Dental Characters Used in Phylogenetic Analyses of Mammals Show Higher Rates of Evolution, but Not Reduced Independence

Neil Brocklehurst[1], Gemma Louise Benevento[2]

[1] *Department of Earth Sciences, University of Oxford, South Parks Road, Oxford, UK, OX1 3QR*

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

ABSTRACT

Accurate reconstructions of phylogeny are essential for studying the evolution of a clade. Morphological characters are necessarily used for the reconstruction of the relationships of fossil organism ~~relationships.~~ ~~H~~however variation in their evolutionary modes (for example rate variation and character non-independence) not accounted for in analyses may be leading to unreliable phylogenies. A recent study suggested that phylogenetic analyses of mammals may be suffering from a dominance of dental characters, which were shown to have lower phylogenetic signal than osteological characters and produced phylogenies less congruent with molecularly-derived benchmarks. Here we build on this previous work by testing ~~seven~~ five additional morphological partitions for phylogenetic signal and examining what aspects of dental and other character evolution may be affecting this, by fitting models of discrete character evolution to phylogenies inferred and time calibrated using molecular data. Results indicate that the phylogenetic signal of discrete characters correlate most strongly with rates of evolution, with increased rates driving increased homoplasy. ~~I~~in the total mammal dataset, ~~Dental~~ dental characters have higher rates of evolution than other partitions. They do not, however, fit a model of independent character evolution any worse than other regions. Primates and ~~Marsupials~~ marsupials show different patterns to other mammal clades, with dental characters evolving at slower rates and being more heavily integrated (less independent). While the dominance of dental characters in analysis of mammals could be leading to inaccurate phylogenies, the issue is not unique to dental characters, and the results are not consistent across datasets. Molecular benchmarks (being entirely independent of the character data) provide a framework for examining each dataset individually to assess the evolution of the characters used.

Keywords: Evolutionary Rates; Homoplasy; Independence; Phylogeny; Mammalia

Commented [R1]: What exactly is the "total mammal dataset"? Is this the Bi et al. matix?

INTRODUCTION

Accurate reconstructions of phylogenetic relationships are essential for studying the evolutionary history of a clade, with hypotheses being based on molecular or morphological data, or both. While it is comparatively straightforward to observe patterns of evolution in molecular sequence data and therefore develop models more closely representing the evolutionary processes, this is more difficult in the case of morphological characteristics due to a poorer understanding of how novel morphology is evolved from ancestral traits. Nonetheless, morphological data ~~is~~ are our only means of reconstructing the phylogenetic relationships of fossil organisms that are too old to preserve DNA or usable proteins. It is therefore imperative that we strive to better understand the evolutionary modes of morphological traits. In recent years many studies have examined how variation in ~~their~~ evolutionary patterns of discrete morphological traits, not accounted for by current analyses, may be affecting phylogenetic inferences (e.g. O'Keefe & Wagner, 2001; Scotland et al., 2003; DeGusta, 2004~~;~~, Sansom et al. 2017; Billet & Bardin, 2018).

The high percentage of dental characters used in the reconstruction of fossil mammal phylogenies has become a particular cause for concern. Numerous studies have highlighted issues such as the non-independent evolution of dental characters (Kangas et al., 2004; Kavanagh et al., 2007; Harjunmaa et al., 2014; Dávalos et al., 2014; Billet & Bardin, 2018) and increased convergence relative to other character partitions due to ecological selective pressures (Evans et al., 2007; Kavanagh et al., 2007). In a recent meta-analysis, Sansom et al. (2017) examined the phylogenetic signal of tooth and osteological character partitions, using phylogenies derived from molecular data as a benchmark. This study found that osteological characters were more consistent with the ~~molecular~~molecularly-derived phylogenies and

contained greater phylogenetic signal than dental characters. Further, ~~, while~~ parsimony analyses with only dental characters produced results less similar to the molecularly phylogenies than analyses where the same number of characters were selected at random from both partitions (Sansom et al. 2017).

This paper builds on the work of Sansom et al. (2017) in two ~~principle~~principal ways. Sansom et al. (2017) employed two partitions, dental and osteological, to assess ~~whether~~ the performance of dental characters ~~performed more poorly than~~relative to osteological characters in phylogenetic analyses. As such, while dental characters have been demonstrated to ~~potentially~~ be problematic, an understanding of whether this problem was limited to them, or whether it ~~is seen~~extends to ~~in~~ other partitions, is lacking. We therefore examine phylogenetic signal in ~~eight~~ six morphological partitions in mammals in order to establish whether any other skeletal regions may be a poor indicator of phylogeny.

Secondly, we also aim to understand why dental characters may be producing phylogenies less congruent with molecularly-derived benchmarks. ~~It is becoming more well~~Many studies have established that morphological characters frequently violate at least some of the ~~principle~~principal assumptions of parsimony (see below): between-character rate homogeneity (all characters being just as likely to transition), within-character rate homogeneity (all character states within the same character being similarly likely to transition than others), and character independence ~~(see below)~~. We test each morphological character partition for variation in ~~rates of state transition within characters, variation in rates of evolution between characters, and character independence.~~these parameters.

In most published phylogenetic analyses performed using parsimony, the characters are weighted equally (Käl~~l~~ersjö et al., 1999; Kluge, 2005; Goloboff et al., 2008). Under such a scheme, a change in any character is given equal ~~emphasis~~ influence in determining tree length. However, such a scheme only produces ~~reliable~~ robustly supported results when the

characters are all equally likely to change. If, however, there is variation in the rates of character evolution, certain characters will change more frequently and are more likely to show homoplasy (Felsenstein, 1981; Goloboff, 1993). While parsimony analysis does not incorporate an explicit evolutionary model, an equal weights analysis does rely on equal between-character rates for its accuracy.

Furthermore, in most published phylogenetic analyses, transitions between different combinations of character states are given equal weight (i.e. a transition from state 0 to state 1 is just as likely as a transition from state 1 to state 0; an assumption of within-character rate homogeneity). This assumption may be relaxed by incorporating step matrices which give greater weight to particular transitions (Sankoff & ~~Kedergren~~Cedergren, 1983), or by ordering (Fitch, 1971), an extreme modification of step matrices, setting the possibility of ~~most~~ non-adjacent transitions to 0. However, such modifications are rarely employed (see Marjanović & Laurin, 2019 for summary of their history) and most ~~analysis~~ analyses assume equality of within-character rates.

Finally, all methods of phylogenetic analysi~~e~~s (parsimony, Bayesian, and likelihood ~~methods~~), treat all characters as independent of one another (i.e. an assumption that a change in one character will have no effect on the transition probability in another character~~an assumption that a change in one character will have no effect on another character transition probability~~). Extensive study has shown this assumption of independence to be frequently violated (e.g. Kangas et al., 2004; Kavanagh et al., 2007; Harjunma~~a~~ et al., 2014; Dávalos et al., 2014; Billet and Bardin, 2018), with many traits or regions forming integrated modules that change as a unit (Goswami, 2006, 2007; Goswami & Polly, 2010).

By analysing phylogenetic signal, between- and within-character rates, and character independence across ~~eight~~ six morphological partitions, within mammals as a whole and four mammalian subclades, we aim to better understand how morphological characters can be

Commented [R2]: I would agree that step-matrices are rarely employed, because they're quite fiddly to set up (and current versions of MrBayes can't implement step matrices). However, character ordering is regularly employed, and there are methodological/philosophical justifications for doing so – see especially Wiens (2001 – Systematic Biology) "Character Analysis in Morphological Phylogenetics: Problems and Solutions"

selected and formulated during phylogenetic analyses of mammals~~we aim to better understand how the use of mammalian morphological characters can be optimised during phylogenetic analyses~~. The results should provide future studies that intend to reconstruct the relationships of fossil mammals with a framework to enable more evidence-based decisions about which characters are more reliable for use in phylogenetic analyses.

MATERIALS AND METHODS

Data

This study builds on the protocol established by Sansom et al. (2017), where molecularly-derived phylogenies are used as the framework over which morphological evolution may be analysed. This allows the evolutionary patterns of the characters to be examined over a phylogeny produced and time calibrated from data entirely independent of those characters. For mammals, the time-scaled ~~molecular~~molecularly-derived phylogeny was taken from ~~Meredith~~ Dos Reis et al. (~~2011~~2012), and the morphological data from Bi et al. (2014), both recent and comprehensive datasets. Although the Bi et al. (2014) matrix was focussed on Mesozoic mammals, it contains a broad sampling of modern clades, including from ~~the~~ monotremes, marsupials and placentals. The morphological characters were divided between ~~eight~~ six partitions: dental, cranial, axial, ~~pectoral girdle, pelvic girdle,~~ forelimb (including pectoral girdle), hindlimb (including pelvic gridle), and soft tissue. Taxa not present in both the morphological matrix and ~~molecular~~molecularly-derived tree were dropped. If, after doing so, a character showed no variation in score among the remaining taxa, that character was also dropped from subsequent analyses.

As well as the global analysis of mammals, ~~three~~ four subclades were subjected to the same analyses to test for variation in the macroevolutionary patterns within Mammalia. The clades chosen were as follows: ~~Cetartiodactyla~~ Artiodactyla (~~Molecular~~Molecularly-derived

tree from Hassanin et al. [2012], Morphological matrix from ~~O'Leary & Gatesy [2008]~~Spaulding et al. [2009]), Carnivora (~~(Molecular~~Molecularly-derived tree from Eizirk et al. [2010], Morphological matrix from Tomiya [2011]~~Pattinson et al. [2015]), (),~~ Primates (~~Molecular~~Molecularly-derived tree from Perelman et al. [2011], Morphological matrix from ~~Pattinson et al~~Ni et al. [~~2015~~2013]) and Marsupialia (~~Molecular~~Molecularly-derived tree from Mitchell et al. [2014], Morphological matrix from Beck [2017]). These clades were chosen for the following reasons: 1) they have been analysed using morphological character matrices containing characters from all ~~eight~~ six of the morphological partitions; 2) there exist time calibrated ~~molecular~~molecularly-derived phylogenies with substantial taxonomic overlap with the morphological matri~~c~~ces; 3) the character list, data matrix and time calibrated phylogeny were available in usable formats; and 4) they are morphologically and ecologically diverse lineages, and therefore the morphological characters have the potential to be heavily influenced by functional and ecological constraints.

Phylogenetic Signal

Levels of homoplasy relative to the ~~molecular~~molecularly-derived phylogeny were used as an estimate of the phylogenetic signal of the characters, measured using Pagel's lambda (Pagel, 1999), a metric shown to perform well under simulations (Münkemüller et al., 2012). This statistic produces a value between 0 and 1, where 0 indicates that character states are distributed independent of phylogeny (no phylogenetic signal). Other methods of calculating phylogenetic signal in discrete characters, for example Moran's I (Gittleman & Kot, 1990) or Fritz & Purvis's D (Fritz & Purvis, 2010), were not used as they are only suitable for binary characters and would require a large proportion of the characters to be dropped. For each character, taxa scored as unknown were dropped from the tree. If more than a quarter of the taxa were scored as unknown, the character was not considered in this or

subsequent analyses. Pagel's lambda was calculated in R version 3.3.2 (R ~~core~~ Core ~~team~~Team, 2016) using the *fitDiscrete* function in the package Geiger (Harmon et al., 2007).

Testing the Assumptions of Phylogenetic Analysis

Within-character rate homogeneity was tested by fitting models of discrete character evolution to the observed phylogeny and trait values using the function *fitDiscrete* in the R package Geiger. This method calculates the likelihood of a particular model based on the data, and also estimates the values of variable parameters within that model that best fit~~ting~~ the observed data (Pennell and Harmon 2013, Pennell et al. 2014). Two models were compared: an equal rates (ER) model, where every possible character state transformation has the same rate, and an all-rates-different (ARD) model, where every possible character state transformation is allowed a different rate. The models are compared using the Akaike information criterion, which penalises the parameter-rich ARD model. The Akaike weights of the ER model are used as a metric to assess how well a character obeys the assumption of within-character rate homogeneity.

The *fitDiscrete* function also allows testing of between-character rate homogeneity. As mentioned above, a~~A~~s well as identifying the model of discrete character evolution that best fits the trait and phylogeny, it also identifies the rates of character-state transformation that best fits the observed data. A higher rate of change means a character is more likely to change multiple times by convergence. If a character was found to best fit the ER model in the above analysis, then the single rate of change was assigned to the character. If the ARD model was found to fit best, the rate assigned to that character was the mean of all rates assigned to each possible transformation, weighted by the number of times each transformation occurred over the phylogeny. The number of transitions was inferred by stochastically mapping the character over the phylogeny 1000 times using the *make.simmap*

function in in the R package phytools (Revell, 2012), and calculating the mean frequency of each possible transition.

To test character independence, the method of Pagel (1994) was applied to pairwise comparisons of characters. This is again a model-fitting approach, where non-independeant and independent models of character evolution are fit to pairs of traits and the observed phylogeny. Under the non-independeant model, the rate of character change in trait 1 will depend on which character state is observed in trait 2, and vice versa. Under the independent model, both characters change state independently of each other. Again, the two models may be compared via the Akaike information criterion, and the Akaike weights of the independent model may be used as a metric for how well a pair of characters obeys the assumption of independent evolution. Unfortunately, this method is only applicable to binary characters, so non-binary characters were not considered in this section of the analyses. The analysis was implemented using the function *fitPagel* in phytools.

Statistical Comparisons

Pagel's lambda values for each character partition were compared using generalised least squares (GLS), using the R package nlme (Pinheiro et al 2017). For each partition, a null model where all the phylogenetic signal of all partitions comes from the same distribution, was compared to a model where only the partition of interest had a different phylogenetic signal to the others (H1). The Akaike weights was used to infer which best fit the data. Partitions that better fit the H1 model were deemed to have significantly different phylogenetic signals than the other partitions, with the GLS coefficient used to identify whether higher or lower. The same method was also applied to the rate values, the support for the ER model, and support for the independent model of evolution.

The rate of character change for each character, and the Akaike weight for the ER model for each character, were both compared to Pagel's lambda using the Kendall's tau correlation coefficient, a non-parametric method that does not assume normality of the data. This latter test could not be applied to the Akaike weights values of the independent model of evolution because these represent pairwise comparisons of characters rather than individual characters.

RESULTS

Results from the Total Mammalian Dataset

The median phylogenetic signal calculated from the Bi et al. (2014) character matrix (the total Mammalia dataset) was 1 for all partitions (white point, Fig. 1A). This indicates that at least half of the characters in each partition are synapomorphies for a single clade. The dental characters do show a larger range and interquartile range of lambda values than ~~most of~~ the other partitions. However, the range of values observed for cranial characters is ~~even wider,~~similar ~~indicating that for Mammalia the cranium possesses the largest number of characters with reduced phylogenetic signal~~. In the GLS analysis, cranial characters are the only partition to not fit the null model best; instead they are found to have significantly lower phylogenetic signal than other partitions (Table 1).

Dental characters show no evidence of increased within-character rate heterogeneity than do the other partitions (Fig. 1B). In fact, the Akaike weights of the equal rates (ER) model are the highest of all the partitions, and in the GLS analysis the dental partition is the only one have significantly better support for the ER model than other partitions (Table 2). Dental characters also show no evidence of increased non-independence relative to other partitions (Fig. 1C). Only the ~~pectoral girdle~~forelimb partition was found to have significantly

worse support for the independent model of evolution than other partitions (Table 3). The ~~forelimb~~ hindlimb was found to have significantly better support for the independent model.

However, dental characters have the highest median rates of evolution compared to all other partitions (Fig. 1D), and the increase in rates is significant according to the GLS analysis (Table 4). ~~The pectoral girdle was found to have reduced rates of evolution relative to other partitions, albeit only a marginally significant reduction.~~No other partitions were found to have a significant difference in rate relative to the null.

Commented [R3]: null **model**?

Results from Mammalian Subclade Datasets

The ~~Cetartiodactyla~~ Artiodactyla datasets produced similar results to those of mammals overall, albeit with considerably more variation in phylogenetic signal from the vertebral, forelimb and soft tissue characters (Fig. 2). The dental characters are the only partition where the GLS analysis found phylogenetic signal to be significantly reduced relative to other partitions (Table S1). Rates of dental evolution are again significantly higher than for other partitions (Fig. 2, Table S4). There is no significant difference found between the Akaike weights support for the ER model of evolution in teeth (Table S2), nor the support for the independent model of character evolution, compared to other partitions (Table S3). The skull partition shows better support for the independent model, while the forelimb shows statistically significantly reduced independence.

The carnivoran dataset also found dental characters to have significantly lower phylogenetic signal than other partitions (Fig 3A, Table S7). In this clade the dental character partition has higher rates than all other partitions except the Forelimb (for which there is only one character) (Fig 3D)

Commented [R4]: I don't think it's appropriate now for me to make you do further analyses, but in future you might want to consider Spaulding's papers, which have included lots of postcranial characters for carnivorans and related taxa

The primate and marsupial datasets produced results conflicting with the other two subclades and mammals as a whole (Figs 4,5). The dental partitions in primates has

significantly higher phylogenetic signal and significantly lower rates of evolution than other partitions (Fig 4, Tables S9, S12). The dental partition also had significantly better support for the equal rates model of evolution than other partitions. However, primate characters suffer from being highly integrated: all partitions other than the vertebrae and soft tissue characters show a significantly low fit to the independent model of evolution (Fig. 4C, Table S11).

~~The primate dataset showed less variation in the performance of the various character partitions compared to the cetartiodatyl dataset (Fig. 3). The dental characters again show significantly lower phylogenetic signal than other partitions (Table S5). The range of Pagel's lambda values obtained for the dental characters was wider than other partitions, as was that of forelimb characters (Fig. 3A). However, there is no significant difference in their support for an ER model of evolution compared to other partitions, and their fit to the independent model of evolution is actually significantly better than other partitions. (Tables S6-S7). Rates of evolution in primate dental characters are faster than most other partitions, but the difference is not significant. The only partition to show significantly high rates of character evolution is the pectoral girdle (Table S8)~~

~~The marsupial dataset produced results conflicting with the other subclades (Fig. 4).~~ In marsupials, w~~W~~hile many of the character partitions, including dentition, show a wide range of Pagel's lambda values, the lambda values of the tooth characters are more concentrated towards higher values ~~than~~ compared to other partitions (Fig 5A. The tooth characters show no significant difference in their phylogenetic signal ~~than~~ relative to other partitions (Table ~~S9~~S13). The dental characters showed no significant difference from any other partitions in support for the ER model of evolution (Table ~~S10~~S14), and no significant difference in rates (Fig. 54D). In contrast to the other datasets, however, the marsupial dataset does support increased character non-independence of dental characters relative to other

partitions, with median Akaike weights support for the independent model of evolution lower than all other partitions ~~except the pelvic girdle~~ (Fig. 5~~4~~C; Table S15~~1~~).

Correlation Tests

In all ~~four~~ five datasets, there is a ~~significant~~ negative correlation between lambda and rate of character evolution that is significant in all except Carnivora (Table 5). The correlation between the lambda values and Akaike weights of the ER model is weaker in all four, but in some is still significant. None of the parameters tested correlated significantly with the number of characters in each partition (Table 6)

Discussion

Mammalian tooth characters have been a source of much discussion over the last two decades, due in part to their dominance of the character lists used in morphological phylogenetic analyses of mammals, itself ~~largely~~ to an extent a product of their dominance in the mammalian fossil record. Teeth have been shown to suffer from issues such as large amounts of homoplasy (Evans et al., 2007; Davalos et al., 2014) and non-independence (Kangas et al., 2004; Harjunmaa et al., 2014). While these issues clearly do impact on the utility of dental characters in phylogenetic analysis, what has received less attention is whether dental characters are in fact worse affected than other body partitions in these regards. The majority of studies cited above focus solely on teeth, but issues of homoplasy due to ecological and functional constraints might be expected to affect other character partitions (e.g. limb characters being functionally linked to locomotion). Indeed, ecological constraint and developmental linkage has been demonstrated in cranial and limb characters across various tetrapod groups, including mammals (Ruvinsky & Gibson-Brown, 2000; Young & Hallgr~~i~~mson, 2005; Sadleir & Mackovicky, 2008). The same argument could be

made for the issue of character non-independence: while this has been demonstrated to be a problem with mammal dentition, recent work on modularity and integration highlights that this issue might just as strongly impact on non-dental characters (Goswami 2006, 2007; Goswami & Polly 2010).

Our analyses suggest that increased homoplasy driven by increased rates of evolution may affect dental characters to a greater extent than other partitions. Dental characters from the total ~~mammalian~~ Mammalia dataset and the a~~cetar~~tiodactyl and carnivoran dataset~~s~~ are found to evolve at faster rates than the other character partitions, and so are more likely to transition multiple times. T~~Moreover, t~~he strong and significant inverse correlation~~s~~ between phylogenetic signal and rates of evolution in all tested datasets indicates that rate variation is likely to be the main driving force behind loss of phylogenetic signal, more so than within-character rate heterogeneity. However, this signal is not consistent across all the tested clades. In the marsupial and primate dataset~~s, for example~~, dental characters have lower rates (and higher phylogenetic signal) than most other partitions.

Moreover, while the results obtained here seem to suggest that dental characters have lower phylogenetic signal than some other characters when optimised over a molecular-based phylogeny, they are not alone in this respect. The total ~~mammal~~ Mammalia dataset indicates that cranial characters also ~~produce~~ have low phylogenetic signal. In both primates and marsupials, the ~~forelimb~~ soft tissue characters have ~~a similar range of~~ Pagel's lambda values ~~to the dental characters~~ than any other partition (Fig ~~4~~3A, 5A) and in carnivorans both limb partitions perform poorly in this respect (Fig 3A)~~, and in cetartiodactyls the same may be said for hindlimb characters (Fig 2A). One might take this as an indication that, while it is not unreasonable to expect dental characters to contain a strong ecological signal, such a signal is likely to be found in other regions. The limbs of cetartiodactyls, for example, will be heavily constrained by locomotor type and, in particular, the restrictions placed on the hindlimb by a~~

Commented [R5]: Elsewhere you refer to this as the "total mammal" dataset – either refer to it as the "total mammal dataset" OR "total Mammalia dataset", but be consistent throughout!

Commented [R6]: I assume this should be lower or worse Pagel's lambda values – there is a word missing here anyway!

~~cursorial lifestyle may be responsible for the reduced phylogenetic signal of hindlimb characters. The hindlimbs in cursorial artiodactyls, as well as in cursorial perissodactyls, have been shown to be responsible for providing the majority of the driving force for such locomotion (Merkens et al., 1993; Dutto et al., 2004; Vaughn et al., 2011). The architecture of the limbs in both clades independently reflects this, with more limited ranges of stance and planes of movement (Liem et al., 2001) and increased muscle mass relative to length (Crook et al., 2008). However, as a counter-point to the suggestion that the constraints of cursoriality are responsible for the reduced phylogenetic signal in cetartiodactyl hindlimbs, one might ask why it is only the hindlimbs that are affected in this way. The forelimbs, for example, while not as important in driving locomotion, should be constrained by the need to "catch" the weight of the animal as it lands (McGuigan & Wilson, 2003; Witte et al., 2004; Vaughn et al., 2011), and so their architecture is constrained by the need to support greater forces. A potential area of future study is to examine whether forelimbs or hindlimbs in cursorial mammals show greater ranges of morphological variability or convergence.~~

The results observed in a~~cet~~artiodactyls raise a possibility that might warrant future study: the increase in rates of dental evolution observed might be due to the dominance of herbivores in this dataset. Herbivory has been suggested to be a driver of dental disparity in mammals (~~Jernvell~~ Jernvall et al., 1996, 2000) as their~~y~~ morphology tracks a constantly changing resource (plants). Since the functional requirements of eating meat has not changed over time, carnivorous mammals show reduced dental disparity and less evolutionary change (Van Valkenburgh, 1988; Wesley-Hunt et al., 2005). In an analysis of diversification patterns across all mammals, herbivores showed significantly higher diversification rates than carnivores or omnivores (Price et al. ~~& Hopkins~~, ~~2015~~2012). While this analysis focussed on lineage diversification, the authors cited increased specialisation and niche-subdivision as a

Commented [R7]: This point needs explaining better. How exactly have plants changed more than "meat" in terms of functional requirements? I am not an expert in this, but I would assume that the functional requirements for eating a leaf would be the same regardless if it was a Palaeocene or Pliocene leaf. However… clearly the evolution of abrasive grasses does represent a major change that likely led to a major change in dental structure of mammals, with the evolution of increasingly hypsodont and in some cases hypselodont teeth. Having said that, some authors have argued that exogenous grit has been the major driver of hypsodonty in mammals (see e.g. Madden's book "Hypsodonty in Mammals"). Regardless, this point needs to be expanded upon"

potential driving force behind diversification patterns, and morphological diversification patterns should respond to these drivers in the same way.

It is finally worth noting that in the total-mammal dataset and the two of the ~~two~~ three placental subclades tested, there is little evidence that tooth characters are affected by non-independence to any greater extent than the other morphological partitions. The primate and marsupial dataset~~s is~~ are the exception, with dental characters showing a weaker fit to the independent model than ~~almost~~ all other partitions~~, with the exception of pectoral girdle~~. ~~That pectoral characters are strongly affected by character non-independence in marsupials is unsurprising due to the developmental constraints placed on this girdle and the forelimb; the need for neonatal marsupials, born extremely early in their development, to crawl to the pouch requires these structures to develop precocially, and therefore potentially from a more integrated module (Sears, 2004; Cooper & Steppan, 2010)~~. The integration of the dental characters and their low rates of evolution in marsupials ~~is likely~~may be due to ~~similar~~ the unusual development~~constraints~~: neonatal marsupials, born extremely early in their development~~;~~ ~~the~~ need to attach to the teat, ~~leads~~ leading to precocial development of the jaw and facial region in marsupials (Smith, 1996, 2006). ~~and they do show reduced dental disparity relative to placentals (Werdlin, 1987)~~This could lead to this region evolving as a more integrated module. Alternatively, it may be a result of character selection: the Beck (2017) dataset contains large numbers of characters relating to the presence or absence or particular dental loci in both upper and lower jaws, which are likely to be heavily integrated.

The concept, pioneered by Sansom et al. (2017), of testing morphological discrete characters over a molecular benchmark~~,~~ is a powerful tool, and it would be highly recommended that researchers studying clades where ~~molecular~~molecularly-derived phylogenies exist examine the performance of their characters in this manner. But given the extremely wide variation in results found by this study, where different partitions produced

different relative phylogenetic signals (with the primates and marsupials in particular producing results conflicting strongly with the other datasets studied), one should perhaps be cautious of basing assumptions of character quality on the results of large meta-analyses. While the latter are useful for identifying broad-scale patterns, it is necessary that each dataset be examined individually, and decisions made based on the macroevolutionary patterns observed in that clade.

A fair and comprehensive sampling of characters across partitions should be the aim; experiments incorporating random sampling of characters show that sapling across partitions leads to a more reliable estimation of phylogenetic relationships than sampling within single partitions (Pattinson et al., 2015). While dental characters have been shown to suffer from issues of homology and non-independence (Kangas et al., 2004; Evans et al., 2007; Harjunmaa et al., 2014), the comparison of the dental characters to finer partitions of data presented here demonstrates that these issues are not unique to teeth. In fact, in some cases other regions perform even worse, and ~~that~~ the nature of these issues varies from clade to clade.

ACKNOWLEDGMENTS

We would like to thank Roger Benson and Robert Sansom for helpful ~~comments and~~ discussion. Robert Asher, Robin Beck and an anonymous reviewer provided insightful comments on an early draft of the manuscript. NB's research is funded by Deutsche Forschungsgemeinschaft grant number BR 5724/1-1. GLB's research is funded by a NERC studentship from the Oxford DTP in Environmental Research (NE/L0021612/1), alongside additional support provided by the European Research Council (grant agreement 637483).

REFERENCES

Beck R. M. 2017. The skull of *Eepidolops ameghinoi* from the early Eocene Itaboraí fauna, southeastern Brazil, and the affinities of the extinct marsupialiform order Polydolopimorphia. *Journal of Mammalian Evolution, 24*, 373–414

Bi S., Wang Y., Guan J., & Meng J. 2014.Three new Jurassic euharamiyidan species reinforce early divergence of mammals. *Nature*, *514*, 579–584.

Billet, G. & Bardin, J. 2018. Serial homology and correlated characters in morphological phylogenetics: modelling the evolution of dental crests in placentals. *Systematic Biology, In Press*

Dávalos L.M., Velazco P.M., Warsi O.M., Smits P.D., & Simmons N.B. 2014. Integrating incomplete fossils by isolating conflicting signal in saturated and non-independent morphological characters. *Systematic Biology*, *63*, 582–600.

DeGusta D. 2004. A method for estimating the relative importance of characters in cladistic analyses. *Systematic Biology*, *53*, 529–532

Eizirik E, Murphy WJ, Koepfli K-P, Johnson WE, Dragoo JW, Wayne RK, O'Brien SJ. 2010. Pattern and timing of diversification of the mammalian order Carnivora inferred from multiple nuclear gene sequences. *Molecular Phylogenetics and Evolution*, 56, 49-63

Evans A.R., Wilson G.P., Fortelius M., & Jernvall J. 2007. High-level similarity of dentitions in carnivorans and rodents. *Nature*, *445*, 78–81.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution*, *17*, 368-376

Fitch W.M. 1971. Toward defining the course of evolution: minimum change for a specific tree topology. *Systematic Biology*, *20*, 406-416

446 Fritz S. A., & Purvis A. 2010. Selectivity in mammalian extinction risk and threat types: a
447 new measure of phylogenetic signal strength in binary traits. *Conservation Biology*,
448 *24*, 1042-1051.

449 Gittleman J. L. & Jot, M. 1990. Adaptaion-statistics and a null model for estimating
450 phylogenetic effects. *Systematic Zoology*, *39*, 227-241

451 Goloboff, P.A. 1993. Estimating character weights during tree search. *Cladistics*, *9*, 83-91

452 Goloboff, P.A, Carpenter, J.M., Arias, J.S. & Esquivel, D.R.M. 2008. Weighting against
453 homoplasy improves phylogenetic analysis of morphological datasets. *Cladistics*, *24*,
454 758-773.

455 Goswami A. 2006. Cranial modularity shifts during mammalian evolution. *American*
456 *Naturalist*, *168*, 270–280.

457 Goswami, A. 2007. Cranial modularity and sequence heterochrony in mammals. *Evolution*
458 *and Development*, *9*, 290–298.

459 Goswami A. & Polly P. D. 2010. The influence of modularity on cranial morphological
460 disparity in Carnivora and Primates (Mammalia). *PLoS One*, *5*, e9517.

461 Harjunmaa E., Seidel K., Häkinen T., Renvoisé, Corfe I.J., Kallonen A., Zhang, Z.Q., Evans
462 A.R., Mikkola M.L., Klein O.D., & Jernvall, J. 2014. Replaying evolutionary
463 transitions from the dental fossil record. *Nature*, *512*, 44–48.

464 Harmon L. J., Weir J. T., Brock C. D., Glor R. E., & Challenger W. 2007. GEIGER:
465 investigating evolutionary radiations. *Bioinformatics*, *24*, 129–131

466 Hassanin A., Delsuc F., Ropiquet A., Hammer C., van Vuuren B. J., Matthee C., Ruiz-Garcia
467 M., Catzeflis F., Areskoug V., Nguyen T. T., & Couloux A. 2012. Pattern and timing
468 of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a

comprehensive analysis of mitochondrial genomes. *Comptes Rendus Biologie, 335,* 32–50.

Jernvall J., Hunter J. P., & Fortelius M. 1996. Molar tooth diversity, disparity, and ecology in Cenozoic ungulate radiations. *Science*, *274*, 1489-1492.

Jernvall J., Hunter J. P., & Fortelius M. 2000. Trends in the evolution of molar crown types in ungulate mammals: evidence from the Northern Hemisphere. In Teaford M. F., Smith M. M., & Ferguson M. W. J. (eds) *Development, Function and Evolution of Teeth* (p.p. 269-281). Cambridge University Press, Cambridge.

Kälersjö, M., Albert, V.A., & Farris, J.S. 1999. Homoplasy increases phylogenetic structure. *Cladistics, 15*, 91-93

Kavanagh, K.D., Evans, A.R. & Jernvall, J., 2007. Predicting evolutionary patterns of mammalian teeth from development. *Nature, 449*, 427–432.

Kangas A.T., Evans A.R., Thesleff I., & Jernvall J. 2004. Nonindependence of mammalian dental characters. *Nature, 432*, 211–214.

Kluge A.G. 2005. What is the rationale for 'Ockham's razor' (aka parsimony) in phylogenetic inference? In Albert, V.A. (ed) *Parsimony, Phylogeny and Genomics* (pp. 15-42). Oxford University Press, Oxford

Marjanović, D., & Laurin, M. 2019. Phylogeny of Palaeozoic limbed vertebrates reassessed through revision and expansion of the largest published relevant data matrix. *PeerJ, 6*, e5565

~~Meredith R.W., Janecka J.E., Gatesy J., Ryder O.A., Fisher C.A., Teeling E.C., Goodbla A., Eizirik E., Simão T.L., Stadler T., & Rabosky D.L. 2011. Impacts of the Cretaceous~~

~~Terrestrial Revolution and KPg extinction on mammal diversification. *Science*, *334*, 521–524~~

Mitchell K. J., Pratt R. C., Watson L. N., Gibb G. C., Llamas B., Kasper M., Edson J., Hopwood B., Male D., Armstrong K. N., Meyer M., Hofreiter M., Austin J., Donnellan S. C., Lee M. S. Y., Philips M. J., & Cooper A. 2014. Molecular phylogeny, biogeography, and habitat preference evolution of marsupials. *Molecular Biology and Evolution*, *31*, 2322–2330.

Münkemüller T, Lavergne S, Bzeznik B, Dray S, Jombart T, Schiffers K, Thuiller W. 2012. How to measure and test phylogenetic signal. *Methods in Ecology and Evolution*, *3*, 743-56.

Ni X, Gebo DL, Dagosto M, Meng J, Tafforeau P, Flynn JJ, Beard KC. 2013 The oldest known primate skeleton and early haplorhine evolution. *Nature*, *498*, 60.

O'Keefe F.R., & Wagner P.K. 2001. Inferring and testing hypotheses of cladistic character dependence by using character compatibility. *Systematic Biology*, 50, 657–675.

~~O'Leary M.A., & Gatesy, J. 2008. Impact of increased character sampling on the phylogeny of Cetartiodactyla (Mammalia): combined analysis including fossils. *Cladistics*, *24*, 397–442~~

Pagel M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society B*, *255*, 37–45

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

513 Pattinson D. J., Thompson R. S., Piotrowski A. K., & Asher R. J. 2015. Phylogeny,
514 paleontology, and primates: do incomplete fossils bias the Tree of Life? *Systematic*
515 *Biology*, *64*,169–186

516 Pennell MW, Harmon MJ. 2013. An integrative view of phylogenetic comparative methods:
517 connections to population genetics, community ecology and paleobiology. Annals of
518 the New York Academy of Sciences, *1289*, 90-105

519 Pennell MW, Eastman JM, Slater GJ, Brown Jw, Uyeda JC, FitzJohn RG, Alfaro ME,
520 Harmon LJ. 2014. Geiger v2.0: an expanded suite of methods for fitting
521 macroevolutionary models to phylogenetic trees. *Bioinformatics*, 30, 2216-2218.

522 Perelman P., Johnson W. E., Roos, C., Seuánez H. N., Horvath J. E., Moreira M. A. M.,
523 Kessing B., Pontius J., Roelke M., Rumpler Y., Schneider M. P. C., Siva A., O'Brien
524 S. J., & Pecon-Slattery J. 2011. A molecular phylogeny of living primates. *PLoS*
525 *Genetics*, *7*, e1001342

526 Pinheiro J, Bates D, DebRoy S, Sarkar D, Heisterkamp S, Van Willigen B. 2017. Package
527 'nlme'. Linear and Nonlinear Mixed Effects Models, version 3.1

528 Price S. A., Hopkins S. S. B., Smith K. K. & Roth, V. L. 2012. Tempo of trophic evolution
529 and its impact on mammalian diversification. *Proceedings of the National Academy of*
530 *Science*, *109*, 7008-7012.

531 R Core Team. 2016. *R: a language and environment for statistical computing*. Vienna,
532 Austria.

533 Revell L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other
534 things. *Methods in Ecology and Evolution*, *3*, 217–223

Ruvinksy I., & Gibson-Brown J. J. 2000. Genetic and developmental bases of serial homology in vertebrate limb evolution. *Development, 127*, 5233–5244.

Sadleir R.W., & Makovicky P.J. 2008. Cranial shape and correlated characters in crocodilian evolution. *Journal of Evolutionary Biology, 21*, 1578–1596.

Sankoff D., & Cedergren R.J. 1983. Simultaneous comparison of three or more sequences related by a tree. In Sankoff, D., Kruskal, J.B. (eds). *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison* (pp. 253-263). Addison-Wesley, Reading, Massachusetts

Sansom, R. S., Wills, M. A., & Williams, T. 2017. Dental data perform relatively poorly in reconstructing mammal phylogenies: morphological partitions evaluated with molecular benchmarks. *Systematic Biology, 66*, 813–822.

Scotland R.W., Olmstead R.G., & Bennett J.R. 2003. Phylogeny reconstruction, the role of morphology. *Systematic Biology, 52*, 529–548.

Smith KK. 1996. Integration of craniofacial structures during development in mammals. *American Zoologist, 26*, 70-79

Smith KK. 2006. Craniofacial development in marsupial mammals: developmental origins of evolutionary change. *Developmental dynamics, 235*, 1181-1193

Spaulding M, O'Leary MA, Gatesy J. 2009. Relationships of Cetacea (Artiodactyla) among mammals: increased taxon sampling alters interpretations of key fossils and character evolution. *PLoS one, 4*, e7062.

Tomiya S. 2010. A new basal caniform (Mammalia: Carnivora) from the middle Eocene of North America and remarks on the phylogeny of early carnivorans. *PLoS ONE*, 6, e24146

Van Valkenburgh B. 1988. Trophic diversity in past and present guilds of large predatory mammals. *Paleobiology, 14*, 155-173

Wesley-Hunt G.D. 2005. The morphological diversification of carnivores in North America. *Paleobiology, 31*, 35-55

Young N., & Hallgrímsson B. 2005. Serial homology and the evolution of mammalian limb covariation structure. *Evolution, 59*, 2691–2704.

FIGURE CAPTIONS

Figure 1: Violin plots illustrating results ~~Results~~ from the Bi et al. (2014) character matrix (total Mammalia). A) Pagel's lambda values (phylogenetic signal) of each character. A value of 0 indicates no phylogenetic signal, while a value of 1 indicates high phylogenetic signal. B) Akaike weights support for the ER model of evolution of each character. Characters with an Akaike weights score of 1 have equal rates of within-character evolution between each state, while characters with a score of 0 display unequal rates of within-character state evolution. C) Akaike weights support for the independent model of evolution of all pairwise comparisons of characters in each partition. Pairwise comparisons that have an Akaike weights score of 1 evolve independently of one another, while pairwise comparisons with a score of 0 display character non-independence. D) Rates of character evolution of each character (log ~~scale~~transformed). For each partition, the horizontal spread of the violin plot represents the density of data at each point on the y-axis. Box plots with a white point representing the median are plotted within each violin plot. The heatmap is a visual representation of the y-axis.

Commented [R8]: Log10 or natural log?

Figure 2: Violin plots illustrating results ~~Results~~ from Spaulding *et al.* (2009) ~~O'Leary & Gatesy (2008)~~ matrix (A~~Ceta~~rtiodactyla). A) Pagel's lambda values (phylogenetic signal) of each character. B) Akaike weights support for the ER model of evolution of each character. C) Akaike weights support for the independent model of evolution of all pairwise comparisons of characters in each partition. D) Rates of character evolution of each character (log ~~scale~~transformed).

Commented [R9]: Log10 or natural log?

Figure 3: Results from the Tomiya (2011) matrix (Carnivora). A) Pagel's lambda values (phylogenetic signal) of each character. B) Akaike weights support for the ER model of evolution of each character. C) Akaike weights support for the independent model of

evolution of all pairwise comparisons of characters in each partition. D) Rates of character evolution of each character (log transformed).

Commented [R10]: Log10 or natural log?

Figure ~~3~~4: Violin plots illustrating results ~~Results~~ from the ~~Pattinson~~ Ni et al. (~~2005~~2010) matrix (Primates). A) Pagel's lambda values (phylogenetic signal) of each character. B) Akaike weights support for the ER model of evolution of each character. C) Akaike weights support for the independent model of evolution of all pairwise comparisons of characters in each partition. D) Rates of character evolution of each character (log ~~scale~~transformed).

Commented [R11]: Log10 or natural log?

Figure ~~4~~5: Violin plots illustrating results ~~Results~~ from the Beck (2017) matrix (Marsupialia). A) Pagel's lambda values (phylogenetic signal) of each character. B) Akaike weights support for the ER model of evolution of each character. C) Akaike weights support for the independent model of evolution of all pairwise comparisons of characters in each partition. D) Rates of character evolution of each character (log ~~scale~~transformed).

Commented [R12]: Log10 or natural log?

TABLES

Table 1. Results of GLS analyses of Pagel's λ (phylogenetic signal of character partitions) in mammals. Rows coloured are those where the partition best fits the H1 model (partition has a different lambda value to all others); blue indicates lower phylogenetic signal, red indicates higher phylogenetic signal.~~.~~

Partition	Median λ	GLS Coefficient	lnL (null)	lnL (H1)	AIC (null)	AIC (H1)	Akaike weights (null)	Akaike weights (H1)
Teeth	1	-0.15	-103.02	-102.6	210.0	211.1	0.63	0.37
Skull	1	-0.11	-103.02	-101.4	210.0	208.8	0.35	0.65
Vertebrae	1	0.15	-103.02	-103.5	210.0	212.9	0.81	0.19
Forelimb	1	0.16	-103.02	-101.3	210.0	208.5	0.32	0.68
Hindlimb	1	0.07	-103.02	-103.8	210.0	213.6	0.86	0.14
Soft tissue	1	0.15	-103.02	-103.4	210.0	2129	0.81	0.19

Table 2. Results of GLS analyses of Akaike weight support for the equal rates (ER) model of character evolution in mammals. ~~Rows coloured are those where the partition best fits the H1 model (partition has a different rate value to all others); red indicates higher support for equal rates.~~

Partition	Median weight	GLS Coefficient	lnL (null)	lnL (H1)	AIC (null)	AIC (H1)	Akaike weights (null)	Akaike weights (H1)
Teeth	0.78	0.02	26.77	24.75	-49.55	-43.50	0.95	0.05
Skull	0.71	0.02	26.77	24.45	-49.55	-42.89	0.97	0.03
Vertebrae	0.62	-0.13	26.77	26.60	-49.55	-47.19	0.76	0.24
Forelimb	0.62	-0.09	26.77	26.91	-49.55	-47.82	0.70	0.30
Hindlimb	0.72	0.04	26.77	24.85	-49.55	-43.70	0.95	0.05
Soft tissue	0.67	0.05	26.77	25.69	-49.55	-45.38	0.89	0.11

Table 3. Results of GLS analyses of Akaike weight support for the independent model of character evolution in mammals. Rows coloured are those where the partition best fits the H1 model (partition has a different rate value to all others); blue indicates lower Akaike weights, red indicates higher.

Partition	Median weight	GLS Coefficient	lnL (H0)	lnL (H1)	AIC (H0)	AIC (H1)	Akaike weights (H0)	Akaike weights (H1)
Teeth	0.69	0.012	1043.0	1040.5	-2082	-2075	0.97	0.10
Skull	0.73	0.013	1043.0	1040.8	-2082	-2076	0.76	0.22
Vertebrae	0.50	-0.034	1043.0	1041.1	-2082	-2076	0.95	0.04
Forelimb	0.57	-0.077	1043.0	1073.1	-2082	-2140	~0	~1
Hindlimb	0.75	0.039	1043.0	1050.7	-2082	-2095	0.001	0.999
Soft tissue	0.76	0.100	1043.0	1042.4	-2082	-2079	0.83	0.16

Table 4. Results of GLS analyses of rates of character evolution in mammals. Rows coloured are those where the partition best fits the H1 model (partition has a different rate value to all others); ~~blue indicates lower rate,~~ red indicates higher rate.

Partition	Median rate	GLS Coefficient	lnL (null)	lnL (H1)	AIC (null)	AIC (H1)	Akaike weights (null)	Akaike weights (H1)
Teeth	0.0016	0.29	-53.01	-43.63	110.03	93.27	0.0002	0.99
Skull	0.0010	0.02	-53.01	-55.19	110.03	116.38	0.96	0.04
Vertebrae	0.0006	-0.20	-53.01	-52.59	110.03	111.19	0.64	0.36
Forelimb	0.0006	-0.08	-53.01	-53.76	110.03	113.52	0.85	0.15
Hindlimb	0.0007	-0.07	-53.01	-53.85	110.03	113.70	0.96	0.04
Soft tissue	0.0006	-0.21	-53.01	-53.12	110.03	112.24	0.75	0.25

Table 5. Results of Kendal's tau correlation tests between phylogenetic signal, rates of evolution, and support for the equal rates model

	Pagel's lambda vs Rates of character evolution	Pagel's lambda vs Akaike weight support for ER model of character evolution
Total mammal dataset	-0.22 ($p = 3.67 \times 10^{-6}$)	-0.050 (p = 0.2996)
Cetartiodactyl dataset	-0.24 ($p=3.49 \times 10^{-10}$)	0.15 ($p=1.05 \times 10^{-4}$)
Carnivoran dataset	-0.1 (p=0.4435)	-0.04 (p=0.5701)
Primate Dataset	-0.22 ($p=<2.2 \times 10^{-16}$)	-0.012 (p=0.56)
Marsupial dataset	-0.22 ($p=2 \times 10^{-5}$)	0.11 (p=0.025)

Table 6. Results of Kendal's tau correlation tests between number of characters in the partitions and phylogenetic signal, support for the equal rates and independent models, and rates of evolution.

Correlation test	Kendall's tau	P value
Number of characters in dataset partition ~ Median Pagel's lambda	-0.009	0.95
Number of characters in dataset partition ~ Median Akaike weights (ER model)	0.31	0.10
Number of characters in dataset partition ~ Median Akaike weights (independent model)	0.26	0.17
Number of characters in dataset partition ~ Median rate	0.28	0.12