evolution starting from phylogenetic tree and phenotypic data. One obvious advantage of *RRphylo* over existing PCMs is that it allows testing for significant rate shifts even at the level of species, rather

than entire clades. One additional advantage is that by being free from any strong a priori assumption about the tempo and mode of phenotypic evolution, *RRphylo* takes full advantage from the inclusion of fossil forms in the tree, which is important because phylogenies limited to extant species give a poor (and phylogenetically restricted) depiction of the diversity of any animal group with a rich fossil record.

search.shift proved well suited to find shifts in the rate of phenotypic evolution. The simulations demonstrated it has low rates of both Type I and Type II errors, even under mild evolutionary rate change. Importantly, it works with variable tree sizes, and, under most circumstances, when the proportion of species affected by the rate shift is as low as 10%. Crucially, *search.shift* performs well even when the deviation in the phenotypic rate affects disparate, unrelated species in the phylogeny (the "sparse" condition) that is something impossible to test with any of the currently available PCMs. This is further important because rate shifts are often reported to act at the level of species (e.g. accelerated body size evolution with insularity, Lister, 1989; Millien, 2006), which is best understood with phylogenetic and fossil information at hand (Schnitzler et al., 2017; Slater et al., 2012).

I confirmed the idea that flight is paralleled by a significant shift in the body size evolutionary rate in vertebrates. I additionally went through the idea that locomotory type, in general, does affect the rate of evolution. I was moved from the simple observation that bipedal animals tend to be smaller sized than quadrupedal species, and that quadrupeds themselves might experience limits on body size evolution stirring ad hoc adaptations, such as the acquisition of graviportal legs as in proboscideans and sauropods. Ornithopods are especially well suited to study such effects, since the clade includes the whole gamut of possible locomotory types for limbed animals (exclusive of flight), and all the categories are present in more than one clade, making the application of the "sparse" condition necessary. I found that bipeds did experience slower rate of body size evolution than flying animals, but both categories' body size evolves faster than in quadrupedal animals.

Unsurprisingly, the most pervasive instances of significantly high rates in mammals regard Afrotheria, the clade including both the small bodied dassies (Hyracoidae) and the only mammals with graviportal legs, elephants, and the South American folivores of the Pilosa clade, which includes animals as small as pigmy anteaters (genus *Cyclopes*) to the end Pleistocene 6-tons ground sloths (which are present in the tree).

The results for mammals are entirely consistent with those for Mesozoic reptiles. Bats as a whole confirm that flight propelled accelerated rates of body size evolution. Yet, significantly small rates accrue to vesper bats, which have been already noted in the past (Dzeverin, 2008), while large rates were calculated for Yinpterochiroptera (Pteropodiformes) exclusive of megabats, which is conceivable given they are more variable than megabats in terms of body size. A similar case regards the rodent family. Although there are cases of >100 kg rodents in the past (e.g. blunt toothed giant jutía *Amblyrhiza inundata*, McFarlane et al., 1998), body size evolution in rodents might be constrained by gut anatomy, and by competition to foregut fermenting herbivores (Demment & Van Soest, 1985). Yet, there are clades, such as rice mice (Oryzomyni) which have been shown to have had a very dynamic body size evolution (Avaria-Llautureo et al., 2012). Perhaps unsurprisingly, I found a significant rate increase pertaining to the sigmodontini, the rodent clade including rice mice.

The results confirm the idea that rate of body size evolution is not faster on islands (Raia & Meiri, 2011; Thomas et al., 2009), which conflates on the observation that islands contain no more "extreme-sized" species per clade than expected by chance (Meiri et al., 2011).

Raia et al. 2018 – Scientific Reports

Primates are a large group of mainly arboreal, mostly tropical mammals, ranging in body size from 30 g in Berthe's mouse lemur (*Microcebus berthae*) to 200 kg in male gorilla. In terms of diet, primates are nearly equally variable, being adapted to feed on insects, honey, fruits, leaves,

seeds, nuts, and even vertebrate meat. Such wide dietary ambit reflects in the primate mandible and teeth variation. The extent to which diet actually influences the masticatory apparatus in Primates is the subject of intense investigation. It is now well recognized that variation in both mandibular shape and body size were the primary pathways for ecological diversification in fossil, as well as in living primates (Marroig & Cheverud, 2005), with diet acting primarily at high taxonomic level, while size has stronger effects between closely related species (Meloro et al., 2015). Hominins (which include the species belonging to either *Homo* or to the so-called 'australopiths') make no exception to this pattern. Members of the hominin clade have been long noted for their peculiar mandible shape, with short and deep corpus (the horizontal part that bears the tooth-row), low-cusped molars, and reduced incisors and canines. This suite of features is said to allow for a diet including tough food items such as roots and seeds (Sponheimer et al., 2013), and is linked to the reduced importance of food processing by the anterior dentition, as compared to fellow apes. This habitus is common to many, but by no means to all of the australopiths (Sponheimer et al., 2013), and reached its extreme in the Early Pleistocene hominin *Paranthropus boisei*, consistently with the lifestyle in the grasslands the late australopiths adapted to (Cerling et al., 2011). While living in open-habitats was common to *Homo* as well, species in our own genus have smaller, thinner-enamel cheek teeth, less robust mandible and zygomatic arches, reduced masticatory muscles and bite force, and decreased protrusion of the dental arcade (i.e. prognathism). Most of the differences between *Homo* and the australopiths are believed to relate to the evolution of an extremely large brain in *Homo*, which is responsible for ever increasing technological abilities and, later, for the control over fire. This would have eventually released adaptive pressures on the mandible and teeth, by endowing efficient mechanical food processing before chewing (Zink & Lieberman, 2016). As such, while the evolution of a mandible shape responsive to a new lifestyle and diet in australopiths should make them no different from the other primates, the robust relationship between mandible shape and diet presumably faded out in *Homo*, with the expected consequence of low evolutionary rate of change in *Homo* mandibles.

To verify this hypothesis, we analyzed mandibular shape variation in a large sample of primates, ranging from Paleogene 'plesiadapids' to living species, by applying geometric morphometrics (GMM) and the `RRphylo` toolkit. We assembled a dataset of 731 primate mandible images belonging to 211 different species and built a phylogenetic tree for those. Then I applied *RRphylo* and *search.shift* to the multivariate shape data ordinated via GMM to retrieve the rate of shape evolution for all the branches in the tree, and verified the existence of shifts in the rate of evolutionary change among clades.

We found the entire hominin clade to stand out among Primates, accounting for a disproportionately large share of the clade mandibular shape variation. More importantly, hominins represent the only instance of (multivariate) rate shift in mandibular shape evolution in primates, either according to *RRphylo*, or by using the more traditional, multivariate Brownian rate variation approach. Since body size variation accounts for a large share of ecological diversification within primates (Marroig & Cheverud, 2005), and is significantly related to shape variation, I also repeated the analyses after factoring out the effect of size on shape, by using the centroid size of the landmark configuration as a proxy for size. Again, only hominins stand out for having exceptionally large rates.

The results show that mandibular shape in hominins evolved faster than in any other primate clade. Contrary to our expectations, the rate of evolution in *Homo* is not smaller than in the australopiths, and the direction of the shape change velocity is one and the same for the two hominin clades. This means that the reason for the unexpected pattern of rapid mandible shape evolution observed across hominins has to be found among the characteristics shared by the australopiths and *Homo*. According to a large corpus of available data, the australopiths and *Homo* differ from each other in terms of habitat preferences, body size, patterns of sexual dimorphism, diet and food processing behavior (Foley et al., 2016). However, tool use has been hypothesized to occur in all early hominids, including australopiths (Skinner et al., 2015). Such emphasis on mechanical food processing might have caused parallel evolutionary changes in the mandible of

hominins. Relevant dental features shared by all hominins are the reduction of maxillary canines crown height, reduced sexual dimorphism, and loss of the honing capacity of the C/P3 complex (Haile-Selassie et al., 2004), which by contrast represents a nearly ubiquitous and stable adaptation in nonhuman anthropoids. As compared to the greater apes, all hominins evolved after *A. anamensis* also share a derived temporomandibular joint, that allows for a peculiar forward translation and rotation of the mandible during mouth opening in increase gape, and show strongly reduced anterior dentition (incisors and canines), shorter mandibular corpus with more divergent rami and an increase in the absolute and relative size and complexity of the post-canine dentition. The evolutionary emergence of these features has been related to dietary shifts, sexual selection, or a combination of both.

We propose the reshaping of the mandible, shared by the australopiths and *Homo*, was startled by both biomechanical and "structural" events such as the loss of a functioning of the C/P3 honing complex (Haile-Selassie et al., 2004). This exaptive condition occurred early in hominin evolution and generated "cascading effects" that were recruited for a number of different adaptations along and across the history of the human clade, in response to the rapid environmental changes recorded in Africa from the Upper Miocene through the Plio-Pleistocene.

Mondanaro et al. 2020 – iScience

Human occupation of Northern Europe and Western Siberia traces back to Middle-Pleistocene, when the establishment of full glacial cycles was making global temperatures colder than ever before during the history of the genus. The occupation of such northern temperate and boreal zones presents a number of notable challenges. Not only was the cold itself challenging for hominins physiologically adapted to African climates but also seasonality imposes extreme annual resource fluctuations, which imply a reliance on hunted meat for survival. Adaptations facilitating survival in cold environments may have included the use of fire, shelters or clothing, weapons

useful to bring down large game species, as well as extended social networks, with vulnerable infants being particularly susceptible to mortality (Martin et al., 2020).

Unfortunately, clothing manufacturing leaves very little in the way of fossil remains. Only the two most recent human species, *H. neanderthalensis* and *H. sapiens*, left incontrovertible evidence that they were able to produce complex, cold-proof clothing at that time. To make things more complex, in the particular case of *H. neanderthalensis*, biological adaptation, besides material culture, was possibly involved in their ability to withstand the cold. *H. neanderthalensis* possessed relatively short limbs, and a large midface and nasal cavity proposed to be specific cold adaptations, to heat and humidify inspired air, although the issue is far from resolved and there is evidence for the contrary (Wroe et al., 2018). In contrast to any other *Homo*, *H. sapiens* is considered the only species in the genus able to occupy cold regions through a genuinely cultural process, driven by our technology, including the mastering of fire, ever improving clothing craftsmanship, and construction of shelters (Boivin et al., 2016). This view sets *H. sapiens* apart from any other human species in terms of cognitive skills and implicitly rejects the idea that older *Homo* may have had sufficiently modern material culture to overcome climatic harshness (Roberts & Stewart, 2018). With such a poor fossil record of clothes and tools to produce them and because of great uncertainty about deep past local paleoclimates and human dispersal timing and direction, the issue of when humans first became cognitively and culturally able to extend their climatic tolerance beyond their physiological limits remains very difficult to decipher.

In this paper, we addressed the more restricted issue of when during the history of *Homo* the limits of climatic tolerance expanded and which species were involved. We modelled the evolution of climatic tolerance (i.e. niche) limits in the *Homo* genus by associating paleoclimatic values with fossil occurrences in the archaeological record. I used *RRphylo* to estimate the rate of change of climatic tolerance limits across the human phylogenetic tree, and *search.shift* to search for possible shifts in such rate. In the present context, shifts in the rate of evolution of climatic tolerance that accrue to the clade including *H. heidelbergensis*, plus *H. neanderthalensis*, and *H. sapiens* (modern

Homo species, MHS, hereafter) would indicate these hominins were the first to acquire the capacity to develop cold climate-related technological skills and cultural adaptations. Conversely, if either no rate shift occurs or the rate shift coincides with different clades (e.g. early *Homo* species, EHS, hereafter), the colonization of Northern habitats would not be indicative of any sudden increase in the ability to face environmental harshness.

We found that the clade identified by *H. heidelbergensis*, *H. neanderthalensis*, and *H. sapiens* and their common ancestor experienced a significant evolutionary rate shift toward wider climatic tolerance. The rate shift does not depend on the specific phylogenetic hypothesis (tree topology) assumed, neither does it depend on the selection of species we used. Randomly changing the tree node ages (to account for dating uncertainty) and species positions in the hominin tree (to account for phylogenetic uncertainty) 100 times, the shift appears for this clade 95 times. Subsampling the most abundant species (randomly selecting no more than 100 fossil occurrences per species) to account for sampling differences between species, the shift appears 91 times out of a hundred.

The estimated values of realized climatic niche limits at nodes in the hominin phylogeny suggest that the rate shift in the climatic niche limits for the MHS clade was not an exclusively biological process. At the root of the hominin tree the predicted range in annual temperatures spans from 20°C (coldest quarter of the year) to 29.9°C (warmest quarter) and in mean rainfall from 12 mm (driest quarter) to 512 mm (wettest quarter). This is entirely consistent with today's African savannah environment. At the node subtending the pair *H. ergaster* plus *H. erectus* (which is the first hominin to disperse over Southern Eurasia), the corresponding figures are 0.7°C–31.9°C for temperature range and from 4.8 mm to 1080 mm for precipitation range. These estimates are reasonable considering both the range expansion into temperate regions and the colonization of warm and humid environments (Indonesia) by *H. erectus* (Rizal et al., 2020). Yet, at the common ancestor to the three MHS, the estimates for annual temperature extremes span from minus 21.1°C to plus 31.4°C and for annual precipitation from 0.7 mm to 905 mm. Although the common

ancestor to MHS was an African species which probably never experienced these extreme climates (Profico et al., 2016), the values agree qualitatively with the notion that a sudden widening of climatic niche limits occurred with the advent of this ancestor, whose offspring lived after the onset of fully glacial Pleistocene conditions. The massive increase in the estimated range of thermal conditions suitable for the MHS clade taxa does not depend on the phylogenetic hypothesis we applied and surpasses what is expected by a random process of increased phenotypic variance over time (namely the Brownian motion model of evolution). Using 100 different tree topologies and branch lengths to account for phylogenetic uncertainty, we found a significant trend in the temperature of the coldest season realized by hominins 97 times, whereas no trend was found in the maximum temperatures of the warmest season. We found that in African species and ancestors, the average temperature of the coldest quarter of the year was no less than 9.4°C, meaning that the winter chill is unlikely to have been a problem for them. In contrast, within the range of temperatures experienced by *H. heidelbergensis*, the coldest quarter of the year was as cold as −12.3°C, suggesting specific technological and cultural adaptations were needed to fend off the risk of hypothermia and to live in the highly seasonal, cold northern environments. These adaptations may have included fitted clothing (Amanzougaghene et al., 2020), thrown spears or adhesives (Cârciumaru et al., 2012), and enhanced healthcare practices (Spikins et al., 2019).

Although the real consequences of any individual cultural or technological adaptation introduced by MHS will almost certainly be a matter for debate for some time, our results indicate that these hominins were able to overcome the challenges imposed by life in northern habitats by a non-biological process, suggesting that behavioral modernity, interpreted as the capacity to use technology and culture to overcome the constraints imposed by natural climate variability on the geographic distribution, is not limited to *H. sapiens*.

Sansalone et al. 2020 – Proceeding of the Royal Society B: Biological Sciences

Body size exerts pervasive effects on animal morphology and behavior (Calder 1984). Macroevolutionary trends in body size are thus expected to play a major role in the evolution of morphological structures, influencing the tempo and mode of phenotypic change (Pélabon et al., 2014; Marroig & Cheverud, 2005).

Among primate clades, some have followed no clear pattern regarding body size over time, but others have consistently increased in size (Cope's rule; Soligo & Martin, 2006). By contrast, the evolution of a large brain, in both absolute and relative terms, has been a consistent trend (Neubauer & Hublin, 2012), culminating in the emergence of large-brained apes and humans. Progressive encephalization has had profound physiological, ecological and social consequences, linked to diet, home range size and activity period, mating system, life-history traits, dexterity and sociality.

In addition to brain size, a diversity of ecological and behavioral characteristics has also been attributed to changes in the shape, position and orientation of individual brain structures. Several studies have suggested that brain reorganization in primates follows a mosaic pattern (Smaers & Soligo, 2013), with separate cortical areas responding differently to different selective pressures. It has been argued that this has produced a greater diversity in brain shape than would be predicted on the basis of absolute or relative brain sizes alone (Smaers & Vanier, 2019).

However, although it is clear that body mass influences brain size, whether the relationship between brain size and brain shape themselves is characterized by allometry is not known. Yet, allometric scaling may occur at different evolutionary rates and previous studies have highlighted how brain size can evolve at different rates in different primate lineages (Smaers & Vanier, 2019).

Herein, we asked whether the relationship between primate brain size and brain shape is characterized by allometry, and whether any such relationship may reflect shared macroevolutionary trends in primate brain shape. Identifying the brain components associated with

major shifts in rates of morphological evolution will contribute to the understanding of how primates and humans evolved their varied and complex behavioral repertoires.

Our study is based on 386 skull endocasts (including both males and females for each species where available) representing 151 primate species. We assessed allometric relationships between brain size and brain shape in a phylogenetically explicit context and investigated whether size has a global (affecting brain shape as a whole) or a localized (to specific cortical areas) effect on brain shape reorganization using the novel 'shape integration' concept proposed by Bookstein (2015). I used *RRphylo* and *search.shift* to determine the presence of shifts in the evolutionary rates across primate history in a phylogenetic context marking brain shape discontinuity across the primate tree.

The PGLS regression between shape variables and the lnCS (the logarithm of centroid size, lnCS, a measure of brain volume) is highly significant ($r2 = 0.17$, p-value = 0.01). Yet, phylogenetic MANCOVA (multivariate analysis of covariance) indicates that allometric slopes differ among clades (p-value = 0.001). Per-clade PGLS regressions revealed that significant allometric effects are entirely restricted to Hominoidea and Cercopithecinae, although their allometric trajectories diverge from one another (difference at small size = 0.09, difference at large size = 0.21; p-value = 0.008).

Bookstein's integration test allows evaluation of the degree to which spatial scale the brain size is impacting brain shape. In Hominoidea and Cercopithecinae, the impact is great. Unlike Hominoidea and Cercopithecinae, in Platyrrhini and Colobinae, allometry drives no discernible effect. In Strepsirrhini (Lemuriformes and Lorisiformes), brain size influences brain shape only at the local scale.

When investigating the rates of morphological evolution, I found contrasting results depending on whether allometry was present or not. In fact, using a two-tailed permutation test of rates across the tree, I found instances of phenotypic rate acceleration in the clades showing strong allometric effects, Hominoidea (p-value = 0.99) and Papionini (within Cercopithecinae; p-value =

1.00) and two significant occurrences of rate deceleration in clades which did not show allometry: Strepsirrhini (p-value = 0.003) and Platyrrhini (p-value = 0.001). These results are not a consequence of the tree topology we adopted, remaining stable after accounting for phylogenetic uncertainty by randomly swapping tree branches and node ages to account for phylogenetic uncertainty.

Our study demonstrates that the evolution of brain shape in primates cannot be explained by any single, simple evolutionary model. The trajectories and rates of brain shape evolution in different primate clades are strongly influenced by the presence or absence of allometry between brain size and shape with different primate clades occupying distinct regions of morphospace along a brain size gradient (the small-brained strepsirrhines and the large-brained hominoids at opposite extremes). The presence of allometry significantly impacted evolutionary rates and triggered the diversification of brain shape in hominoids and papionins along two distinct trajectories. We propose that faster rates of evolution driven by the presence of allometry may help to explain how Hominoidea and Papionini evolved more versatile and complex range of behaviors (such as multi-level social interactions), possibly reflecting a shift in their ecological and positional behavior (Lefebvre, 2013). By contrast, strepsirrhines and platyrrhines both exhibit slowdowns in the rate of brain shape evolution and, in New World monkeys, reduced magnitudes of shape change, consistent with the conservative ecological habits of platyrrhines. In strepsirrhines, brain shape shows no allometry and a ‘dis-integrated' pattern of shape variation possibly related to the higher eco-morphological diversity of this clade. Eventually, in platyrrhines neither mosaicism nor allometry are present, in compliance with the predictions of the adaptive radiation model and with the observation that New World monkeys possess less diverse lifestyles, lack terrestrial species and converge towards common adaptive peaks (Meloro et al., 2015).

Castiglione et al. 2019 – PloS one

In this paper I presented *search.trend*, a new implementation of the `RRphylo` toolkit which takes advantages of both branch-wise evolutionary rates and phenotypic estimates at nodes to detect simultaneously temporal trends in phenotypic rates and means, and to compare trends among different clades within the phylogeny. I tested *search.trend* performance by using simulations, measuring both Type I and Type II error rates.

Materials and Methods

The *search.trend* algorithm retrieves branch-wise phenotypic evolutionary rates from an object produced by *RRphylo* and regresses the absolute value of such rates against their ages, meant as the distance of the branch from the tree root. A second, separate regression is performed between the vector of phenotypes (obtained by collating ancestral phenotypic estimates to trait values at the tree tips) and their ages. Throughout the rest of the paragraph, I refer to the former as the regression to test for the existence of a 'trend' in the rates, and to the latter as a test for the 'drift' in the phenotypic mean, over time. For both regression slopes (i.e. trend and drift), significance is assessed as the probability that the actual slopes differ from a family of 100 regression slopes (BM_{slopes}) generated according to the Brownian motion model of evolution.

The Brownian motion has two free parameters, the phenotypic value at the tree root (herein named *rootV*) and the Brownian rate σ^2. *RRphylo* estimates ancestral states (including *rootV*) as the products of the matrix of branch lengths multiplied to the vector of rates (the latter are normalized as to avoid extreme rate values, which makes ridge regression different from ordinary least squares regression). By default, *rootV* is computed as the average value of the 10% most ancient tips in the tree, weighted by their squared distance from the root (meaning that older species have more influence on *rootV* estimation). This means that the estimation of unknown phenotypes is entirely dependent on the tree known (tip) phenotypes, rather than depending on the assumption of a

particular mode of evolution, such as BM. Rather than a single rate such as in the Brownian motion, *RRphylo* assigns a rate to each branch of the tree, estimated via phylogenetic ridge regression (Kratsch & McHardy, 2014). Such 'rates' actually represent regression coefficients, describing the pace of phenotypic change between two consecutive nodes in the tree. As such, rates represent the phenotypic change per unit time between consecutive nodes in the tree. Hence, with a phenotype evolving according to the Brownian motion model, the magnitude of *RRphylo* rates increases with the distance from the root (i.e. towards the present) in keeping with the increase in phenotypic variance.

Ideally, when a positive trend in the rate of evolution towards the present applies, the slope of the rates versus age regression would be larger than any BM_{slopes}. However, three sources of uncertainty impinge upon this ideal situation. First, since rates are proportional to the actual the phenotypic values rates must be rescaled into the 0–1 range before running regressions to make the real regression slope entirely comparable to BM_{slopes}. Secondly, in the presence of a temporal trend in the evolutionary rates, *rootV* might assume an extreme value within the distribution of phenotypes, which will generate a heavily skewed distribution of *RRphylo* rates. Thirdly, the distribution of rates is influenced by variation in branching times across the phylogeny. Since closely related species tend to have similar phenotypes, evolutionary rates will be small where the tree is dense with species. This could produce a declining slope of the trend regression line if the dense part of the tree coincides with a single, large, recent clade, even under a regime of increasing rates over time. To account for these caveats, the rates versus age regression in *search.trend* runs with logged data (to reduce the skewness of the rate distribution). Still, to properly assess the direction of rate variation through time, *search.trend* checks whether the standard deviation of the rates in the branches falling in the first (older) half of the tree is significantly smaller than the corresponding figure for the second (more recent) half of the tree, as compared to BM simulations, which is expected to occur if a positive trend in the evolutionary rate is present in the data.

With the drift regression (phenotypes versus age), the original phenotypes as well as those producing the BM_{slopes} should be rescaled to the 0–1 range as well, to account for the actual value range of the individual variables. For all the BM_{slopes}, regardless of whether rates or phenotypes are used, simulated phenotypes are produced by imposing $\sigma^2 = 1$.

Differently from any existing PCM, *search.trend* is designed to contrast different clades within the tree to find significant differences in the pattern of phenotypic evolutionary rate and phenotypic mean change over time. In the case of phenotypic drift, individual clades are tested for the hypothesis the drift slopes do not depart from the Brownian motion expectation. However, in the case of trend regressions the actual regression slope depends on the relative position (age) of the focal nodes respective to the root, given the exponential nature of phenotypic variance change in time. Because of this, for the trend case *search.trend* compares estimated marginal means predictions from the linear regressions (of the rate versus age regression).

I performed a number of simulation experiments to assess the incidence of both Type I and Type II errors for *search.trend* as performed either on the entire tree or on individual clades and to establish the 95% confidence intervals of significance. The general approach was to transform a BM phenotype produced on a random tree according to a trend of exponential increase or decrease of the phenotypic variance over time (representing time-dependent changes in the phenotypic evolutionary rate) and according to a phenotypic drift, by randomly varying the intensity of the patterns. The resulting 'essig' (for the trend in rate) and 'dssig' for the trend in phenotype, thus represent the intensity of the pattern at which *search.trend* sensitivity reaches 95%.

Analogous procedure was applied to generate temporal trends in rates or phenotypes for individual clades. In this case, the BM phenotypic values for the species within the clades were transformed to have twice as much trend as essig (for rates) or dssig (for phenotypes). At comparing trend/drift at specific clades, I selected two non-overlapping nodes in the tree subtending to at least 20 species each. In both the single-clade and two clades modified experiments, the sign of the

transformation was random, meaning the that focal clades phenotypes might be altered either in the same, as well as in opposing directions.

Finally, I compared the performance of the method to a competing model-based one.

Results

search.trend performs well in terms of finding the simulated patterns. Type I error rate is close to the nominal level (5%) in both the trend and the drift cases. The 95% level of 'dssig' is 0.25 (-0.25). This corresponds to 0.18 standard deviations from the root value per unit time, which is a modest phenotypic drift. The 95% level of 'essig' corresponds to the ratio between the range of phenotypic values and the range of such values halfway along the tree height of 84% to 112% of the same figure under the Brownian motion model.

search.trend applied to nodes evolving at different regimes from the rest (Brownian motion) of the tree is able to recognize the simulated patterns. When a single node is transformed, *search.trend* successfully recognizes the phenotypic transformation in > 80% of the cases. With the trend case, this percentage rises above 90%. The corresponding Type I error rates (instances of reportedly significant phenotypic change on nodes which were left, in fact, untransformed) is as low as 6 (trend) and 2% (drift). When two different, non-overlapping clades are transformed at one time, the function power still remains close to 90% if the two nodes are transformed in opposite directions (i.e. by applying transformations with different signs) but becomes much less powerful if the two clades are transformed in the same direction. Under all conditions, the Type I error rate remains close to the nominal alpha level. This is particularly robust considering individual node heights vary in between 20 to 97% of the tree height and might include as many as 20 to 84 species.

As compared to existing methods, *search.trend* is at least as powerful and accurate.

Discussion

The method I propose here, *search.trend*, provides a way to estimate straight away the existence of secular patterns in specific traits, and to compare individual clades within the tree to each other. The *search.trend* function demonstrated to have high power and consistently low Type I error, and to be at least as accurate as competing method. *search.trend* provides information about the direction and intensity of the two patterns, writes figures as pdf files to let the experimenter gauge the exact meaning and distribution of the patterns which are found, provides confidence intervals around the estimates for both rates and phenotypes, restitutes the phenotypes and rates per age and per branch (which allows further inspection of the distribution of such metrics per clade and per age and, if desired, the application of regression models different from the linear model which the function uses by default) and, on top of all, *search.trend* allows comparing directly individual clades within the phylogeny with each other and against the rest of the tree. This latter, fundamental feature means that individual clades can be compared to each other for the existence of either drifts or trends in the phenotypic mean and variance, respectively, even when the actual phenotype is a complex admixture of different evolutionary regimes. I found that *search.trend* has good power in finding the designed pattern and shows small Type I error rates even under small deviation from Brownian motion. The group (clade) comparison module of the function provides evidence that it effectively recognizes whether two clades in the tree evolve into different directions (either in terms of phenotypic mean or change in the evolutionary rate) when they are designed to be. When the two selected clades do evolve into the same direction, the function power to detect deviations of these clades from the Brownian motion decreases dramatically. However, rather than a limitation, this depends on the fact that when two clades in a tree are simulated to evolve according to a certain pattern in the phenotype (or in the rate either) the original BM phenotypic pattern of the tree as a whole is erased altogether.

Fossil information provides unique opportunity to look at phenotypic variation in the past and its change over time. While the relevance of such information to the proper understanding of trait evolution is well known, its full integration to the vast and powerful array of PCMs is limited by the unease of most PCMs to deal with paleontological phylogenies, and their limited description of macroevolutionary patterns. Here I provide a new, powerful addition to the existing PCM toolbox, which is appropriate to use when a full description of macroevolutionary patterns as captured by paleontological data and tree, and the simultaneous comparison between clades within the tree is the goal.

Serio et al. 2019 – Evolutionary Biology

In this paper I presented the multiple regression version of *RRphylo* and a new function to test for phylogenetic uncertainty and sampling effect. Both methods are presented as applications to study the evolution of relative brain size in cetaceans.

Cetaceans, toothed whales in particular, are commonly noted for their extraordinary behavioural plasticity. A number of studies indicate several species evolved learning, communication, social and cultural skills which have no parallel in non-human species, to the extent that they have been described as the 'apes better than apes' (Whiten, 2001). Most odontocete cetaceans live in complex groups with extremely differentiated relationships, long-term bonds, cooperative networks and alliances, which strongly depend on their ability to recognize and understand others, and to communicate with a wide range of vocal, visual, and behavioural signals (Lusseau, 2006). They are also able to recognize themselves and their own body parts, use tools, and transmit learned behaviours.

The explanation for such astonishing abilities likely lies in cetacean brain size and complexity (Marino et al., 2007). Cetacean brains are the largest among vertebrates and are extremely gyrified as compared to the brain of terrestrial mammals. Anatomical investigations of

Odontocete brains revealed their high Encephalization Quotient (EQ, the brain to body relative size; Gingerich, 2016), and that the number of neocortical neurons and glial cells is consistently larger than in any other large-brained mammal species, including humans (Mortensen et al., 2014). It is commonly assumed that the large brain in odontocetes evolved in response to the cognitive demands associated to their social and behavioural complexity. It was possibly accompanied by the acquisition of echolocation (Marino et al., 2007; Mortensen et al., 2014) from the inception of the clade (Park et al., 2016) and persisted through the initial body size reduction the group underwent (Montgomery et al., 2016). However, some studies suggested the cetacean brain size is unrelated to cognitive abilities. Others have shown that the higher relative brain mass in Odontoceti as compared to Mysticeti is just the result of higher rate of body size evolution in the latter, and that toothed whales relative brain mass did not increase through time. These results might imply the high encephalization quotient in odontocetes could be neutral with respect to cognitive abilities (Montgomery et al., 2013).

Herein, we investigated upon the timing, rate and trend of encephalization in odontocetes and cetaceans as a whole. Traditional approaches to study the evolution of encephalization involve the calculation of the typical brain-body allometry (i.e. the encephalization quotient, EQ; or the residuals of brain to body mass regression, RBS), which can be problematic as they do not represent a true evolutionary approach (Shultz & Dunbar, 2010), and because residuals have undesirable statistical properties under both non-phylogenetic and phylogenetic contexts (Freckleton, 2009). Therefore, I developed a new version of *RRphylo* to perform phylogenetic multiple regression between brain size and body size.

Materials and Methods

Materials – The cetacean phylogenetic tree is an informal supertree assembled from the backbone phylogenies in Montgomery et al. (2013) and Marx and Fordyce (2015). The tree includes 89 species, thirty-three of them are extinct (9 archaeocetes, 23 odontocetes, 1 mysticete).

Body size estimates (in grams, converted in ln grams), brain volumes (in cubic centimetres) and brain size (in grams) were taken from several sources.

We tested for the potential influence of diet on encephalization by ascribing each species to a feeding category. We used two alternative categorization schemes. First, we partitioned species into filter feeders (FF), suction feeders (SF), and raptorial feeders (R), according to the feeding category scheme adopted in Uhen (2004), Johnson and Berta (2011) and Berta et al. (2016). In the second feeding category scheme, we further divided raptorial species into brevirostrine (BR), and longirostrine (LR), forms, which are known to differ from each other in terms of prey selection (McCurry et al., 2017). To assign a species to either BR or LR we used the relative length of the rostrum (from the tip of the muzzle to the rearmost point of the fronto-nasal suture). To compare brain size among different feeding categories while keeping into account the allometric effects and group size, we computed the estimated marginal means (the equally weighted group mean predictions) of brain versus body size regression per group and performed pairwise comparisons between feeding category groups.

Phylogenetic Multiple Regression with RRphylo – Since *RRphylo* rates are in fact phylogenetic ridge regression coefficients, their magnitude depends on the absolute values of the phenotypes being regressed, that means large phenotypic values will originate large rates even with small phenotypic change. To standardize the rates, under *RRphylo* it is advisable to use the phenotype itself as a covariate. For instance, it is possible to use body size as a covariate to calculate the mass-specific rate of body size evolution for each branch in the tree. In the case of brain size, the evolutionary rates would strongly depend on allometric effects (so that large-brained species would show large evolutionary rate values, whether or not their brains are larger than expected by their body size). Using encephalization quotients (EQs) or the residuals of the brain to body size regression (RBS) helps fixing the problem, but ratios and residuals have undesirable statistical properties that make their use questionable (Freckleton, 2009). Because of this, I

developed a new version of *RRphylo*, that allows calculating the evolutionary rates from a multiple regression (in the present context, brain size is the response variable, the phylogeny and body size are the predictors).

In the multiple regression *RRphylo* version rates are calculated as:

$$\hat{\beta} = \left(L'^{T} L' + \lambda I\right)^{-1} L'^{T} y$$

Where L' is the tip to node path of branch lengths matrix (which represents phylogeny as a predictor) supplemented with body size as its last column (representing body size as the additional predictor) and y is the vector of species brain sizes. Lambda (λ) is the normalization factor optimized to avoid abnormal rate values within clades and $\boldsymbol{I}$ is the identity matrix. This way, the vector of rates $\hat{\beta}$ is calculated for all the branches in the tree and the last element of $\hat{\beta}$ represents the partial phylogenetic ridge regression coefficient of the additional predictor (here body size). I tested the appropriateness of this procedure by virtue of simulation experiments and demonstrated multiple *RRphylo* correctly represents an unbiased estimator of ancestral states and rates just as *RRphylo*.

We applied *RRphylo* on body size and the multiple *RRphylo* version on brain size in cetaceans, then performed *search.shift* on both results. The latter was used under the "clade" setting to automatically detect significantly shifting clades, and under the "sparse" setting to test for difference in brain size evolutionary rates among dietary categories.

After testing for the existence of rate shifts, we applied *search.trend*. As with *RRphylo*, I modified *search.trend* to deal with evolutionary rates derived from a multiple phylogenetic ridge regression. In particular, for each simulated Brownian motion phenotype (used to produce the 100 random BM_{slopes} to assess significance) the new *search.trend* adds a predictor as with the original multiple regression *RRphylo* design. The predictor is simulated and then modified to have the same correlation structure with the response variable as the original data. This is achieved by

transforming both the simulated response and the simulated predictor by using Cholesky decomposition.

We applied *search.trend* on body and brain size data and rates, and contrasted both the crown group and the stem group (extinct representatives of the clade with no living descendants) Odontocetes to the rest of the cetacean tree.

Accounting for Phylogenetic Uncertainty – To assess for potential biases as introduced by phylogenetic uncertainty, I developed and applied the newly-implemented `RRphylo` function *overfitRR* to test the effect of sampling, tree topology and branch lengths on results produced by *search.trend* and *search.shift*. This function randomly removes a number of tips corresponding to 25% of the tree size and swaps species phylogenetic position (thereby accounting for sampling effects) by using the `RRphylo` function *swapONE*. Under *swapONE*, each tip might change its position on the tree by up to two nodes. For instance, a topology of the kind ((A,B),C) might change to ((C,B),A) or ((A,C),B). In addition, each node might change in age in between the age of its ancestor and the age of its daughter node. I set one tenth of the tips to be swapped across nodes and one tenth of the nodes to be changed in age at each iteration. Then, it performs *search.trend* and *search.shift* on pruned tree and data. The procedure is repeated 100 times and the percentage of significant results returned. In this case, we specified the Odontoceti clade to be tested for temporal trends in phenotypic (body size) mean and rates. It must be noted that with such modest level of swapping the original topology (which is used as the reference for producing the random trees) is not altered significantly. This helps avoiding testing unreliable topological arrangements which could be unrealistic or otherwise unsupported in the scientific literature.

Results

Body Size Evolution – By applying *search.shift* under the automatic mode we found a positive and significant shift in body size evolutionary rates pertaining to the clade including Platanistidae and Eurhinodelphinidae (average rate difference = 0.941, p = 0.005). The result was

confirmed in 78 out of 100 random trees derived from the original topology via tree swapping (i.e. by *overfitRR*).

We did not find evidence for increase in body size (Cope's rule) in Cetacea as a whole ($p = 0.470$; 0% significant instances by *overfitRR*). The same is true of toothed whales ($p_{crown} = 0.370$, $p_{stem} = 0.490$). Unsurprisingly, estimated marginal means in body size in Odontoceti are significantly smaller than for the other species. Results for odontocetes are robust to phylogenetic uncertainty and sampling effect.

The temporal trend in body size absolute evolutionary rate is not significant for the entire tree ($p = 0.370$; 2% significant instances by *overfitRR*). A trend for increased rates is present in crown Odontoceti ($p = 0.004$) but not including the stem group ($p = 0.089$). The same pattern is significant in 55 and 26 out 100 random trees for crown and stem group, respectively. The estimated marginal means of the rates of body size evolution versus time regression in Odontoceti are no different from the rest of the cetacean tree for the original tree ($p_{crown} = 0.118$, $p_{stem} = 0.378$), while resulting negative and significant in 13% (stem) and 24% (crown) of the random trees.

Brain Size Evolution – By scanning the phylogeny for significant shifts in brain size evolutionary rates, we found significantly higher absolute rates for Balaenopteridae (83% significant instances by *overfitRR*) and Ziphiidae (77%), and significantly smaller rates pertaining to the clades including Platanistidae and Eurhinodelphinidae (78%), and Physeteroidea (sperm whales; 23%), respectively.

Cetaceans as a whole show nearly significant (at the nominal alpha level = 0.05) trend for increasing brain size over time ($p = 0.060$; 37% significant instances by *overfitRR*). The rate of brain size evolution significantly increases trough time in the original tree ($p < 0.001$) and in 90 out of 100 random trees.

There is no evidence for a macroevolutionary trend in brain size in stem Odontoceti ($p = 0.260$; 1% significant instances by *overfitRR*) nor for crown Odontoceti ($p = 0.230$; 0%). However,

both stem and crown group Odontoceti show significantly higher estimated marginal means in the brain size versus time regression as compared to the other cetaceans. This is also true in 100% random trees for stem and in 2% for crown odontocetes.

The rate of brain size evolution does not change over time in toothed whales, either considering the stem ($p = 0.297$; 8% significant instances by *overfitRR*) or crown ($p = 0.092$; 8%) Odontoceti (Fig. 7). However, estimated marginal means in brain size evolutionary rates for both stem and crown Odontoceti are significantly lower than for other cetaceans (Fig. 7). The same applies to 93 (stem) and 95 (crown) out of 100 random trees.

Feeding Category – We computed differences between feeding categories in terms of brain size considering either 3 (Feed3) or 4 (Feed4) different categories. The group comparison produced no significant differences between feeding categories.

Comparing rates by state by means of *search.shift* performed under "sparse" condition for the Feed3 category, we found negative andsignificant shifts in rates of brain size evolution in raptorial feeders compared to the rest of the tree and to filter feeders ($p = 0.023$ and 0.024 respectively). The same analysis performed on Feed4 indicates filter feeders have significantly higher rates than longirostrine species ($p = 0.022$).

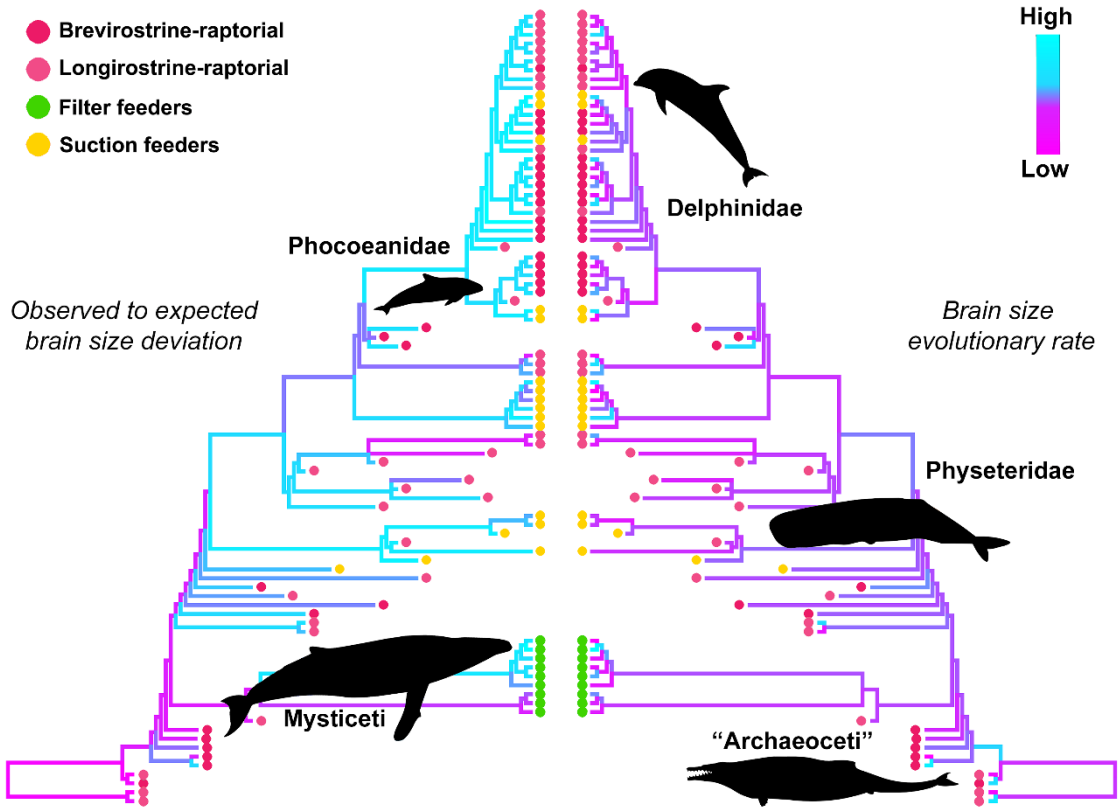

Figure 7. Cetacean brain size evolution along phylogenetic tree. Tree branches are colored according to the difference between observed brain sizes and the corresponding values predicted by regressing brain size versus body size (left), and to brain size evolutionary rate (right). Colored dots at tips represent the feeding categories. Image from Serio et al. 2019.

Discussion

We found evidence that Cetacean relative brain size increased through time. Since we found no evidence for body size increase over time (Cope's rule) for Cetacea, these results indicate a genuine tendency for increased encephalization in these marine mammals (Marino et al., 2007; Shultz & Dunbar, 2010; Gingerich, 2016).

Odontoceti show no trend for increased encephalization through time, yet they have distinctly higher brain size than expected by their body size as compared to other cetaceans. Although our findings comply with numerous reports attributing an outstanding encephalization

grade to odontocetes (using EQ), our data further indicate the absence of any significant trend in the rate of brain size evolution in this clade, and negative and significant shifts in the rate pertaining to river dolphins and sperm whales. In the former we further found a significant and positive shift in body size.

These results clearly indicate that the history of Odontoceti, and especially so stem Odontoceti, is characterized by high and conservative relative encephalization ever since the inception of the clade. Thus, rather than a macroevolutionary pattern, the large brain of toothed whales is best described as a plesiomorphic feature of the group.

The results are not explained by differences in feeding habits among cetaceans. In terms of evolutionary rates, raptorial longirostrine species tend to have low rates, and filter feeders tend to have high rates. This suggests that longirostrine species (which mostly include delphinidae plus a number of extinct species of several, unrelated clades such as *Albertocetus* and *Dalanistes*) started with relatively large, but conservative degrees of encephalization, that is the general pattern found among odontocetes, while Mysticeti tend to have comparatively lower but more variable brains sizes. We classified most river dolphins (*Inia, Pontoporia, Platanista, Lipotes*) as longirostrine (as in McCurry et al., 2017), and the filter feeder category coincides with Chaeomysticeti. Thus, the low brain size evolutionary rates in longirostrine species is entirely coherent with the negative shifts in the rate of brain size evolution we found in platanistoid Odontoceti and probably is a phylogenetic, rather than purely dietary, pattern.

Prominent reasons advanced for the evolution of comparatively large brain size in odontocete cetaceans are the development of complex intraspecific interaction in their social groups (Marino et al., 2007; Shultz & Dunbar, 2010) and echolocation (Mortensen et al., 2014). Establishing the degree of sociality for fossil species is challenging. Whereas some extinct odontocete species were reported as social (e.g. *Kentriodon*, *Dorudon*,), the lack of information for most extinct taxa prevents a formal analysis of the link between social group size and relative brain

size. Most river dolphins and rorquals are indeed solitary species (Shultz & Dunbar, 2010). Yet, cultural transmission via social bonds is present in some baleen whales at least (e.g. *Megaptera*).

Echolocation has been proposed as an explanation for the acquisition of large brain mass in Odontoceti (Mortensen et al., 2014; Churchill et al., 2018). Our results partially support this notion. We found consistent evidence that odontoceti are characterized by high and stable relative brain size. Importantly, the analysis of macroevolutionary rates indicate odontoceti are more variable in terms of body rather than brain size. Echolocation appears within stem Odontoceti (the earliest toothed whale able to echolocate was the Oligocene *Cotylocara macei*), suggesting echolocation might have prompted, or has at least coincided, with a consistent increase in relative brain size in Odontoceti (Churchill et al., 2018).

Our results suggest that Cetacea as a whole experienced a pattern for increased encephalization, while odontocetes always have had distinctly higher relative brain mass, soon acquired at the inception of the group. The absence of a sensible trend in body size for both Cetacea as a whole and for Odontoceti indicates the slow rate of brain size evolution in the latter and the coincident evolution of extremely large body size in Chaeomysticeti is not a viable explanation for the different relative brain size in the two groups. Echolocation remains the most probable candidate alternative to explain the outstandingly large odontocete brains.

McCurry et al. 2021 – Biological Journal of the Linnean Society, in press

After Serio et al. 2019, I joined a foreign team to expand the topic of brain size evolution in cetaceans by including brain and body size data for several archaic baleen whales. We first compared relative brain size (meant as log brain size versus log body size regression residuals) between stem cetaceans, odontocetes, and mysticetes via phylogenetic ANOVA. Next, I tested for temporal trends in either relative brain size per se or its evolutionary rates pertaining all cetaceans, mysticetes, and odontocetes by means of *search.trend*. Then, I applied *search.shift* twice, first to

automatically scan the phylogeny for possible shifts in relative brain size evolutionary rates occurring at clade level, then to detect possible shifts related to foraging complexity and sociality (used as category). For both *search.shift* and *search.trend*, rates were calculated by applying the multiple regression version of *RRphylo* on absolute brain size estimates, while taking into account body size as an additional predictor. Finally, I tested the robustness of our results to phylogenetic uncertainty and sampling via *overfitRR*.

We found both mysticetes and odontocetes have notably larger brains than stem cetaceans, but phylogenetic ANOVA produced no significant results ($p = 0.120$). There are significant trends towards greater encephalisation across all cetaceans ($p = 0.02$; 89% significant instances by *overfitRR*), mysticetes ($p = 0.05$; 49% significant instances by *overfitRR*), and odontocetes ($p = 0.05$; 42% significant instances by *overfitRR*), but for the latter two only if stem species are included (crown mysticetes $p = 0.24$; crown odontocetes $p = 0.16$). Evolutionary rates also increase across cetaceans, but marginally so and with poor support ($p = 0.05$; 18% significant instances by *overfitRR*). Better-supported increases characterize odontocetes ($p << 0.001$; 99% significant instances by *overfitRR*) and crown odontocetes ($p << 0.001$; 98% significant instances by *overfitRR*). By contrast, mysticetes show a poorly supported decrease in rates ($p < 0.01$; 66% significant instances by *overfitRR*), which again disappears when stem species are excluded ($p = 0.398$; 1% significant instances by *overfitRR*).

In 'clade' mode, *search.shift* identified a significant decrease in the rate of brain size evolution in crown ($p = 0.001$; 100% significant instances by *overfitRR*), and an increase within oceanic dolphins at the node uniting *Steno*, *Globicephala*, *Pseudorca*, *Grampus*, *Sotalia*, *Delphinus*, *Stenella* and *Tursiops* ($p << 0.001$; 95% significant instances by *overfitRR*). There are no significant shifts at or near the base of Odontoceti. In 'sparse' mode, there is no significant difference between riverine vs oceanic species ($p = 0.206$). By contrast, species aggregating in stable pods do show significantly higher rates of brain size evolution than those forming more ephemeral associations ($p = 0.998$; 88% significant instances by *overfitRR*).

Baleen and toothed whales are both more encephalized than stem cetaceans, but their respective origins are not marked by a significant increase in either relative brain size or the rate of brain size evolution. In addition, some of the latest archaeocetes (*Basilosaurus* and *Dorudon*) were under selective pressure towards greater encephalization. Large brains thus preceded, and seemingly were not influenced by, the appearance of echolocation in early odontocetes. Acoustic processing may still explain independent trends towards greater encephalization within some toothed whales – especially delphinoids, as suggested by their enlarged temporal lobe.

Dietary breadth is another possible driver of brain size (Muller & Montgomery, 2019), yet difficult to reconcile with the observation that stem cetaceans retained relatively small brains despite their seemingly varied diets (Fahlke et al., 2013). Nevertheless, encephalization at times may still reflect concurrent shifts in feeding mode. Thus, long-snouted snap-feeders like *Inia*, *Lipotes*, *Pontoporia* and *Platanista* and skim-feeders like *Balaena* and *Caperea* have relatively small brains, whereas those of rorquals increased in tandem with the evolution of lunge feeding. Such patterns may reflect differences in energy uptake between feeding strategies, which could plausibly constrain brain size.

Social complexity is a defining feature of many cetaceans, and along with behavioral richness may have coevolved with their large brains. Our results support this hypothesis insofar as species forming medium to large pods – all of them odontocetes, and most of them delphinids – appear to experience significantly higher rates of brain size evolution. Although mysticetes are usually regarded as more solitary, they are still capable of complex social interactions and behaviors. Likewise, long-term bonds between individuals and/or complex socio-sexual behaviors occur in sperm whales and at least some beaked whales and river. This widespread distribution suggests that sociality is deeply rooted in cetacean evolution, and as such plausibly contributed to the emergence of their large brains. Conversely, reductions in relative brain size might reflect secondary decreases in social complexity, perhaps correlated with feeding strategy (e.g. skim feeding in right whales and *Caperea*) or habitat (e.g. in ‘river dolphins).

The relative importance of factors like sociality and diet are still actively debated in primates, ungulates and carnivorans. As brains are responsible for a complex variety of functions, they are also under a complex array of selective pressures. This matches our results, which suggest that cetacean encephalization cannot be attributed to a single driver. Instead, echolocation, foraging complexity and sociality may all have played a role, albeit at different times and to varying degrees.

Melchionna et al. 2020 – Frontiers in Earth Science

The evolution of the human brain is one of the most intensely investigated topics in anthropology. Most studies on the subject matter focus on the achievement of our outstanding brain size; fewer more focus on brain shape, which is intrinsically hard to study given it takes producing skull endocasts of our ancestors that come short in numbers and are not always easily accessible (Holloway, 2018). Recent developments in virtual anthropology are now making fossil human endocasts a less rare commodity (Bruner et al., 2018), so that we are gaining scientific knowledge on our brain evolution at an unprecedented rate.

Recent studies on endocranial volume (i.e., the best proxy for brain size) in hominins invariably point to the presence of phenotypic leaps coinciding with the appearance of *Homo* (Du et al., 2018), although several lines of evidence indicate that not all *Homo* species belong to this "unusually big-brained" class of species, the latter being restricted to *Homo heidelbergensis*, *Homo neanderthalensis*, and *Homo sapiens* (Profico et al., 2017; Diniz-Filho et al., 2019a). These studies point to a non-gradual process of brain increase along the hominin lineage, probably prompted by the causal association between speciation and brain size (Du et al., 2018).

Our understanding of the evolution of brain shape in the human lineage might be experiencing a reverse trend. Despite logical enthusiasm around early findings illustrating an exquisitely human brain shape and level of brain asymmetry, it has been later noted that the typical brain shape in *H. sapiens*, which is characterized by a strong left-occipital right-frontal asymmetry

known as the Yakovlevian torque or occipital bending (Balzeau et al., 2013), is present to a degree in both fossil human species and great apes (Neubauer et al., 2020). This casts doubt on the link between brain asymmetry and properly human cognitive abilities and still suggests that the evolution of human brain shape is best viewed as a gradual process toward exaggerated asymmetry and large size (Neubauer et al., 2018). However, the observation by Xiang et al. (2019) that torque magnitude is independent of brain size variation within *H. sapiens* and repeated findings that brain asymmetry is more variable in humans than in apes (Neubauer et al., 2020) challenge this view and point to either a punctuational evolutionary event (Xiang et al., 2019) or a shift in the rate of evolution adding variability in humans (Neubauer et al., 2020).

Here, we used 3D geometric morphometrics to study brain size and shape evolution in Hominoidea using 123 cranial endocasts belonging to 19 different extant and fossil species, including *Australopithecus africanus*, *Homo ergaster*, *Homo erectus*, *H. heidelbergensis*, and *H. neanderthalensis*. We built a phylogenetic tree for the species in the study sample and applied *RRphylo* and *search.shift* to compute the rate of brain shape evolution and search for possible rate shifts. We investigated how rates of brain shape evolution correlate to levels of brain asymmetry and eventually tested for a possible significant allometric component in brain shape variation.

We found a negative rate shift in brain volume (per unit mass) coinciding with the genus *Hylobates* ($p_{two-tailed} < 0.001$). Interestingly, brain shape asymmetry shows a positive rate shift regarding the species of the genus *Homo* ($p_{two-tailed} = 0.979$). When the average brain volume is considered as a predictor in *RRphylo*, there is a positive and significant rate shift in brain shape evolution for the clade including *H. heidelbergensis*, *H. neanderthalensis*, and *H. sapiens* ($p_{two-tailed} = 0.999$) and a negative and significant rate shift for the Hylobatidae clade ($p_{two-tailed} < 0.001$). When brain asymmetry is considered as a predictor in *RRphylo*, the results do not change; there is a positive and significant rate shift in brain shape evolution for the clade including *H. heidelbergensis*, *H. neanderthalensis*, and *H. sapiens* ($p_{two-tailed} = 0.999$) and a negative and

significant rate shift for the Hylobatidae clade ($p_{two-tailed} < 0.001$). Eventually, the same two rate shifts still apply when shape is analyzed on its own, ignoring the allometric component.

There is positive allometric scaling between the brain shape and the endocast volume, either using the Brownian motion ($p = 0.001$) or allowing the evolutionary rates to change across the tree with *RRphylo* ($p = 0.001$), and significant correlation between endocast shape and asymmetry.

The pattern of low evolutionary rates in hylobatids is consistent with their evolutionary history and ecologies. The origin of Hylobatidae is placed at 21.8 Ma, and the radiation of the clade occurred between 6.4 and 8.0 Ma (Israfil et al., 2011). Israfil et al. (2011) suggest that such rapid radiation could be paired with biogeographic factors as the population dispersal and variations in the density of the forestal habitat (vicariant speciation). The low degree of body size differentiation and relatively minor ecological diversity when compared to great apes might be responsible for the apparently slow rate of brain shape evolution in this group.

Along the human lineage, the shifts in evolutionary rate in brain size evolution (Du et al., 2018; Diniz-Filho et al., 2019a) and increased rates of brain shape evolution (this study) coincide, thus suggesting that brain evolution followed a distinctive path joining *H. heidelbergensis*, *H. neanderthalensis*, and *H. sapiens*. This is entirely consistent with the archeological record of human behavior. Numerous studies on the anterior dentition in *H. heidelbergensis*, Neanderthals, and Paleolithic modern humans show a predominant right-hand frequency (> 90%), similar to the pattern for living humans attributed to right-handers. Studies of experimental production of lithic stone tools demonstrate that brain lateralization is involved in the production of late Acheulean tools or when participants watch Acheulean tool production, but not when participants make or watch others making Oldowan tools (Stout & Khreisheh, 2015). Uomini and Meyer (2013) suggested that the same areas for stone tool manufacture, of the Acheulean in particular, are lateralized and located on the left side in brain areas also involved in language processing and social learning.

Castiglione et al. 2020 – Evolutionary Biology

In this paper I presented an implementation of *RRphylo* specifically meant to estimate ancestral states taking advantage of explicit fossil information. In contrast to most other methods, under this new implementation, named *RRphylo-noder*, there is no expected distribution of trait changes during evolution (e.g. the normal distribution under BM), so that I expect *RRphylo-noder* could outperform competing methods in the presence of complex phenotypes. I tested this hypothesis by means of extensive simulations and real-case applications. I show that *RRphylo-noder* is faster and increasingly more accurate than competing methods in reconstructing ancestral states as phenotypic complexity increases.

Materials and Methods

RRphylo-noder integrates the phenotypic information at internal nodes in the estimation of evolutionary rates and ancestral character states. Given a vector ***n*** of phenotypic values known in advance to be placed at internal nodes (***fossil.states***), a vector of false tips ***ftips*** of length ***n*** is added to the tree. Each i_{th} element of ***ftips*** is phenotypically identical to the corresponding $\boldsymbol{fossil.states_i}$ and is attached to the tree at the position of $\boldsymbol{fossil.states_i}$ with a branch of length = 0. Then, the vector of regression coefficients $(\vec{\beta})$ is estimated by means of *RRphylo* by using the modified tree and phenotype (which include both ***ftips*** and the real tips). Since the branch lengths of ***ftips*** are equal to zero, the phenotypic rate between each $\boldsymbol{ftips_i}$ and the corresponding node is zero, which means the ***fossil.states*** and their corresponding ***ftips*** will have the same phenotypic estimates. After β coefficients are estimated, the vector of phenotypic values at nodes $\vec{a}$ is calculated as usual as:

$$\vec{a} = L'\vec{\beta}$$

where each row of *L'* represents the path of branch lengths moving from a specific node in the tree. The final step of the algorithm consists in removing ***ftips*** from the tree, and from the rate and phenotypic vectors.

I tested *RRphylo-noder* accuracy and compared it to other available methods for ancestral states estimation. The goals of the simulations were to assess 1) the accuracy of single-rate (i.e. fitting a single rate parameter for the entire phylogeny) and variable-rates (i.e. allowing the rate to vary across the tree) methods to ancestral state estimation with complex phenotypes (i.e. resulting from different phenotypic patterns applied to different parts of the tree), 2) the effect of sampling on ancestral state estimation and 3) the impact of using fossil phenotypes as ancestral character states known in advance. I used the function *ace* from the R package `ape` (Paradis & Schliep, 2019) to represent the simplest, most straightforward method for ancestral character state estimation under BM. The function *fastAnc* in `phytools` (Revell, 2012) is based on BM as *ace*, but additionally allows specifying phenotypic states at nodes. Elliot and Mooers' *StableTraits* (Elliot & Mooers, 2014) estimates ancestral states using a generalization of BM to the stable random walk, represented by a symmetrical, zero-centered distribution of phenotypic increments during the evolutionary time defined by the parameters α (the index of stability) and c (the scale). *StableTraits* performs the Bayesian estimation of ancestral states fitting the α and c parameters and allows the comparison with BM. As in *RRphylo-noder*, node priors can be used by grafting zero-branch length false tips to specified nodes. I tested both *RRphylo* and *RRphylo-noder* along with all of these other methods. In sum, I used two methods which do not allow to specify phenotypic values known in advance at nodes (*ace* and *RRphylo*), two single-rate methods (*ace* and *fastAnc*), and four methods which allow the evolutionary rate to change across the tree (which we collectively refer to as 'variable-rate' models: *StableTraits-Brownian, StableTraits, RRphylo* and *RRphylo-noder*). I tested such methods on forty different kinds of phenotypes to assess their performance under phenotypes differing in terms of complexity and under different sampling regimes.

Testing Ancestral State Reconstruction Methods Accuracy – The ultimate goal of any ancestral state estimation method is to predict phenotypes at nodes. Under perfect prediction, regressing the simulated phenotypes at nodes against their predictions originates a regression slope = 1 and intercept = 0. For each method, I calculated the slope and the intercept of the regression between simulated and fitted values, and the root mean squared error (*rmse*) of the regression. Under *fastAnc*, ancestral values are not fitted but given as phenotypic values at nodes known in advance. This means the more ancestral values are provided the lower *fastAnc*'s *rmse* will be. Thus, rather than a measure of the goodness of fit of *fastAnc*, its corresponding *rmse* depends on how many ancestral values are provided. For this reason, beyond *rmse* I calculated the *rmse* over the fitted nodes only (*reduced rmse*) in order to compare *fastAnc* to the other methods. Since *ace* and *RRphylo* make no use of ancestral values, I used only *rmse* to compare the ancestral state estimates of these methods. Mean *rmse* between methods was compared by means of repeated-measures ANOVA taking the kind of phenotype tested as the random effect.

To assess how simulated phenotypic complexity affects methods' performance, I averaged the 10 *rmse* (and *reduced rmse* as well) per type of phenotype, and collated *rmse*s for the different phenotypes from the smallest to the largest *rmse* as predicted by `ape`'s function *ace*. In this way, the forty *rmse* estimates were effectively ordered from the most similar, to the most dissimilar from BM, which is the evolutionary model *ace* is based upon. Then, I used categorical regression to calculate the slope of methods' *rmse* (the response variable) against the phenotypic kinds collated from the most similar to Brownian Motion (BM) to the most dissimilar used as the predictor variable. This way I estimated how *rmse* grows away from BM, for each method. I compared the regression slopes and estimated marginal means per method (Lenth, 2019). Ideally, the shallower the slope of the regression the less sensitive to the phenotype type a method is. Similarly, lower estimated marginal means indicate better prediction accuracy across phenotypes. This same procedure was repeated on *reduced rmse*, using *StableTraits-Brownian* predictions to collate phenotypes from the simples (i.e. most similar to BM) to the most complex.

Testing the Importance of Sampling and Phylogenetic Uncertainty – A perfect prediction of ancestral states values (i.e. corresponding to the slope = 1 and intercept = 0), could indicate that a method is particularly accurate but could also depend on a method being overfit. Overfitting is the major drawback for overparametrized methods such as *RRphylo* and *RRphylo-noder*. An overfit method may appear superior to other methods when assessed for prediction accuracy based on simulated data but could fail to capture the fundamental processes that led to the observed patterns in real data, which represent a subset of the real diversity of the clades, providing much reduced prediction accuracy.

To evaluate the potential for overfit, I applied ANOVA and post-hoc TukeyHSD test to assess whether the slope and intercept of the regression between observed and estimated ancestral states differ among sampling schemes. I similarly assessed whether the phenotypic deviation between known (simulated) and fitted ancestral states estimates change per sampling scheme. Phenotypic deviation was calculated as average percent deviation of the fitted versus simulated ancestral states.

I further measured the ability of *RRphylo-noder* to capture the processes producing the observed patterns in real data. In particular, we analysed the ability of *RRphylo-noder* to reveal the existence of phenotypic drift for the clades that were designed to be so, in spite of sampling. To this aim, for each clade selected, I calculated the intercept of the regression between the clade phenotypes and time (i.e. the distance of each tip from the tree root) and assumed these intercepts as the ancestral values for each clade, to be passed on to *RRphylo-noder* as known ancestral values. Then, I performed *search.trend* on the results of *RRphylo-noder* by indicating selected clades as individual nodes to be tested. Overfit should result in reduced power to retrieve the imposed phenotypic drift under subsampling. At the same time, this allows to test whether the use of known ancestral values increases the power of *RRphylo-noder* to retrieve the true evolutionary process simulated on the tree.

Real Cases – I tested the *RRphylo-noder* method on the evolution of body size in mysticete cetaceans.

Baleen whales are among the largest species ever lived. Yet, early Mysticeti include much smaller representatives (Fitzgerald, 2012). The sister group to baleen whales, the Oligocene Aetiocetidae, were toothed whales up to 8 m in body length. I assembled a cetacean phylogenetic tree from the backbone phylogenies in Montgomery et al. (2013) and Marx and Fordyce (2015). The composite phylogeny includes 116 species we had body size estimates for. Thirty-six species in the tree are extinct (10 archaeoceti, 23 odontoceti, 3 mysticeti). I tested whether baleen whale body size increased over time, in keeping with Cope's rule (Raia et al., 2012) by using *search.trend*. Then, I used *RRphylo-noder* setting the recently discovered *Mystacodon selenensis* (Lambert et al., 2017) body size as the mysticete most recent common ancestor prior. *Mystacodon selenensis* was almost the size of a bottlenose dolphin. It is considered a stem mysticete, perhaps sister to Llanocetidae (Fordyce & Marx, 2018). I therefore settled *Mystacodon* body size at 150 kg, which is typical for a bottlenose dolphin. I performed *RRphylo* and *search.trend* on the phylogeny as a whole, either with *Mystacodon* as ancestor of mysticetes, or without it. In addition, I performed, for the sake of comparison, both *RRphylo* and *search.trend* on the same tree and data, but removing extinct mysticeti. Body size estimates for the mysticete most recent common ancestor were further estimated by means of the functions *ace*, *fastAnc*, *StableTraits, StableTraits-Brownian* and *RRphylo*, always the full tree (i.e. with fossil species) without using *Mystacodon* body size as the ancestral value to Mysticeti.

To assess the potential for overfit, I applied *overfitRR* to test the effect of sampling on results produced by *search.trend* by specifying the Mysticeti clade to be tested for temporal trends in phenotypic (body size) mean and rates.

Results

Ancestral State Reconstruction Methods Accuracy – The regression of simulated versus fitted ancestral states gives slope consistently close to 1 and intercept close to 0 for all methods. ANOVA indicates that for 16 out of 40 different kinds of phenotypes there are significant differences among methods in terms of *rmse*. Among them, TukeyHSD indicates *StableTraits-Brownian*, *RRphylo-noder* and *StableTraits* are the most accurate methods overall, being selected 16, 15 and 15 times respectively among the best models. The methods *ace* (3) and *RRphylo* (1) still figure among the best candidate models. In terms of *reduced rmse*, only 4 times I found significant differences between methods. *StableTraits* was selected as the best method 4 times, *StableTraits-Brownian*, *RRphylo-noder* and *fastAnc* 2 times each. In general, both *StableTraits* methods and *RRphylo-noder* perform equally well, under a variety of sampling conditions and across different kinds of phenotypes. Variable rates methods consistently outperform single-rate methods. Only four times methods without specified ancestral values (ape's *ace* and *RRphylo*) perform as well as methods that allow for their specification.

As the complexity of the evolutionary process generating the simulated phenotypes increases, *RRphylo-noder* and *StableTraits*, in this order, performs best. The estimated marginal means of the regression indicate that all methods with prior phenotypic knowledge about specific nodes in the tree outperform methods with no such information and *RRphylo-noder* is the most accurate method overall, being the least sensitive to change in the type of phenotype simulated.

Assessing the Impact of Sampling – When the tree is subsampled, the slope and intercepts for variable-rates methods remain close to 0 and 1 respectively, when the starting phenotype was simulated according to BM, and slightly less than 0 and 1, respectively, when the starting phenotype was simulated according to BM with trend.

I tested whether sampling affects these parameters as well as the percent phenotypic deviation from the simulated parameters by means of ANOVA and post-hoc testing, using the

sampling intensity (removing 10% or 50% of the species) and sampling type (either by removing species at random or conditioning the sampling to be inversely proportional to species phenotypic value) as factors. The results indicate that slopes, intercepts and percent deviations from the original phenotypes are never statistically different with the random sampling, except for *StableTraits* when sampling intensity is 50%. With intense sampling (i.e. reducing tree size to 50% of the original tree) and the biased sampling design the slope and intercepts differ significantly from the unsampled tree and data for all methods. Importantly, all single-rate models perform worse, both in terms of slope and intercepts change across sampling levels.

I used *search.trend* to test whether sampling affects the probability to retrieve the correct structure in the data. ANOVA and post-hoc tests indicate there is no significant difference per sampling scheme and intensity. However, the use of ancestral values sensibly increases the possibility to find a phenotypic pattern at specific clades when it is real. Under different sampling conditions this increase in power is as high as 82.2% on average.

Real Cases – By applying *search.trend*, I found Cope's rule to apply to mysticetes, regardless of whether ancestral states are indicated as node priors or ignored. Yet, the regression slope increases adding *Mystacodon* body mass as the ancestor of Mysticeti (Fig. 8). The results are robust to the effect of sampling (97% and 74% instances of significant phenotypic trends are found with and without *Mystacodon* body mass as the ancestral value to all Mysticeti, respectively, by removing 25% of the tips randomly with *overfitRR*).

By applying *RRphylo-noder* the cetacean phylogeny produced an estimate of 150.04 Kg for the most recent common ancestor of Mysticeti, which is coincident with the *fossil.state* provided (i.e. the size of *Mystacodon selenensis*). The same estimate as calculated by *RRphylo* without *fossil.state* is 385.36 Kg. I derived for comparison the corresponding values as estimated by *ape* and *fastAnc* (which is 430.70 Kg and 457.85 Kg, respectively), and by *StableTraits*. At 150.02 kg.

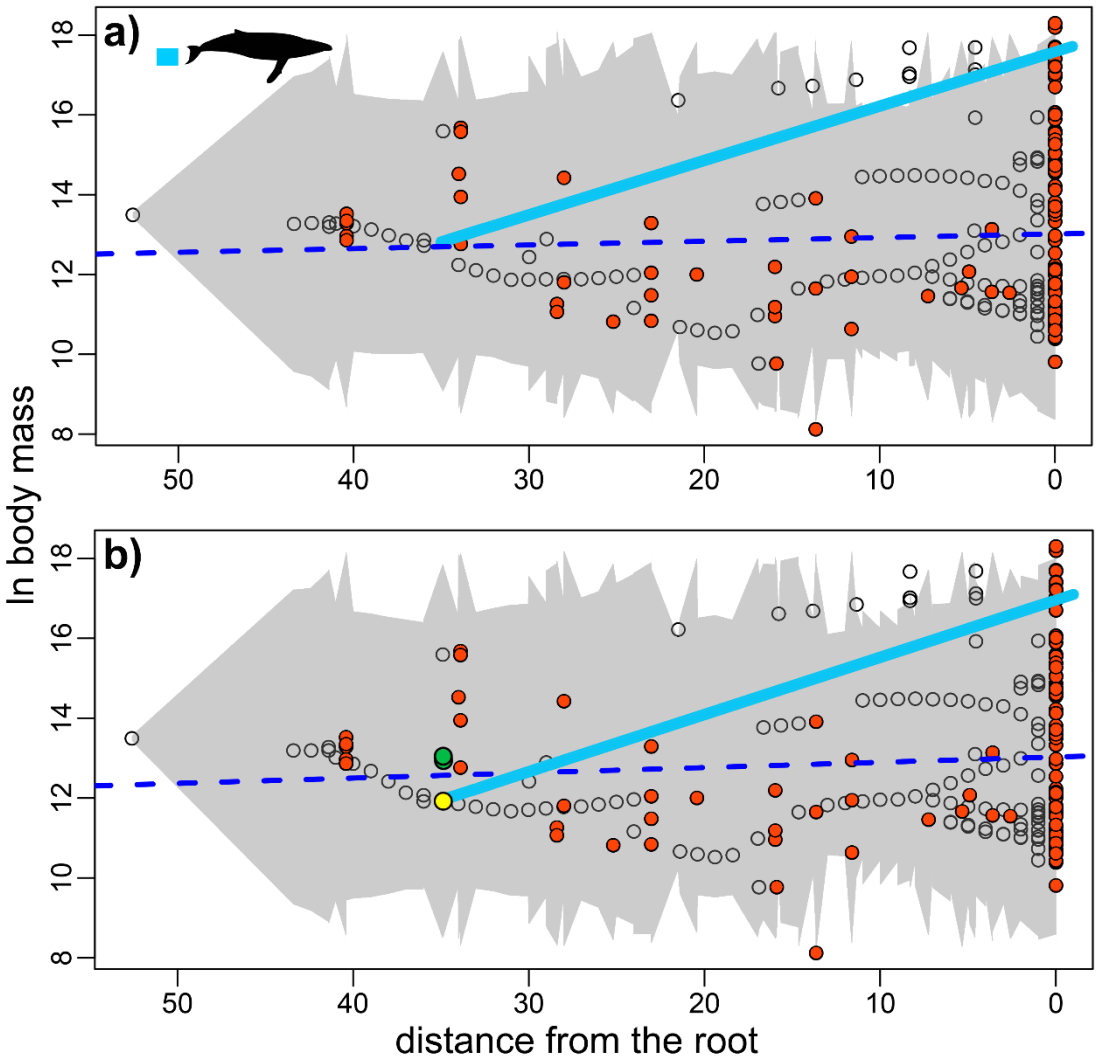

Figure 8. Cetacean body size versus time plots. White dots represent ancestral estimates at internal nodes, orange dots represent species phenotypes. The regression of phenotypes through time for the entire phylogeny is indicated by a blue dashed line. The phenotypic trend through time for Mysticeti is represented by the solid pale blue line. Upper row: cetaceans body size evolution according to *search.trend* as produced by considering fossil mysticetes, ancestral estimates are derived by the *RRphylo* method. Lower row: cetaceans body size evolution according to *search.trend* as produced by considering fossil mysticetes, ancestral estimates are derived by the *RRphylo-noder* method. The yellow dot represents the ancestral character for *Mystacodon* as estimated by *RRphylo-noder* and *StableTraits* (fitted by using the *StableTraits* software), green dots represent ace (fitted by using `ape` function *ace*), and *fastAnc* (fitted by using `phytools` function *fastAnc*). The ace estimate is the lowest. The y-axis is in ln-grams, time (x-axis) represents the distance from the cetacean tree root. Image from Castiglione et al. 2020.

Discussion

Variable-rates methods, that is *RRphylo-noder* and the two *StableTraits* models, consistently outperform all other methods in terms of ancestral states prediction accuracy. However, *RRphylo-noder* is the least sensitive to changes in the complexity of phenotypes and provides the smallest root mean squared error overall. The inclusion of known phenotypic values at internal nodes substantially increases ancestral states estimation accuracy for all methods and regardless of the sampling scheme. With *RRphylo-noder*, this further translates into a nearly twofold increase in the power to detect the phenotypic drift imposed to specific clades within the tree in the simulations.

The results demonstrate that variable rates methods are best suited to cope with tree and data generated under complex phenotypic processes that cannot be captured by abstract evolutionary models (Chira & Thomas, 2016). In *RRphylo* the phenotypic difference between any parent to descendant pair in the tree is fitted as a linear transformation proportional to the time intervening between the two according to a given slope (i.e. the elements of the evolutionary rates vector $\vec{\beta}$) while minimizing rate variation within clades. In contrast, other variable-rate methods refer to a single evolutionary model describing the distribution of phenotypic changes for the whole tree (e.g. the normal distribution in BM, the stable distribution in *StableTraits*). This makes *RRphylo* the least sensible to the actual shape of the distribution of phenotypic change across the tree.

Bayesian estimation of ancestral states, as currently implemented in *StableTraits* (Elliot & Mooers, 2014), accommodates for ancestral state estimation uncertainty and provide credible estimates when node priors are used. Herein, by using both simulations and application to real case I demonstrated *RRphylo-noder* performs at least as well as such Bayesian estimation approaches, and bears the advantage of being much faster (nearly twenty times faster according to our simulations) and less dependent to a specific distribution of phenotypic changes.

One potential problem with methods fitting many parameters at once is overfit. Overfit methods tend to perform very well with given data, but they often bear the potential to provide biased and much less precise estimation when data are subsampled. We applied either mild or strong subsampling to our tree and data in the simulations and to a real case (Mysticeti). We found that both *RRphylo-noder* and *StableTraits* are robust to sampling effects but for the most severe sampling design (i.e. removing half of the tree species proportionally to the species phenotypic values). In phylogenetic ridge regression (Kratsch & McHardy, 2014) a normalization factor λ is applied to avoid abnormally large phenotypic rate estimates, at the expense of prediction accuracy of tip (species) values. In *RRphylo*, λ maximum likelihood estimation is performed as to minimize the variance of rates within clades, so that phenotypes (and rates) tend to show phylogenetic signal. The structure of rate variation is constrained to maintain patterns of phenotypic evolution within clades and rates are treated as phylogenetically non-independent. I guess this is the reason why *RRphylo-noder* is robust to even strong sampling effects.

I found Cope's rule to apply to mysticetes. However, the phenotypic drift becomes much more evident when the bottlenose dolphin-sized *M. selenensis* is placed at the root of the Mysticeti clade (Fig. 8). The application of *RRphylo-noder* to these case study demonstrates the importance of using the fossil record to guide the recognition of phenotypic patterns under a PCM context, which has been pointed out several times in other studies (Hunt & Slater, 2016; Puttick & Thomas, 2015) and remains evident here also with methods other than *RRphylo-noder* (see Fig. 8).

Although ancestral states estimation is usually difficult and generally constrained within the limits of actual phenotypes at the tree tips, new approaches are being developed to provide more sensible estimates. Herein, we demonstrated that *RRphylo-noder* is virtually as powerful in fitting ancestral states as other available methods, and slightly more accurate in terms of fitting the true ancestral states when the description of phenotypic evolution is not reducible to a simple evolutionary model. This better accuracy probably depends on the fact that *RRphylo-noder* does not need to comply to the predictions of any abstract hypothesis about the tempo and mode of

evolutionary change. As such, I believe *RRphylo-noder* is the most appropriate in cases of complex phenotypic distributions. This is especially true by considering that competing methods using Monte Carlo Markov Chain (MCMC) approaches require much longer computational times.

Castiglione et al. 2021 – Biological Journal of the Linnean Society

In this study I presented a further extension of multiple phylogenetic ridge regression able to include additional predictor variables at the same time and to deal with discrete predictors. We applied such implementation to assess the effect of insularity, domestication, and sociality on the evolution of brain size in mammals.

Possessing a large, complex brain is typical of mammals (Herculano-Houzel et al., 2015). The evolution of such large brains is said to be promoted by a variety of factors, including high levels of sociality, living in demanding or rapidly changing habitats, the consumption of highly nutritious food and prolonged gestation length. These factors are commonly seen as selection agents favouring large-brained individuals (Benson-Amram et al., 2016). On the contrary, saving energy during growth by producing smaller brains could be adaptive where the importance of sense organs and anti-predator behaviour is de-emphasized, such as on islands (Köhler & Moyà-Solà, 2004) and in domestic species (Zeder, 2015).

It has been argued that the comparatively small brain of domestic mammals (Zeder, 2015) is acquired through rapid evolution via artificial selection. However, this conventional view has recently been challenged by Geiger et al. (2018), who found no evidence of differences in the rate of evolution in comparisons of domestic pig and dog skulls with those of wild boar and wolves, respectively.

Moreover, although it is often posited that natural selection favors the evolution of smaller brains in the extinct insular goat *Myotragus* and in the fossil hippopotami of Madagascar (Köhler & Moyà-Solà, 2004), insular dwarf elephants appear to have possessed extremely large brains for their

size. Likewise, the proposal that body size in insular mammals evolves quickly (Lister, 1989; Millien, 2011) has been contested on the basis of phylogenetic analyses (Raia et al., 2010a; Raia & Meiri, 2011), lending weight to the assertion that the rate of brain size evolution in insular species might also differ little from that of mainland populations. Consequently, the existence of a trend towards the evolution of smaller size or of faster rates of brain evolution in insular mammals is now contested.

A further factor that might influence the evolution of brain size is an increased level of sociality to facilitate the development of complex networks of relationships with conspecifics. This 'social brain hypothesis' (Dunbar, 2009) has been supported by studies of several mammalian groups (Dunbar & Shultz, 2007; Holekamp et al., 2015) and birds (Lefebvre, 2013). However, there are at least some clear exceptions, e.g. solitary tigers have larger brains compared with the brains of social lions. If the opportunity for and intensity of social learning (presumed to prompt the evolution of larger brains; Lefebvre, 2013) are correct, then we would expect that brain size would scale positively with the number of individuals within social groups (Dunbar & Shultz, 2007). However, the formation of particularly large and unstable herds of hundreds of individuals might provide no opportunity to learn complex behaviors and memorize past interactions with conspecifics. Consequently, brain size evolution might be a non-linear function of sociality, with comparatively larger brains and faster rates of evolution accruing to species with intermediate levels of sociality.

We ask whether domestication, insularity, and sociality have increased or slowed down the pace of brain size evolution in mammals and what the effects of social level are on it, focusing on both the rate and the direction of evolution of brain size.

Materials and Methods

Materials – We collected data from published articles to create a dataset inclusive of brain and body mass for 426 taxa: 178 artiodactyls (six extinct, 172 extant), 26 perissodactyls (four

extinct, 22 extant), ten proboscideans (eight extinct, two extant) and 212 carnivores (all extant). Seventy-two taxa represent domestic breeds, including pig, horse, cow, sheep, goat, cat and dog breeds. The 354 remaining taxa are wild species.

To perform phylogenetic ridge regression, we modified the backbone phylogeny published by Raia & Meiri (2011), adding domestic breeds to their wild relatives as polytomies. The phylogenetic tree was calibrated by using the function *scaleTree* in the R package `RRphylo`. *scaleTree* allows the tree branch lengths to be tuned by imposing specific ages at given nodes and for the terminal leaves (species).

We considered as insular species those occurring exclusively on islands, provided the island was smaller than Australia, i.e. up to the size of New Guinea (785 753 km^2). The insularity category includes 19 species, both extinct (i.e. Elephas falconeri) and extant, and 407 mainland species (Raia et al. 2010a, Raia & Meiri 2011).

Species in the tree were classified further according to four categories, representing increasing levels of sociality according to Walker's Mammals of the World (Nowak & Walker, 1999). Species whose individuals spend most of their lifetime as solitary (or as solitary mothers and their offspring) were classified as 'solitary'. Species living in groups of up to ten individuals for most of their lives were placed in the 'family' category. Species whose individuals live in groups of 11–30 individuals were categorized as members of a 'group'. Finally, species whose individuals live in groups of > 30 individuals were considered as members of a 'herd'. The criteria we used for categorizing social groups are somewhat arbitrary, given that group size is not fixed and can vary depending on the breeding season and changes in the availability of resources. Thus, we produced a second categorization, whereby species living in pairs (of a breeding male and a female) staying together outside the breeding season are tallied as 'pair'. One further potential problem with the sociality classification is that domestic animals live in unnatural groups confined to closed areas for husbandry. Therefore, we repeated the sociality analyses excluding domestic breeds.

Phylogenetic Multiple Regression with RRphylo – I implemented *RRphylo* to include additional predictor variables at the same time and to deal with categorical predictors. To test the accuracy of *RRphylo* at deriving sensible estimates of ancestral states for categorical variables, I compared the outcomes of *RRphylo* with existing methods to estimate ancestral categorical states.

Multiple regression *RRphylo* computes the rates of brain size evolution accounting for body size. Although this is sound to study rates (i.e. the evolutionary tempo), the method does not account for the effect of the predictor on brain size (i.e. the evolutionary mode). With this aim, for each factor, we assessed differences in brain size among categories while accounting for phylogenetic effects by means of phylogenetic ANOVA (Revell, 2012).

Differences in evolutionary rates of brain size among categories within individual factors were assessed by means of *search.shift*. For each factor, we computed rates of evolution of brain size by performing the multiple regression version of *RRphylo* using the natural logarithm of brain size as the response and the natural logarithm of body size along with the factor as the predictors. The result of each *RRphylo* was fed to *search.shift* to look for differences in rate among different categories for each factor (i.e. domestication, insularity and sociality). In the case of sociality, the analyses were repeated on the tree deprived of domestic species.

Results

Evolutionary Tempo of Brain Size Evolution – The absolute rates of evolution of brain size computed by accounting for the effect of body size and domestication are significantly different between wild and domestic species, with the latter evolving more quickly towards smaller brain size (i.e. negative rate difference, two-tailed p-value = 0.001).

The comparison of rates of evolution of brain size between insularity categories (computed in a multiple regression framework, with body size and insularity as predictors) returned no significant difference (two-tailed p-value = 0.927).

Regarding the rates of evolution of brain size computed by accounting for the effect of body size and sociality, the rate of 'solitary' species is significantly lower than the rate obtained for the rest of the tree (two-tailed p-value = 0.001). The reverse characterizes species living in family groups (two-tailed p-value = 1). Species belonging to the 'group' category shows no significant absolute difference in rate from the rest of the tree (two-tailed p-value = 0.502). The absolute brain size evolutionary rates for 'herd' species are not significantly different than the rates of the rest of the tree (two-tailed p-value = 0.427).

Comparisons of absolute rates between social categories revealed that 'solitary' species evolve more slowly and 'family' species faster, in terms of brain size, compared with the rest of the tree. Brains for the species in the 'family' category also evolved faster than those of the 'group' and 'herd' species. The addition of the 'pairs' level in social category 2 did not change these results. Brains of the species in the 'family' category, however, evolved faster than in 'pairs' species.

When domestic species are excluded, 'family' species remain the fastest (and 'solitary' the slowest) in terms of the rate of evolution of brain size.

Evolutionary Mode of Brain Size Evolution – The phylogenetic ANOVA performed on the natural logarithm of brain size comparing domestic vs. wild species produced no significant difference (p-value = 0.373). The comparison of the natural logarithm of brain size between insular and mainland species is not significant (p-value = 0.946).

Brain size is smaller, on average, for solitary than for social species. Results of phylogenetic ANOVA indicated that there is a significant difference in brain size among social categories. The pairwise comparison between social groups produced significant differences between 'group' and 'herd' species compared with 'solitary' taxa (p-value = 0.006 and 0.020, respectively). By using social category 2, this also applies when comparing 'group' and 'herd' species with 'pairs' taxa (p-value = 0.010 and 0.010, respectively).

When domestic species are excluded, global differences in brain size among social categories are still apparent, meaning that there is an increase in brain size with increasing levels of sociality.

Discussion

Our findings show that domestication and sociality have profound and significant effects on rates of evolution of brain size, whereas insularity does not. To our knowledge, this is the first study to demonstrate quantitatively that artificial selection (domestication) drives faster rates of brain evolution than does natural selection at the macroevolutionary level, after controlling simultaneously for body size and phylogeny. This contrasts with the recent study which found that evolutionary rates did not differ significantly between domestic and wild animals (wild boar/pig and wolf/dog; Geiger et al., 2018).

On the basis of brain size, the median *RRphylo* rate in domestic species is negative (−0.0012). Given that *RRphylo* rates are regression slopes between consecutive nodes in the phylogeny, this indicates that there is an average tendency for brain size to decrease in domesticated mammals, whereas in wild species the median rate is positive (0.0024). Although domestic breeds tend to have smaller brains than their wild relatives, this is not equally evident across all domesticated species, and the reduction is much more apparent in artiodactyls (i.e. livestock; Heck et al., 2018). This inconsistency explains why brain size does differ not significantly between domestic and wild forms, according to phylogenetic ANOVA. One potential factor confounding the difference between wild and domestic species in terms of brain size is sociality and the type of interaction that domestic breeds have with humans. Most livestock gather in herds, exposed to reduced social stimuli, and are sometimes even killed for commercial purposes before maturity, whereas pet animals (i.e. dogs and cats in our dataset) are commonly raised in continuous interaction with their owners, implying richer psychological experiences during growth for these carnivores, in comparison to the herbivorous species (Heck et al., 2018). However, other studies

have found that, with the exception of pigs, the reduction in brain size in domestic animals does not differ between carnivores and herbivores (Zeder, 2015). Furthermore, the extent of reduction of brain size in domestic sheep, goats and horses is difficult to determine with our data. This is because, for these clades, it is difficult to identify 'wild' representatives for which gene flow with domestic individuals was limited. Consequently, although our results are consistent with the traditional view that the rate of evolution of brain size is faster in domestic species, the lack of significant phenotypic difference might be influenced by limitations in the dataset.

Mammalian species confined on islands often live in species-poor ecosystems and tend to be herbivores. An apparently frequent set of phenotypic transformations shared by these insular 'ungulates' is the loss of anti-predator behaviors and reductions in body and brain size (Köhler & Moyà-Solà, 2004). Although clearly common, the tendency toward brain size reduction is reversed in insular dwarf elephants and, most notably, does not apply to Homo floresiensis, the smallest-brained hominid ever described (Diniz-Filho & Raia, 2017; Diniz-Filho et al., 2019b). These results are consistent with recent evidence that insular vertebrates do not evolve to either extreme of body size compared with their mainland relatives (Meiri et al., 2011), nor do they achieve comparable evolutionary rates (Raia & Meiri, 2011).

We found that sociality exerts a strong influence on the trajectories and rates of brain size evolution. 'Family' species consistently present the highest rates of evolution of brain size. Shultz & Dunbar (2006) observed that although variation in brain size in 'ungulates' conforms with the social brain hypothesis, large, commonly ephemeral ungulate herds leave little room for the formation of long-lasting social bonds, perhaps explaining why herd species are the least influenced (together with solitary species) by the 'social brain effect'. Our results are entirely consistent with this proposition and are supportive of the positive link between complexity of social interactions and brain evolution that is at the core of the social brain hypothesis.

Castiglione et al. 2019 – PloS one

In this paper I presented *search.conv*, a method to evaluate instances of evolutionary convergence in a phylogenetically explicit context. I showed through simulations that *search.conv* is remarkably powerful and fast. It does not require the convergent clades to be phenotypically unusual as compared to the rest of the tree. In addition, it has low Type I error rates (false positives).

We applied *search.conv* to three well-supported cases of morphological convergence, namely the independent adaptation to grazing in perissodactyl and artiodactyl mandibles, the evolution of the sabertooth morphology in machairodont cats and barbourofelids, and the evolution of distinct ecomorphs by Caribbean Anolis.

Materials and Methods

Dealing with multivariate data, each species at the tree tips is represented by a phenotypic vector, including one entry value for each variable. Naming **A** and **B** the phenotypic vectors of a given pair of species in the tree, the angle θ between them is computed as the inverse cosine of the ratio between the dot product of **A** and **B**, and the product of vectors sizes:

$$\theta = arccos \frac{A \bullet B}{|A||B|}$$

The cosine of angle θ actually represents the correlation coefficient between the two vectors (Zelditch et al., 2012). As such, it exemplifies a measure of phenotypic resemblance (Adams & Collyer, 2009). Possible θ values span from 0 to 180 degrees. Small angles (i.e. close to 0°) imply similar phenotypes. At around 90° the phenotypes are dissimilar, whereas towards 180° the two phenotypic vectors point in opposing directions (i.e. the two phenotypes have contrasting values for each variable). For a phenotype with *n* variables, the two vectors intersect at a vector of *n* zeros.

Under the Brownian Motion (BM) model of evolution, the phenotypic dissimilarity between any two species in the tree (hence the θ angle between them) is expected to be proportional to the

age of their most recent common ancestor. Under convergence, this expectation is violated and the angle between species should be shallower than expected by their phylogenetic distance. I developed *search.conv* specifically to calculate θ values and to test whether actual θs between groups of species are smaller than expected by their phylogenetic distance. The function tests for convergence in either entire clades or species grouped under different evolutionary 'states' (Fig. 9).

Given two monophyletic clades (subtrees) **C1** and **C2**, *search.conv* computes the mean angle θ_{real} over all possible combinations of pairs of species taking one species per clade. This θ_{real} is divided by the patristic (i.e. the sum of branch lengths) distance between the most recent common ancestors (mrcas) to **C1** and **C2**, **mrcaC1** and **mrcaC2**, respectively (Fig. 9), to account for the fact that the mean angle (hence the phenotypic distance) is expected to increase, on average, with phylogenetic distance. To assess significance, *search.conv* randomly takes a pair of tips from the tree (*t1* and *t2*), computes the angle θ_{random} between their phenotypes and divides θ_{random} by the distance between *t1* and *t2* respective immediate ancestors (i.e. the distance between the first node *N1* above *t1*, and the first node *N2* above *t2*). This procedure is repeated 1,000 times generating θ_{random} per unit time values, directly from the tree and data. The θ_{random} per unit time distribution is used to test whether θ_{real} divided by the distance between **mrcaC1** and **mrcaC2** is statistically significant, meaning it is smaller than 5% of θ_{random} values the two clades are said to converge.

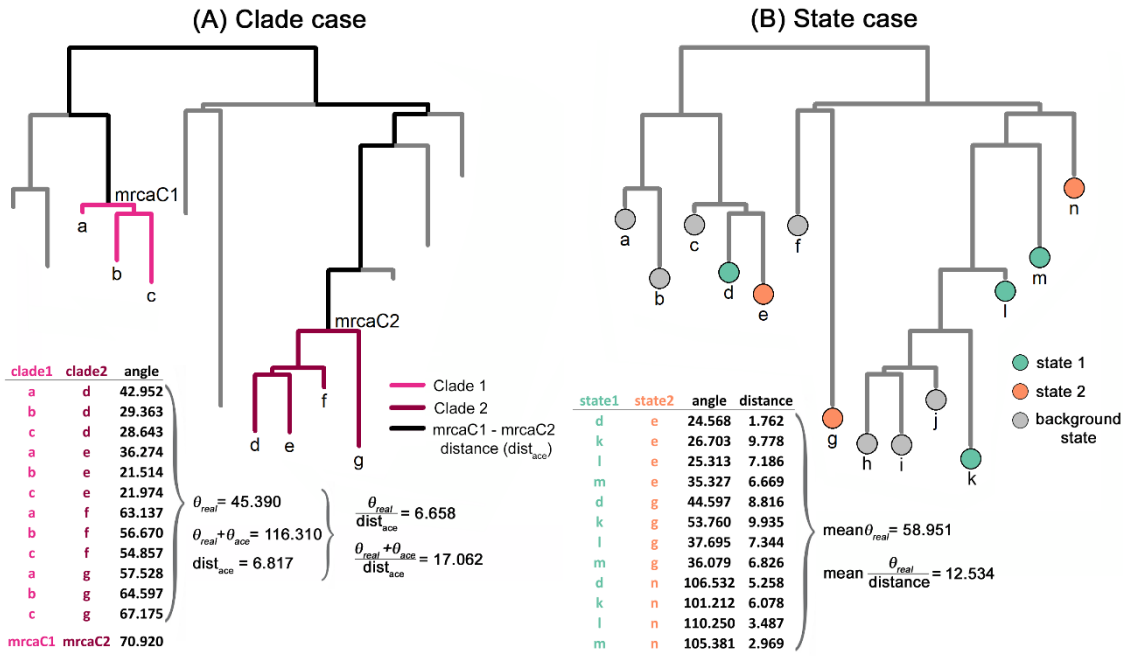

Figure 9. Hypothetical example illustrating how the *search.conv* function algorithm works. In the clade case (A), given any two monophyletic clades in the tree, the mean angle averaged over all possible combinations of two species (one per clade) is computed. This θ_{real} angle is divided by the distance between the most recent common ancestors to the respective clades, mrca1 and mrca2. Significance is assessed by comparing the result of this procedure to 1,000 randomly generated angles θ_{random} computed between species extracted by chance from the tree, divided by their respective distances. Angles are further computed between phenotypes at the mrcas. These θ_{ace} angles are summed to the corresponding θ_{real} to test whether convergence was already present at the beginning of clade history. Ancestral phenotypes are either estimated by *RRphylo* or provided by the user according to the fossil record. In the state case (B), θ_{real} are computed as in the clade case, but taking the mean angle between each combination of pairs of species (taken one per state), divided by their distance. Image from Castiglione et al. 2019.

When testing convergence, researchers typically have species phenotypes and, ideally, a phylogenetic tree representing their relationships. This means that while it is usually possible to test convergence among species, it is generally not possible to identify entire clades evolving under convergence. In the real world, these clades actually coincide with **mrcaC1** and **mrcaC2** and their descendants. In *seach.conv*, I devised a strategy to identify **mrcaC1** and **mrcaC2**. In practice, given a pair of candidate nodes tested for the initiation of convergence, the phenotypes at **mrcaC1** and

mrcaC2 are estimated by *RRphylo*, and the angle between the ancestral states (θ_{ace}) is calculated. Then, θ_{ace} is added to θ_{real} and the resulting sum divided by the distance between **mrcaC1** and **mrcaC2**. The sum $\theta_{ace} + \theta_{real}$ should be small for clades evolving from similar ancestors towards similar daughter phenotypes. Importantly, a small θ_{ace} means similar phenotypes at the mrcas of the two clades, whereas a small θ_{real} implies similar phenotypes between their descendants. It does not mean, though, that the mrcas have to be similar to their own descendants. Two clades might, in principle, start with certain phenotypes and both evolve towards a similar phenotype which is different from the initial shape. This means that the two clades literally evolve along parallel trajectories. Under *search.conv*, simple convergence is distinguished by such instances of convergence with parallel evolution. The former is tested by looking at the significance of θ_{real} . The latter is assessed by testing whether the quantity $\theta_{ace} + \theta_{real}$ is small (at alpha = 0.05) compared to the distribution of the same quantity generated by summing the θ_{random} calculated for each randomly selected pair of species *t1* and *t2* plus the angle between the phenotypic estimates at their respective ancestors *N1* and *N2* divided by their distance.

As with many other methods concerned with testing convergence (e.g. Harmon et al., 2005; Revell et al., 2007), the *search.conv* function suffers from the problem that ancestral states estimation entirely depends on the phylogenetic tree and data at hand and the evolutionary model used to fit the states. To help addressing this issue, under *search.conv* phenotypes at the nodes can be indicated directly by the user, when there is a specific hypothesis (i.e. real fossil specimens) about the phenotype of the most recent common ancestor to the clades. This is useful since the inclusion of fossil information increases power and reliability of comparative analyses of trait evolution (Slater et al., 2012).

Under *search.conv*, instances of convergence may be either assessed under the 'automatic mode' or specifying candidate node pairs. By default, *search.conv* runs the former, testing all clade pairs which are at least as distant as a one tenth of the tree size, counted as the number of nodes between their most recent common ancestors (i.e. clades 10 nodes apart for a 100 species tree).

Alternatively, a time, rather than number of nodes, distance could be specified. Although any minimum distance can be specified, it must be reminded that testing too many node pairs at once potentially introduces Type I error inflation. We empirically found that this just becomes a problem by testing very small clades in very large trees. With the default option (i.e. nodes that are at least as distant as a one tenth of the tree size) Type I error inflation is negligible. As detailed below, I assessed the effect of phylogenetic distance on *search.conv* Type I and Type II error rates. The expectation is that the closer the clades are on the tree, the harder it becomes to find convergence, as phenotypic similarity is best explained in this case by phylogenetic proximity.

Several candidate node pairs could map on the same region of the tree, because phenotypic values in close nodes are strongly autocorrelated (for instance, a candidate node pair could be represented by nodes n1 and n2, and another by the pair of nodes immediately bracketing n1 and n2). For each candidate node pair representing a statistically significant signal for convergence, *search.conv* performs the analysis of multivariate homogeneity of group dispersions, calculates the average distance from group centroids for individual species in the clades, and orders candidate and significant node pairs (if they are > 1 in number) from the least variable to the most. The rationale is that under convergent evolution, species phenotypes are expected to deviate the least from group centroids, at least when the convergent states represent evolutionary attractors (Beaulieu et al., 2012).

The clade-wise approach I have described so far ignores instances of phenotypic convergence that occur at the level of species rather than clades. The *search.conv* function is also designed to deal with this case. To do that, the user must specify distinctive 'states' for the species presumed to converge. The function will test convergence between any pair of given states. The species ascribed to a given state may belong anywhere on the tree or be grouped in two separate regions of it, in which case two states are indicated, one for each region. The former design facilitates testing questions such as whether all hypsodont ungulates converge on similar shapes,

while latter aids in testing questions such as whether hypsodont artiodactyls converge on hypsodont perissodactyls.

If provided with such 'states' *search.conv* will calculate the mean θ_{real} between all possible species pairs evolving under a given state (or between the species in the two states presumed to converge on each other). The θ_{random} angles are calculated by shuffling the states 1,000 times across the tree tips. Both θ_{real} and individual θ_{random} are divided by the distance between the respective tips.

Testing Procedures – I assessed the power of *search.conv* using both simulation experiments and real cases. The first set of simulations reproduces the existence of phenotypically similar clades or species in distant regions of the tree. This corresponds to the traditional observation of entire clades converging towards similar ecomorphologies (e.g. adaptation to durophagy in the mandible and skull of borophagine canids and modern hyaenids, body shape in ichthyosaurs and dolphins). I performed *search.conv* on such kind of phenotypes under the 'automatic mode' first, and then by indicating the target nodes. A similar set of simulations was performed by reducing the distance between clades set to converge. In this latter case, the power of *search.conv* is expected to decrease because the phenotypic similarity between clades so close to each other is best explained by phylogenetic proximity rather than by phenotypic convergence.

Another set of simulations was devised to test for convergence among groups of species evolving under a single state or between states.

We tested three real cases for possible instances of morphological convergence. They represent well-supported instances of morphological convergence during the evolution of the mammalian mandible (cases 1–2) and the colonization of the Caribbean islands by the lizard genus *Anolis* (case 3).

The first case concerns felids. Felids fall in two major ecotypes. Pantherine and feline cats possess robust, conical upper canines. A second ecotype was present in two extinct clades within the cat family, i.e. machairodonts and barbourofelids. The latter is the sister group to true felids.

Machairodonts include three tribes, one with short and not particularly flattened upper canines, the Metailurini, a second with long, flattened upper canines often possessing crenulated margins, the Homotheriini, and a last tribe with exceptionally long, extremely flat upper canines with smooth margins, the Smilodontini. Smilodontini are the sister clade to Metailurini. Both Homotheriini and Smilodontini are "true" sabertooths (Christiansen, 2007). The true sabertooth cats and barbourofelids present highly derived mandibular morphologies, specialized to confer these cats their unique killing behavior, including reduced dentition, low coronoid and condyle processes and protruding incisors (Christiansen, 2008). We tested whether mandibular shape in the extinct sabertooth cat clade Machairodontini converges on mandibular shape in Barbourofelidae (the sister clade to all felids which is usually referred to as 'false' sabertooth cats). We used geometric morphometric data and the tree published in Piras et al. (2018). The geometric morphometrics data included 83 species and 711 specimens, and we chose 10 landmarks and 23 semi-landmarks to record the mandibular shape. We ran this experiment with the 'automatic' procedure of *search.conv* (i.e. without specifying which clades to be tested). We further explored the potential effect of specifying ancestral states in finding morphological convergence by applying *search.conv*. To this aim, we repeated the analysis by setting the ancestral mandibular phenotype of barbourofelids and machairodonts to be equal to *Barbourofelis fricki* and *Smilodon fatalis*, respectively.

We compared *seach.conv* to an existing method sought to address morphological convergence embedded in the R package `convevol` (Stayton, 2018). To this aim, we performed both *search.conv* (under the 'state' condition) and *convratsig* (Stayton, 2018) by collapsing barbourofelids and sabertoothed cats under a single state.

The second case study was based on hooved mammals (Ungulatomorpha). Hooved mammals fall into two major feeding categories, that is browsing on soft vegetable matter, and grazing on harder vegetable material, typically grasses, whose leaves are rich in silica and therefore result in increased wear rate of the molar tooth crowns. Browsing is typical of most Palaeocene and Eocene 'ungulates' and persists today in most deer, tragulids and other small-bodied forms. With

the emergence of grasslands and the spread of grasses, the inclusion of grasses in the diet became widespread in herbivorous mammals (Figueirido et al., 2012) and is responsible for the rapid diversification of grazing artiodactyls. In morphology, the dietary shift from soft (browsing) to hard and fibrous (grazing) plant material is accompanied by profound changes in the skull and mandible, including the acquisition of high-crowned (hypsodont) molars, longer snout, and deeper mandible (Raia et al., 2010b). This pattern is present in equids, and also appeared several times among Pecora. Nonetheless, true grazing is restricted to a minority of species, most of them being properly defined as mixed-feeders consuming both grasses and soft material.

The data were obtained from 353 images in lateral view taken from the scientific literature or directly from specimens, representing 205 species. On each image we recorded nine landmarks to register mandibular shape and analyzed shapes by means of geometric morphometrics. We considered individual species as either grazing artiodactyls, grazing perissodactyls, or "others" (i.e. non convergent) depending on their molar morphology (i.e. degree of hypsodonty) and tested whether grazing ungulates from different parts of the tree converged on similar mandibular morphologies by using the 'state' approach.

The third real case pertains to extant lizards of the genus *Anolis*. The genus includes more than 400 species distributed in the Neotropical region and the Caribbean. Insular anoles fall into six distinct ecomorphs which have been intensely studied as a classic example of convergent evolution (Losos, 1992). The data include a 100 species wide tree for *Anolis* lizards living on the main islands of the Greater Antilles, and 11 phylogenetic principal components extracted analyzing lizards body shapes (Mahler et al., 2013). Six species do not fall into any ecomorph category and are therefore not expected to converge.

Results

Despite great variation in simulated convergent clade size, distance and height, under the automatic mode the Type II error (the rate of false negatives) is as low as 6%. Type I error (false positive) rate is similarly low at 4%.

As expected, by repeating the simulations with clades separated by only three nodes, Type I error is 0%, whereas Type II error increases to 54%. These results indicate that *search.conv* does not find convergence between clades that are very close to each other on the tree, whose phenotypic resemblance is best explained by phylogenetic proximity rather than convergence.

The power of *search.conv* to correctly identify the convergent clades when they are specified by the user (i.e. both θ_{real} and $\theta_{real} + \theta_{ace}$ are significant) is 71%. However, considering cases when species phenotypes (θ_{real}) are found to be significantly convergent but $\theta_{real} + \theta_{ace}$ is not, the identified mrcas for the clades found to converge were correct 88% of the time, within 2 nodes distance from the convergent clades' mrcas. *search.conv* often identifies nodes which are very close to the 'real' mrcas rather than the 'real' mrcas themselves. I found this usually depends on the balancing between the clade set to converge and its sister node, and the strong phenotypic autocorrelation between these clades (because a given clade necessarily includes all of the descendants of its daughter node). Whichever exact mrca pair is identified, 97.5% of the species set to converge are, on average, found to do so.

The Type I and Type II error rates of *search.conv* (automatic mode) are little influenced by how the phenotypes are simulated. The Type I error (the percentage of false positives) remains remarkably low.

By using the 'state' specification, the Type I error rate is 5%, either within or between states. Type II error of search.conv is 1% when testing for convergence within a group and 6% testing two different states for convergence on each other.

Felid Mandibles – When testing for convergence between clades, we found two instances of convergent morphological evolution, both pertaining the same clade, Barbourofelidae. The latter includes false saber-toothed cats of the genera *Barbourofelis* and *Albanosmilus*. They were found to be convergent on both Smilodontini and Homotheriini within machairodonts, which represent the true sabertoothed cats. It is noteworthy that *search.conv* effectively failed to find convergence between barbourofelids and Metailurini, which form a clade of machairodont cats sister to Smilodontini but did not possess the full sabertooth morphology. The mean angle between barbourofelids and Smilodontini is 29.93 degrees. The angle between their ancestors is 21.50 degrees. Both θ_{real} and $\theta_{real} + \theta_{ace}$ are statistically smaller than expected by chance ($p = 0.009$ for both). This suggests that the two clades evolved along parallel trajectories. The angle between barbourofelids and Homotheriini is 43.09 degrees, the angle between their reconstructed ancestors is 39.09 degrees, and both θ_{real} and $\theta_{real} + \theta_{ace}$ are statistically significant ($p = 0.019$ and 0.011, respectively). The computational time was 145 seconds.

By using the mandibular shapes of *Barbourofelis fricki* and *Smilodon fatalis* as the ancestral states to all barbourofelids and machairodonts, respectively, the results are similar to those obtained without specifying phenotypes at the mrca nodes, and this may help explaining the good performance of *search.conv* in finding the correct position, hence the true identity, of converging clades.

By performing the analysis collapsing machairodonts and barbourofelids under a single state, *search.conv* produced a small and significant mean angle (19.93 degrees, $p = 0.001$) between convergent species. The computational time was 44 seconds. This latter analysis performed by using *convratsig* (Stayton, 2018) produced significant results for all the measures. The computational time was 21h 48’ 7”.

Grazing Ungulate Mandibles – We performed *search.conv* once taking grazers as a single group, then considering grazing artiodactyls and grazing perissodactyls separately.

The mean angle between all grazers collapsed under a single state is 69.62 degrees. This is significant at p = 0.041. The mean angle between grazing artiodactyls and grazing perissodactyls is 77.53 degrees. Although large, we found this angle is less than expected by chance (p = 0.001). In fact, the angle θ_{real} increases by 0.51 degrees per million year between grazing artiodactyls and grazing perissodactyls (which are separated by some 152 million years of independent evolution on the 'ungulate' tree, i.e. at least twice as much as the inferred age of the most recent common ancestor to all 'Ungulatomorpha'). This same figure is 0.71 degrees per million year between grazing perissodactyls and "others" and 0.65 between grazing artiodactyls and "others".

Caribbean Anolis – By using *search.conv*, we found significant convergence in 5 out of the 6 ecomorphs traditionally recognized for insular anoles.

We found convergence in 5 out of 6 different ecomorphs, the only exception being 'trunk' anoles. The *Anolis* species that cannot be ascribed to any ecomorphs are, unsurprisingly, not found to converge. By using the C1 metric, Stayton (2018) found 4 of 6 ecomorphs converging. By using the metric C5, convergence is found in 3 ecomorphs. Species not ascribed to an ecomorph were not found to converge for either of the metrics.

Discussion

I demonstrated *search.conv* is robust, has low Type I and Type II error rates, and is very fast even with reasonably large trees. Although the mrcas set to converge are not always found with precision under the automatic mode, the species actually set to converge are correctly identified up to 97.5% of the time, further demonstrating the selection of clade pairs is reasonably precise.

We successfully applied *search.conv* to mandibular shape evolution in mammals in two different real cases and to Caribbean islands anole ecomorphs. The first real case study regards the evolution of mandibular shapes in felids. We found "true" sabertooths (Homotheriini and Smilodontini) independently converge on barbourofelids in their mandible morphology. Intriguingly, Metailurini (i.e. "false" sabertooths) which is nested within the machairodont family,

were not found to converge on barbourofelids under the automatic mode. This means *search.conv* successfully excluded the false sabertooths from the convergence pattern despite their phylogenetic position close to other "true" sabertoothed machairodont cats (Meachen-Samuels, 2012).

We used the felid data to compare *search.conv* to `convevol`'s *convratsig* function. While both functions recognize the same pattern, *search.conv* was found to be three orders of magnitude faster, which could be crucial when it comes to assessing convergence with uncertain state categorization, or to taking the effect of phylogenetic uncertainty into account, as this implies repeating the analyses dozens of times by using different phylogenetic hypotheses.

The second real case application, performed with the 'state' approach, relates to the evolution of hypsodonty due to grass feeding in 'ungulates'. Grazing adaptations in the mandible evolved independently in horses (genus *Equus*) and several bovid lineages, most notably among antelopes. We found evidence for convergent evolution between *Equus* and strictly grazing bovids, such as *Bison*, *Bos*, and *Alcelaphus*. This is especially noteworthy considering that the paleontological tree we used includes a number of non-grazing equids, such as hipparionoid horses and browsing anchitheriine equids, plus several extinct rhinos and tapirs which were all browsers. This demonstrates the method was able to find convergence among grazers despite the effect of phylogeny and body size on mandibular shape variation (Raia et al., 2010b).

The final real case pertains to *Anolis* ecomorphs. We found evidence for convergence in all of them but the 'trunk' ecomorph species. Intriguingly, five of the six 'trunk' groups belong to a single monophyletic clade, indicating that the trunk ecomorph evolved only twice, once for a single clade only present on Hispaniola and then again when Cuban *Anolis loysiana* converged on them.

Compared to other statistical procedures used to test for morphological convergence, *search.conv* offers the possibility to test convergence between entire clades, and allows testing specific 'states' sparsely distributed across the tree. In addition, being much faster than alternative approaches, *seach.conv* allows exploring the potential effect of phylogenetic uncertainty and use of fossil phenotypes as ancestral states, that can be crucial in the presence of non-Brownian processes.

It must however be noted that not all cases of "convergence" may be explored best with *search.conv*. There are several instances reported in literature of convergence between closely related clades and even single species with close phylogenetic proximity. I provided a test which is useful to find instances of large-scale morphological resemblance between distant clades that are generally referred at as either 'convergent' or just cases of iterative evolution. Caution must be applied to the choice of the ancestral phenotype in the presence of strong phenotypic drift.

Melchionna et al. 2021 – Palaeontology, in press

The evolution of convergent phenotypes can be assessed by applying a variety of statistical methods. Despite the pattern is exquisitely morphological in essence, none of these methods is suited to visualize convergence, taking the usually unstated assumption that the biological objects under study either converge or not as a whole, whereas different parts of them may in fact either arise or contrast convergence, and morphologically similar parts may behave differently, affecting the functioning of the structures. We proposed a new method sought to chart patterns of convergence on three-dimensional digital models. Under this method, named *conv.map*, convergence between pairs of models is mapped on them so that the operator is given the immediate opportunity to figure out where convergence applies, and which structures are involved. We applied *conv.map* to a well-known case-study, the evolution of the saber-tooth morphotype in carnivorous mammals. We found that convergence among saber-tooths mainly applies to the rostral part of the skull (particularly around the anterior dentition and upper molars) and around the strongly developed nuchal crest. This mapping suggests that saber-tooths shared killing behavior was characterized by a stabbing bite by powerful neck muscles.

Sansalone et al. 2020 – Frontiers in Earth Science

Fossorial mammals have often been reported as a textbook example of convergent evolution (Hildebrand et al., 1985). The subterranean environment is relatively predictable in time, and requires a high degree of specialization, which in turn can lead to comparable morphological forms (Sansalone et al., 2019). Moreover, locomotion underground requires extreme morphological specialization and digging kinematic is likely to have a strong impact on other functional traits (Piras et al., 2015). Here we measured the extent of functional and morphological convergence among ecologically convergent fossorial mammals ranging in degree of fossorial specialization (cursorial, semi-fossorial, fossorial) and from across Mammalia (monotremes, marsupials and placentals) including a number of extinct forms. When comparing distantly related groups, the presence of similar morphological features could be interpreted as the result of phylogenetic relatedness (e.g., parallelisms) rather than convergence. For this reason, it is highly recommended to consider the phylogenetic history of the species examined. Here, we combined 3D geometric morphometrics (GMM), finite elements analysis (FEA) and *search.conv* to quantify the extent of functional and morphological convergence in the humeri of fossorial mammals.

We focused on a range of mammals exhibiting putative fossorial adaptations of the humerus. Among Talpidae we included highly fossorial taxa (Scalopini and Talpini), semi-aquatic (e.g., desmans) and semifossorial/ambulatory (e.g., shrew moles) species (Piras et al., 2015), including both extinct and living species. We applied *search.conv* under both “clade” and “state” cases on shape values (PC scores), by classifying species under fossorial/non-fossorial lifestyle. We repeated the analyses by collating the values of functional variables (i.e., angle and von Mises stress values) to PC values for each species and taking such “composite” vectors as the *search.conv* entry.

By applying *search.conv* on shape data (PC scores) alone we found no significant instance of convergence, neither under “clade” or under “state” case. When integrating functional variables, we found Japanese moles (i.e., *Mogera* species exclusive of *Mogera insularis*) to significantly

converge on the clade including the north American moles within *Scapanus* plus *Scalopus* genera (i.e., Scalopini), starting from their respective common ancestors.

As for the "state" case, the mean angle between species within the "fossorial" category becomes 0.404 degrees/myr by adding performance variables, which is significant ($p = 0.042$).

Shape analysis showed the presence of a mixture of taxonomic and functional signals. All the functional groups occupied different regions of the morphospace, suggesting the absence of a significant convergent signal in the humeral shape of fossorial mammals. This was confirmed by *search.conv* which did not recover evidence of convergence across our morphological sample. This result suggests that fossorial mammals were able to find different solutions to a similar adaptive challenge: the exploitation of the underground. In this case natural selection has produced some of the most divergent and highly derived humeral morphologies among vertebrates (Piras et al., 2015).

We found only one instance of morpho-functional convergence accounting for trade-off measures between the Japanese and North-American fossorial moles (*Mogera*, *Scalopus*, and *Scapanus* genera). This result strongly suggests that fossorial mammals evolved around multiple different optima rather than around a single adaptive peak. It is likely that natural selection favored the evolution of humeral morphologies characterized by high performance (low stress) rather than promoting the optimization of humeral mobility. However, the trade-off between humeral strength and mobility resulted in different morphologies having similar fitness in the subterranean environment (Losos, 2011).

Many examples of convergent evolution are based on simplistic descriptions of both function and phenotypes, but accounting for trade-off measures between different behaviors under similar selective contexts may reveal more complex scenarios. Our investigations highlight the fact that natural selection, even operating in similar selective conditions, may fail to produce convergent phenotypes in different evolutionary lineages.

COMPARATIVE METHODS AND TAXONOMIC DIVERSIFICATION

Clade diversity differs in orders of magnitude in the animal kingdom. While some clades such as hyenas are represented by a manifold of species, others like passerine birds and bugs count to the thousands. Paleontologists are confronted with even more striking patterns, as clades often present *within* their history, nearly as much diversity variation as *between* clades today. Early attempts to model the process producing standing diversity, that is taxonomic diversification, trace back to the 80s, thanks to the effort of Jack Sepkoski (1984). As diversification is the net gain of species, it should be modelled as the difference between speciation and extinction rate, that the latter in turn be derived from appearance and disappearance from the fossil record, along stratigraphic intervals. Sampling, though, is more than an issue in paleontology, so that modern-day versions of Sepkoski's early attempts, strive to include sampling quality within the calculation of diversification rate. Foote (2000) developed the computation of instantaneous per-capita speciation and extinction rates by measuring the proportion of genera surviving from the start to the end of a time interval. Alroy (2008) calculated an interval-specific sampling probability by taking into account the number of species that are present and sampled in three consecutive intervals and those sampled only in the first and last interval (thereby presumably missing because of sampling in the middle interval). Yet, neither of them accounted for taxa confined to a specific time bin (Foote's 'singletons' or Alroy's 'one-timers') because their relative frequency is disproportionately affected by incomplete preservation, producing biased estimates of speciation and extinction rates.

A more recent attempt to use occurrence in stratigraphic intervals to derive diversification parameters relies on Pradel models (Pradel, 1996). They belong to Jolly-Seber family of capture-mark-recapture (CMR) models, commonly used in analysis of population dynamics of extant species. Starting from species occurrences split into constant-length time intervals, Pradel models jointly estimate survival probability, seniority (the probability for a species extant in a given interval to be already present in the previous interval), growth rate, and sampling probability. Such

metrics were successfully applied to paleontological data to compute extinction probability (as the complement of survival), speciation probability (as the complement of seniority), net per capita diversification rate (starting from growth rate), and sampling probability (Liow & Finarelli, 2014; Finarelli & Liow, 2016). The common approach consists in fitting a number of different Pradel models by keeping the parameters constant or allowing them to vary through time or according to a covariate, and comparing them by means of AIC. The advantage of Pradel models over traditional approaches such as Foote's and Alroy's is that absences are interpreted as either real absences or as failed-to recognize presence, so that sampling probability is estimated jointly with the other parameters. For this reason the method is recommended with paleontological data, especially with heterogeneous and incomplete sampling (Liow & Finarelli, 2014).

PyRate (Silvestro et al., 2014) is a computer program that allows performing macroevolutionary analyses based on paleontological data using a Bayesian framework. Under this approach, fossil occurrences are modeled as the result of sampling and species diversification processes. Sampling includes all historical and geological conditions that led to the preservation of an organism in the paleontological record and its subsequent sampling, description and identification. Species diversification represents temporal changes in species richness as result from the interaction between speciation and extinction rates through time. The two processes are combined in a hierarchical Bayesian framework that jointly estimates the times of speciation and extinction of each species and derives the parameters of the underlying birth-death model. The advantages of this method as compared to other existing approaches are several. First, it utilizes all available fossil occurrences, including the "singleton" taxa (having only a single occurrence). Secondly, it allows for continuous time and avoids the use of predefined, discrete time bins that do not necessarily correspond to biological processes as they are dictated by the geological record. The ages of fossil occurrences are randomly drawn from their temporal ranges, and the procedure is repeated to account for age uncertainty. Additionally, it explicitly models the temporal heterogeneity of sampling for each species in the record. The sampled occurrences of any fossil

species follow a canonical pattern, decreasing at both the start and the end of the species lifetime. Thus, considering only the first and last appearances could severely underestimate the speciation and extinction time. The framework models this sampling pattern to obtain confident estimates for each species in the dataset. Third, in the Birth-Death model, extinction can exceed speciation, both rates are allowed to vary through time, and their changes are not assumed to be temporally linked.

All of these methods provide often robust estimates of diversification rates and its components. Yet, they have a fundamental shortfall, they ignore the phylogenetic component. This is detrimental since diversification rates depend on trait values, such as body size, that map explicitly on the phylogeny, being their variation correlated to phylogenetic distance. For instance, it hardly makes sense comparing clades with any of the above, since differences in computed diversification metrics may just reflect difference in trait values. Further, important limitation of phylogeny-free methods are that they usually (1) use discrete time bins, (2) analyze only first and last appearances of a taxon, ignoring other occurrences, (3) lack the ability to perform model testing against over parameterization, and (4) do not include extant species in the analysis. These features make the speciation and extinction rate estimates difficult to compare to those obtained from molecular phylogenies.

Phylogenetic comparative methods applied to taxonomic diversification typically derive as birth (***b***) and death (***d***) parameters for the entire tree from the branching process (Yule, 1925; Nee, 2006). More recent approaches allow to model rate heterogeneity by imposing different ***b*** and ***d*** parameters for different clades. Additionally to the aforementioned BAMM (Rabosky, 2014), MEDUSA (Modelling Evolutionary Diversification Using Stepwise AIC, Alfaro et al., 2009) is a comparative method that integrates phylogenetic information about the timing of the splits along a phylogenetic tree with taxonomic richness data to estimate rates of speciation and extinction. It fits a number of Birth-Death models to the combined dataset imposing different ***b*** and ***d*** to some branches on the tree and compares them by means of maximum likelihood estimation. It finally

returns the simplest model possible (i.e. with the lowest number of shifts in ***b*** and ***d*** parameters) that represents the diversification pattern of the clade.

Again, the main limitation of such methods is that they only work with phylogenies of extant species, which makes it challenging or even impossible to evaluate meaningful extinction rates (Rabosky, 2010).

A popular model used to jointly estimate time-calibrated phylogeny of living and extinct taxa together with the rates of speciation and extinction is the fossilized birth-death model (FBD; Stadler, 2010). It defines the probability density of a phylogenetic tree including extant species and fossil samples conditioned on speciation (branching event), extinction, fossil sampling (the rate of producing fossil sample), and extant sampling (extant species probability sampling) rate parameters of the model. This probability density directly allows to infer the parameters of the model given a phylogenetic tree, using maximum likelihood or Bayesian inference methods. Despite being widely tested through simulations (Heath et al., 2014; Gavryushkina et al., 2014) and successfully applied to a range of datasets on different taxonomic levels (Heath et al., 2014; Gavryushkina et al., 2017), FBD model was proved to underestimate net diversification rate, turnover, and fossil-sampling proportion when sampling of species and fossils is not complete or random but selective (Matschiner, 2019)

The species-lineage diversification rate (DR, Jetz et al. 2012, Cantalapiedra et al., 2017) is based on the mean equal splits (ES) measure of evolutionary isolation (Redding & Mooers, 2006). Such measure divides the evolutionary time represented by a branch equally among its daughter branches (Fig. 10), so that the sum of the ES values for each species (or any taxonomic level) is the estimated amount of evolutionary time embodied by the species itself.

Since species in rapidly-diversifying clades have short edge lengths shared among many species and low ES values, while isolated species on a tree have no evidence of recent diversification and large ES values, the inverse of this metric is used as a measure of the splitting rate of the path to a tip, that is the Diversification Rate (DR). DR was successfully applied to study

diversification dynamics in phylogenies including both extinct and living species (Cantalapiedra et al., 2017) or living species only (Cooney et al., 2017; Price Waldman et al., 2020).

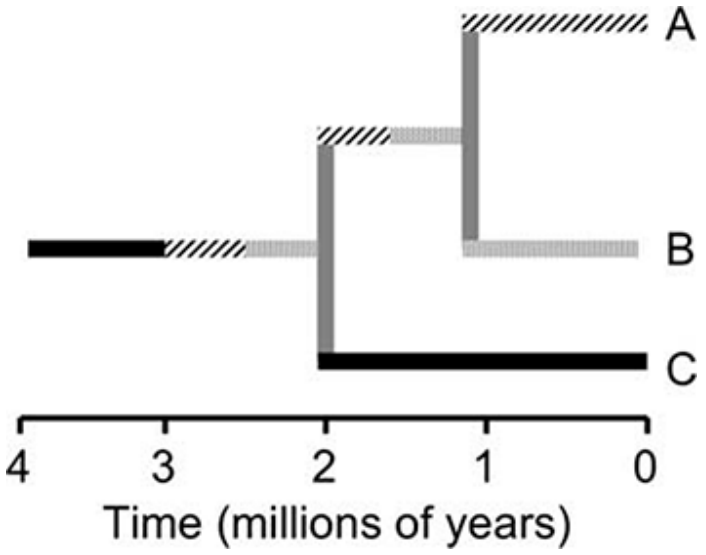

Figure 10. The equal-splits approach is used to apportion the total evolutionary history of this tree (7 million years) among the three constituent species in the tree (A, B, C). Species A has its own branch length equal to 1 time unit, plus half of the branch length for the common ancestor of the clade AB that is 0.5, plus one quarter of the root branch that is 0.5. The same for species B. Species C has its own branch length of 2, plus half of the root branch that is 1. Summing up the values for each species, the equal-splits value for species A is 2 MY, for species B 2 MY, and for species C 3 MY. The sum of these values equals the total evolutionary history of the clade. Image from Redding & Mooers, 2006.

I took advantage of DR being a lineage-specific measure of taxonomic diversification to implement a new method meant to identify clade-associated and character-associated diversification. I used DR and phylogenetic evolutionary rates produced by *RRphylo* to locate clade-wise shifts in taxonomic diversification rates coinciding with shifts in phenotypic evolutionary rates or the association between phenotypic values and DR for the tree as a whole.

Castiglione et al. 2017 – Frontiers in Ecology and Evolution

The extent of the geographical range is one of the most important biological attributes of a species. Species occupying large ranges may have better dispersal ability (Lester et al., 2007), usually have larger populations (Blackburn & Gaston, 2003; Carotenuto et al., 2010), and engender wider niche (Slatyer et al., 2013), as compared to small-ranged species. Large population size, good dispersal ability and large range size, in turn, significantly decrease the chance of stochastic extinction. Therefore, perhaps unsurprisingly, large-ranged species have been repeatedly demonstrated to endure lower extinction risk, or longer duration in the fossil record (Cardillo et al., 2005). The relationship between range size and speciation rate is much less obvious. On the one hand, large-ranged species may have greater chance to produce daughter species (Wagner & Erwin, 1995), either by peripheral isolation or microvicariance. However, assuming most originations occur by allopatric speciation, the likelihood that a barrier is large enough to break a species' range into smaller pieces is arguably low for the largest-ranged taxa (Gaston, 2003). This suggests that even under allopatry, speciation probability might be a peaked function of species range size (Jablonski & Roy, 2003). However, it must be noted that since species with large range might be expected to live for longer (because of the positive effect of range size on survival), they will likely leave more descendants in the fossil record (Wagner & Erwin, 1995).

Such diversification dynamics significantly affect the range size-frequency distribution (Gaston, 1998, 2003; Blackburn & Gaston, 2003). In most animal groups, the RFD is strongly right-skewed, meaning that most species within a clade have restricted ranges, and a very few of them are extremely widespread (Blackburn & Gaston, 2003). The evolutionary process producing RFD at the clade level ultimately depends on the distribution of speciation and extinction rates along the range size spectrum. If small-to-medium ranged species produce more daughter species, most speciation occurs by vicariance (Gaston, 1998, 2003), and the right hand of the RFD is populated by the few

long-living species escaping range division (i.e., by allopatric speciation). This would imply a negative, or no relationship either, between speciation rate and range size. On the other hand, if speciation rate increases with range size, the left hand of the RFD must be populated with species deriving from highly asymmetrical range splits in large ranged species, according to the peripheral isolation speciation mode. Under this scenario, the relationship between range size and speciation rate should be positive and significant. Finally, if most speciation occurs by allopatric speciation, extinction rate should significantly decrease with range size for the RFD distribution to hold a constant shape, because small-ranged species are bound to appear more often than large-ranged species to maintain the right skew.

In this study, we computed range size at species and clade level, as well as per capita speciation and extinction rates for 21 extinct clades of marine metazoans. We tested whether speciation and extinction rates were significantly affected by the average range size of species within their clades, partitioned into consecutive, non-overlapping temporal bins. Then, we inferred from the shape of the speciation rate vs. range size curve whether the dominant mode of speciation in the fossil record has been either peripheral isolation or simple vicariance. This is feasible because with the former, speciation rate should be higher in larger ranged species, whereas under the latter, the rate should peak at intermediate ranges, or show no relationship with range size, either.

We used data from Raia et al. (2016) regarding 21 clades of marine metazoans belonging to five different animal phyla (Arthropoda, Brachiopoda, Bryozoa, Cnidaria, Mollusca). We computed the species and clades geographic range sizes per interval by the minimum convex polygon approach (MCP, Carotenuto et al., 2010) and applied Pradel models (Pradel, 1996) to derive speciation and extinction rates per time interval. Since speciation and extinction rates were calculated for each clade within each time bin, we had to use an "average" representation of the species geographic range per time bin in order to assess the effect of range size on the rates. To this aim, we took the average range size computed over all of the species within a clade and a given time bin, and divided this average area by the clade range size, obtaining a "standard average range

size" (hereafter simply standard range, SR). To test the statistical relationship between diversification rates estimated by Pradel models and SR, we used generalized linear mixed models (GLMM). As response variables, we used alternatively the two rates (i.e., extinction and speciation) as estimated by Pradel's models and SR as explanatory variable. In both regression models, linear and quadratic terms between SR and the response variables were calculated.

Regression models reported a significant, negative relationship between extinction rate and SR. Both linear and quadratic terms were significant. A significant, positive relationship occurs between speciation rate and SR (both linear and quadratic terms were significant). The regressions including only linear terms showed similar results, with the extinction rate being significantly and negatively related to SR, and speciation rate reporting a significant, positive relationship with SR. For both rate metrics, models including both linear and quadratic terms reported lower AIC values than those including only linear terms. Eventually, the purely quadratic models show the least fit to the data.

The geographic distribution of a species depends on how individuals are organized in space, how much they tolerate environmental variation, and ultimately how they interact with each other to occupy the territory. The geographic range size is intimately linked to dispersal ability, which, in turn, is influenced by organisms' physiological tolerance. The range size frequency distribution thus depends on how all of these features vary across species within a clade (Blackburn & Gaston, 2003). The RFD is a right-skewed unimodal distribution when using raw data for most of the studied living and fossil clades. This means that clades include mostly species with small range size, and just a few species that were environmentally tolerant and good at dispersing enough as to occupy a large space.

By using a hefty, extinct metazoan clades dataset, spanning over more than 400 Myr of evolution, we were able to demonstrate that speciation rate is positively related to average range size, and extinction rate negatively associated to such average range size, especially when a quadratic term is added to the equation. The latter result is nothing but a confirmation of the

strongly hold notion that having a wide geographic range favors species survival (Gaston, 2003; Cardillo et al., 2005). Our measure of range size, SR, is not range size per se, but a metric that correlates to the presence of (many) extremely small and (few) extremely large ranged species within the RFD. According to the simulations we performed, low SR means that extremely small-ranged species are overrepresented in the RFD, which provides a plain ecological explanation for the relationship between SR and extinction.

The relationship between SR and speciation rate is positive and significant, but more difficult to interpret. According to the peripheral isolation mode of allopatric speciation (Gaston, 1998), the rate should increase linearly with range size. However, although at high skew values species with extremely large ranges are expected to appear, the number of such large-ranged species in the RFD decreases, which means that the total number of daughter species produced by such large-ranged species is not expected to peak at the highest range class. It is important to notice this does not mean speciation is a peaked function of range size (as expected under the vicariance model). First, if species with small to medium range originate more species (Jablonski & Roy, 2003), the speciation rate to range size regression should be negative or non-significant, whereas we obtained a significant and positive relationship. Secondly, the peak in speciation rate is at some 44% of the SR distribution, which is much higher than the modal SR class. Again, this suggests that large, but not the largest size class produces more daughter species.

According to the peripheral isolation mode we support here, species with wide geographical extension may originate daughters at the boundary of their distribution, either because of small barriers intervening to break the continuity of the parental species range at its borders, or by waif dispersal. Under this model, the formation of geographic barriers or the dispersal ability of propagules are pivotal to genetic isolation. If a range is large enough, the likelihood that barriers may divide it into halves is much smaller than the probability of a small barrier occurring at the range periphery. However, a widespread species is also likely to be tolerant, which could prevent the formation of demographic isolates (Jablonski & Roy, 2003).

Silvestro et al. 2020 – Ecology Letters

Leigh Van Valen's Law of Constant Extinction (Van Valen, 1973) states that under constant ecological conditions the probability for a species to go extinct is independent from its age, that is, the time passed since its origination. This idea has been the focus of a number of paleontological studies providing evidence for increased extinction risk with stratigraphic duration (Pearson, 1992), for the opposite (Finnegan et al., 2016), or even giving full support to the law of constant extinction (Quental & Marshall, 2013; Alroy, 2014). Unfortunately, although the law of constant extinction has enormous implications to understand the evolution of diversity and the mechanisms driving species extinct, testing it empirically is difficult, due to sampling biases (Wiltshire et al., 2014), scale issues (Barnosky, 2010), and because overall fluctuations in the extinction rate occur in time (Quental & Marshall, 2013).

In this paper we developed a method to infer age-dependent extinction (ADE) rates from fossil occurrence data using deep learning within a neural network framework (hereafter ADE-NN). Differently from previous efforts (Quental & Marshall, 2013; Alroy, 2014), ADE-NN evaluates whether the extinction rate changes with species age while accounting for the incompleteness of the fossil record and for the limited temporal resolution characterizing the dating of most fossil remains. We used ADE-NN to classify fossil datasets among five discrete categories of age-dependent extinction rates representing (1) strongly (strongD) and (2) moderately (mildD) decreasing rates with age, (3) stochastically constant (constant) rates (i.e. extinction rate independent of age), and (4) moderately (mildI) or (5) strongly (strongI) increasing extinction rates with age.

In modern day as well as in past ecosystems, both species diversity and extinction risk have been shown to generally change with latitude. Hence, to account for possible latitudinal effects, we partitioned the record into distinctive clades and then the species within the clades into low-latitude

(within 23° in latitude from the equator), mid-latitude (from 23° to 46° in latitude from the equator), and high-latitude (> 46° away from the equator) species, depending on the position of the weighted center of their geographic distribution on Earth during their existence. This generated 56 subgroups (21 clades x 3 latitudinal bands minus 7 subgroups relative to clades that are not present in the northern latitudinal band). Finally, since mass extinction events represent moments of extremely high extinction rates (Jablonski, 2001), we checked whether the extinction rates differ when comparing species that lived during times of background extinction rate to species lived during the build-up of a mass extinction.

We found that species extinction rate significantly decreases with species age in 51 out of 56 subgroups (91.1%) when the clade species are partitioned by latitudinal band. In the low-latitude band, the extinction rate decreases with species stratigraphic duration 21 out of 21 times (16 with strong, 5 with mildD). In the mid-latitude band, the extinction risk decreases with species stratigraphic duration 19 out of 21 times (17 strongD, 2 mildD) and remains constant into two instances. Finally, in the high-latitude band the extinction risk decreases with species stratigraphic duration 11 out of 14 times (7 strongD, 4 mildD), remains constant twice, and weakly increases with age once (Cystoporida bryozoans).

The analysis of entire clades (irrespective of geographic range) show decreasing extinction rate with species age as the dominant mode 21 out of 21 times. Among these, strongD is the most probable model in 19 cases. Restricting the analysis to time intervals of background extinction (i.e. after removing either stages or epochs bracketing mass extinctions) did not change considerably these results, with significant evidence for decreasing extinction rate with species age still found across all clades. The dominant mode of age dependent extinction remained strongD, which was the preferred mode of extinction in 14 cases. Finally, separate analyses of individual time intervals of background extinction showed that there are no obvious temporal trends in extinction modes, with instances of strongD and mildD extinction similarly scattered across all intervals.

The prevalence of strongD as the best-fitting model indicates that most species within clades live short lives (because of the extremely high extinction risk for very young species), whereas those which survive the early stages are likely to experience asymptotically decreasing extinction risk. This observation is consistent with the idea that species behave as units of selection. Under species selection (Jablonski, 2008), taxa possessing biotic traits conferring better survival, such as large geographical range or generalist ecological niche, are positively selected, meaning that they endure higher speciation or lower extinction rates, on average, over time (Jablonski, 2008). At high latitudes species tend to be tolerant to climatic variation, a pattern that has been observed even in deep-time fossil records (Blackburn & Gaston, 2003). This greater climatic tolerance, together with the presumably lower intensity of biotic interactions (because of the low diversity) and slower diversification dynamics, makes high-latitude areas the least susceptible to species selection. In contrast, the higher standing diversity toward the tropics makes the network of biotic interactions more complex. It is thus unsurprising that strongD is more prevalent in the tropics, where species diversity and the potential for intense biotic interactions are higher.

The ubiquity of decreasing age-dependent extinction rates across the vast majority of clades indicates that decreasing extinction rates throughout species lifespans represent a general rule governing species survival. While finding a definitive explanation for the mechanisms determining this extinction pattern might be unfeasible by using the fossil record, the latitudinal trend towards stronger age dependency and higher species richness towards the low latitude suggests that biotic interactions might play an important role in shaping spatial extinction patterns. Our results are thus not inconsistent with the Red Queen hypothesis, stating that biotic interactions shape the pace of evolution (Van Valen, 1973; Quental & Marshall, 2013). Rather, we argue the effect on survival rates of such interactions is predictably variable across space, being more intense towards the tropics, and generating a temporal pattern of decreasing extinction rate within clades, favored by the selection of species bearing survival-related traits.

Piras et al. 2018 – Palaeogeography, Palaeoclimatology, Palaeoecology

Felids (Mammalia, Carnivora) form a morphologically homogenous, monophyletic clade, including strictly carnivorous species. In contrast to other meat-eating mammals, felids only retain the anterior, slicing portion in their lower molars, while the crushing part (the talonid) is lost (Van Valkenburgh, 2007). Felids (Felidae plus Barbourofelidae families) can be ecomorphologically subdivided into two categories: conical-toothed cats and sabertooths (Van Valkenburgh, 2007). The former borrow their name from the shape of their canines in cross section. They include the modern cat genera such as Felis, Panthera, and Acinonyx. Sabertooth cats were characterized by laterally-compressed, extremely long upper canines, procumbent incisors, reduced coronoid process, and low glenoid fossa (Christiansen, 2008). All of these features conferred on sabertooths a unique killing behavior. The success of the sabertooth morphology is testified by its iterative evolution among meat eating mammals (Van Valkenburgh, 2007). Within Carnivora, the sabertooth morphology appeared in the Nimravidae family, which emerged in late Eocene, the Barbourofelidae family (known from the early Miocene), and in the true cat subfamily Machairodontinae, which radiated between Miocene and Late Pleistocene (Werdelin et al., 2010).

Sabertooths' highly derived cranial morphology and the extremely long yet fragile upper canines, did not allow the exploitation of a wide prey spectrum (Mondanaro et al., 2017). The peculiar morphology and narrow feeding niche make sabertooths the most specialized among mammalian carnivores (Randau et al., 2013). In ecological terms, specialization usually translates in competitive advantages (or reduced competition) over other guild members. Yet, it may also be associated with increased extinction risk (Cardillo et al., 2005; Slatyer et al., 2013).

In this paper, we collected and analyzed the largest felid plus barbourofelid mandible collection to date in order to better understand the evolutionary processes leading to the sabertooth specialization. We used Geometric Morphometrics to retrieve information on mandible shape and

mechanical performance. Then, we inferred a felid plus barbourofelid phylogenetic hypothesis to compute rates of morphological evolution, and tested for phenotypic rate shifts in the tree by means of *RRphylo* and *search.shift*, under the hypothesis that the sabertooth character represents a morphological discontinuity in felid evolution. We further presumed that the acquisition of the sabertooth morphology increased extinction rates (Liow, 2004; Van Valkenburgh, 2007; Raia et al., 2011). To test for this hypothesis, for each of the two felid ecotypes (conical-, and saber-toothed) we calculated diversification rates and its components (i.e. rates of speciation and extinction) directly from the fossil record by using PyRate (Silvestro et al., 2014), and contrasted sabertooth's versus conical-toothed's rates.

We found a significant shift in the rate of mandible shape evolution to apply to scimitar-toothed cats (homotheriini) within machairodonts. Yet, in terms of physical loadings on the mandible (stress variables), a rate shift applies in coincidence with the clade including the most extreme, dirk-toothed sabertooths such as *Smilodon*. Interestingly, the third machairodont clade, metailurini, does not show a significant rate shift in either shape or stress data. Most scholars agree that metailurini converged on modern pantherine cats in morphology and behaviour (Barycka, 2007). In addition, Metailurini had relatively shorter canines as compared to other sabertooths.

The sabertooth character possibly evolved (or was recruited) for preying upon megaherbivores (Randau et al., 2013). Such narrow feeding niche (as well as the derived morphology) suggests sabertooths were highly specialized taxa. The virtue of specialization is that it confers immediate ecological advantages over competitors, usually by means of large body size (Raia et al., 2012). The effects of specialization on diversification rates are not as easy to predict. On the one hand, it might reduce speciation rates (Rabosky & Hurlbert, 2015). On the other, specialization by evolutionary novelty might promote diversification by expanding clades in novel regions of the morphospace where there is little competition (Jønsson et al., 2012). However, it is much more probable that specialization increases extinction risk (Slatyer et al., 2013).

Our diversification rate analyses indicate that both felids and barbourofelids diversified under essentially constant speciation rates throughout the past >30 Myr. Overall extinction rates were also fairly stable until the Pliocene and smaller than speciation rates, thus yielding positive net diversification. Towards the recent, and in particular since the late Pleistocene, the extinction rates have increased dramatically from 0.35 (95% CI: 0.26–0.43) in the early Pliocene to 0.99 (95% CI: 0.58–1.34) in the late Pleistocene and Holocene. Yet, regardless of which clade they actually belong to, our analysis of the fossil record indicates that the average extinction rate in sabertooths is 50% higher than in conical tooths. Since there is no statistically significant difference in speciation rate between the two ecotypes, diversification rate should necessarily be depressed by the sabertooth character. This might help explaining why there never were sabertooths of different clades living in the same ecogeographical region for long, and why sabertooth species are rare in the fossil record overall.

Melchionna et al. 2020 – Biological Journal of the Linnean Society

The last common ancestor of primates lived some 71–63 Mya (Springer et al., 2012). The oldest known stem Primates, the possibly polyphyletic plesiadapiforms, appeared in North America around the Palaeocene/Eocene boundary (Bloch et al., 2007). From the tiny, strictly arboreal pleasiadapiforms to today's larger, ecologically versatile Strepsirhini and Haplorhini, primates have experienced evolution towards larger brains, in terms of both absolute brain volume and relative to body mass (Montgomery et al., 2016). Such a high degree of encephalization is known to influence physiological, ecological and anatomical attributes in extant species (Neaux et al., 2018).

The large costs incurred by increased encephalization appear to be offset by either expanded cognitive abilities in social species (Dunbar, 2009), enhanced capacities for psychological manipulation and female control in monogamous species, improved dexterity, or better foraging abilities linked to frugivory. In addition, the evolution of larger brains (in absolute terms) could also

be a mere byproduct of a macroevolutionary trend for increased body mass. Still at the macroevolutionary level, a significant association between increased brain mass and speciation and extinction dynamics is often reported in the scientific literature (Isler & van Schaik, 2009; Aristide et al., 2016). Large brain mass correlates with low extinction risk in birds (Sol et al., 2005) and mammals (Isler & van Schaik, 2009). A significant increase in brain mass is linked to a positive shift in diversification rate (i.e. increased difference between speciation and extinction rates) in carnivores (Finarelli & Flynn, 2009). Clade-level endocranial volume patterns are associated with origination and extinction in hominids (Du et al., 2018). There is substantial evidence for an increase in speciation rates over time in Primates (Arbour & Santana, 2017). Yet, how these rates relate to relative brain mass and to body mass remains unknown.

In this paper, we hypothesized that Primates follow (1) a trend for increased body mass over time and (2) a trend for increased encephalization, and we tested (3) whether these trends are associated with each other and with diversification dynamics. To test these hypotheses, we used a large palaeontological phylogeny comprising 317 primate species for which endocranial volumes and body mass estimates are available. I applied the multiple version of *RRphylo*, *search.shift*, and *search.trend* to search for possible shifts in the evolutionary rate of encephalization across the Primates tree and assess whether relative brain mass has increased over time. Finally, I computed interval-to-interval speciation and extinction rates by applying an implementation of Pradel (Pradel, 1996) models on the palaeontological data. Per-interval rates were used to assess whether increased encephalization is associated with species diversification in Primates.

Results for body mass show an increase in average rates in *Colobus* + *Pliocolobus* + *Procolobus* monkeys (rate difference = 0.419, p-value = 0.018), whereas Pitheciinae + Callibecinae (rate difference = -0.490, p-value = 0.001) and Indriidae (rate difference = -0.286, p-value = 0.020) show an opposite pattern. There is a general positive trend in body mass along the evolutionary history of Primates (slope = 0.003, p-value = 0.047). We found significantly lower average rates of ECV increase in Hylobatidae (rate difference = -0.016, p-value = 0.018), *Macaca* (rate difference =

-0.013, p-value = 0.012), the clade including *Cercopithecus* + *Erythrocebus* + *Chlorocebus* monkeys (rate difference = -0.014, p-value = 0.003), Lorisoidea (rate difference = -0.018, p-value = 0.001) and *Saguinus* (rate difference = -0.022, p-value << 0.001). Only hominins showed significantly higher average rates than the rest of the tree (rate difference = 0.091, p-value << 0.001). Macroevolutionary trends in ECV and absolute rates of ECV for Primates as a whole are both significantly positive (ECV: slope = 0.006 p << 0.001; ECV rate: slope = 0.006 p << 0.001).

The best Pradel model is the one in which parameter estimates are a function of ECV and sampling is allowed to change from one temporal bin to the next. We extracted regression coefficients from model results and computed extinction and speciation rates for each ECV value, and finally for each species. We found an increasing trend in speciation rate and a decrease in the extinction rate for ECV. We also calculated the average rate values for each time interval. We found an increase in speciation rate and a corresponding decrease in extinction rate starting from the Early Oligocene. The distance between the curves increases at the beginning of the Miocene.

During their evolutionary history, Primates experienced trends for (1) larger body mass, (2) increased encephalization (in keeping with our first hypothesis), (3) lower extinction rate and (4) higher speciation rate, and (5) hominins stand out for their exceptionally large brain and more intense encephalization pattern. Cope's rule does not produce the observed encephalization pattern. Although both body and brain mass increased through time in the Primate clade, relative encephalization itself increased, meaning that the primate brain became proportionally larger. This trend became evident with the appearance of crown Primates and is associated with an increased speciation rate in this clade, starting during the Oligocene. Unfortunately, we had to calculate speciation rates interval by interval, and the trend in ECV as continuous. This means that we cannot comment on the apparent relationship between the two variables beyond their pure association.

Although the finding that hominins have exceedingly large brains is not novel, it is intriguing that australopiths and *Homo* share a rapid evolution of high ECV values and remain significantly different from any other Primate clade with regard to several additional aspects,

including tool use (Skinner et al., 2015), full bipedalism (Dowdeswell et al., 2017) and mandible evolutionary rates. The evolution of a larger brain in hominins is usually discussed as a gradual process punctuated by stepwise encephalization pulses (Du et al., 2018; Diniz-Filho et al., 2019a). In agreement with this scenario, we found that hominins, as a whole, represent an instance of positive rate shift in ECV among Primates, yet within this lineage the progressive encephalization appears to be a gradual process starting around the appearance of *A. africanus*.

Looking at Primates as a whole, our results show that the positive relationship between ECV and speciation rates becomes evident from the early Oligocene. The Oligocene marked a period of intense climatic change and major turnover in the history of mammals in general, and primates in particular (Prothero, 2012). The global climatic cooling near the Eocene–Oligocene boundary (Zachos et al., 2001) was the major driver of such intense species turnover, and of a major peak in the extinction rate curve. Thereafter, there was an increase in diversification rates in Primates (Arbour & Santana, 2017). This is consistent with our conclusions, which point to a peak in speciation rate during the Miocene, and with genetic analyses linking the intense Miocene primate diversification to higher global mean temperatures (Springer et al., 2012).

Castiglione et al. 2021 – Systematic Biology, under review

It is generally expected that the rate of taxonomic diversification could be positively correlated to the rate of morphological evolution (Rabosky et al., 2013). The expectation makes intuitive sense because a positive association between the two metrics is expected during an adaptive radiation, according to the punctuated equilibrium theory, because genome evolution is accelerated by diversification, and since climatic change and key innovations may spur both new morphological adaptations and the appearance of new species (Raia et al. 2011). Despite these strong theoretical underpinnings, many studies failed to support the expectation (Adams et al.

2009). This might depend on the fact that diversification rate could not grow unbounded (Rabosky & Adams, 2012), that adaptive radiation is limited in time to the early phase of clade history (Pincheira-Donoso et al. 2015), that the pace of trait change might be correlated to extinction, rather than speciation (Lanfear et al. 2010), and that trait-associated diversification should not occur during non-adaptive radiation (Rundell and Price 2009). Most of these processes could actually erase the link between speciation and morphological differentiation even when it is real. For instance, extinction might remove the signature of adaptive radiation in phylogenetic trees deprived of fossil species (Slater et al. 2010). The inclusion of fossils allows better understanding of trait evolution and timing (Slater et al. 2012; Schnitzler et al. 2017), and permits integrating the effect of extinction in the study of phenotypic diversification (Quental and Marshall 2010). Still, for some groups, the information available for living species is poor subsidiary of clade (and hence trait) history [e.g. equids, (Cantalapiedra et al., 2017); archosaurs, (Benson et al., 2014); cetaceans, (Slater et al., 2010)]. Up to now, the inclusion of fossil information in phylogenetically explicit methods to study diversification has been limited. Most phylogenetic comparative methods (PCMs) do not allow the inclusion of fossil species (Rabosky 2014; Maliet et al., 2019). Because of this, paleontological investigations upon the phenotypic diversification of clades and its relationship to taxonomic diversification usually do not include any phylogenetic information (Jablonski & Roy, 2003; Liow, 2004; Liow & Finarelli, 2014; Silvestro et al., 2015) which is restrictive given the overarching effect that shared inheritance might exert on the diversification of species and phenotypes, and the inherently richer information on trait evolution and diversification fossil information provides.

The rate of taxonomic diversification (DR) could simply be computed as the inverse of equal splits, that is the proportion of unique evolutionary time (i.e. the evolutionary distinctiveness) each lineage represents (Redding & Mooers, 2006). Since its debut, DR has been successfully applied to study diversification with paleontological (Cantalapiedra et al., 2017) and living species phylogenies (Cooney et al. 2017; Price Waldman et al. 2020). Title and Rabosky (2019)

demonstrated DR to be the most suitable metrics to study the correlation between species and phenotypic diversifications, as it is prone to low false positive (Type I) error rate. In addition, DR quantifies a single diversification rate value to each lineage in the tree, meaning that the correlation between DR and phenotypic diversification may be inspected lineage by lineage, and distinct lineages be compared to each other.

Here, I developed a new method to test clade-associated and character-associated (treewise) diversification by deriving rates of morphological and taxonomic evolution directly from phylogenetic tree and phenotypic data. Under this new implementation, two different patterns are investigated: 1) shifts in phenotypic evolutionary rates coinciding with shifts in taxonomic diversification rates pertaining to specific clades in the tree or 2) the association between phenotypic trait values and taxonomic diversification rates for the tree as a whole. I investigated the power and accuracy of the new method using both simulation experiments and by applying it to two real cases, the diversification of species and body size in Ornithopoda, and the effect of body size on diversification in mammals.

Materials and Methods

I developed a new R function, *search.shiftDR*, which uses a phylogenetic tree and phenotypic data to identify instances of clade-associated and character-associated diversification.

Clade-Associated Diversification – To recognize clade-associated diversification shifts, *search.shiftDR* scans the phylogeny to find clades showing significantly higher/lower mean DR values associated to significantly faster/slower rates of phenotypic evolution, as compared to the rest of the tree. Under this 'automatic' mode, all the clades ranging in size from one tenth to one half of the tree size are selected for potential shifts testing, although minimum clade size argument could be specified by the user. It must be kept in mind that too small clades are prone to spurious trait-associated diversification results (Adams et al., 2009), hence the minimum clade size argument should carefully be selected. The function starts by computing DR values at the tips. DRs are

calculated as the inverse of the equal splits (ES; Redding & Mooers, 2006; Jetz et al., 2012), a measure of the tree total evolutionary time exclusive to individual lineages. For each clade, *search.shiftDR* computes the difference between the mean DR calculated across all the species within the clade and the DRs averaged over the rest of the species in the tree. Significance is tested by randomization, comparing the difference of DR means $\boldsymbol{m}$ to a null distribution of *n* random differences $\boldsymbol{m}_{random}$ obtained by sampling each time as many DRs as the number of species in the clade under testing and then calculating the *ith* $\boldsymbol{m}_{random}$ difference of means. The metric $\boldsymbol{m}$ is significant if it occurs less than 5% of the times in the $\boldsymbol{m}_{random}$ vector. By default, *n* is set to 1000.

Clades showing significant $\boldsymbol{m}$ are then tested for potential shift in the rate of phenotypic evolution by using the rates obtained applying *RRphylo*. For each selected clade, *search.shiftDR* calculates the difference between the mean absolute *RRrates* of the branches (i.e. both leaves and internal branches) belonging to the clade, and the same figure calculated over the rest of the tree. Significance is assessed via randomization as with DR. Clades showing significant rate shifts in both DR and *RRrates* are taken as candidate instances of clade-associated diversification.

To investigate the phenotypic rate shifts, the user may use Brownian rates (*BMrates*) alternatively to *RRrates* and compare phenotypic rates among clades by means of the function *compare.evol.rates* from the R package `geomorph` (Adams & Otárola Castillo, 2013). *compare.evol.rates* computes rates of morphological evolution for two or more groups of species (that is each selected clade and the rest of the tree, in this case) in a phylogeny, under the Brownian motion model of evolution. The function calculates the ratio of maximum to minimum rates (here *BMrates* ratio) and assesses significance through simulations. Within the *search.shiftDR* machinery, each clade is compared individually (setting it as a separate group) to the rest of the tree under a two-rate model framework. I verified empirically that differences in phenotypic evolutionary rates may appear as a by-product of differences in DR, even when the phenotypic trait is simulated under the Brownian motion (BM). To account for this, regardless of whether *RRrates* or *BMrates* are used, *search.shiftDR* internally calculates the *BMrates* ratio of the candidate clades, then simulates a

trait evolving under BM, and calculates the maximum *BMrates* ratio (***bm.lim***) realized among the candidate clades for such BM phenotype. Eventually, it takes ***2 * bm.lim*** as a threshold to exclude candidate clades with low *BMrates* deviations in the real data. This approach minimizes the number of false positives while having little influence on Type II error. Performing these procedures, a number of candidate clades showing associated rate shifts in the phenotypic and taxonomic diversification are selected, either by using *RRrates* or *BMrates*. These candidates are subjected to a further two steps selection process. First, they are partitioned between those showing positive deviation in DR (i.e. increased diversification) and those showing negative (i.e. decreased diversification) deviation in DR. The candidate clades showing the same direction (positive or negative) in DR shift and laying along a single path (i.e. joined along a single lineage) are contrasted to each other in terms of *RRrates* difference (or *BMrates* ratio), in order to select a single clade per path, picking the node reporting the largest absolute *RRrates* difference (or *BMrates* ratio). This procedure is necessary since nested clades possess strongly correlated patterns in DR and phenotypic rate evolution.

Secondly, the randomization process I apply to select significant ***m*** values (difference in DR means) for a given clade implies a somewhat high probability to find a significant ***m*** of the opposite sign in a different portion of the tree. For this reason, *search.shiftDR* tests each positive-shifting clade by computing the DR means difference and significance after excluding from the vector of DRs any species belonging to negative-shifting clades, and vice-versa. Finally, candidate clades whose shift in DR is eventually confirmed are tested again to confirm the phenotypic evolutionary rate shifts. If *RRphylo* is used, this latter procedure is performed by computing the absolute *RRrates* difference between each clade and the rest of the tree after excluding from the vector of the *RRrates* the values occurring on the branches belonging to other (presumably) shifting clades. Alternatively, if *BMrates* are used, the clades showing DR shifts are fed collectively, each as a distinct group, to *compare.evol.rates* in order to compare their *BMrates* to the rate of the rest of the tree in a multi-rate BM model framework.

In *search.shiftDR*, the user may test specific clades for character-associated diversification. Under such 'node' condition, the function analyzes the focal clade alone by comparing its DR values and *RRrates* (*BMrates*) to the rest of the tree as described above. Then, *search.shiftDR* selects two nodes above and below along the path of the focal clade, and tests them as well by applying the same selection procedures described for the 'automatic' mode. The rationale is that the clade-associated rate shift may actually belong to the same 'region' of the tree as the user-specified clade, without coinciding exactly with it, that is something *search.shiftDR* does not leave unrecognized.

Character-Associated Diversification – To test the hypothesis of character-associated diversification (i.e. a linear relationship between DR and phenotypic trait values), *search.shiftDR* regresses the vector of phenotypes versus DR values computed for each lineage. To assess significance, the regression slope is compared to a random distribution of slopes obtained by regressing 1000 random BM phenotypes against the DR vector. Random phenotypic vectors are produced on the phylogeny according to the BM model by using the function *fastBM* within the R package `phytools` (Revell, 2012) setting σ^2=1 and the phenotype at the root a=0. In order to make slopes comparable, both real data and simulated phenotypes are rescaled in the 0-1 range before regression.

Accounting for the Effect of Sampling – Although being simple and intuitive, the DR metric is sensitive to the effects of sampling (Title & Rabosky, 2019; Louca & Pennell, 2020). Still, a fundamental difference between extant and extinct species is that some of the latter will actually represent chronospecies along a single, non-splitting lineage, hence pseudo-diversification events (i.e. false speciation and false extinction). This may affect the calculation of diversification rates (Silvestro et al., 2018) as much as incomplete sampling does (Etienne et al., 2012). To take into account the potential effects of anagenesis and incomplete sampling on *search.shiftDR* results, I developed the function *overfitDR*. It works by adding or removing a certain (user-specified)

proportion of species from the tree to simulate instances of over-sampling or anagenesis, respectively. In the former case (adding tips to cope with incomplete sampling), the function randomly samples 20% (by default) of the tree species and adds a new 'phantom' tip (species) halfway along the tree branch of each selected species (Fig. 11a-b). 'Phantom' species are identical to their sisters in terms of distance from the tree root. To reconstruct phenotypic values for the 'phantom' species, *overfitDR* uses the function *phylopars* from the R package `Rphylopars` (Goolsby et al., 2017). Given a phylogenetic tree, *phylopars* fits a BM model of evolution on dataset with missing observations and returns BM parameter estimates along with predicted species means for missing phenotypes.

As regards anagenesis, *overfitDR* incorporates a single, extinct species to the branch leading to its sister clade pretending the removed species gradually evolves along the branch into its sister clade ancestor (Fig. 11a-c). This is accomplished by randomly sampling a certain proportion (20% by default) of species among those going extinct before the onset of their sister clade, and then removing them from both phylogeny and phenotypic data.

The user may choose to account for both sampling and anagenetic effects sequentially. In this case, the function adds as many 'phantom' species as 20% of the tree size to account for incomplete sampling, and then produces anagenesis by randomly removing 20% of the species of this enlarged tree.

The procedure to simulate either incomplete sampling, or anagenetic effects, or both, is replicated 10 times to derive 10 sets of modified phylogenetic trees and phenotypic vectors, which are fed as input data to 10 additional *search.shiftDR* runs. At each repetition, the results derived from the *ith search.shiftDR* run are compared to the output produced by the original *search.shiftDR*, recording whether the (possible) clade- or character-associated diversification patterns appear in the *ith* run. The function further stores the proportion of tips falling in the possible clade-associated shift in the original *search.shiftDR* run compared to the *search.shiftDR* runs performed in *overfitDR*. This percentage is 0 if the possible clade-associated shift found in the *ith search.shiftDR*

run in *overfitDR* pertains to a different, non-overlapping clade as compared to the original *search.shiftDR* run and 1 if the two clades coincide. 'Phantom' and anagenetically-removed tips are excluded from the calculation of such percentages.

Effect of Polytomies – Being DR the inverse of the evolutionary time exclusive to a given lineage, as derived from the tree branch lengths, one potential problem with the calculation of DRs is the presence of soft polytomies and zero-length branches. Soft polytomies are usually designed to represent lack in knowledge about the topology of a specific clade. The use of a soft polytomy to represent a clade topology thus forces artificially long terminal branches for all the species in the clade, increasing their evolutionary time and therefore artificially decreasing DR (Fig. 11a-d) To tackle this issue I developed a new function, *fix.poly*, which resolves polytomies randomly, adding nodes within the polytomous clade while keeping fixed the time distance from the root for individual tips and their common ancestor. Differently from other functions designed to resolve polytomies though, the additional nodes are not subtended by zero-length branches, which implies *fix.poly* decreases the total evolutionary time captured by the soft-polytomous clade.

Testing Procedures – I devised two simulation experiments to test the power of *seach.shiftDR* at correctly identifying patterns of clade-associated and character-associated diversification. I also assessed the effect of sampling on *search.shiftDR* results by the means of *overfitDR*.

Clade-associated diversification was simulated to represent all the four possible association between higher/lower DR and higher/lower phenotypic evolutionary rate for a given clade as compared to the rest of the tree. Character-associated diversification was simulated to produce a significant (non-zero) relationship between DR and the corresponding phenotypic values for a continuous trait.

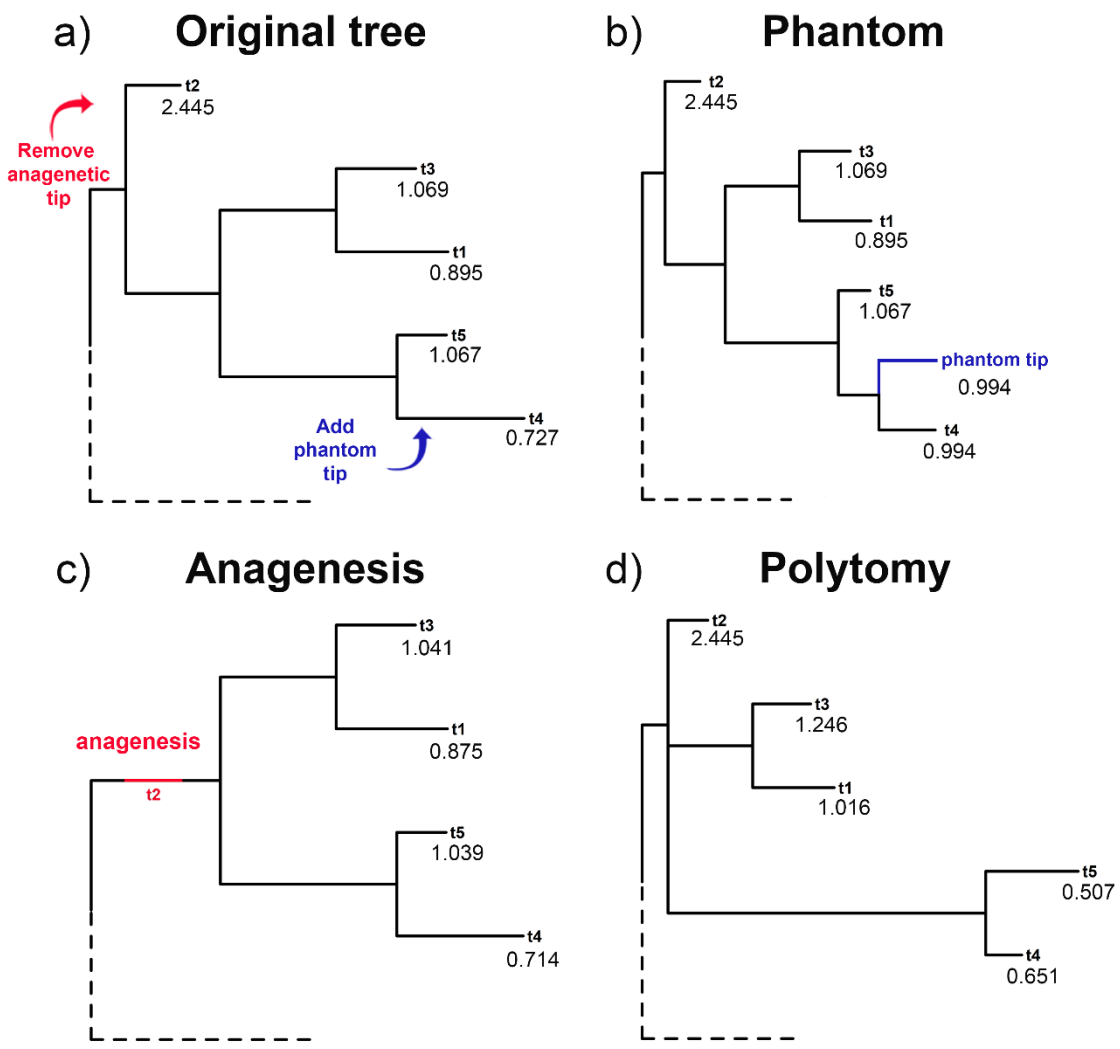

Figure 11. Effect of incomplete sampling, anagenetic speciation, and soft polytomies on Diversification Rates (DR). The values below each tip label represent the DR computed for the lineage. a) the original tree; b) the original tree modified by adding a 'phantom' tip to simulate the effect of incomplete sampling; c) the original tree modified by removing a tip to simulate the effect of anagenetic speciation; d) the original tree modified to simulate the effect of a soft polytomy. Image from Castiglione et al. under review.

In addition to the simulation experiments, I tested the performance of *search.shiftDR* on two real cases commonly reported to represent instances of clade-associated and character-associated diversification, respectively.

The diversification of dinosaurs is commonly described as a two-steps process, marked by bursts of cladogenesis in sauropodomorphs and then in theropods and armored ornithischians

between the Late Triassic and Early Jurassic (Brusatte et al., 2008). Within this large-scale pattern, ornithopods (duck-billed dinosaurs) are recognized as one of the most exceptionally diverse groups, experiencing high taxonomic diversification during Early Cretaceous times. Strickson et al. (2016) found several instances of significant increase in diversification rate in ornithopods, first among iguanodontians in the Middle to Late Jurassic, and then among Hadrosauridae later in the Cretaceous. However, they also noted that the great diversity peaks of these ornithopod clades is accompanied by a limited array of morphologies of jaws and teeth, suggesting taxonomic and phenotypic diversity might be associated in an inverse relationship.

I tested the association between DRs and the rates of body size evolution as occurring in ornithodirans (i.e. clade-associated diversification pattern). Since body size and dental morphologies are tightly linked by shared allometric patterns governing feeding habits, I expected clade-associated diversification in ornithopods, with positive taxonomic and negative phenotypic diversification rates. To test this hypothesis, I derived the ornithodirans phylogenetic tree and body mass data for 419 species, including 93 pterosaurs and 326 dinosaurs from Castiglione et al. (2018).

The observation that small mammals are taxonomically more diverse than large mammals, originating the familiar 'right-skewed' distribution of species richness as a function of body mass, is pervasive in the scientific literature (Martin, 2016). Although the relationship received little to no support in phylogenetic analyses (Isaac et al., 2015), with some studies even recording higher turnover rates for larger-bodied species (Liow et al., 2008), it still suggests that rates of diversification are influenced by body size in mammals. I tested the relationship between DRs and body size (i.e. character-associated diversification) for a 929 species-wide phylogenetic tree of large mammals derived from Raia et al. (2013), by means of both *search.shiftDR* and ES-sim.

Both the ornithodirans and mammalian phylogenies included soft polytomies, possibly affecting the calculation of DRs. Therefore, I tested both real cases under a routine function, *FSO*, which performs *search.shiftDR* and *overfitDR* after randomly dichotomizing the tree by means of *fix.poly*. Since *fix.poly* fixes the polytomies at random, the procedure is replicated a user-specified

number of times as to account for the effect of different polytomies resolutions. I applied *FSO* on both ornithodirans and mammals phylogenetic trees and phenotypic data once by using *RRrates* and then by using *BMrates*, performing 10 replications and setting the *overfitDR* runs at one.Eventually, I tested the body size-associated diversification in mammals by performing *ES-sim* first on the phylogenetic trees derived by using *fix.poly* within *FSO*, then on the tree produced by *multi2di*.

Results

Simulations – Using *RRphylo* under the automatic mode I found 88.12% of simulations producing significant results for DR+RR+ model. In 96.81% of them the model was correctly identified. There was no instance of Type 1 error. For DR+RR-, I found 86.13% simulations producing significant results, with 99.02% of them correctly identifying the model. I recorded no instance of Type 1 error. As for DR-RR+, 62.86% simulations produced significant results. In 92% of them the model was correctly identified. The Type 1 error was 0.58%. I found 53.86% simulations producing significant results for DR-RR-. In 94.54% of them the model was correctly identified. I recorded 0.55% instances of Type 1 error.

Results of simulations performed by using BMrates to identify phenotypic evolutionary rates shift, produced 90.5% significant results for DR+BR+. The model was correctly identified in 97.61% of cases. The incidence of Type 1 error rate was 0.43%. I found 87.4% significant simulations for DR+BR-, with 99.04% correctly identifying the model. There was no instance of Type 1 error. For DR-BR+ simulations, the function produced significant results in 66.24% of cases. In 91.25% of cases the model was correctly identified. I found 0.39% instances of Type 1 error. As for DR-BR-, I found 51.65% significant simulations. The model was correctly identified in 95.02%. The incidence of Type 1 error was 0.73%.

Accounting for the effect of both incomplete sampling and anagenesis on *search.shiftDR* as applied by using the *RRrates*, the percentage of simulations including at least one significant *search.shiftDR* result among the 10 runs within *overfitDR* was 99.4%. The mean percentage of

significant runs per simulation was 93.97%. I computed the percentage of tips belonging to the clade-associated shift found by the original *search.shiftDR* which were also found to shift in *overfitDR* runs, at each simulation. The average value for overall simulations was 80.55%. The incidence of Type 1 error rate was 0.68%. By using *BMrates*, 98.54% *overfitDR* simulations included one significant run at least, with 91.41% significant runs per simulation on average. The mean percentage of tips found by both the original *search.shiftDR* and the runs within *overfitDR* over all simulations was 78.57%. I found 1.81% instances of Type 1 error.

search.shiftDR proved powerful in finding either positive (83.5%) or negative (77.5%) instances of character-associated diversification. The percentage of Type 1 error rate was 5.05% for either simulation type. For *ES-sim* the type 1 error rate was identical to *search.shiftDR*. The power was 74.7% and 70.2% for positive and negative association between DR and phenotype, respectively.

By applying *overfitDR* to account for the effect of both incomplete sampling and anagenesis, I found 90.2% of positive simulations to include one significant *search.shiftDR* run at least. The same figure for negative simulations was 90%. Applying false discovery rate correction of the unequal number of tests, these figures became 76.8% for the positive and 77.8% for the negative relationships, indicating *search.shiftDR* is robust to the effects of sampling and anagenesis. The mean percentage of significant *search.shiftDR* runs per simulation was 81.46% (that is 8.15 times out of ten runs) and 81.66% (that is 8.17 times out of ten runs) for simulated positive and negative relationships, respectively. The incidence of Type 1 error rate on *overfitDR* results was 4.5% overall (after accounting for false discovery rate), with 2.85% (that is 0.28 times out of ten runs) significant runs per simulation on average.

Real Cases – By testing clade-associated diversification in Ornithodirans, I found support for a positive shift in DR coincident to a negative shift in body size evolutionary rate (regardless of whether *RRrates* or *BMrates* were used) in Hadrosauroidea (Fig. 12). By using *RRrates*, 50%

search.shiftDR as performed on the dichotomized versions of the original tree produced significant results. In all of them, *Dryomorpha* belonged to the shifting clade. In 30% cases the genus *Tenontosaurus* (sister to *Dryomorpha*) was also included in the shift. *overfitDR* proved the results were robust to sampling. I found 40% significant shift in *overfitDR* including the clade Hadrosauriphormes (Fig. 12a). In 30% of the cases the shift extended to Dryomorpha + genus *Tenontosaurus*, in 20% it involved the entire Iguanodontia excluding *Anabisetia saldiviai*. Finally, in 10% of the *FSO* the shift extended to stem Iguanodontia (i.e. including *Gasparinisaura cicosaltensis* and *Hypsilophodon foxii*).

By using *BMrates*, 20% *search.shiftDR* performed within *FSO* produced significant results. Hadrosauriphormes were always included in the shifting clade. In 10% of cases the shift pertained to Hadrosauriphormes only, while in 10% it extended to Dryomorpha. *overfiDR* returned 60% of the cases indicating a shift for Hadrosauroidea (Fig. 12b). In 30% of the cases the shift pertained to Hadrosauriphormes either including or excluding a single species within the clade.

I found limited support for body size-associated diversification pattern in Mammals. *search.shiftDR* and *overfitDR* performed within *FSO* produced both 20% significant and negative slopes of DRs versus body size regression. *ES-sim* produced no significant results.

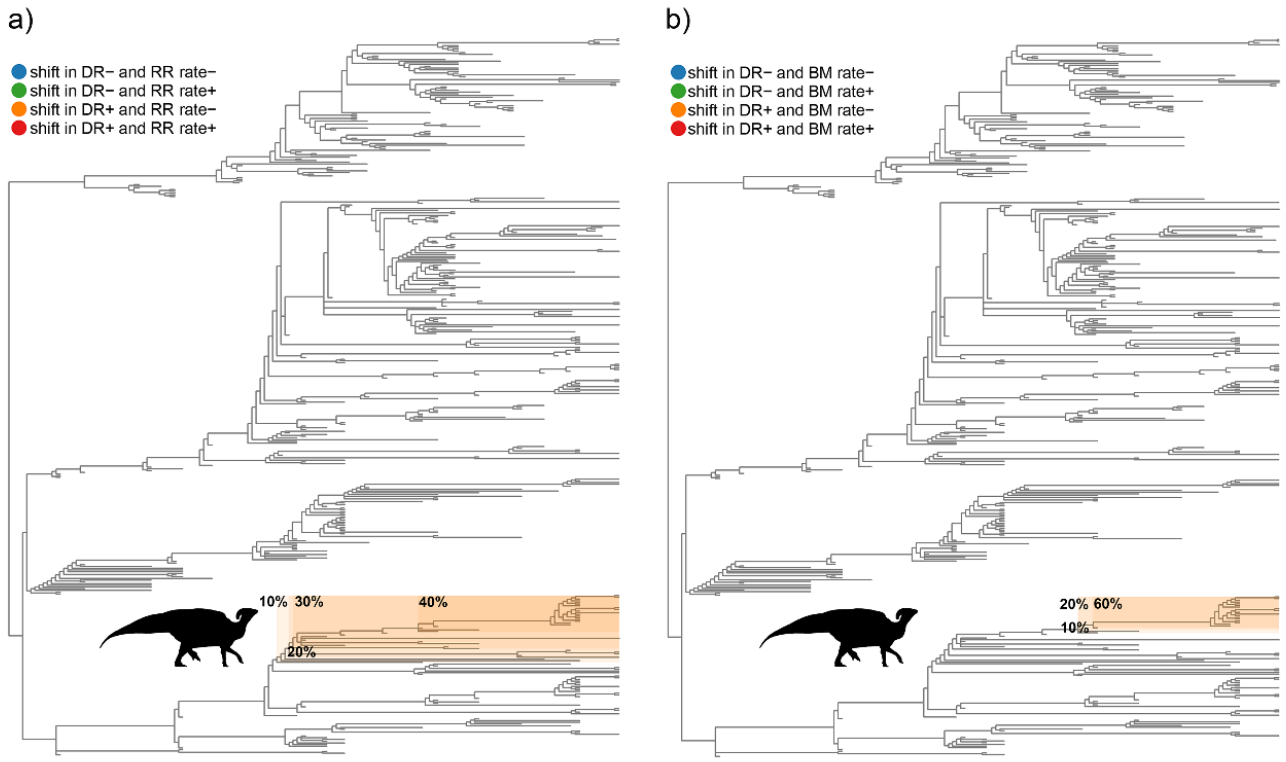

Figure 12. Results of 10 *FSO* replications performed on the Ornithodirans tree and log body size data. The shaded rectangles overlaying the clades indicate the direction of the shift in both DR and RR/BM rate. The percentages specify the proportion of *overfitDR* runs producing significant result for each clade. a) *FSO* performed by using *RRrates.* 40% significant shift in *overfitDR* include the clade Hadrosauriphormes, in 30% of the cases the shift extends to Dryomorpha + the sister genus *Tenontosaurus*, in 20% it involves the entire Iguanodontia excluding *Anabisetia saldiviai*, in 10% the shift extends to stem Iguanodontia (i.e. including *Gasparinisaura cicosaltensis* and *Hypsilophodon foxii*). b) *FSO* performed by using *BMrates.* 60% cases indicate a shift for Hadrosauroidea, in 30% of the cases the shift extends to Hadrosauriphormes either including or excluding a single species within the clade.

Image from Castiglione et al. under review.

Discussion

Studying the relationship between morphological evolution and taxonomic diversification is of paramount importance to understand the diversity of life. The rate of evolution of the two should be correlated during ecological speciation, that is when adaptation prompts speciation (Adams et al., 2009; Anderson & Weir, 2020), or when adaptation increases survival rate (Otto, 2018). At the

macroevolutionary scale, ecological adaptation during adaptive radiation should equally originate a positive association between rates of taxonomic and phenotypic diversification (Schluter, 2001; Harmon et al., 2003; Adams et al., 2009) whereas non-adaptive radiation (Rundell & Price, 2009) and density-dependent taxonomic diversification disrupt the pattern or make it dependent on the age of the clade under scrutiny (Rabosky & Lovette, 2008). A major problem with the study of a trait-associated diversification pattern is understanding its timing under a phylogenetically explicit context (Etienne and Haegeman 2012). Unfortunately, few statistical tools allow performing the task, and most of them cannot include explicit fossil evidence (Rabosky, 2014; Maliet et a,l. 2019) which is a limit given the enormous importance of such piece of information in terms of correctly understanding morphological and taxonomic evolution of clades (Raia et al., 2013; Schnitzler et al., 2017). Here I fill this gap, by developing a new method to study the association between taxonomic and phenotypic diversification under a phylogenetically informed framework, that makes full use of fossil information. The method seeks for instances of significant trait-associated diversification pertaining to specific clades in the tree, and character-associated diversification, that is the significant association between trait values and taxonomic diversification.

I used simulation experiments to test the method. In testing clade-associated diversification, the method provides low false positives rate and good power. In terms of character-associated diversification, the method is as powerful and as prone to Type I error as an existing method (ES-sim, Harvey & Rabosky, 2017). I applied the new method to two real cases, that is clade-associated diversification in ornithodirans (the clade including pterosaurs and dinosaurs) and character-associated diversification in mammals. In both cases, body size was the phenotype under testing. Major increases in diversification rate were related in the past to the diversification of sauropods, maniraptoran theropods and ornitischians among dinosaurs (Benson et al., 2014; Strickson et al., 2016). I deliberately excluded avian dinosaurs from the ornithodiran tree, which is I cut the tree to the Late Cretaceous. Hence, trait-associated diversification in maniraptorans (the carnivorous dinosaur clade including the birds' lineage) is not directly tested. I found significant evidence that

hadrosaurid (duck-billed) dinosaurs underwent a major taxonomic radiation accompanied by little body size radiation. Strickson et al. (2016) found major diversification bursts pertain to Campanian (Cretaceous) hadrosaurids, which were at one time greatly diversifying in terms of species, and yet represented by a very limited array of different dental morphologies. Because of the tight link between body size and diet, I argue the close association between low rates of body size evolution and high rates of taxonomic diversification in hadrosaurids I found here is entirely consistent with Strickson et al.'s (2016) findings. The second real-case study I dealt with was the potential association between taxonomic diversification rate and body size in mammals. Although intuitive because of the higher metabolic rate and short intergenerational time in smaller animals (Calder, 1984), a negative relationship between body size and the rate of taxonomic diversification was not strongly supported in literature (Feldman et al., 2015). In mammals, either no support was found for the relationship (Isaac et al., 2015), or even the opposite pattern (i.e. higher rates in lineage of large bodies taxa) was found (Liow et al., 2008). I found weak evidence that, in large mammals at the least, diversification rate is inversely related to body size.

A potential limitation with the use of DR as a diversification metric is that it cannot distinguish between the relative contribution of speciation and extinction to diversification (Title & Rabosky, 2019; Maliet et al., 2019) and is sensitive to the relative proportion of extant taxa which are included (Louca & Pennell, 2020). However, the method works at least as well as competing existing methods in terms of estimating taxonomic diversification (Harvey & Rabosky, 2017), and is among the few feasible to apply to phylogenies including fossils (Cantalapiedra et al., 2017). I developed a specific function, *overfitDR*, to deal with the great uncertainty in taxonomic diversification estimates which comes with sampling uncertainty (Etienne et al., 2012) and the presence of anagenetic lineages in the tree (Silvestro et al., 2018). I found that s*earch.shiftDR* is robust to moderate effects of sampling uncertainty and anagenesis, even tested in combination. It is important to remark that *search.shiftDR* cannot test for shifts in either taxonomic or phenotypic diversification rate individually, but only to the association between the two. The results of

simulation experiments seem nonetheless to suggest that *search.shiftDR* is sensible to relatively small shifts in trait-associated diversification, and is prone to low Type I error rate. However, it must be taken in mind *search.shiftDR* does not derive from explicit modelling of variation in the birth-death process (Maliet et al., 2019), and as such must be understood in terms of finding patterns of trait-associated diversification, rather than to figure out the process behind them.

LIST OF RESEARCH PRODUCTS

1. Castiglione S., Tesone G., Piccolo M., Melchionna M., Mondanaro A., Serio C., Di Febbraro M., & Raia P. (2018). A new method for testing evolutionary rate variation and shifts in phenotypic evolution. *Methods in Ecology and Evolution, 9* (4), 974-983. DOI:10.1111/2041-210X.12954

   The paper was cited 27 times (Web of Science). The method was used to study the evolution of red blood cell shape in fishes (Oliveira Martins et al., 2020, Journal of Evolutionary Biology), the evolution of distal forelimbs in horses (MacLaren, 2021, Proceedings of the Royal Society B), the role of GC content in accumulation of deletion mutations in the coding region of angiosperm plastomes (Yu et al., 2021, Journal of Molecular Evolution), the difference in evolutionary rates in life-history traits between insular and mainland palms (Cassia-Silva et al., 2020, Scientific Reports), butterfly dichromatism (van der Bijl et al., 2020, Evolution Letters), plumage color complexity in the largest family of songbirds (Price-Waldman et al., 2020, Evolution), the evolvability and craniofacial diversification in genus *Homo* (Baab, 2018, Evolution).

2. Raia P., Boggioni M., Carotenuto F., Castiglione S., Di Febbraro M., Di Vincenzo F., Melchionna M., Mondanaro A., Papini A., Profico A., Serio C., Veneziano A., Vero V.A., Rook L., Meloro C., & Manzi G. (2018). Unexpectedly rapid evolution of mandibular shape in hominins. *Scientific Reports, 8* (1), 7340. DOI:10.1038/s41598-018-25309-8
3. Mondanaro A., Melchionna M., Di Febbraro M., Castiglione S., Holden P.B., Edwards N.R., Carotenuto F., Maiorano L., Modafferi M., Serio C., Diniz-Filho J.A.F., Rangel T.F., Rook L., O'Higgins P., Spikins P., Profico A., & Raia P. (2020). A major change in rate of climate niche envelope evolution during hominid history. *iScience, 23*(11), 101693. DOI:10.1016/j.isci.2020.101693

4. Sansalone G., Allen K., Ledogar J.A., Ledogar S., Mitchell D.R., Profico A., Castiglione S., Melchionna M., Serio C., Mondanaro A., Raia P., Wroe S. (2020). Variation in the strength of allometry drives rates of evolution in primate brain shape. *Proceedings of the Royal Society B: Biological Sciences, 287*(1930), 20200807. DOI:10.1098/rspb.2020.0807
5. Castiglione S., Serio C., Mondanaro A., Di Febbraro M., Profico A., Girardi G., Raia P. (2019). Simultaneous detection of macroevolutionary patterns in phenotypic means and rate of change with and within phylogenetic trees including extinct species. *PLoS ONE, 14* (1), e0210101. DOI:10.1371/journal.pone.0210101

   The paper was cited 6 times (Web of Science). The method was used to investigate progressive evolution of genome sequence complexity in Cyanobacteria (Moya et al., 2020, Scientific Reports) and Crocodylomorph cranial shape evolution (Godoy, 2020, Journal of Evolutionary Biology)

6. Serio C., Castiglione S., Tesone G., Piccolo M., Melchionna M., Mondanaro A., Di Febbraro M., Raia P. (2019). Macroevolution of Toothed Whales Exceptional Relative Brain Size. *Evolutionary Biology, 46* (4), 332-342. DOI:10.1007/s11692-019-09485-7
7. McCurry M.R., Marx F.G., Evans A.R., Park T., Pyenson N.D., Kohno N., Castiglione S., Fitzgerald E.M.G. (2021). Brain size evolution in whales and dolphins: new data from fossil mysticetes. *Biological Journal of the Linnean Society*, in press.
8. Melchionna M., Profico A., Castiglione S., Sansalone G., Serio C., Mondanaro A., Di Febbraro M., Rook L., Pandolfi L., Di Vincenzo F., Manzi G., Raia P. (2020). From Smart Apes to Human Brain Boxes. A Uniquely Derived Brain Shape in Late Hominins Clade. *Frontiers in Earth Science, 8*, 273. DOI:10.3389/feart.2020.00273
9. Castiglione S., Serio C., Mondanaro A., Melchionna M., Carotenuto F., Di Febbraro M., Profico A., Tamagnini D., Raia P. (2020). Ancestral State Estimation with Phylogenetic Ridge Regression. *Evolutionary Biology, 47* (3), 220-232. DOI:10.1007/s11692-020-09505-x

10. Castiglione S., Serio C., Piccolo M., Mondanaro A., Melchionna M., Di Febbraro M., Sansalone G., Wroe S., Raia P. (2020). The influence of domestication, insularity and sociality on the tempo and mode of brain size evolution in mammals. *Biological Journal of the Linnean Society, 132*, 121-131. DOI:10.1093/biolinnean/blaa186
11. Castiglione S., Serio C., Tamagnini D., Melchionna M., Mondanaro A., Di Febbraro M., Profico A., Piras P., Barattolo F., Raia P. (2019). A new, fast method to search for morphological convergence with shape data. *PLoS ONE, 14* (12), e0226949. DOI:10.1371/journal.pone.0226949

    The paper was cited 7 times (Web of Science). The method was used to study the macroevolutionary landscape of short-necked plesiosaurs (Fischer et al., 2020, Scientific Reports) and the evolution of scapula shape in bats (Gaudioso et al., 2020, Journal of Zoological Systematics and Evolutionary Research)

12. Melchionna M., Profico A., Castiglione S., Serio C., Mondanaro A., Modafferi M., Tamagnini D., Maiorano L., Raia P., Witmer L.M., Wroe S., Sansalone G. (2021). A method for mapping morphological convergence on three-dimensional digital models: the case of the mammalian saber-tooth. *Palaeontology, in press*
13. Sansalone G., Castiglione S., Raia P., Archer M., Dickson B., Hand S., Piras P., Profico A., Wroe S. (2020). Decoupling Functional and Morphological Convergence, the Study Case of Fossorial Mammalia. *Frontiers in Earth Science, 8*, 112. DOI:10.3389/feart.2020.00112
14. Castiglione S., Mondanaro A., Melchionna M., Serio C., Di Febbraro M., Carotenuto F., Raia P. (2017). Diversification rates and the evolution of species range size frequency distribution. *Frontiers in Ecology and Evolution, 5*, 147. DOI:10.3389/fevo.2017.00147
15. Piras P., Silvestro D., Carotenuto F., Castiglione S., Kotsakis A., Maiorino L., Melchionna M., Mondanaro A., Sansalone G., Serio C., Vero V.A., Raia P. (2018). Evolution of the sabertooth

mandible: A deadly ecomorphological specialization. *Palaeogeography, Palaeoclimatology, Palaeoecology, 496*, 166-174. DOI:10.1016/j.palaeo.2018.01.034

16. Melchionna M., Mondanaro A., Serio C., Castiglione S., Di Febbraro M., Rook L., Diniz-Filho J.A.F., Manzi G., Profico A., Sansalone G., Raia P. (2020). Macroevolutionary trends of brain mass in Primates. *Biological Journal of the Linnean Society, 129*, 14-25. DOI:10.1093/biolinnean/blz161
17. Silvestro D., Castiglione S., Mondanaro A., Serio C., Melchionna M., Piras P., Di Febbraro M., Carotenuto F., Rook L., Raia P. (2020). A 450 million years long latitudinal gradient in age-dependent extinction. *Ecology Letters, 23* (3), 439-446. DOI:10.1111/ele.13441
18. Castiglione S., Serio C., Mondanaro A., Melchionna M., Di Febbraro M., Meloro C., Raia P. Identifying Trait-Associated Diversification with Fossil Phylogenies. *Systematic Biology, under review.*

REFERENCES

1. Adams, D. C., & Collyer, M. L. (2009). A general framework for the analysis of phenotypic trajectories in evolutionary studies. *Evolution*, *63*(5), 1143-1154.
2. Adams, D. C., & Otárola-Castillo, E. (2013). geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution, 4*(4), 393-399.
3. Adams, D. C., Berns, C. M., Kozak, K. H., & Wiens, J. J. (2009). Are rates of species diversification correlated with rates of morphological evolution?. *Proceedings of the Royal Society B: Biological Sciences, 276*(1668), 2729-2738.
4. Alexander, R. (1998). All-time giants: the largest animals and their problems. *Palaeontology, 41*(6), 1231-1246.
5. Alfaro, M. E., Santini, F., Brock, C., Alamillo, H., Dornburg, A., Rabosky, D. L., ... & Harmon, L. J. (2009). Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences*, *106*(32), 13410-13414.
6. Alroy, J. (2008). Dynamics of origination and extinction in the marine fossil record. *Proceedings of the National Academy of Sciences, 105*, 11536-11542.
7. Alroy, J. (2014). Accurate and precise estimates of origination and extinction rates. *Paleobiology*, *40*(3), 374-397.
8. Amanzougaghene, N., Fenollar, F., Raoult, D., & Mediannikov, O. (2020). Where are we with human lice? A review of the current state of knowledge. *Frontiers in cellular and infection microbiology, 9*, 474.
9. Anderson, S. A., & Weir, J. T. (2020). A Comparative Test for Divergent Adaptation: Inferring Speciation Drivers from Functional Trait Divergence. *The American Naturalist*, *196*(4), 429-442.

10. Arbour, J. H., & Santana, S. E. (2017). A major shift in diversification rate helps explain macroevolutionary patterns in primate species diversity. *Evolution*, *71*(6), 1600-1613.

11. Aristide, L., Dos Reis, S. F., Machado, A. C., Lima, I., Lopes, R. T., & Perez, S. I. (2016). Brain shape convergence in the adaptive radiation of New World monkeys. *Proceedings of the National Academy of Sciences*, *113*(8), 2158-2163.

12. Avaria-Llautureo, J., Hernández, C. E., Boric-Bargetto, D., Canales-Aguirre, C. B., Morales-Pallero, B., & Rodríguez-Serrano, E. (2012). Body size evolution in extant Oryzomyini rodents: Cope's rule or miniaturization?. *PLoS One*, *7*(4), e34654.

13. Balzeau, A., Grimaud-Hervé, D., Détroit, F., Holloway, R. L., Combès, B., & Prima, S. (2013). First description of the Cro-Magnon 1 endocast and study of brain variation and evolution in anatomically modern Homo sapiens. *Bulletins et Mémoires de la Société d'Anthropologie de Paris*, *25*(1-2), 1-18.

14. Barnosky, A. D. (2001). Distinguishing the effects of the Red Queen and Court Jester on Miocene mammal evolution in the northern Rocky Mountains. *Journal of Vertebrate Paleontology*, *21*(1), 172-185.

15. Barycka, E. (2007). Evolution and systematics of the feliform Carnivora. *Mammalian Biology, 72*(5), 257-282.

16. Beaulieu, J. M., Jhwueng, D. C., Boettiger, C., & O'Meara, B. C. (2012). Modeling stabilizing selection: expanding the Ornstein–Uhlenbeck model of adaptive evolution. *Evolution*, *66*(8), 2369-2383.

17. 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*(5), e1001853.

18. Benson-Amram, S., Dantzer, B., Stricker, G., Swanson, E. M., & Holekamp, K. E. (2016). Brain size predicts problem-solving ability in mammalian carnivores. *Proceedings of the National Academy of Sciences*, *113*(9), 2532-2537.

19. Berta, A., Lanzetti, A., Ekdale, E. G., & Deméré, T. A. (2016). From teeth to baleen and raptorial to bulk filter feeding in mysticete cetaceans: the role of paleontological, genetic, and geochemical data in feeding evolution and ecology. *Integrative and Comparative Biology*, *56*(6), 1271-1284.

20. Blackburn, T. M., & Gaston, K. J. (Eds.). (2003). *Macroecology: concepts and consequences: 43rd symposium of the British Ecological Society (Vol. 43).* Cambridge University Press.

21. Bloch, J. I., Silcox, M. T., Boyer, D. M., & Sargis, E. J. (2007). New Paleocene skeletons and the relationship of plesiadapiforms to crown-clade primates. *Proceedings of the National Academy of Sciences*, *104*(4), 1159-1164.

22. Boivin, N. L., Zeder, M. A., Fuller, D. Q., Crowther, A., Larson, G., Erlandson, J. M., ... & Petraglia, M. D. (2016). Ecological consequences of human niche construction: Examining long-term anthropogenic shaping of global species distributions. *Proceedings of the National Academy of Sciences*, *113*(23), 6388-6396.

23. Bookstein, F. L. (2015). Integration, disintegration, and self-similarity: characterizing the scales of shape variation in landmark data. *Evolutionary biology*, *42*(4), 395-426.

24. Bruner, E., Ogihara, N., & Tanabe, H. C. (2018). *Digital Endocasts: From Skulls to Brains*. Tokyo: Springer.

25. Brusatte, S. L., Benton, M. J., Lloyd, G. T., Ruta, M., & Wang, S. C. (2010). Macroevolutionary patterns in the evolutionary radiation of archosaurs (Tetrapoda: Diapsida). *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, *101*(3/4), 367.

26. Butler, M. A., & King, A. A. (2004). Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *The American Naturalist, 164*(6), 683-695.

27. Calder, W. A. (1984). *Size, function, and life history.* Cambridge, Massachusetts: Harvard University Press.

28. Cantalapiedra, J. L., Prado, J. L., Fernández, M. H., & Alberdi, M. T. (2017). Decoupled

ecomorphological evolution and diversification in Neogene-Quaternary horses. *Science, 355*(6325), 627-630.

29. Cârciumaru, M., Ion, R. M., Niţu, E. C., & Ştefănescu, R. (2012). New evidence of adhesive as hafting material on Middle and Upper Palaeolithic artefacts from Gura Cheii-Râşnov Cave (Romania). *Journal of Archaeological Science*, *39*(7), 1942-1950.
30. Cardillo, M., Mace, G. M., Jones, K. E., Bielby, J., Bininda-Emonds, O. R., Sechrest, W., ... & Purvis, A. (2005). Multiple causes of high extinction risk in large mammal species. *Science, 309*(5738), 1239-1241.
31. Carotenuto, F., Barbera, C. & Raia, P. (2010). Occupancy, range size, and phylogeny in Eurasian Pliocene to Recent large mammals. *Paleobiology, 36*, 399–414.
32. Cerling, T. E., Mbua, E., Kirera, F. M., Manthi, F. K., Grine, F. E., Leakey, M. G., ... & Uno, K. T. (2011). Diet of Paranthropus boisei in the early Pleistocene of East Africa. *Proceedings of the National Academy of Sciences*, *108*(23), 9337-9341.
33. Chira, A. M., & Thomas, G. H. (2016). The impact of rate heterogeneity on inference of phylogenetic models of trait evolution. *Journal of evolutionary biology*, *29*(12), 2502-2518.
34. Christiansen, P. (2008). Evolutionary convergence of primitive sabertooth craniomandibular morphology: the clouded leopard (Neofelis nebulosa) and Paramachairodus ogygia compared. *Journal of Mammalian Evolution*, *15*(3), 155-179.
35. Christiansen, P. E. R. (2007). Canine morphology in the larger Felidae: implications for feeding ecology. *Biological Journal of the Linnean Society*, *91*(4), 573-592.
36. Churchill, M., Geisler, J. H., Beatty, B. L., & Goswami, A. (2018). Evolution of cranial telescoping in echolocating whales (Cetacea: Odontoceti). *Evolution*, *72*(5), 1092-1108.
37. Cooney, C. R., Bright, J. A., Capp, E. J. R., Chira, A. M., Hughes, E. C., Moody, C.J.A., … & Thomas, G. H. (2017). Mega-evolutionary dynamics of the adaptive radiation of birds. *Nature, 542*, 344–347.

38. Cooper, N., Thomas, G. H., Venditti, C., Meade, A., & Freckleton, R. P. (2016). A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. *Biological Journal of the Linnean Society, 118*(1), 64-77.

39. Demment, M. W., & Van Soest, P. J. (1985). A nutritional explanation for body-size patterns of ruminant and nonruminant herbivores. *The American Naturalist*, *125*(5), 641-672.

40. Diniz-Filho, J. A. F., & Raia, P. (2017). Island Rule, quantitative genetics and brain–body size evolution in Homo floresiensis. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1857), 20171065.

41. Diniz-Filho, J. A. F., Jardim, L., Mondanaro, A., & Raia, P. (2019a). Multiple components of phylogenetic non-stationarity in the evolution of brain size in fossil hominins. *Evolutionary Biology*, *46*(1), 47-59.

42. Diniz-Filho, J. A. F., Jardim, L., Rangel, T. F., Holden, P. B., Edwards, N. R., Hortal, J., ... & Raia, P. (2019b). Quantitative genetics of body size evolution on islands: an individual-based simulation approach. *Biology Letters*, *15*(10), 20190481.

43. Dowdeswell, M. R., Jashashvili, T., Patel, B. A., Lebrun, R., Susman, R. L., Lordkipanidze, D., & Carlson, K. J. (2017). Adaptation to bipedal gait and fifth metatarsal structural properties in Australopithecus, Paranthropus, and Homo. *Comptes Rendus Palevol*, *16*(5-6), 585-599.

44. Du, A., Zipkin, A. M., Hatala, K. G., Renner, E., Baker, J. L., Bianchi, S., ... & Wood, B. A. (2018). Pattern and process in hominin brain size evolution are scale-dependent. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1873), 20172738.

45. Dunbar, R. I. (2009). The social brain hypothesis and its implications for social evolution. *Annals of human biology*, *36*(5), 562-572.

46. Dunbar, R. I., & Shultz, S. (2007). Evolution in the social brain. *Science*, *317*(5843), 1344-1347.

47. Dzeverin, I. (2008). The stasis and possible patterns of selection in evolution of a group of related species from the bat genus Myotis (Chiroptera, Vespertilionidae). *Journal of Mammalian Evolution*, *15*(2), 123-142.

48. 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*(12), 3578-3589.

49. Elliot, M. G., & Mooers, A. Ø. (2014). Inferring ancestral states without assuming neutrality or gradualism using a stable model of continuous character evolution. *BMC evolutionary biology*, *14*(1), 226.

50. Etienne, R. S., & Haegeman, B. (2012). A conceptual and statistical framework for adaptive radiations with a key role for diversity dependence. *The American Naturalist*, *180*(4), E75-E89.

51. Etienne, R.S., Haegeman, B., Stadler, T., Aze, T., Pearson, P.N., Purvis, A., & Phillimore, A. B. (2012). Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proceedings of the Royal Society B: Biol. Sci., 279*, 1300–1309.

52. Fahlke, J. M., Bastl, K. A., Semprebon, G. M., & Gingerich, P. D. (2013). Paleoecology of archaeocete whales throughout the Eocene: dietary adaptations revealed by microwear analysis. *Palaeogeography, Palaeoclimatology, Palaeoecology, 386*, 690-701.

53. Feldman, A., Sabath, N., Pyron, R. A., Mayrose, I., & Meiri, S. (2016). Body sizes and diversification rates of lizards, snakes, amphisbaenians and the tuatara. *Global Ecology and Biogeography*, *25*(2), 187-197.

54. Felsenstein, J. (1985). Phylogenies and the comparative method. *The American Naturalist, 125*(1), 1-15.

55. Figueirido, B., Janis, C. M., Pérez-Claros, J. A., De Renzi, M., & Palmqvist, P. (2012). Cenozoic climate change influences mammalian evolutionary dynamics. *Proceedings of the National Academy of Sciences*, *109*(3), 722-727.

56. Finarelli, J. A., & Flynn, J. J. (2009). Brain-size evolution and sociality in Carnivora. *Proceedings of the National Academy of Sciences*, *106*(23), 9345-9349.

57. Finarelli, J. A., & Liow, L. H. (2016). Diversification histories for North American and Eurasian carnivorans. *Biological Journal of the Linnean Society, 118*, 26–38.

58. Finnegan, S., Payne, J. L., & Wang, S. C. (2008). The Red Queen revisited: reevaluating the age selectivity of Phanerozoic marine genus extinctions. *Paleobiology*, *34*(3), 318-341.

59. Fitzgerald, E. M. (2012). Archaeocete-like jaws in a baleen whale. *Biology Letters*, *8*(1), 94-96.

60. Foley, R. A., Martin, L., Mirazón Lahr, M., & Stringer, C. (2016). Major transitions in human evolution.

61. Foote, M. (2000). Origination and extinction components of taxonomic diversity: Paleozoic and post-Paleozoic dynamics. *Paleobiology, 26*(4), 578-605.

62. Foote, M., Crampton, J. S., Beu, A. G., Marshall, B. A., Cooper, R. A., Maxwell, P. A., & Matcham, I. (2007). Rise and fall of species occupancy in Cenozoic fossil mollusks. *Science*, *318*(5853), 1131-1134.

63. Fordyce, R. E., & Marx, F. G. (2018). Gigantism precedes filter feeding in baleen whale evolution. *Current Biology*, *28*(10), 1670-1676.

64. Freckleton, R. P. (2009). The seven deadly sins of comparative analysis. *Journal of evolutionary biology*, *22*(7), 1367-1375.

65. Gaston, K. J. (1998). Species-range size distributions: products of speciation, extinction and transformation. *Philosophical Transactions of the Royal Society of London, 353*(1366), 219-230.

66. Gaston, K. J. (2003). *The Structure and Dynamics of Geographic Ranges.* Oxford: Oxford University Press.

67. Gavryushkina, A., Heath, T. A., Ksepka, D. T., Stadler, T., Welch, D., & Drummond, A. J. (2017). Bayesian total-evidence dating reveals the recent crown radiation of penguins.

Systematic Biology, 66, 57–73.

68. Gavryushkina, A., Welch, D., Stadler, T., & Drummond, A. J. (2014). Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Computational Biology, 10*, e1003919.

69. Geiger, M., & Sánchez-Villagra, M. R. (2018). Similar rates of morphological evolution in domesticated and wild pigs and dogs. *Frontiers in zoology*, *15*(1), 23.

70. Gingerich, P. D. (2016). Body weight and relative brain size (encephalization) in Eocene Archaeoceti (Cetacea). *Journal of Mammalian Evolution*, *23*(1), 17-31.

71. Goolsby, E. W., Bruggeman, J., & Ane, C. (2017). Rphylopars: fast multivariate phylogenetic comparative methods for missing data and within-species variation. *Methods in Ecology and Evolution*, *8*(1), 22-27.

72. Gubry-Rangin, C., Kratsch, C., Williams, T. A., McHardy, A. C., Embley, T. M., Prosser, J. I., & Macqueen, D. J. (2015). Coupling of diversification and pH adaptation during the evolution of terrestrial Thaumarchaeota. *Proceedings of the National Academy of Sciences*, *112*(30), 9370-9375.

73. Haile-Selassie, Y., Suwa, G., & White, T. D. (2004). Late Miocene teeth from Middle Awash, Ethiopia, and early hominid dental evolution. *Science*, *303*(5663), 1503-1505.

74. Hansen, T. F. (1997). Stabilizing selection and the comparative analysis of adaptation. *Evolution, 51*(5), 1341-1351.

75. Harmon, L. J. (2019). Phylogenetic comparative methods. CreateSpace Independent Publishing Platform. Retrieved from https://lukejharmon.github.io/pcm/

76. Harmon, L. J., Kolbe, J. J., Cheverud, J. M., & Losos, J. B. (2005). Convergence and the multidimensional niche. *Evolution*, *59*(2), 409-421.

77. Harmon, L. J., Schulte, J. A., Larson, A., & Losos, J. B. (2003). Tempo and mode of evolutionary radiation in iguanian lizards. *Science, 301*(5635), 961-964.

78. Harvey, M. G., & Rabosky, D. L. (2018). Continuous traits and speciation rates: Alternatives to state-dependent diversification models. *Methods in Ecology and Evolution*, *9*(4), 984-993.

79. Heath, T. A., Huelsenbeck, J. P., & Stadler, T. (2014). The fossilized birth-death process for coherent calibration of divergence-time estimates. Proceedings of the National Academy of Sciences, 111(29), E2957-E2966.

80. Heck, L., Wilson, L. A., Evin, A., Stange, M., & Sánchez-Villagra, M. R. (2018). Shape variation and modularity of skull and teeth in domesticated horses and wild equids. *Frontiers in zoology*, *15*(1), 14.

81. Herculano-Houzel, S., Catania, K., Manger, P. R., & Kaas, J. H. (2015). Mammalian brains are made of these: a dataset of the numbers and densities of neuronal and nonneuronal cells in the brain of glires, primates, scandentia, eulipotyphlans, afrotherians and artiodactyls, and their relationship with body mass. *Brain, Behavior and Evolution*, *86*(3-4), 145-163.

82. Hildebrand, M., Bramble, D. M., Liem, K. F., & Wake, D. B. (1985). Functional Vertebrate Morphology. Cambridge, MA: Belknap Press of Harvard University Press.

83. Holekamp, K. E., Dantzer, B., Stricker, G., Yoshida, K. C. S., & Benson-Amram, S. (2015). Brains, brawn and sociality: a hyaena's tale. *Animal behaviour*, *103*, 237-248.

84. Holloway, R. L. (2018). On the making of endocasts: The new and the old in paleoneurology. In *Digital endocasts* (pp. 1-8). Springer, Tokyo.

85. Hunt, G., & Slater, G. (2016). Integrating Paleontological and Phylogenetic Approaches to Macroevolution. *Annual Review of Ecology, Evolution, and Systematics, 47*(1), 189–213.

86. Isaac, N. J., Jones, K. E., Gittleman, J. L., & Purvis, A. (2005). Correlates of species richness in mammals: body size, life history, and ecology. *The American Naturalist*, *165*(5), 600-607.

87. Isler, K., & Van Schaik, C. P. (2006). Metabolic costs of brain size evolution. *Biology letters*, *2*(4), 557-560.

88. Israfil, H., Zehr, S. M., Mootnick, A. R., Ruvolo, M., & Steiper, M. E. (2011). Unresolved molecular phylogenies of gibbons and siamangs (Family: Hylobatidae) based on

mitochondrial, Y-linked, and X-linked loci indicate a rapid Miocene radiation or sudden vicariance event. *Molecular Phylogenetics and Evolution*, *58*(3), 447-455.

89. Jablonski, D. (2001). Lessons from the past: evolutionary impacts of mass extinctions. *Proceedings of the Royal Society B: Biological Sciences,* 98: 5393–5398.

90. Jablonski, D. (2008). Species selection: theory and data. *Annual Review of Ecology, Evolution, and Systematics, 39*, 501–524.

91. Jablonski, D., & Roy, K. (2003). Geographical range and speciation in fossil and living molluscs. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 270*(1513), 401-406.

92. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning.* New York, NY: Springer Science & Business Media.

93. Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. Ø. (2012). The global diversity of birds in space and time. *Nature, 491*, 444–448.

94. Johnston, C., & Berta, A. (2011). Comparative anatomy and evolutionary history of suction feeding in cetaceans. *Marine Mammal Science*, *27*(3), 493-513.

95. Jones, K. E., Angielczyk, K. D., & Pierce, S. E. (2019). Stepwise shifts underlie evolutionary trends in morphological complexity of the mammalian vertebral column. *Nature communications, 10*, 5071.

96. Jønsson, K. A., Fabre, P. H., Fritz, S. A., Etienne, R. S., Ricklefs, R. E., Jørgensen, T. B., ... & Irestedt, M. (2012). Ecological and evolutionary determinants for the adaptive radiation of the Madagascan vangas. *Proceedings of the National Academy of Sciences, 109*(17), 6620-6625.

97. Köhler, M., & Moyà-Solà, S. (2004). Reduction of brain and sense organs in the fossil insular bovid Myotragus. *Brain, Behavior and Evolution*, *63*(3), 125-140.

98. Kratsch, C., & McHardy, A. C. (2014). RidgeRace: ridge regression for continuous ancestral character estimation on phylogenetic trees. *Bioinformatics*, *30*(17), i527-i533.

99. Lambert, O., Martínez-Cáceres, M., Bianucci, G., Di Celma, C., Salas-Gismondi, R., Steurbaut, E., ... & de Muizon, C. (2017). Earliest mysticete from the Late Eocene of Peru sheds new light on the origin of baleen whales. *Current Biology*, *27*(10), 1535-1541.

100. Lanfear, R., Ho, S. Y., Love, D., & Bromham, L. (2010). Mutation rate is linked to diversification in birds. *Proceedings of the National Academy of Sciences, 107*(47), 20423-20428.

101. Lee, M. S., Cau, A., Naish, D., & Dyke, G. J. (2014a). Morphological clocks in paleontology, and a mid-Cretaceous origin of crown Aves. *Systematic Biology*, *63*(3), 442-449.

102. Lee, M. S., Cau, A., Naish, D., & Dyke, G. J. (2014b). Sustained miniaturization and anatomical innovation in the dinosaurian ancestors of birds. *Science*, *345*(6196), 562-566.

103. Lefebvre, L. (2013). Brains, innovations, tools and cultural transmission in birds, non-human primates, and fossil hominins. *Frontiers in human neuroscience*, *7*, 245.

104. Lenth, R. (2019). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.4.3.01. https://CRAN.R-project.org/package=emmeans.

105. Lester, S. E., Ruttenberg, B. I., Gaines, S. D., & Kinlan, B. P. (2007). The relationship between dispersal ability and geographic range size. *Ecology Letters*, *10*(8), 745-758.

106. Liow, L. H. (2004). A test of Simpson's "rule of the survival of the relatively unspecialized" using fossil crinoids. *The American Naturalist, 164*(4), 431-443.

107. Liow, L. H., & Finarelli, J. A. (2014). A dynamic global equilibrium in carnivoran diversification over 20 million years. *Proceedings of the Royal Society of London B, 281*(1778), 20132312.

108. Liow, L. H., Fortelius, M., Bingham, E., Lintulaakso, K., Mannila, H., Flynn, L., & Stenseth, N. C. (2008). Higher origination and extinction rates in larger mammals. *Proceedings of the National Academy of Sciences*, *105*(16), 6097-6102.

109. Lister, A. M. (1989). Rapid dwarfing of red deer on Jersey in the last interglacial. *Nature*, *342*(6249), 539-542.

110. Losos, J. B. (1992). The evolution of convergent structure in Caribbean Anolis communities. *Systematic biology*, *41*(4), 403-420.

111. Losos, J. B. (2011). Convergence, adaptation, and constraint. *Evolution*, *65*(7), 1827-1840.

112. Louca, S., & Pennell, M. W. (2020). Extant timetrees are consistent with a myriad of diversification histories. *Nature*, *580*(7804), 502-505.

113. Lusseau, D. (2006). Why do dolphins jump? Interpreting the behavioural repertoire of bottlenose dolphins (Tursiops sp.) in Doubtful Sound, New Zealand. *Behavioural processes*, *73*(3), 257-265.

114. Mahler, D. L., Ingram, T., Revell, L. J., & Losos, J. B. (2013). Exceptional convergence on the macroevolutionary landscape in island lizard radiations. *Science*, *341*(6143), 292-295.

115. Maidment, S. C., Linton, D. H., Upchurch, P., & Barrett, P. M. (2012). Limb-bone scaling indicates diverse stance and gait in quadrupedal ornithischian dinosaurs. *PLOS one*, *7*(5), e36904.

116. Maliet, O., Hartig, F., & Morlon, H. (2019). A model with many small shifts for estimating species-specific diversification rates. *Nature ecology & evolution, 3*(7), 1086-1092.

117. Marino, L., Connor, R. C., Fordyce, R. E., Herman, L. M., Hof, P. R., Lefebvre, L., ... & Whitehead, H. (2007). Cetaceans have complex brains for complex cognition. *PLoS Biology*, *5*(5), e139.

118. Marroig, G., & Cheverud, J. M. (2005). Size as a line of least evolutionary resistance: diet and adaptive morphological radiation in New World monkeys. *Evolution*, *59*(5), 1128-1142.

119. Martin, J. S., Ringen, E. J., Duda, P., & Jaeggi, A. V. (2020). Harsh environments promote alloparental care across human societies. *Proceedings of the Royal Society B*, *287*(1933), 20200758.

120. Marx, F. G., & Fordyce, R. E. (2015). Baleen boom and bust: a synthesis of mysticete phylogeny, diversity and disparity. *Royal Society Open Science*, *2*(4), 140434.

121. Matschiner, M. (2019). Selective sampling of species and fossils influences age estimates

under the fossilized birth–death model. *Frontiers in genetics, 10*, 1064.

122. McCurry, M. R., Fitzgerald, E. M., Evans, A. R., Adams, J. W., & McHenry, C. R. (2017). Skull shape reflects prey size niche in toothed whales. *Biological Journal of the Linnean Society*, *121*(4), 936-946.
123. McFarlane, D. A., MacPhee, R. D., & Ford, D. C. (1998). Body Size Variability and a Sangamonian Extinction Model for Amblyrhiza, a West Indian Megafaunal Rodent. *Quaternary Research*, *50*(1), 80-89.
124. Meachen-Samuels, J. A. (2012). Morphological convergence of the prey-killing arsenal of sabertooth predators. *Paleobiology*, *38*(1), 1-14.
125. Meiri, S., Raia, P., & Phillimore, A. B. (2011). Slaying dragons: limited evidence for unusual body size evolution on islands. *Journal of Biogeography*, *38*(1), 89-100.
126. Meloro, C., Cáceres, N. C., Carotenuto, F., Sponchiado, J., Melo, G. L., Passaro, F., & Raia, P. (2015). Chewing on the trees: Constraints and adaptation in the evolution of the primate mandible. *Evolution*, *69*(7), 1690-1700.
127. Meyer, A. L., Román-Palacios, C., & Wiens, J. J. (2018). BAMM gives misleading rate estimates in simulated and empirical datasets. *Evolution, 72*(10), 2257-2266.
128. Millien, V. (2011). Mammals evolve faster on smaller islands. *Evolution*, *65*(7), 1935-1944.
129. Mondanaro, A., Castiglione, S., Melchionna, M., Di Febbraro, M., Vitagliano, G., Serio, C., Vero, V.A., Carotenuto, F., & Raia, P. (2017). Living with the elephant in the room: topdown control in Eurasian large mammal diversity over the last 22 million years. *Palaeogeogrpahy, Palaeoclimatology, Palaeoecology, 485*, 956–962.
130. Montgomery, S. H., Geisler, J. H., McGowen, M. R., Fox, C., Marino, L., & Gatesy, J. (2013). The evolutionary history of cetacean brain and body size. *Evolution, 67(11)*, 3339–3353.

131. Montgomery, S. H., Mundy, N. I., & Barton, R. A. (2016). Brain evolution and development: adaptation, allometry and constraint. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1838), 20160433.

132. Mortensen, H. S., Pakkenberg, B., Dam, M., Dietz, R., Sonne, C., Mikkelsen, B., & Eriksen, N. (2014). Quantitative relationships in delphinid neocortex. *Frontiers in neuroanatomy*, *8*, 132.

133. Muller, A. S., & Montgomery, S. H. (2019). Co-evolution of cerebral and cerebellar expansion in cetaceans. *Journal of Evolutionary Biology*, published online.

134. Neaux, D., Sansalone, G., Ledogar, J. A., Ledogar, S. H., Luk, T. H., & Wroe, S. (2018). Basicranium and face: assessing the impact of morphological integration on primate evolution. *Journal of human evolution*, *118*, 43-55.

135. Nee, S. (2006). Birth-death models in macroevolution. *Annual Review of Ecology, Evolution, and Systematics, 37*, 1-17.

136. Neubauer, S., & Hublin, J. J. (2012). The evolution of human brain development. *Evolutionary Biology*, *39*(4), 568-586.

137. Neubauer, S., Gunz, P., Scott, N. A., Hublin, J. J., & Mitteroecker, P. (2020). Evolution of brain lateralization: A shared hominid pattern of endocranial asymmetry is much more variable in humans than in great apes. *Science advances*, *6*(7), eaax9935.

138. Neubauer, S., Hublin, J. J., & Gunz, P. (2018). The evolution of modern human brain shape. *Science advances*, *4*(1), eaao5961.

139. Nowak, R. M., & Walker, E. P. (1999). *Walker's Mammals of the World (Vol. 1)*. Baltimore, Maryland: JHU press.

140. O'Meara, B. C. (2012). Evolutionary inferences from phylogenies: a review of methods. *Annual Review of Ecology, Evolution, and Systematics*, *43*, 267-285.

141. O'Meara, B. C., Ané, C., Sanderson, M. J., & Wainwright, P. C. (2006). Testing for different rates of continuous trait evolution using likelihood. *Evolution, 60*(5), 922-933.

142. Otto, S. P. (2018). Adaptation, speciation and extinction in the Anthropocene. *Proceedings of the Royal Society B*, *285*(1891), 20182047.

143. Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature, 401*(6756),877.

144. Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526-528.

145. Park, T., Fitzgerald, E. M., & Evans, A. R. (2016). Ultrasonic hearing and echolocation in the earliest toothed whales. *Biology Letters*, *12*(4), 20160060.

146. Pearson, P. N. (1992). Survivorship analysis of fossil taxa when real-time extinction rates vary: the Paleogene planktonic foraminifera. *Paleobiology*, 115-131.

147. Pélabon, C., Firmat, C. J. P., Bolstad, G. H., Voje, K. L., Houle, D., Cassara, J., ... & Hansen, T. F. (2014). Evolution of morphological allometry. *Annals of the New York Academy of Sciences, 1320*, 58–75.

148. Pennell, M. W., Harmon, L. J., & Uyeda, J. C. (2014). Is there room for punctuated equilibrium in macroevolution?. *Trends in ecology & evolution, 29*(1), 23-32.

149. Pincheira-Donoso, D., Harvey, L. P., & Ruta, M. (2015). What defines an adaptive radiation? Macroevolutionary diversification dynamics of an exceptionally species-rich continental lizard radiation. *BMC Evolutionary Biology, 15*(1), 1-13.

150. Piras, P., Sansalone, G., Teresi, L., Moscato, M., Profico, A., Eng, R., ... & Kotsakis, T. (2015). Digging adaptation in insectivorous subterranean eutherians. The enigma of Mesoscalops montanensis unveiled by geometric morphometrics and finite element analysis. *Journal of Morphology*, *276*(10), 1157-1171.

151. Pradel, R. (1996). Utilization of capture-mark-recapture for the study of recruitment and population growth rate. *Biometrics, 52*, 703-709.

152. Price Waldman, R. M., Shultz, A. J., & Burns, K. J. (2020). Speciation rates are correlated with changes in plumage color complexity in the largest family of songbirds. *Evolution*, 25, 471–15.

153. Profico, A., Di Vincenzo, F., Gagliardi, L., Piperno, M., & Manzi, G. (2016). Filling the gap. Human cranial remains from Gombore II (Melka Kunture, Ethiopia; ca. 850 ka) and the origin of Homo heidelbergensis. *Journal of Anthropological Sciences*, *94*, 1-24.

154. Profico, A., Piras, P., Buzi, C., Di Vincenzo, F., Lattarini, F., Melchionna, M., ... & Manzi, G. (2017). The evolution of cranial base and face in Cercopithecoidea and Hominoidea: Modularity and morphological integration. *American journal of primatology*, *79*(12), e22721.

155. Prothero, D. (2012). Cenozoic mammals and climate change: the contrast between coarse-scale versus high-resolution studies explained by species sorting. *Geosciences*, *2*(2), 25-41.

156. Puttick, M. N., & Thomas, G. H. (2015). Fossils and living taxa agree on patterns of body mass evolution: a case study with Afrotheria. *Proceedings of the Royal Society B: Biological Sciences, 282*(1821), 20152023–20152029.

157. Quental, T. B., & Marshall, C. R. (2013). How the Red Queen drives terrestrial mammals to extinction. *Science, 341*, 290–292.

158. Rabosky D., & Slater G. (2014). Macroevolutionary Rates. *Oxford Bibliographies* in Evolutionary Biology.

159. Rabosky, D. L. (2010). Extinction rates should not be estimated from molecular phylogenies. *Evolution, 64*(6), 1816-1824.

160. Rabosky, D. L., & Adams, D. C. (2012). Rates of morphological evolution are correlated with species richness in salamanders. *Evolution, 66*(6), 1807-1818.

161. Rabosky, D. L., & Hurlbert, A. H. (2015). Species richness at continental scales is dominated by ecological limits. *The American Naturalist, 185*(5), 572-583.

162. Rabosky, D. L., & Lovette, I. J. (2008). Density-dependent diversification in North American wood warblers. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1649), 2363-2371.

163. Rabosky, D. L., Santini, F., Eastman, J., Smith, S. A., Sidlauskas, B., Chang, J., & Alfaro, M. E. (2013). Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature communications, 4*(1), 1-8.

164. Rabosky, D.L. (2014) Automatic Detection of Key Innovations, Rate Shifts, and Diversity-Dependence on Phylogenetic Trees. *PLoS ONE, 9*(2), e89543.

165. Raia, P. (2016). Macroevolution. *Oxford Bibliographies* in Evolutionary Biology.

166. Raia, P., & Meiri, S. (2011). The tempo and mode of evolution: body sizes of island mammals. *Evolution*, *65*(7), 1927-1934.

167. Raia, P., Carotenuto, F., & Meiri, S. (2010a). One size does not fit all: no evidence for an optimal body size on islands. *Global Ecology and Biogeography*, *19*(4), 475-484.

168. Raia, P., Carotenuto, F., Eronen, J. T., & Fortelius, M. (2011). Longer in the tooth, shorter in the record? The evolutionary correlates of hypsodonty in Neogene ruminants. *Proceedings of the Royal Society B: Biological Sciences, 278*(1724), 3474-3481.

169. Raia, P., Carotenuto, F., Meloro, C., Piras, P., & Pushkina, D. (2010b). The shape of contention: adaptation, history, and contingency in ungulate mandibles. *Evolution*, *64*(5), 1489-1503.

170. Raia, P., Carotenuto, F., Passaro, F., Fulgione, D., & Fortelius, M. (2012). Ecological specialization in fossil mammals explains Cope's rule. *The American Naturalist*, *179*(3), 328-337.

171. Raia, P., Carotenuto, F., Passaro, F., Piras, P., Fulgione, D., Werdelin, L., ... & Fortelius, M. (2013). Rapid action in the Palaeogene, the relationship between phenotypic and taxonomic diversification in Coenozoic mammals. *Proceedings of the Royal Society B: Biological Sciences, 280*(1750), 20122244.

172. Raia, P., Carotenuto, F., Mondanaro, A., Castiglione, S., Passaro, F., Saggese, F., ... & Fortelius, M. (2016). Progress to extinction: increased specialisation causes the demise of animal clades. *Scientific reports, 6*(1), 1-10.

173. Randau, M., Carbone, C., & Turvey, S.T. (2013). Canine evolution in sabretoothed carnivores: natural selection or sexual selection? *PLoS One, 8*, e72868.

174. Redding, D. W., & Mooers A.O. (2006). Incorporating evolutionary measures into conservation prioritization. *Conservation Biology, 20,* 1670-1678.

175. Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in ecology and evolution*, *3*(2), 217-223.

176. Revell, L. J., Johnson, M. A., Schulte, J. A., Kolbe, J. J., & Losos, J. B. (2007). A phylogenetic test for adaptive convergence in rock-dwelling lizards. *Evolution*, *61*(12), 2898-2912.

177. Rizal, Y., Westaway, K. E., Zaim, Y., van den Bergh, G. D., Bettis, E. A., Morwood, M. J., ... & Ciochon, R. L. (2020). Last appearance of Homo erectus at Ngandong, Java, 117,000–108,000 years ago. *Nature*, *577*(7790), 381-385.

178. Roberts, P., & Stewart, B. A. (2018). Defining the 'generalist specialist'niche for Pleistocene Homo sapiens. *Nature Human Behaviour*, *2*(8), 542-550.

179. Rundell, R.J. & Price, T.D. (2009). Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology Evolution, 24*, 394–399.

180. Sander, P. M., Christian, A., Clauss, M., Fechner, R., Gee, C. T., Griebeler, E. M., ... & Witzel, U. (2011). Biology of the sauropod dinosaurs: the evolution of gigantism. *Biological Reviews*, *86*(1), 117-155.

181. Sansalone, G., Colangelo, P., Loy, A., Raia, P., Wroe, S., & Piras, P. (2019). Impact of transition to a subterranean lifestyle on morphological disparity and integration in talpid moles (Mammalia, Talpidae). *BMC evolutionary biology*, *19*(1), 179.

182. Schluter, D. (2001). Ecology and the origin of species. *Trends in Ecology and Evolution, 16*, 372–380.

183. Schnitzler, J., Theis, C., Polly, P. D., & Eronen, J. T. (2017). Fossils matter–understanding modes and rates of trait evolution in Musteloidea (Carnivora). *Evolutionary Ecology Research*, *18*(2), 187-200.

184. Sepkoski, J. J. (1984). A kinetic model of Phanerozoic taxonomic diversity. III. Post-Paleozoic families and mass extinctions. *Paleobiology, 10*, 246-267.

185. Shultz, S., & Dunbar, R. (2010). Encephalization is not a universal macroevolutionary phenomenon in mammals but is associated with sociality. *Proceedings of the National Academy of Sciences*, *107*(50), 21582-21586.

186. Shultz, S., & Dunbar, R. I. M. (2006). Both social and ecological factors predict ungulate brain size. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1583), 207-215.

187. Silvestro, D., Salamin, N., & Schnitzler, J. (2014). PyRate: a new program to estimate speciation and extinction rates from incomplete fossil data. *Methods in Ecology and Evolution, 5*(10), 1126-1131.

188. Silvestro, D., Warnock, R. C., Gavryushkina, A., & Stadler, T. (2018). Closing the gap between palaeontological and neontological speciation and extinction rate estimates. *Nature Communications*, *9*(1), 1-14.

189. Skinner, M. M., Stephens, N. B., Tsegai, Z. J., Foote, A. C., Nguyen, N. H., Gross, T., ... & Kivell, T. L. (2015). Human-like hand use in Australopithecus africanus. *Science*, *347*(6220), 395-399.

190. Skinner, M. M., Stephens, N. B., Tsegai, Z. J., Foote, A. C., Nguyen, N. H., Gross, T., ... & Kivell, T. L. (2015). Human-like hand use in Australopithecus africanus. *Science*, *347*(6220), 395-399.

191. Slater, G. J., Harmon, L. J., & Alfaro, M. E. (2012). Integrating fossils with molecular phylogenies improves inference of trait evolution. *Evolution*, *66*(12), 3931-3944.

192. Slater, G. J., Price, S. A., Santini, F., & Alfaro, M. E. (2010). Diversity versus disparity and the radiation of modern cetaceans. *Proceedings of the Royal Society B: Biological Sciences, 277*(1697), 3097-3104.

193. Slatyer, R. A., Hirst, M., & Sexton, J. P. (2013). Niche breadth predicts geographical range size: a general ecological pattern. *Ecology letters, 16*(8), 1104-1114.

194. Smaers, J. B., & Soligo, C. (2013). Brain reorganization, not relative brain size, primarily characterizes anthropoid brain evolution. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1759), 20130269.

195. Smaers, J. B., & Vanier, D. R. (2019). Brain size expansion in primates and humans is explained by a selective modular expansion of the cortico-cerebellar system. *Cortex*, *118*, 292-305.

196. Sol, D., Duncan, R. P., Blackburn, T. M., Cassey, P., & Lefebvre, L. (2005). Big brains, enhanced cognition, and response of birds to novel environments. *Proceedings of the National Academy of Sciences*, *102*(15), 5460-5465.

197. Soligo, C., & Martin, R. D. (2006). Adaptive origins of primates revisited. *Journal of Human Evolution*, *50*(4), 414-430.

198. Spikins, P., Needham, A., Wright, B., Dytham, C., Gatta, M., & Hitchens, G. (2019). Living to fight another day: The ecological and evolutionary significance of Neanderthal healthcare. *Quaternary Science Reviews*, *217*, 98-118.

199. Sponheimer, M., Alemseged, Z., Cerling, T. E., Grine, F. E., Kimbel, W. H., Leakey, M. G., ... & Wynn, J. G. (2013). Isotopic evidence of early hominin diets. *Proceedings of the National Academy of Sciences*, *110*(26), 10513-10518.

200. Springer, M. S., Meredith, R. W., Gatesy, J., Emerling, C. A., Park, J., Rabosky, D. L., ... & Murphy, W. J. (2012). Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. *PloS one*, *7*(11), e49521.

201. Stadler, T. (2010). Sampling-through-time in birth-death tree. *Journal of Theoretical Biology,*

267 (3), pp. 396-40

202. Stayton, CT. (2018). convevol: Analysis of Convergent Evolution. R package version 1.3. https://CRAN.R-project.org/package=convevol

203. Stout, D., & Khreisheh, N. (2015). Skill learning and human brain evolution: An experimental approach. *Cambridge Archaeological Journal*, *25*(4), 867.

204. Strickson, E., Prieto-Márquez, A., Benton, M. J., & Stubbs, T. L. (2016). Dynamics of dental evolution in ornithopod dinosaurs. *Scientific Reports*, *6*(1), 1-11.

205. Thomas, G. H., & Freckleton, R. P. (2012). MOTMOT: models of trait macroevolution on trees. *Methods in Ecology and Evolution, 3*(1), 145-151.

206. Thomas, G. H., Meiri, S., & Phillimore, A. B. (2009). Body size diversification in Anolis: novel environment and island effects. *Evolution*, *63*(8), 2017-2030.

207. Title, P. O., & Rabosky, D. L. (2019). Tip rates, phylogenies and diversification: What are we estimating, and how good are the estimates?. *Methods in Ecology and Evolution, 10*(6), 821-834.

208. Uhen, M. D. (2004). Form, function, and anatomy of Dorudon atrox (Mammalia, Cetacea): An Archaeocete from the Middle to Late Eocene of Egypt. *Papers on Paleontology, 34*, 1–222.

209. Uomini, N., & Meyer, G. (2013). Shared cerebral blood flow lateralization patterns in language and stone tool production. *PLoS ONE*, *8*(8), e72693.

210. Van Valen, L. (1973). A new evolutionary law. *Evolutionary Theory*, *1*, 1-30.

211. Van Valkenburgh, B. (2007). Déjà vu: the evolution of feeding morphologies in the Carnivora. *Integrative and Comparative Biology, 47*(1), 147-163.

212. Villalobos, F., Olalla-Tarraga, M. A., Vieira, C. M., Mazzei, N. D., & Bini, L. M. (2017). Spatial dimension of body size evolution in Pterosauria: Bergmann's rule does not drive Cope's rule. *Evolutionary Ecology Research*, *18*(2), 169-186.

213. Wagner, P. J., & Erwin, D. H. (1995). “Phylogenetic patterns as tests of speciation models” in *New Approaches to Speciation in the Fossil Record*, eds D. Erwin and R. L. Anstey (New York, NY: Columbia University Press).

214. Werdelin, L., Yamaguchi, N., Johnson, W.E., & O'Brien, S.J., (2010). “Phylogeny and evolution of cats” in *Biology and Conservation of Wild Felids*, eds. D.W. MacDonald and A. J. Loveridge (Oxford University Press, New York).

215. Whiten, A. (2001). Imitation and cultural transmission in apes and cetaceans. *Behavioral and Brain Sciences*, *24*(2), 359.

216. Wiltshire, J., Huffer, F. W., & Parker, W. C. (2014). A general class of test statistics for Van Valen's Red Queen hypothesis. *Journal of applied statistics*, *41*(9), 2028-2043.

217. Wroe, S., Parr, W. C., Ledogar, J. A., Bourke, J., Evans, S. P., Fiorenza, L., ... & Yokley, T. R. (2018). Computer simulations show that Neanderthal facial morphology represents adaptation to cold and high energy demands, but not heavy biting. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1876), 20180085.

218. Xiang, L., Crow, T., & Roberts, N. (2019). Cerebral torque is human specific and unrelated to brain size. *Brain Structure and Function*, *224*(3), 1141-1150.

219. Yule, G. U. (1925). A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FRS. *Philosophical transactions of the Royal Society of London. Series B*, *213*, 21-87.

220. Zachos, J., Pagani, M., Sloan, L., Thomas, E., & Billups, K. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. *science*, *292*(5517), 686-693.

221. Zeder, M. A. (2015). The domestication of animals. *Journal of anthropological research*, *68*(2), 161-190.

222. Zelditch, M.L., Swiderski, D.L., & Sheets, H.D. (2012). Geometric Morphometrics for Biologists. Academic Press.

223. Zink, K. D., & Lieberman, D. E. (2016). Impact of meat and Lower Palaeolithic food processing techniques on chewing in humans. *Nature*, *531*(7595), 500-503.