Comparing Morphology and Molecules to Evaluate Phylogenetic Inference

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy
in the Faculty of Science and Engineering

2019

Leah M. Callender-Crowe

Department of Earth and Environmental Sciences

Contents

Abstract....3
Lay Abstract....4
Declaration....5
Copyright Statement....5
Acknowledgements....7
Curriculum Vitae....8
Introduction: Phylogenetic Conflict and Integration of Morphology with Molecules....14
1. Osteological Characters Show Greater Congruence with Molecular Phylogenies than Soft Characters in Avian and Reptilian Morphological Datasets....60
2. The Prevalence of Correlated Morphological Characters and their Effect on Phylogenetic Reconstruction....98
3. Cliques of Morphological Characters in Avian and Squamate Datasets...142
4. Convergence, Correlations and Cliques: An Analysis of the Relative Performance of Morphological Character Partitions in Mammals....170
General Discussion and Summary....213
Appendices....231

Thesis word count (including tables): 43899

Abstract

Phylogenetics, the reconstruction of evolutionary relationships between species, underpins evolutionary biology. These relationships are most often inferred from either or both morphology and molecules, but molecular data are often preferred because of the objective atomization and large number of characters, the relative neutrality of some genomic regions and the sophisticated statistical techniques that these benefits allow. Morphology, by contrast, remains relatively neglected. However, morphological data remain essential, particularly for the phylogenetic placement of fossil species for which no molecular data are available. Morphology nevertheless suffers from issues such convergence and non-independence of traits, the extent and distribution of which remain largely unknown.

In this thesis I compare osteological, dental and other partitions of morphological data in tetrapods, using molecular trees as a benchmark. I assess differences in levels of homoplasy, ages of character transition, tree-based correlations and internal consistency. To do this, I compute the retention index of characters and partitions on molecular trees, perform ancestral state reconstructions to estimate character transition ages, apply correlated and uncorrelated models of character pair evolution, perform cluster analyses, and build trees using subsets of data.

I find heterogeneity between these partitions, both with and without respect to molecular trees. Specifically, I find that osteological characters are more homoplasious and transition earlier than either dental or soft characters. Further, characters are more correlated within partitions than between partitions. These results highlight the importance of partitions, implying differences in convergence, evolutionary rates and integration between different morphological subsets. As well as shedding light on the nature of morphological evolution in tetrapods, these results have important implications for phylogenetic inference, and suggest the need for careful consideration of the properties of morphological data when reconstructing evolutionary history. Specifically, composite coding and partitioning may be necessary in model-based approaches.

Lay Abstract

Evolutionary trees are built by coding biological data and then using mathematical techniques to arrive at an estimate of the relationships between species. These trees are most often built from either or both organismal form, known as morphology, and DNA sequences, usually called molecular data. In the age of genomics, sequence data is often preferred as it is objective to code, abundant and can be relatively neutral in comparison to morphology, which is both directly affected by natural selection and harder to objectively code. However, morphological data remain important in evolutionary biology, not least because DNA usually cannot be extracted from fossils, meaning that morphological information is the only information available for these species.

Trees built from these different sources of information sometimes conflict, which raises the question of which parts of morphology, if any, are best for reconstructing evolutionary relationships. Different parts of morphology, for example teeth, bones, plumage and scales, evolve differently as a result of different evolutionary origins and pressures. This may manifest as faster or slower evolution, more or less convergence (where unrelated species evolve similar morphologies as a consequence of sharing a similar habitat or ecological niche), or different levels of correlated or concerted evolution (meaning the evolution of multiple traits together resulting from shared function). These properties can all effect the estimation of evolutionary relationships.

Here I compare regions of morphology with molecules, and find that skeletal data, that is data from bones, display less convergence and are also older than most other forms of morphological data. Bones are also slightly less likely to display correlated evolution. In short, bones reconstruct evolutionary relationships in a similar way to DNA. These results all imply that skeletal data may be the best form of morphological data for building reliable evolutionary trees. These results shed light on the evolution of different aspects of organismal form, as well as being useful for researchers, particularly palaeontologists, building evolutionary trees from morphology.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420), in any

relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in The University's policy on Presentation of Theses

Acknowledgements

I dedicate this thesis to my sweet little sisters, Emilia and Olivia. I hope I can inspire them to love learning.

I firstly thank my primary supervisor, Dr Robert Sansom, for steadfast guidance, encouragement and support. I also thank my co-supervisors, Professors David Robertson and Chris Klingenberg, and my advisor Professor Susanne Shultz.

I thank many researchers at the University of Manchester, both current and former, including Dr Robert Brocklehurst, Dr Joe Keating, Dr Emma Randle, Dr Russell Garwood, Professor James McInerney, Dr Chris Knight and others for various forms of academic and personal support. I thank various colleagues here and at other institutions, far too many to name individually, for guidance and friendship.

I thank my partner Dean, my parents Nick and Naomi, my nannas Sylvia and Sarah, and all my family for all their support. I thank all my friends in London and elsewhere, especially those who visited me and who I visited at various times while I was in need of company. My cat Tsar must also receive a mention for his constant affection, and his naughtiness, for keeping me on my toes. I thank all my senseis, coaches and training partners - martial arts helped keep my mind healthy, and has been and remains an extremely important source of stress relief.

Curriculum Vitae

Ms Leah Callender-Crowe
leah.callender-crowe@postgrad.manchester.ac.uk
Michael Smith Building, The University of Manchester, Dover St, Manchester M13 9PT

PROFILE

I am a BBSRC-funded PhD candidate researching evolutionary biology at The University of Manchester. My research involves comparing the properties of morphological and molecular data in the context of phylogenetic analysis, examining the prevalence and effects of homoplasy and correlated evolution in morphological datasets. I have presented my work at several conferences including SVPCA and the 5th International Palaeontological Congress. I have experience teaching undergraduates. I also have some experience with collections, having volunteered at the Natural History Museum during my MSc.

KEY SKILLS

- **Phylogenetics** – using R, TNT, Mesquite and other software
- **Statistics and coding** – using R
- **Writing** – for both academic and lay audiences
- **Public speaking, engagement, and dissemination of research**

EDUCATION

The University of Manchester, Oxford Road, Manchester, UK, *Sept 2015-present*

PhD Integrating Morphology, Fossils and Molecules to Evaluate Major Evolutionary Events

Supervised by Dr Robert Sansom

During my PhD I have investigated properties of partitions of morphological data (hard and soft characters) in birds and reptiles with respect to concerted evolution, integration, homoplasy and transition ages. I have used molecular trees as a benchmark.

Imperial College London/The Natural History Museum, South Kensington, UK, *Sept 2013 – Sept 2014*

MSc Taxonomy and Biodiversity, Distinction

Dissertation: The evolutionary reality of higher taxa and conservation of extinction threat in birds

Supervised by Dr Aelys Humphreys and Professor Tim Barraclough

I studied modules including Palaeontology, Cladistics, Molecular Systematics, Morphometrics, Statistics, Applied Biodiversity and Concepts in Biodiversity as well as a field course and a research project.

For my dissertation I used the generalised mixed Yule coalescent method (GMYC), a tree-based statistical method typically used to delimit species, to test for higher evolutionarily significant units (hESUs) in birds. I then used IUCN threat data to determine whether hESUs corresponded with extinction risk.

Swansea University, Singleton Park, Swansea, UK, *Sept 2009 – July 2012*

BSc Zoology, 1st Class Honours

Dissertation: Biodiversity and latitude: is variation among populations of invertebrate species influenced by latitude?

Supervised by Dr Patricia Lee

For my dissertation I used the fixation index (F_{st}) to compare genetic divergence between high and low latititude populations of a number of invertebrate species to determine whether intraspecific diversity was higher at lower latitudes.

Hornsey School for Girls, Haringey, London, *Sept 2002 – July 2009*

SELECTED WORK EXPERIENCE

Sense about Science, 14a Clerkenwell Green, EC1R 0DP, London, *01/2017-04/2017, 05/2019 - present*

Intern, Ask for Evidence Ambassador

I worked as an intern at Sense about Science during the second year of my PhD. I helped to organise speakers for the Evidence Matters event in European parliament, helped organise the John Maddox Prize, which is awarded for standing up for science, researched and proofread a publication about nuclear energy and wrote for the Ask for Evidence blog. Additional duties included taking calls, organising other events, helping with social media and general admin duties. I now volunteer as an ambassador, supporting and promoting the Ask for Evidence campaign. I particularly enjoy discussing scientific issues of political and social importance such as climate change, GMOs, vaccines and other health related topics.

The University of Manchester, Oxford Road, M13 9PL, Manchester, *01/2016-08/2018*

Teaching Assistant

I worked as a student demonstrator on a casual basis at the University of Manchester. My first role was assisting undergraduates in practical sessions in a human anatomy module, which involved performing heart dissections and discussing comparative vertebrate heart anatomy in an evolutionary context, followed by sessions on DNA manipulation and plant physiology. I have also invigilated exams.

The Manchester Museum, Oxford Road, M13 9PL, Manchester, *10/2015-05/2017*

Student Demonstrator

I worked as a student demonstrator on a casual basis at the Manchester Museum. I was trained to and helped deliver Diversity of Life and Human Evolution sessions for secondary school pupils.

The Natural History Museum, Cromwell Road, SW7 5BD, London, *10/2013-02/2014*

Assistant Curator

During my MSc I did some part-time voluntary work with the Hymentoptera collection at the Natural History Museum. I mounted delicate and historically important parasitoid specimens using various techniques. This gave me an appreciation for the importance of natural history collections and improved my understanding of the taxonomy of ants, bees and wasps.

PUBLICATIONS

Humphreys, A. M., Rydin, C., Jønsson, K. A., Alsop, D., Callender-Crowe, L. M., & Barraclough, T. G. (2016). Detecting evolutionarily significant units above the species level using the generalized mixed Yule coalescent method. *Methods in Ecology and Evolution*, *7*(11), 1366–1375.

Callender-Crowe, L. M., & Sansom, R. (2019). Homoplasy of morphological partitions of birds and squamates relative to molecular trees. (submitted)

Callender-Crowe, L. M., & Sansom, R. (2019). The presence of correlated morphological characters and their effect on phylogenetic reconstruction. (in prep)

CONFERENCES

Callender-Crowe, L. M., & Sansom, R. (2018). Homoplasy and correlation of morphological characters in birds and reptiles relative to molecular phylogeny [oral presentation]. 5th International Palaeontological Congress, Paris

Callender-Crowe, L. M., & Sansom, R. (2018). Evaluating the performance and correlation of morphological characters in reconstructing avian and squamate evolutionary history [oral presentation]. SVPCA, Manchester

Callender-Crowe, L. M., & Sansom, R. (2018). Evaluating the performance of morphological partitions in recovering Avian evolutionary history [oral presentation]. BBSRC DTP, Manchester

Callender-Crowe, L. M., & Sansom, R. (2017). Differences between the morphological partitions of birds and reptiles used to build evolutionary trees [oral presentation]. SEES PGRC, Manchester

Callender-Crowe, L. M., & Sansom, R. (2017). Evaluating the performance of morphological partitions in recovering Avian evolutionary history [poster presentation]. SVPCA, Birmingham

Callender-Crowe, L. M., & Sansom, R. (2017). Evaluating the performance of morphological partitions in recovering Avian evolutionary history [poster presentation]. BBSRC/MRC DTP Conference, Manchester

Callender-Crowe, L. M, Choate, P., & Sansom, R. (2016). Evaluating the phylogenetic consistency of morphological data for birds using molecular trees [poster presentation]. Palaeontological Association Annual Meeting, Lyon

FUNDING AND AWARDS

First Year Symposium Oral Presentation (2nd place), *The University of Manchester*, 2016

BBSRC DTP Studentship, *Bioscience and Biotechnology Research Council*, 2015

Nuffield Science Bursary, *The Nuffield Foundation*, 2008

Introduction: Phylogenetic Conflict and Integration of Morphology with Molecules

Abstract

The topology and dating of evolutionary trees are both dependent on a number of factors such as the quality and quantity of data, phylogenetic inference method and model choice. Unsurprisingly, a great source of conflict between different hypotheses of evolutionary history lies in the kind of data used. The most salient example of this is the disparity between the estimates of evolutionary relationships offered by morphological and molecular data. This problem has plagued researchers since the advent of molecular systematics, and continues to do so today. The differences between these sources of data may be the result of differences between molecular and morphological evolution, such as the prevalence of modularity, integration, correlation and homoplasy in morphological characters. These processes pose problems for morphological data, and thus the conflict between these data types is often framed as a problem of morphology. However, morphology continues to be an indispensible source of phylogenetic data, for example when placing fossil species, and as such it is important to evaluate how morphology performs in phylogenetic inference. One way to do this is to analyse subsets of morphological data, for example against existing independent phylogenetic trees, in order to estimate the relative prevalence of these processes in these different character types. Ultimately, neither morphological nor molecular data is without its biases that can potentially mislead phylogenetic analyses. Nevertheless, testing and accounting for these processes in morphological data, and highlighting areas of consilience between morphology and molecules, will help to shed light on sources of conflict in phylogenetic data and potentially aid in the creation of more accurate model-building in phylogenetic inference. Further, these insights may be of use for researchers interested in the trends and dynamics of morphological trait evolution. This chapter discusses the sources of conflict and consilience between trees

constructed between different data types, examining different trends in different subsets of morphological data.

1. Introduction

Phylogenetics is the subfield of evolutionary biology concerned with reconstructing the evolutionary relationships between species. Phylogenies, or evolutionary trees, cannot be known with complete confidence. This is because, with the exception of simulated and experimental evolution, we cannot directly observe historical lineage splitting. To infer phylogeny, we therefore must extract information from species and employ statistical methods to produce the most likely estimates of evolutionary history given the data. The lack of direct observation means that independent sources of information must ideally be used in order to corroborate phylogenetic hypotheses (e.g. Beutel *et al.*, 2011; Field *et al.*, 2014). However, there is often conflict between phylogenetic datasets, both within and between morphology and molecules owing to differences in evolutionary rates and dynamics (e.g. Flynn & Nedbal, 1998; Pisani, Benton, & Wilkinson, 2007; Bibi, 2013; Sharma *et al.*, 2014; Reddy *et al.*, 2017).

While it is well established that different datasets belonging to the same broad data type produce different tree topologies for the same clades (Flynn & Nedbal, 1998; Sharma *et al.*, 2014; Reddy *et al.*, 2017), trees built from different data types, i.e. molecular and morphological data, are usually more different to each other than to trees built from the same data type both in topology (Pisani *et al.*, 2007) and in the timing of branching events (Benton, 1999).

Sometimes molecular and morphological datasets each strongly support different hypotheses of evolutionary history, and this disparity is often consistent across different studies of particular clades. Salient examples of these recurring incongruences between molecular and morphological trees include in birds (Torres & van Tuinen, 2013) and squamate reptiles (Reeder *et al.*, 2015). For example, morphological analyses have traditionally placed

starlings in a group including the crows and ravens, whereas later molecular analyses have placed them as the closest relatives of the mockingbirds (Sibley, Ahlquist, & Monroe, 1988). Further, sequence data places flamingoes and grebes together, contradicting previous morphology-based hypotheses (Van Tuinen *et al.*, 2001). These and other important discrepancies imply that the morphological data for these birds have consistently misled analyses (Torres & van Tuinen, 2013), allowing that the molecular data track evolutionary history relatively accurately in these cases.

While these incongruences are well documented, the question of why these data types disagree so strongly in their resultant topology is still outstanding. One view is that morphological characters are relatively sparse and ambiguous compared with the high number of characters produced by sequence data (Scotland, Olmstead, & Bennett, 2003) and therefore that differences between morphological and molecular phylogenies simply reflect the relative inability of morphological data to produce accurate and well-supported phylogenies. However, it is also true that individual morphological characters often contain higher information content than molecular characters.

These incongruences may be also the result of differences between molecular and morphological evolution; i.e. morphological trees are not simply incorrect as a result of inadequacy, subjectivity or lack of statistical power in morphological datasets, but in fact often capture the signature of ecological specialization repeated in unrelated clades. Morphological-molecular conflicts are often reflective of morphological convergence, where shared morphological traits between unrelated clades have misled phylogenetic analysis.

A striking and well-documented case of this is in the *Anolis* lizards of the Caribbean islands, where multiple cases of convergence have resulted in equivalent morphotypes arising several times independently (Losos, 1998), and cluster analysis of morphological data places these morphotypes together. Molecular evidence shows that these morphotypes have often arisen independently on each island (Figure 1).

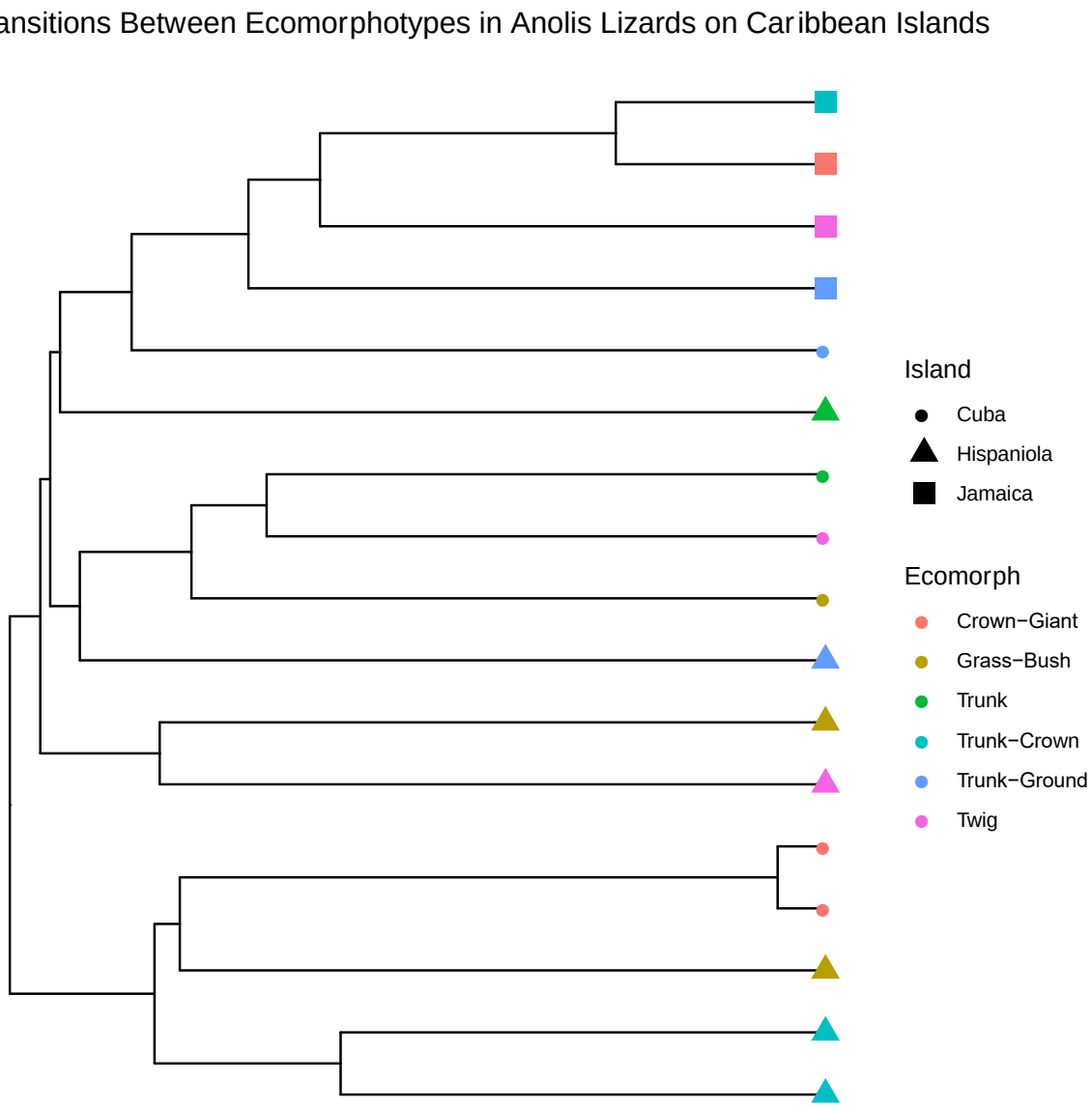


Figure. 1 *Phylogeny of 17* Anolis *lizard species belonging to 6 ecomorphotypes on 3 Caribbean islands. Each tip represents a different species. Tip colours indicate ecomorphotypes and tip shapes indicate the island on which the species is found. Often, different ecomorphotypes on the same island are more closely related to each other than the same ecomorphotypes on different islands, demonstrating multiple evolutionary transitions between ecomorphotypes. There is also some movement between islands. The phylogeny is adapted from* Zheng & Wiens, (2016) *squamate phylogeny.*

Ideally, molecular and morphological datasets should converge on a similar topology. This is because if one can get two or more independent datasets to agree, it is possible to have greater confidence that the tree is a good estimate of evolutionary history (e.g. Beutel *et al.*, 2011; Borsch *et al.*, 2018).

The agreement between multiple lines of evidence is known as consilience, and is a commonly used criterion for confidence in hypotheses in evolutionary biology (Field *et al.*, 2014; McInerney, O'Connell, & Pisani, 2014). Multiple morphological or multiple molecular phylogenies may broadly agree with each other, but these phylogenies may be based on the same or similar data, or, even if they are not, may each suffer from the same biases and systematic errors as the other.

Morphology therefore continues to be important in phylogenetics as a point of comparison with phylogenies built from molecular data which otherwise would have no corroborating evidence. Additionally, combined analyses where both molecular and morphological data are permitted to influence topology can sometimes produce better-resolved trees than either type of data used alone (e.g. Beutel *et al.*, 2011). However, a difficulty with combined analyses is the circularity of subsequently attempting to map character changes onto the tree in studies of morphological evolution and character distribution if these characters were used to build the tree in the first instance (de Queiroz, 1996).

The increase in the use of molecular data in phylogenetics analysis raises the question of how to better exploit morphological data to elucidate evolutionary history, including past branching events and phenotypic change over time. This involves interrogating morphological datasets, and may include, for example, partitioning morphological data to identify areas of morphology conveying conflicting signal (Sansom, Wills, & Williams, 2016; Sansom & Wills, 2017), and identifying correlated character pairs (Leslie *et al.*, 2015; Sauquet *et al.*, 2017) and cliques (Holland *et al.*, 2010; Blanke *et al.*, 2013).

2. Phylogenetic Data Types

2.1. Morphological Data

Organismal form is a very important area of research in evolutionary biology, and without cladistics or morphometric data only limited insight into the evolution of organisms is possible. Prior to the advent of modern phylogenetics, traditional taxonomic studies based on morphology alone placed organisms into hierarchically patterned groups broadly congruent with how they are classified today. Taxonomic groupings are typically based on real and biologically significant morphological differences that are discontinuous and non-randomly distributed. Indeed, discrete and independently evolving groups are possible even above the species level, possibly the result of broad niche occupation and turnover within clades (Barraclough, 2010; Humphreys & Barraclough, 2014). Much of the discrete nature of organismal form is driven by modularity, i.e. the association and correlation of traits (Goswami *et al.*, 2014; Felice, Randau, & Goswami, 2018), a consequence of the existence of local optima within morphospace (Mitteroecker & Huttegger, 2009; Dumont *et al.*, 2014; Button, Barrett, & Rayfield, 2017), which are often convergent (Losos, 1998; Goswami & Polly, 2010a; Goswami, Milne, & Wroe, 2011; Friedman *et al.*, 2016; Davis & Betancur-R, 2017), differ with the environment (Hadfield, 2016), and outside of which groups of traits are maladaptive. The discontinuity of morphological form is captured by both geometric morphometric and cladistic data (Hetherington *et al.*, 2015), and therefore raises questions about the validity and use of morphological data in phylogenetics since they are affected by the correlated evolution of phenotypic traits.

The relative utility of morphological data in phylogenetic analysis, particularly as more molecular data have become available, has been widely discussed in the literature (Hillis, 1987; Scotland *et al.*, 2003; Jenner, 2004; Wiens, 2004; Giribet, 2015; Lee & Palci, 2015), with some authors suggesting we disregard morphology altogether. However, morphological information may always be necessary for phylogenetic reconstruction considering that many organisms, such as fossil species and some rare species have no molecular information available with which to build phylogenies. In fact, there is no molecular information available for the vast majority of life that has ever existed. The only phylogenetic information we

have for these species, if any, must therefore necessarily be derived from morphology. This is especially critical since the correct placement of fossil species is necessary for the calibration of time-trees (Springer *et al.*, 2001). Secondly, only through morphological information can character state changes and the evolution of body plans be mapped over time (Giribet, 2015). Even aside from these two functions for which morphology is obviously necessary in evolutionary biology, it may also be the case that even the relatively small number of characters yielded by morphological data influence tree topology in combined phylogenetic analysis (Gatesy *et al.*, 2003), give stronger statistical support (Lee & Camens, 2009) and result in better-resolved trees (Beutel *et al.*, 2011). The addition of fossil taxa in particular can affect topology and produce greater congruence with independent trees (Asher *et al.*, 2019). Even when morphology does not actually improve topological accuracy, good resolution of phylogenetic trees is sometimes a goal in its own right, since many analyses such as tree distance methods (Robinson & Foulds, 1981) and correlations (Billet & Bardin, 2019) require or perform better with bifurcating trees.

However, morphology in phylogenetic analysis suffers from many pitfalls, one of which is the problem of how to objectively identify and codify homologous characters. There are several related definitions of homology, but in this context it is generally defined as structures belonging to different species but sharing common ancestry, and therefore often sharing functional or superficial similarities (Wiley & Lieberman, 2011). One of the issues here is that to define a character or set of characters is to propose a hypothesis of homology, and therefore of common ancestry, and is therefore arguably circular when subsequently using this data to reconstruct evolutionary history (see Cartmill, 1994 for a critique of homology as a concept in morphological phylogenetics). This is not a problem if homologous structures are correctly identified, but when characters are incorrectly judged to be homologous, potentially as a result of the bias of individual researchers, or the historical burden of past taxonomic judgments, this misleads phylogenetic analysis. Nevertheless, most systematists use operational homology, where there is high similarity

between characters thought to be homologous in terms of composition and structure (Wiley & Lieberman, 2011).

This also raises the issue of evolutionary convergence, whereby analogous structures arise in different species as a consequence of similar evolutionary regimes acting on them in similar environments (Losos, 1998, 2011; Schluter, 2000; Moen, Morlon, & Wiens, 2016; Gheerbrant, Filippo, & Schmitt, 2016; Mugleston *et al.*, 2018). Distantly related species exploiting the same niche will often display phenotypic convergence reflecting their shared ecology. When homoplasy as a result of convergence is mistaken for homology, relationships between species can spuriously appear closer than they really are.

A related but less-discussed issue is the role of character non-independence on phylogenetic reconstruction (Emerson & Hastings, 1998; Sadleir & Makovicky, 2008; Goswami & Polly, 2010a; Guillerme & Brazeau, 2018; Billet & Bardin, 2019). Simply put, non-independence occurs when change in one character is accompanied by change in another (Pagel, 1994; Beaulieu, O'Meara, & Donoghue, 2013). Often, whole suites of characters will evolve together in semi-autonomous blocks (Holland *et al.*, 2010; Blanke *et al.*, 2013).

Modularity (the association of several traits) has several potential influences on morphological evolution, including the prevention of canalisation and either the facilitation or restriction of phenotypic divergence – in other words, it has implications for the 'evolvability' of structures (Wagner & Altenberg, 1996; Goswami & Polly, 2010b; Goswami *et al.*, 2014). The influence of modularity on morphological evolution has been explored by Goswami & Polly, (2010b). They tested two hypotheses, namely 'facilitation', whereby the presence of modules increases disparity between species, and 'constraint' whereby the presence of modules limits morphological disparity between species, with mixed results but finding that strong integration in primate and carnivore skulls generally has either no effect on disparity or constrains it.

One important explanation for modularity is the correlated evolution of traits as a response to environmental pressures. Harmon *et al.*, (2005)

discuss this idea of multidimensional convergence, and stress the distinction between the coordinated response of traits to a single environmental variable, and that of traits independently evolving in response to many environmental variables. To return to the example of *Anolis* lizards, Harmon *et al.*, (2005) found that the variation between traits within certain sets of characters (e.g. cranial and postcranial morphology) in these reptiles is independent of other character sets, implying some degree of the latter kind of multi-dimensional convergence. Within these structures, however, characters are evolving in concert, implying concerted convergence (Patterson & Givnish, 2002) in some subsets of characters.

Importantly, modularity has the potential to influence phylogenetic reconstruction. Goswami & Polly, (2010a) used morphometric methods to demonstrate that some modules in the cranium of carnivores are not correlated with phylogeny. Pairwise Procrustes distances were calculated for each module for a number of species, and each resulting distance matrix was compared to a molecular phylogenetic distance matrix using a matrix correlation analysis to test for correlation between module shape and phylogeny. Additional matrices for similarity of integration were generated and also compared with the phylogenetic patristic distance matrix to test for correlations between integration patterns and phylogeny. In feliforms, only the orbit shape was significantly correlated with phylogeny, and overall the basicranium and zygomatic-pterygoid modules had the strongest phylogenetic signal. Although morphological analyses of Carnivora largely rely on the basicranium, this reliance on only a few anatomical regions could pose a problem in cases where the characters relied upon are integrated but do not reflect evolutionary history.

2.2. Molecular Data

Although phylogenetic inference has historically used morphological information as its primary source of data, as DNA sequence data and greater computational power have become available over the last several decades,

the use of morphology in phylogenetics has been largely replaced by molecular systematics (see Hillis, 1987 for an early review). Molecular phylogenies often corroborate previously accepted relationships based on morphology (Springer *et al.*, 2004). This is unsurprising, since we expect evolutionary history to produce congruent patterns of morphological and molecular similarity. Sometimes, however, molecular data are at odds with previous morphological phylogenies, often demonstrating that phylogenies based on morphology alone have been mislead by phenotypic convergence (Hedges & Sibley, 1994; Van Tuinen *et al.*, 2001; Dunn *et al.*, 2008; Alter, Brown, & Stiassny, 2015; Peters *et al.*, 2018; Ran *et al.*, 2018; Mugleston *et al.*, 2018). Additionally, molecular data can often resolve relationships for which there is conflicting morphological support (Lee & Camens, 2009; Torres & van Tuinen, 2013; Dávalos *et al.*, 2014; Cruaud *et al.*, 2019). This may happen when some characters or modules appear to reflect one evolutionary history while others appear to reflect another.

There are several properties of molecular data which promote the assumption that they are more robust than morphological data for inferring evolutionary history. These include a) the high number of characters that are made available by DNA sequencing, leading to greater statistical power and the opportunity for complex statistical analysis involving a choice of models and tree-building methods, b) the relatively objective and unambiguous nature of molecular sequence data (i.e. the four nucleotides), c) the relative lack of evolutionary pressures acting on certain, neutral parts of the genome (Kimura, 1983), and d) the relative independence of molecular characters (Huelsenbeck & Nielsen, 1999). Thus, molecular data can act as an appropriate, independent benchmark against which to assess morphological data.

However, molecular data also suffer from biases and pitfalls. For example, recent work has suggested that genomes can be subject to some convergence, as in the cases of marine and echolocating mammals (Parker *et al.*, 2013; Foote *et al.*, 2015), squamate reptiles (Castoe *et al.*, 2009) and fish (Brown *et al.*, 2019). When undetected, genetic convergence has the potential to be a serious issue in the phylogenetic reconstruction of certain

groups. A further problem with genetic convergence is that it is likely to be accompanied by morphological convergence if the convergence is ecological or functional (Castoe *et al.*, 2009; Parker *et al.*, 2013; Foote *et al.*, 2015; Brown *et al.*, 2019) and thus it could be difficult to detect the convergence in these cases, and to subsequently tease out true signal reflecting the evolutionary history in these clades. In the case of squamates, an integrated analysis involving both morphology and molecules was able to resolve some of these conflicts (Reeder *et al.*, 2015).

Disagreement between individual genes as a result of, for example, incomplete lineage sorting or horizontal gene transfer presents further problems for phylogenetic analysis using molecular data. There are often disagreements between concatenated and coalescent trees (Lambert, Reeder, & Wiens, 2015).

The difficulties involved in the proper alignment of DNA sequences presents a further problem in molecular systematics as a result of insertions and deletions (indels) in genetic data, one that is analogous to the problem of correctly judging homology in morphological phylogenetics (Kumar & Filipski, 2007). The treatment of gaps produced by indels during multiple sequence alignment varies among studies, resulting in discordant tree topologies. There are also programmes that infer trees with unaligned sequences, by aligning the sequences and inferring the tree concurrently (Varón, Vinh, & Wheeler, 2010).

Tree building methods such as Maximum Likelihood (ML) and Bayesian inference require both phylogenetic data (i.e. DNA sequence data) and a realistic model of molecular evolution. Model misspecification is known to mislead tree topology (Rodríguez-Ezpeleta *et al.*, 2007; Naser-Khdour *et al.*, 2019) and evolutionary rates (Beaulieu *et al.*, 2015; dos Reis, Donoghue, & Yang, 2015) in molecular systematics, since correct parametrization is required in order for models to accurately reflect the nature of molecular evolution.

Systematic errors amplify with the number of characters, which in molecular systematics is usually very high, especially when compared with the characters used in morphological phylogenetics. High statistical support

even for erroneous results owing to incorrect model choice makes these errors difficult to spot (Rodríguez-Ezpeleta *et al.*, 2007). Although the use of the Akaike Information Criterion (Akaike, 1973) or some other comparison method, can reduce the risk of including too few or too many parameters, and can help to choose between models, comparative methods can only inform researchers how good a model is relative to other models. Therefore, it is very difficult to know whether a model is really 'correct' or even 'good' except in a relative sense.

Molecular phylogenies often disagree not only with morphological phylogenetics, but also with each other. Estimation of topology can differ depending on which part of the genome is sequenced, the substitution model used and the inference method employed, for example see recent disagreements regarding the phylogenetic placement of the comb jellies (Pisani *et al.*, 2015; Whelan *et al.*, 2015, 2017; Borowiec *et al.*, 2015; Simion *et al.*, 2017).

These sources of conflict within molecular systematics mean that, although molecular data is often considered more reliable than morphological data, we cannot trust it implicitly. We must therefore take care when using molecular data as a benchmark against which we measure the performance of morphology, and furthermore it is prudent, where possible, to use multiple independent molecular phylogenies.

3. Identifying and Addressing Phylogenetic Conflict

3.1 Molecular Versus Morphological Data

There are many examples of conflict between molecular and morphological data when constructing phylogenies, and in such cases the molecular data often usurp morphology (Sibley *et al.*, 1988; Van Tuinen *et al.*, 2001; Bibi, 2013; Torres & van Tuinen, 2013; Peters *et al.*, 2014; Reeder *et al.*, 2015; Alter *et al.*, 2015). Major sources of conflict between phylogenies reconstructed using morphological versus molecular data arise from the

differences in the way that molecules and morphology evolve. For example, phenotypes are under direct selection and are therefore prone to convergence and divergence, whereas large parts of the genome are thought to evolve neutrally (Kimura, 1983).

The relationship between molecules and morphology is complex (Orgogozo, Morizot, & Martin, 2015; Fisch, 2017), for reasons such as phenotypic plasticity and genetic dominance. Further, small genetic changes can often produce large phenotypic changes, for example, in the complex system of genetic switches and *hox* genes (genes which control the expression of other genes during development). This is illustrated by the fact that many characters or organs exhibit 'context insensitivity' (Musser & Wagner, 2015), whereby they retain their traits even if ectopically formed on different parts of the body (Halder, Callaerts, & Gehring, 1995), resulting in a substantial departure from ancestral bauplan with a relatively small genetic change. When occurring in nature, significant changes in phenotype arising from relatively subtle genetic events will produce the appearance of divergence. When used in phylogenetic analysis, this may make species appear erroneously unrelated. This is an opposite problem to the one of convergence, where distantly related species display similar phenotypes.

3.1.1. Convergence

The convergence of morphological traits is one of the most pervasive and well-understood problems in phylogenetic analysis, often occuring as a result of shared evolutionary regimes acting on traits of ecological or functional importance (Schluter, 2000) in response to biotic factors such as predation (Moody & Lozano-Vilano, 2018) and abiotic environmental factors (Moen & Wiens, 2017). Convergence is known to occur across the tree of life, from the molecular to organismal level, and at all taxonomic scales, for example at the generic level (Losos, 1998; Moody & Lozano-Vilano, 2018) up to parallel radiations at superordinal levels (Sibley *et al.*, 1988; Gheerbrant *et al.*, 2016). Convergence results in conflict between the

signal conveyed by traits under selection and the true underlying phylogeny, and can erroneously group unrelated taxa together according to phenotype. The introduction of molecular data, which are less prone to convergence, can reveal instances of convergence in morphological data and overcome this problem to some degree (Sibley *et al.*, 1988; Van Tuinen *et al.*, 2001; Dunn *et al.*, 2008; Alter *et al.*, 2015; Peters *et al.*, 2018; Ran *et al.*, 2018; Mugleston *et al.*, 2018).

A salient example involves the extraordinary phenotypic convergence between the tenrecs, particularly the lesser hedgehog tenrec *Echinops telfairi*, and the Laurasiatherian hedgehogs. These two families were formerly placed together in the order Insectivora, owing to shared characteristics such as a mobile proboscis and spines. Genomic evidence, however, places the tenrecs firmly within the Afrotheria (Stanhope *et al.*, 1998), closer to elephants and hyraxes. This highlights how misleading phenotypic resemblance can be, even when the resemblance is marked. Further, that the molecular evidence is in line with the fossil record and geographical distribution of tenrecs is a striking example of the importance of comparing and reconciling different lines of evidence to reach satisfactory phylogenetic conclusions.

A further example can be found in the relationships among aquatic birds, particularly the surprising phylogenetic placement of the diving grebes with wading flamingoes in the clade Mirandornithes based on molecular data (Van Tuinen *et al.*, 2001; Jarvis *et al.*, 2014). A phenotypically distinct group of diving birds, grebes had previously been placed with loons, another group of foot-propelled diving birds, based on morphological characteristics such as the form of the femur, humerus, tibia, ilium and palate (Cracraft, 1981; Livezey & Zusi, 2007). Mounting molecular evidence overturned this hypothesis, with later studies finding previously overlooked morphological (Mayr, 2004) and new fossil (Grellet-Tinner *et al.*, 2012) evidence supporting the monophyly of Mirandornithes. These synapomorphic characters uniting grebes and flamingoes, which include vertebral, plumage, egg and nest characteristics, highlight the potential role of hidden support in uncovering instances of convergence when some characters are congruent

with molecular data but not other morphological characters (Gatesy & Arctander, 2000; Gatesy *et al.*, 2003).

Other important examples include the placement of legless lizard family Dibamidae as sister to the other squamates (Pyron *et al.*, 2013; Reeder *et al.*, 2015), as opposed to hypotheses placing them with snakes or other squamate groups, and the modern placement of hyraxes, part of the sub-ungulate clade, within Afrotherian mammals based on molecular data (Seiffert, 2007), as opposed to their earlier unstable placement with Laurasiatherian ungulates.

In addition to simply using molecular data to either overturn or corroborate relationships based on morphology, convergence can be explicitly tested for using measures of homoplasy such as the retention index or excess index on cladistic data (Holland *et al.*, 2010; Blanke *et al.*, 2013) or with phylogenetically-corrected multivariate methods on morphometric data (Moen *et al.*, 2016; Thacker & Gkenas, 2019). For example, phylogenetic PCAs have detected convergence in measurements such as head length, toe tip size and foot webbing in frogs, variables each strongly correlated with microhabitat (Moen *et al.*, 2016).

3.1.2. Modularity, Integration and Correlation: Characters as Historical Entities Sharing Common Evolutionary Origins

The hierarchical and discontinuous pattern of physical form between organisms is widely discussed, and partially forms the basis of phylogenetic analysis when using morphology to inform topology. Somewhat less discussed in terms of phylogenetic implications is the evolutionary descent of characters and cell types *within* organisms (see Vickaryous & Hall, 2006; Arendt *et al.*, 2019 for reveiws on the evolution of cell types). It has been noted that morphological characters within organisms are entities which share common descent and should be discussed as such (Musser & Wagner, 2015; Liang *et al.*, 2018; Arendt *et al.*, 2019). This is because morphological characters within an organism exhibit a kind of homology that is distinct

from between-species homology; that of shared ancestry (i.e. related gene regulatory networks) and semi-independent evolution of cell types or organs. This produces the kind of hierarchical relationships similar and analogous to those observed between species (Figure 2).

This is important because it explains the 'species signal' (Musser & Wagner, 2015), whereby cell transcription profiles sometimes cluster by species rather than by tissue type (Pankey *et al.*, 2014; Lin *et al.*, 2014; Tschopp *et al.*, 2014), which we would not expect if cell types evolved wholly independently.

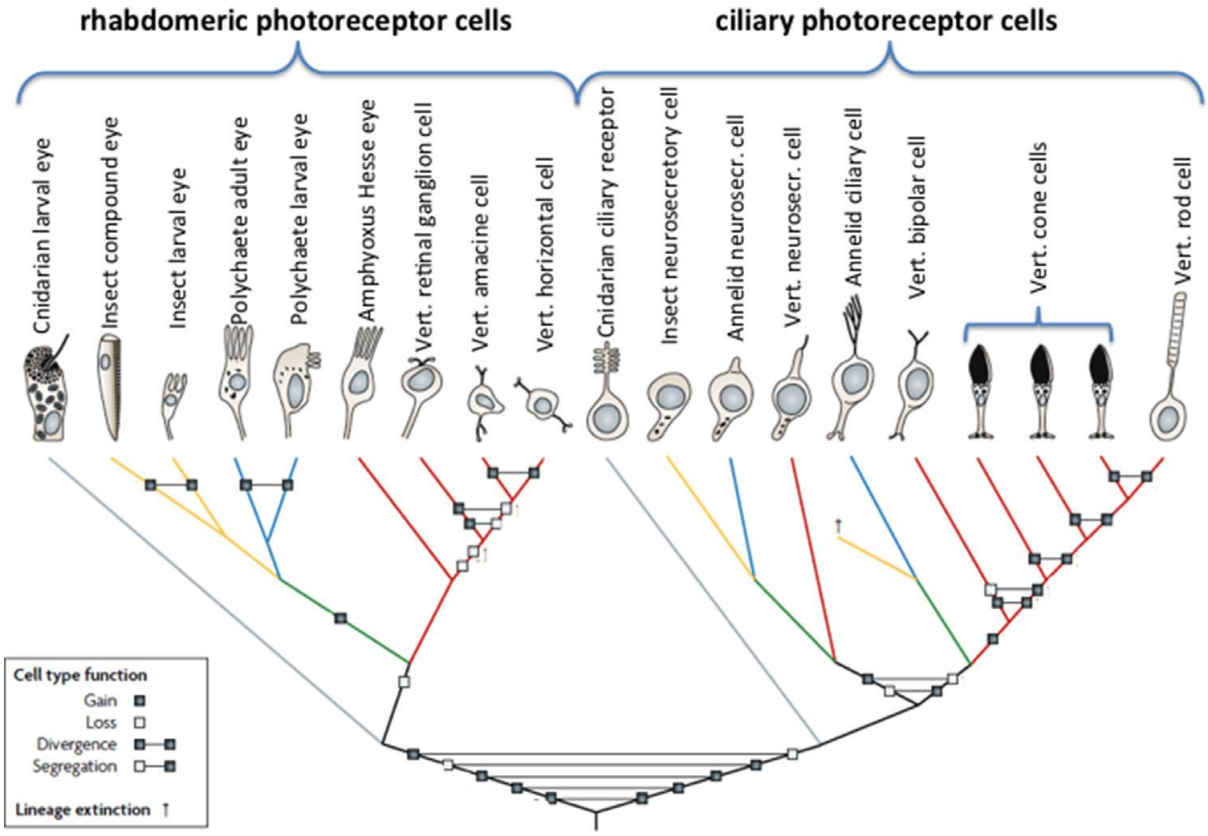


Figure. 2 *'Phylogeny' of photoreceptor cell types, from Wagner et al. 2015. This demonstrates how cells are similar to each other both by cell type and by taxonomic group. Cell types cluster by species as they are often similar because of common genetic networks that affect the development of multiple cells.*

Further, this strongly relates to the modularity observed in many morphometric studies, since if different characters share common gene regulatory networks, then the characters are not independent of each other.

Transcriptome studies have also cast doubt on the assessment of homology in morphological phylogenetics. For example, Wagner & Gauthier, (1999) used transcriptomes to address the conflict between embryology and palaeontology regarding the evolution of avian digits, and revealed that the digit in position 2 of the avian hand is more closely related to the digit at position 1 in the foot. If morphological characters are essentially repeated, since their development is controlled by related gene networks, but are treated as independent because they are spatially, functionally or superficially distinct, this results in a clear disparity between the genetic relatedness between species and the similarity as it may appear when observing particular morphological characters (Musser & Wagner, 2015). This is essentially a case of pseudoreplication, and results in correlated characters being treated as independent. In phylogenetic analysis, this can result in artificially inflated clade support, or spuriously close relationships between species.

There are a number of ways to test for character correlations in phylogenetic data, including the concentrated changes test (Maddison, 1990), the pairwise comparisons test (Read & Nee, 1995) and the character compatibility test (O'Keefe & Wagner, 2001). Briefly, the concentrated changes test calculates the probability that changes in one binary character are significantly concentrated in branches of the tree where a second character is in a given state. This involves calculating the number of ways that the number of observed losses and gains in the character can be distributed on the tree, summing the number of ways the observed gains and losses can be distributed on the branches where the second character is in the state of interest, and dividing the second by the first. This yields the probability that the character changes are more concentrated in particular branches than expected by chance. The pairwise comparisons test instead compares sister clades, and calculates the probability that the state of one character is seen more frequently in clades where a second character is more frequently in a particular state than their sister clades. The character compatibility test is a method that does not require a tree. Instead, two

characters are compatible, or correlated, if a tree exists onto which they can be plotted without homoplasy.

Recently, maximum likelihood and Bayesian approaches have been developed to detect correlated character pairs (Pagel, 1994; Beaulieu *et al.*, 2013), fitting models of trait evolution to character pairs. Additionally, clustering methods can be applied to pairwise metrics of the dissimilarity between characters, revealing instances of concerted evolution (Holland *et al.*, 2010; Blanke *et al.*, 2013).

Modularity and integration are necessary aspects of morphological evolution, as traits requiring coordinated responses to evolutionary pressures (such as parts of the skull or limbs) are often controlled by common genetic switches (Goswami & Polly, 2010a), a process known as pleiotropy (Cheverud, 1996; Klingenberg, 2008). The evolution of more complex body plans may require further fragmentation of parts into smaller modules in order to limit morphological constraints (see Figure 3 for an overview of integration and fragmentation events in the evolution of mammalian limbs and skull (Goswami *et al.*, 2014)). Thus, a hypothetical optimum between complete modularity and total independence of characters must be reached in order for characters to be able to optimally respond to selective pressures, with neither the burden of constraint nor the need for each character to evolve independently.

While some studies suggest that modularity and phylogeny are correlated (Goswami, 2006), modularity retains the potential to mislead analysis. A striking example of the effect of non-independence on phylogenetic analysis is in crocodiles. High levels of homoplasy have been noted in the group, particularly with respect to the gharials (Harshman *et al.*, 2003; Lee & Yates, 2018). Sadleir & Makovicky, (2008) note that this might be partly owing to the possibility of convergent modularity making some slender-snouted species, namely *Gavialis gangeticus* and *Tomistoma schlegelii* difficult to place. Paradoxically, however, these species are united by molecular data, and are not united by morphological data. A combined analysis (Gatesy *et al.*, 2003) groups them together, demonstrating the need for combined analyses in these cases in order to resolve paradoxical relationships such as

this. Sadleir & Makovicky, (2008) also stress the need to investigate whether characters correlating with skull shape mislead phylogenetic analysis in such cases. They go on to test for the existence of character correlations using all three methods discussed above (Maddison, 1990; Read & Nee, 1995; O'Keefe & Wagner, 2001) and conclude that transitions between crocodilian ecomorphs over the course of evolution affect the cranial characters used in morphological phylogenetics.

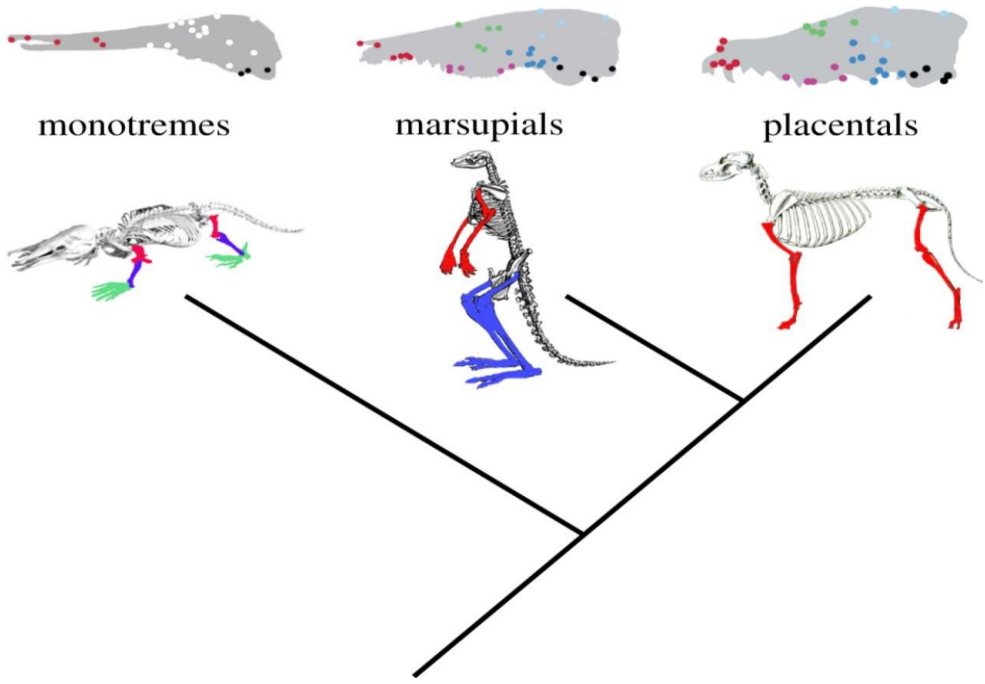


Figure. 3 *The development of character correlations, comparing patterns of modularity between monotremes, marsupials and placental mammals, from Goswami et al. 2014. Colours indicate separate, internally integrated modules. In monotremes, hindlimbs and forelimbs evolve in concert with each other (but with three separate modules comprising each limb) in contrast to marsupials, where forelimbs and hindlimbs are internally integrated but evolve independently. Placental limbs are highly integrated both internally and with each other. Placental and marsupial skulls have a higher number of separately evolving modules than monotreme skulls.*

Sanger *et al.*, (2012) demonstrated that in *Anolis* lizards, modules are not conserved over evolutionary time, and that patterns of integration, as well as skull shape itself, is convergent in these reptiles. Some *Anolis* skull shapes have exhibited the breakdown of modules or the evolution of new modules over relatively short timescales, again demonstrating the potential complications of character correlations on phylogenetic accuracy.

More recently, probabilistic methods have identified modularity in dental characters in placental mammals, and discussed its implications for phylogenetics (Billet & Bardin, 2019). Models of correlated and uncorrelated evolution were fitted to pairwise comparisons between the presence and absence of two traits on three successive molars in placental mammals. The best model was a correlated model that constrained the number of transition paths between pairs of states, thus supporting the use of composite coding in these characters in matrix construction for phylogenetic analysis.

Guillerme & Brazeau, (2018) additionally used simulations to assess the impact of correlations on tree topology. Character matrices ranging from 25-1000 characters were simulated, and from these three new groups of matrices were created: maximized matrices in which characters in the original simulated matrix were retained and duplicated so as to maximize differences between characters, minimized matrices in which characters were retained and duplicated so as to minimize differences between characters, and finally randomized matrices in which characters were retained and duplicated at random. Topologies were then inferred, using various methods, from these four classes of matrices and compared using Robinson-Foulds distances (Robinson & Foulds, 1981). As expected, matrices with minimized character differences, (i.e. high correlation) performed poorly at reconstructing the topology inferred from the original matrices compared with the other matrix classes. This directly demonstrates the impact of duplicated and correlated characters on topology, revealing the need for identifying and accounting for such characters in empirical data.

3.1.3. Rates of Evolutionary Change

The relative rates of evolutionary change over time, across lineages and between loci may impact the inferred topology of phylogenetic trees. In molecular data, fast-evolving characters are prone to homoplasy, while

slowly-evolving characters may convey less phylogenetic information, implying the existence of an optimal rate of sequence evolution for phylogenetic inference (Dornburg, Su, & Townsend, 2019). Sequences that have evolved over time to the extent that substitutions have, by chance, converged on the same nucleotides (known as long branch attraction or LBH) will appear erroneously closely together. The problem of LBH can be partly overcome by the addition of taxa to break up the longer branches or by excluding faster evolving sites from analysis (Bergsten, 2005).
Rapid phenotypic and molecular evolution can make it very difficult to resolve deep branches in some lineages (Whitfield & Kjer, 2008; Suh, 2016). For example in birds, uncertainty and conflict in phylogenetic hypotheses at the ordinal level probably reflect the rapid diversification of Neoaves at the K-Pg boundary (Suh, 2016). In some cases, fossils may be able to resolve these short branches, offering a temporal dimension to phylogenetic analysis that is not offered by molecular data or morphological data from extant species. That faster rates result in greater homoplasy opens avenues for the exploration of the relationship between the two.

3.2. Intra-Morphological Conflict

Morphological data are likely to share properties such as correlation and homoplasy within subsets, or partitions, which may result in conflicting topologies supported by these different data types. For example, dental data and osteological data are known to support different trees in mammals (Sansom, Wills, & Williams, 2017). Dental and other cranial characters are often affected by strong correlation by virtue of the ecological importance of mastication (Goswami *et al.*, 2011), developmental constraints arising from occlusion (Wolsan *et al.*, 2019), and also because individual teeth represent repeated structures, resulting in serial homology (Billet & Bardin, 2019). Further, soft and hard characters have different properties (Sansom & Wills, 2017). For example, plumage characters are evolutionary labile and

characterized by convergence (Omland & Lanyon, 2000; Hofmann, Cronin, & Omland, 2007).

It is therefore important to consider how to identify and treat classes of morphological characters when constructing matrices, or when applying statistical methods to these matrices, in phylogenetic reconstruction. Different coding strategies may be employed, for example treating correlated or serially homologous traits as single characters, known as composite coding (Torres-Montúfar, Borsch, & Ochoterena, 2018). This reduces pseudoreplication, but also overlooks information if structures are highly variable in some taxa. Alternatively, downweighting correlated characters would retain all relevant information while not treating them as independent, equal characters. Another possibility includes improving the modeling of morphological trait evolution in maximum likelihood and Bayesian analyses. The use and modeling of morphological characters in probabilistic methods of phylogeny reconstruction has recently been discussed in the literature in several contexts such as evolutionary rates (Wright, Lloyd, & Hillis, 2016), ontogeny (Bardin, Rouget, & Cecca, 2016) and serial homology (Billet & Bardin, 2019), and thus a discussion on the phylogenetic properties and treatment of such characters is timely.

Where there is conflict between morphological characters, subsets of these characters may nevertheless contain phylogenetic signal, recoverable as hidden support when consilient with molecular data (Lee, 2009; Reeder *et al.*, 2015).

4. Integration of Data for Better Analysis of Evolutionary History

A key question remains in phylogenetics whether morphological characters should be included in phylogenetic inference alongside molecular data (known as the 'total evidence' approach), or whether data types should be analysed separately, and subsequently character states can be mapped onto trees built from molecular data. Characters can be mapped onto trees after

having been used for inference, but the circularity of this method can limit its usefulness in some cases.

The total evidence approach was originally proposed as the best method for phylogenetic inference by Kluge, (1989), who suggests that gaining a consensus from all available data in one dataset is superior to the independent analysis of data, where a subsequently derived consensus from equally-weighted trees could be misleading. Eernisse & Kluge, (1993) compare both methods, and find that total evidence better reproduces the evolutionary history of amniotes, suggesting that the equal weighting of both morphological and molecular data could confound analysis.

The total evidence approach can be justified on statistical grounds in some cases. The simultaneous use of both fast- and slow evolving genes, for example, in a phylogenetic analysis, may help to resolve different parts of the tree (Hillis, 1987). The total evidence method has been used successfully to reconstruct bat echolocation (Springer *et al.*, 2001), and to resolve longstanding disagreements in crocodylian phylogeny (Gatesy *et al.*, 2003) and in squamate phylogeny (Reeder *et al.*, 2015). A problem with this method is that because molecular datasets usually contain many more (often orders of magnitude more) characters than morphological datasets, that any signal in the morphological data may simply be swamped by the molecular data. However, as was the case in the Gatesy *et al.*, (2003) analysis, morphological data can affect the outcome of analysis if its phylogenetic signal is strong enough. This is because it may interact with congruent signal in the molecular data, affirming one of several conflicting signals within the molecular data.

Miyamoto & Fitch, (1995) argue that instead of combined analysis, separate analyses of data into partitions might better resolve evolutionary history, since each partition represents an independent estimate of evolutionary history, and clades should be well-supported by all or most partitions to be considered real. A major problem with this method is determining how data should be partitioned. The partitioning of molecular data into fast and slow evolving loci, and the partitioning of morphological data into modules, or

into subsets based on some other criteria, may be a way to investigate this problem.

Dávalos *et al.*, (2014) used a combination of approaches to integrate morphological data with molecular data, demonstrating the existence not only of non-independence but also saturation in morphological data derived from Phyllostomidae (Mammalia: Chiroptera) fossils and extant species. They then extended the molecular scaffold approach used in many studies in such a way as to allow morphological data to influence topology only if the nodes generated from molecular data are ill supported. This seems to be a reasonable way to reconstruct evolutionary history using both molecular and morphological data, as the topology is mostly informed by the molecular data, unless the molecular signal is weak. The morphological data therefore acts as a reserve in the reconstruction of phylogeny where molecular data fails.

5. Discussion and Conclusion

There are conflicting signals in phylogenetic data. It is unclear what proportion of the data used in phylogenetic inference, whether morphological or molecular, correlates with phylogeny, what proportion is homoplastic or convergent, and what proportion is simply stochastic noise (e.g. Flynn and Nedbal 1998). Homoplasy in phylogenetic data has always plagued phylogenetic analysis, but the interesting question of what proportion of phylogenetic characters converge and why remains. Separating noise from phylogenetic or functional convergent signal is another problem in phylogenetics. However, we would expect convergent signal to at least be congruent with itself (Patterson & Givnish, 2002; Holland *et al.*, 2010; Blanke *et al.*, 2013), whereas random noise will on average not corroborate any strong existing signal.

Although it is widely agreed that molecular data provides more accurate estimates of phylogeny in general, the extent to and manner in which morphological data should be used to elucidate evolutionary history in the

age of genomic phylogenetics remains unclear and widely discussed (Scotland *et al.*, 2003; Wiens, 2004; Pisani *et al.*, 2007; Giribet, 2015; Lee & Palci, 2015; Wanninger, 2015). The use of fossils adds another dimension to phylogenetic analysis, as the unique character combinations provided by fossils can yield otherwise unobtainable insights (e.g. Seiffert, 2007). Although some would decry the assumption of superiority (Livezey, 2011), molecules seem to more faithfully track and therefore reproduce the evolutionary history of species, such as the timing and order of branching events, than do morphological characters. Although tree-building with genetic data is not without its important systematic errors, it is ultimately more objective, statistically powerful, and neutral than morphological data. Morphology, by contrast, is sometimes beset with such problems as functional convergence, rapid radiation, character non-independence and modularity that make it less reliable in terms of accurately reproducing phylogeny.

However, morphological data continues to be important in evolutionary biology as it alone can tell us about character changes over time, past character states, and phenotypic evolution. A recent example of the integration of phenotypic data into the study of evolutionary history has revealed the sequence of evolution of the beaks of Darwin's finches (Lamichhaney *et al.*, 2015), a classic example of phenotypic divergence. The authors scanned the genome to reveal genomic regions of increased divergence, exploring the relationships between genotype and phenotype. Further, the addition of even a small amount of morphological data can produce different topologies to trees built with molecular data alone (Gatesy *et al.*, 2003; Reeder *et al.*, 2015). These two quite different uses for morphology (i.e. to better resolve phylogenetic relationships, and to investigate phenotypic change over time) demonstrate the continuing importance of morphology in phylogenetics.

However, morphology is not always used as effectively as it might be; the correlated evolution of characters and the evolution of modularity can confound analysis if not taken into account during phylogenetic analysis (Sadleir & Makovicky, 2008; Goswami & Polly, 2010a; Guillerme & Brazeau,

2018; Billet & Bardin, 2019), although statistical power may be lost if groups of correlated characters are reduced to one character each. Utilizing more of the morphology of an organism may be a practicable way to circumvent this loss of power, such as the use of characters which can be gained by the use of new imaging techniques (Lee & Palci, 2015), or using continuous characters. This expansion of morphological information may make it easier to find characters that are phylogenetically useful.

Although not addressed here, I note that the added complexities of hybridisation, introgression and horizontal gene transfer may further confound analysis in some particular cases, as species may not be directly descended from one ancestor, but instead may contain genetic information and morphological traits from multiple, genetically distinct ancestors. This will result in discordance between gene trees. For example, attempts to resolve the deeper nodes in the 'tree of life' (i.e. the phylogeny of all life on earth) have probably been confused by gene transfer and gene fusion across lineages (Rivera & Lake, 2004).

Ultimately, the proper evaluation of major evolutionary events requires the concerted utilization of morphology (including fossils) and molecules, for insights into the evolution of body plans, the genetic changes and morphological innovations that thence arise, and ultimately the relationship between genotype and phenotype. Evolutionary history includes, of course, not only the pattern and timing of historical branching events but also genetic changes over time and corresponding changes in morphospace and niche occupation of species. Evolution is a dynamic and complex process of many parts, and as such, any examination into evolutionary history cannot be complete without the integration of genetic, fossil and morphological information.

6. Aims and Methods of the Thesis

Given the above context, and the ongoing discussion surrounding the use of morphological data, the following thesis aims to answer questions regarding

the nature of morphological evolution and patterns, particularly with regards to the partitioning of subsets of data. There are a priori reasons to suspect that different morphological characters display different cladistic patterns, including the existence of different evolutionary pressures acting on different aspects of organismal form. These differences arise from ecological, functional or genetic relatedness between characters of the same type, which are divergent between character types. For example, as discussed above, soft and hard characters (Sansom & Wills, 2017), dental and osteological characters (Sansom *et al.*, 2017) and cranial and postcranial (Mounce, Sansom, & Wills, 2016) all display differing levels of homoplasy and imply different trees. Cranial and dental characters tend to be prone to modularity and correlation owing to occlusion, trophic convergence and serial homology (Goswami, 2007; Sadleir & Makovicky, 2008; Goswami & Polly, 2010b; Kelley & Motani, 2015; Billet & Bardin, 2019), while some soft characters are more homoplasious (Price, Friedman, & Omland, 2007).

6.1. Approach and Methods Overview

The overarching themes of this thesis therefore involve a) evaluating the properties of partitions of morphology in tetrapods, and b) comparing these data against molecular topologies. Specifically, I test for differences in homoplasy, transition ages, tree-dependent correlation and tree-independent correlation between character types in tetrapod morphological datasets. I interrogate differences in these areas between osteological, soft, cranial, postcranial and dental characters. I use a variety of computational phylogenetic methods including the calculation of consistency of characters on molecular trees, character state reconstructions, detection of correlated character pairs using probabilistic methods, dataset and tree simulation, identification of cliques of internally-consistent characters, cluster analysis and phylogenetic comparative methods.

In all analyses, I aim to use the broadest amount of data possible. To this end, I use datasets spanning birds, squamates and mammals. Further, I use

entire datasets where possible, as opposed to subsets of the data. In this way, I investigate the behavior and properties of morphological data in vertebrates.

I do not discriminate between datasets or remove characters on the basis of data quality for 2 reasons: 1. There is no objective way to do this, and I wish to keep my analyses as free of bias as possible, and 2. I aim in part to evaluate not just the nature of morphology, but to evaluate the behaviour of morphological datasets actually published and used for analysis.

For avian datasets and squamate datasets, I systematically searched the literature for morphological matrices (further details on data collection in chapter 1). For analyses of mammalian data, I used the datasets collected by Sansom *et al.*, (2016). For comparison with molecular data, I primarily used the topologies and ages of Jetz *et al.*, (2012) for birds, Zheng & Wiens, (2016) for squamates, and the new tree inferred by Upham, Esselstyn, & Jetz, (2019) for mammals. I additionally use alternative trees for comparison for some analyses.

6.2. Format and Content of Results Chapters

This thesis is presented in journal format. I have chosen to present my work this way as several of my chapters are suitable for and will be submitted for publication, and journal format therefore better reflects the content and aims of my work than traditional format. One of the 4 data chapters has been submitted for publication, with the remaining 3 chapters ready to submit. All chapters are presented in pre-publication, manuscript form. The rest of this section deals with the content and author contribution for each chapter. Briefly, chapter 1 deals with homoplasy and transition ages of character partitions and subpartitions in sauropsids (birds and reptiles), using molecular trees as a benchmark. Chapter 2 deals with the correlation of pairs of characters within and between osteological and non-osteological partitions in sauropsids, again comparing against the molecular data. As this approach relies on a reference tree, this may be termed external correlation.

Chapter 3 deals with the internal correlation, or consistency, of osteological and non-osteological characters in sauropsid datasets, termed cliques, independent of an external topology. In chapter 4 I use the methods of the previous chapters, this time comparing dental with osteological data in mammalian morphological datasets.

6.2.1. Osteological Characters Show Greater Congruence with Molecular Phylogenies than Soft Characters in Avian and Reptilian Morphological Datasets

In this chapter I split 28 avian and squamate morphological datasets into osteological and non-osteological data, and further into cranial, postcranial, integumentary and myological data. I use phylogenetic methods to establish whether a) different partitions of morphological data display different levels of congruence with molecular trees, and b) these partitions transition at different ages on the molecular trees. I do this by computing the retention index for each character on each molecular tree, averaged over 1,000 trees from the posterior distribution of the Jetz *et al.*, (2012) avian phylogeny, and the single Zheng & Wiens, (2016) squamate topology. I additionally computed the ensemble retention indices (i.e., the retention index for a group of characters) for partitions and subpartitions. I compared ensemble retention indices with a t test, and individual character retention indices with both Mann-Whitney *U* tests within datasets and linear mixed effects models including all datasets. I estimate transition ages by performing ancestral state reconstructions on molecular trees, and simulating transition ages for each character. I calculate average transition ages for each partition and compare individual character transition ages using linear mixed effects models. I additionally test whether homoplasy or partition has greater explanatory power for the transition ages.

I performed all data analysis with advice and guidance from my primary supervisor, particularly receiving help for coding in the phylogenetics

program TNT (Goloboff, Farris, & Nixon, 2008). I additionally received advice on statistical analysis using linear mixed effects models from Dr Chris Knight. I produced the text with input from my primary supervisor. I also received helpful comments from two anonymous reviewers, who provided suggestions on data analysis and helped me to improve the clarity of my manuscript. Since I have completed the revisions, I will resubmit this chapter for publication.

6.2.2. The Prevalence of Correlated Morphological Characters and their Effect on Phylogenetic Reconstruction

In this chapter I compare correlated and uncorrelated models of character evolution between all pairs of binary morphological characters in 11 avian and squamate datasets. I compare models using the AICc and computing the AICc weights. I calculate the average AICc weight for character pairs for each partition in each dataset, as well as the average between partition and within partition character pairs. I additionally calculate the proportion of correlated characters, using 0.95 AICc weight as a threshold. I compare AICc weights between individual character pairs using linear mixed effects models. I additionally perform a sensitivity analysis on simulated data to test how robust these methods are to underlying data structure and amount of data present in each dataset.

I performed all data analysis with advice and guidance from my primary supervisor. I received help in parallelizing my scripts in R and using the Computational Shared Facility from Dr Joseph Keating and the IT services Support Centre at the University of Manchester, as well as some additional advice on the use of the R package corHMM (Beaulieu *et al.*, 2013), used for fitting the correlated and uncorrelated models, from Professor Jeremy Beaulieu. I wrote the chapter. This chapter has not yet been through peer-review, but is ready for submission.

6.2.3. Cliques of Morphological Characters in Avian and Squamate Datasets

In this chapter I look at the internal consistency of morphological characters. For all datasets, I calculate the size of the largest internally consistent group of characters, and compare this against a null distribution of shuffled datasets. I further infer pairwise dissimilarity matrices for these characters for each dataset, and perform cluster analyses on these to create trees where characters are clustered by similarity. I perform phylogenetic comparative methods to establish whether a) characters cluster by osteological and non-osteological partitions, and b) whether characters cluster by their fit on molecular trees.

I performed all data analysis with advice and guidance from my primary supervisor. I was provided with Python scripts for the calculation of pairwise excess indices, the shuffling of matrices and calculation of maximum clique size, as well as advice on using Python, from Professor Barbara Holland. I received advice on the use of comparative methods using Phytools (Revell, 2012) from Professor Liam Revell. I wrote the chapter. This chapter has not yet been through peer-review, but is ready for submission.

6.2.4. Convergence, Correlations and Cliques: An Analysis of the Relative Performance of Morphological Character Partitions in Mammals

In this chapter I perform many of the above analyses, including calculating consistency on molecular trees, estimating relative transition ages, fitting correlated and uncorrelated models of character pair evolution, and cluster analyses on mammalian datasets, splitting the data into dental and osteological partitions. In addition, I infer trees for one large dataset from the dental, osteological, and combined data to directly examine phylogenetic

signal contained within these partitions, and compare these with tree-distance metrics.

I performed all data analysis with advice and guidance from my primary supervisor. I wrote the chapter. This chapter has not yet been through peer-review, but is ready for submission.

References

Akaike H. 1973. Maximum Likelihood Identification of Gaussian Autoregressive Moving Average Models. *Biometrika* 60: 255.

Alter SE, Brown B, Stiassny MLJ. 2015. Molecular phylogenetics reveals convergent evolution in lower Congo River spiny eels. *BMC Evolutionary Biology* 15: 224.

Arendt D, Bertucci PY, Achim K, Musser JM. 2019. Evolution of neuronal types and families. *Current Opinion in Neurobiology* 56: 144–152.

Asher RJ, Smith MR, Rankin A, Emry RJ. 2019. Congruence, fossils and the evolutionary tree of rodents and lagomorphs. *Royal Society Open Science* 6: 190387.

Bardin J, Rouget I, Cecca F. 2016. Ontogenetic data analyzed as such in phylogenies. *Systematic Biology* 66: 23–37.

Barraclough TG. 2010. Evolving entities: towards a unified framework for understanding diversity at the species and higher levels. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365: 1801–1813.

Beaulieu JM, O'Meara BC, Crane P, Donoghue MJ. 2015. Heterogeneous rates of molecular evolution and diversification could explain the triassic age estimate for angiosperms. *Systematic Biology* 64: 869–878.

Beaulieu JM, O'Meara BC, Donoghue MJ. 2013. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. *Systematic Biology* 62: 725–737.

Benton MJ. 1999. Early origins of modern birds and mammals: Molecules vs. morphology. *BioEssays*.

Bergsten J. 2005. A review of long-branch attraction. *Cladistics* 21: 163–193.

Beutel RG, Friedrich F, Hörnschemeyer T, Pohl H, Hünefeld F, Beckmann F, Meier R, Misof B, Whiting MF, Vilhelmsen L. 2011. Morphological and molecular evidence converge upon a robust phylogeny of the megadiverse Holometabola. *Cladistics* 27: 341–355.
Bibi F. 2013. A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics. *BMC Evolutionary Biology* 13: 166.
Billet G, Bardin J. 2019. Serial homology and correlated characters in morphological phylogenetics: modeling the evolution of dental crests in Placentals. *Systematic Biology* 68: 267–280.
Blanke A, Greve C, Wipfler B, Beutel RG, Holland BR, Misof B, Forschungsmuseum Z, Koenig A. 2013. The identification of concerted convergence in insect heads corroborates Palaeoptera. *Systematic Biology* 13: 250–263.
Borowiec ML, Lee EK, Chiu JC, Plachetzki DC. 2015. Extracting phylogenetic signal and accounting for bias in whole-genome data sets supports the Ctenophora as sister to remaining Metazoa. *BMC Genomics* 16: 987.
Borsch T, Flores-Olvera H, Zumaya S, Müller K. 2018. Pollen characters and DNA sequence data converge on a monophyletic genus *Iresine* (Amaranthaceae, Caryophyllales) and help to elucidate its species diversity. *Taxon* 67: 944–976.
Brown AP, McGowan KL, Schwarzkopf EJ, Greenway R, Rodriguez LA, Tobler M, Kelley JL. 2019. Local ancestry analysis reveals genomic convergence in extremophile fishes. *Philosophical Transactions of the Royal Society B: Biological Sciences* 374: 20180240.
Button DJ, Barrett PM, Rayfield EJ. 2017. Craniodental functional evolution in sauropodomorph dinosaurs. *Paleobiology* 43: 435–462.
Cartmill M. 1994. A critique of homology as a morphological concept. *American Journal of Physical Anthropology* 94: 115–123.
Castoe TA, de Koning APJ, Kim HM, Gu W, Noonan BP, Naylor G, Jiang ZJ, Parkinson CL, Pollock DD. 2009. Evidence for an ancient adaptive episode of convergent molecular evolution. *Proceedings of the National Academy of Sciences* 106: 8986–8991.

Cheverud JM. 1996. Developmental Integration and the Evolution of Pleiotropy. *American Zoologist* 36: 44–50.
Cracraft J. 1981. Toward a Phylogenetic Classification of the Recent Birds of the World (Class Aves). *The Auk* 98: 681–714.
Cruaud A, Delvare G, Nidelet S, Sauné L, Ratnasingham S, Chartois M, Blaimer BB, Gates M, Brady SG, Faure S, Noort S van, Rossi JP, Rasplus JY. 2019. Ultra-Conserved Elements and morphology reciprocally illuminate conflicting phylogenetic hypotheses in Chalcididae (Hymenoptera, Chalcidoidea). *bioRxiv*: 761874.
Dávalos LM, Velazco PM, Warsi OM, Smits PD, Simmons NB. 2014. Integrating incomplete fossils by isolating conflicting signal in saturated and non-independent morphological characters. *Systematic Biology* 63: 582–600.
Davis AM, Betancur-R R. 2017. Widespread ecomorphological convergence in multiple fish families spanning the marine–freshwater interface. *Proceedings of the Royal Society B: Biological Sciences* 284: 20170565.
Dornburg A, Su Z, Townsend JP. 2019. Optimal Rates for Phylogenetic Inference and Experimental Design in the Era of Genome-Scale Data Sets (R Mueller, Ed.). *Systematic Biology* 68: 145–156.
Dumont ER, Samadevam K, Grosse I, Warsi OM, Baird B, Davalos LM. 2014. Seelction for mechanical advantage underlies multiple cranial optima in new world leaf-nosed bats. *Evolution* 68: 1436–1449.
Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, Sørensen M V., Haddock SHD, Schmidt-Rhaesa A, Okusu A, Kristensen RM, Wheeler WC, Martindale MQ, Giribet G. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452: 745–749.
Eernisse DJ, Kluge AG. 1993. Taxonomic congruence versus total evidence, and amniote phylogeny inferred from fossils, molecules, and morphology. *Molecular Biology and Evolution* 10: 1170–1195.
Emerson SB, Hastings PA. 1998. Morphological correlations in evolution: consequences for phylogenetic analysis. *The Quarterly Review of Biology* 73: 141–162.
Felice RN, Randau M, Goswami A. 2018. A fly in a tube: Macroevolutionary

expectations for integrated phenotypes. *Evolution* 72: 2580–2594.
Field DJ, Gauthier JA, King BL, Pisani D, Lyson TR, Peterson KJ. 2014. Toward consilience in reptile phylogeny: microRNAs support an archosaur, not a lepidosaur affinity for turtles. *Evolution & Development* 16: 189–196.
Fisch GS. 2017. Whither the genotype-phenotype relationship? An historical and methodological appraisal. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* 175: 343–353.
Flynn JJ, Nedbal MA. 1998. Phylogeny of the Carnivora (Mammalia): Congruence vs Incompatibility among Multiple Data Sets. *Molecular Phylogenetics and Evolution* 9: 414–426.
Foote AD, Liu Y, C Thomas GW, Vinař T, Alföldi J, Deng J, Dugan S, van Elk CE, Hunter ME, Joshi V, Khan Z, Kovar C, Lee SL, Lindblad-Toh K, Mancia A, Nielsen R, Qin X, Qu J, Raney BJ, Vijay N, W Wolf JB, Hahn MW, Muzny DM, Worley KC, Thomas Gilbert MP, Gibbs RA. 2015. Convergent evolution of the genomes of marine mammals. *Nature Genetics* 47: 272–275.
Friedman ST, Price SA, Hoey AS, Wainwright PC. 2016. Ecomorphological convergence in planktivorous surgeonfishes. *Journal of Evolutionary Biology* 29: 965–978.
Gatesy J, Amato G, Norell M, Desalle R, Hayashi C. 2003. Combined Support for Wholesale Taxic Atavism in Gavialine Crocodylians. *Syst. Biol* 52: 403–422.
Gatesy J, Arctander P. 2000. Hidden Morphological Support for the Phylogenetic Placement of Pseudoryx nghetinhensis with Bovine Bovids: A Combined Analysis of Gross Anatomical Evidence and DNA Sequences from Five Genes (M Westneat, Ed.). *Systematic Biology* 49: 515–538.
Gheerbrant E, Filippo A, Schmitt A. 2016. Convergence of Afrotherian and Laurasiatherian ungulate-like mammals: first morphological evidence from the Paleocene of Morocco. *PLOS ONE* 11: e0157556.
Giribet G. 2015. Morphology should not be forgotten in the era of genomics-a phylogenetic perspective. *Zoologischer Anzeiger* 256: 96–103.
Goloboff PA, Farris JS, Nixon KC. 2008. TNT, a free program for phylogenetic analysis. *Cladistics* 24: 774–786.
Goswami A. 2006. Morphological integration in the carnivoran skull.

Evolution 60: 169–183.
Goswami A. 2007. Phylogeny, diet, and cranial integration in australodelphian marsupials. *PLoS ONE* 2: e995.
Goswami A, Milne N, Wroe S. 2011. Biting through constraints: cranial morphology, disparity and convergence across living and fossil carnivorous mammals. *Proceedings of the Royal Society B: Biological Sciences* 278: 1831–1839.
Goswami A, Polly DP. 2010a. The influence of character correlations on phylogenetic analyses: a case study of the carnivoran cranium. In: Goswami A,, In: Friscia A, eds. *Carnivoran Evolution: New Views on Phylogeny, Form, and Function*. Cambridge: Cambridge University Press, 141–164.
Goswami A, Polly PD. 2010b. The influence of modularity on cranial morphological disparity in Carnivora and Primates (Mammalia). *PLoS ONE* 5: e9517.
Goswami A, Smaers JB, Soligo C, Polly PD. 2014. The macroevolutionary consequences of phenotypic integration: from development to deep time. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369: 20130254–20130254.
Grellet-Tinner G, Murelaga X, Larrasoaña JC, Silveira LF, Olivares M, Ortega LA, Trimby PW, Pascual A. 2012. The First Occurrence in the Fossil Record of an Aquatic Avian Twig-Nest with Phoenicopteriformes Eggs: Evolutionary Implications (LA Newsom, Ed.). *PLoS ONE* 7: e46972.
Guillerme T, Brazeau MD. 2018. Influence of different modes of morphological character correlation on phylogenetic tree inference. *bioRxiv*: 308742.
Hadfield JD. 2016. The spatial scale of local adaptation in a stochastic environment (O Ovaskainen, Ed.). *Ecology Letters* 19: 780–788.
Halder G, Callaerts P, Gehring W. 1995. Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. *Science* 267: 1788–1792.
Harmon LJ, Kolbe JJ, Cheverud JM, Losos JB. 2005. Convergence and the multidimensional niche. *Evolution* 59: 409–421.
Harshman J, Huddleston CJ, Bollback JP, Parsons TJ, Braun MJ. 2003. True and False Gharials: A Nuclear Gene Phylogeny of Crocodylia (A Baker, Ed.).

Systematic Biology 52: 386–402.
Hedges SB, Sibley CG. 1994. Molecules vs. morphology in avian evolution: the case of the 'pelecaniform' birds. *Proceedings of the National Academy of Sciences of the United States of America* 91: 9861–9865.
Hetherington AJ, Sherratt E, Ruta M, Wilkinson M, Deline B, Donoghue PCJ. 2015. Do cladistic and morphometric data capture common patterns of morphological disparity? *Palaeontology*.
Hillis DM. 1987. Molecular Versus Morphological Approaches to Systematics. *Annual Review of Ecology and Systematics* 18: 23–42.
Hofmann CM, Cronin TW, Omland KE. 2007. Melanin coloration in New World orioles II: ancestral state reconstruction reveals lability in the use of carotenoids and phaeomelanins. *Journal of Avian Biology* 38: 172–181.
Holland BR, Spencer HG, Worthy TH, Kennedy M. 2010. Identifying cliques of convergent characters: concerted evolution in the cormorants and shags. *Systematic Biology* 59: 433–445.
Huelsenbeck JP, Nielsen R. 1999. Effect of nonindependent substitution on phylogenetic accuracy. *Systematic Biology* 48: 317–328.
Humphreys AM, Barraclough TG. 2014. The evolutionary reality of higher taxa in mammals. *Proceedings of the Royal Society B: Biological Sciences* 281: 20132750.
Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SYW, Faircloth BC, Nabholz B, Howard JT, Suh A, Weber CC, da Fonseca RR, Li J, Zhang F, Li H, Zhou L, Narula N, Liu L, Ganapathy G, Boussau B, Bayzid MS, Zavidovych V, Subramanian S, Gabaldón T, Capella-Gutiérrez S, Huerta-Cepas J, Rekepalli B, Munch K, Schierup M, Lindow B, Warren WC, Ray D, Green RE, Bruford MW, Zhan X, Dixon A, Li S, Li N, Huang Y, Derryberry EP, Bertelsen MF, Sheldon FH, Brumfield RT, Mello C V., Lovell P V., Wirthlin M, Schneider MPC, Prosdocimi F, Samaniego JA, Velazquez AMV, Alfaro-Núñez A, Campos PF, Petersen B, Sicheritz-Ponten T, Pas A, Bailey T, Scofield P, Bunce M, Lambert DM, Zhou Q, Perelman P, Driskell AC, Shapiro B, Xiong Z, Zeng Y, Liu S, Li Z, Liu B, Wu K, Xiao J, Yinqi X, Zheng Q, Zhang Y, Yang H, Wang J, Smeds L, Rheindt FE, Braun M, Fjeldsa J, Orlando L, Barker FK, Jønsson KA, Johnson W, Koepfli KP, O'Brien S, Haussler D, Ryder OA, Rahbek C, Willerslev E,

Graves GR, Glenn TC, McCormack J, Burt D, Ellegren H, Alström P, Edwards S V., Stamatakis A, Mindell DP, Cracraft J, Braun EL, Warnow T, Jun W, Gilbert MTP, Zhang G. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346.

Jenner RA. 2004. Points of View Accepting Partnership by Submission? Morphological Phylogenetics in a Molecular Millennium. *Syst. Biol* 53: 333–342.

Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO. 2012. The global diversity of birds in space and time. *Nature* 491: 444–448.

Kelley NP, Motani R. 2015. Trophic convergence drives morphological convergence in marine tetrapods. *Biology Letters* 11: 20140709.

Kimura M. 1983. *The neutral theory of molecular evolution*. Cambridge: Cambridge University Press.

Klingenberg CP. 2008. Morphological integration and developmental modularity. *Annual Review of Ecology, Evolution, and Systematics* 39: 115–132.

Kluge AG. 1989. A Concern for Evidence and a Phylogenetic Hypothesis of Relationships among Epicrates (Boidae, Serpentes). *Systematic Biology* 38: 7–25.

Kumar S, Filipski A. 2007. Multiple sequence alignment: In pursuit of homologous DNA positions. *Genome Research* 17: 127–135.

Lambert SM, Reeder TW, Wiens JJ. 2015. When do species-tree and concatenated estimates disagree? An empirical analysis with higher-level scincid lizard phylogeny. *Molecular Phylogenetics and Evolution* 82: 146–155.

Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, Promerová M, Rubin CJ, Wang C, Zamani N, Grant BR, Grant PR, Webster MT, Andersson L. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing.

Lee MSY. 2009. Hidden support from unpromising data sets strongly unites snakes with anguimorph 'lizards'. *Journal of Evolutionary Biology* 22: 1308–1316.

Lee MSY, Camens AB. 2009. Strong morphological support for the molecular

evolutionary tree of placental mammals. *Journal of Evolutionary Biology*.
Lee MSY, Palci A. 2015. Morphological phylogenetics in the genomic age. *Current Biology* 25: R922–R929.
Lee MSY, Yates AM. 2018. Tip-dating and homoplasy: reconciling the shallow molecular divergences of modern gharials with their long fossil record. *Proceedings of the Royal Society B: Biological Sciences* 285: 20181071.
Leslie AB, Beaulieu JM, Crane PR, Knopf P, Donoghue MJ. 2015. Integration and macroevolutionary patterns in the pollination biology of conifers. *Evolution* 69: 1573–1583.
Liang C, Musser JM, Cloutier A, Prum RO, Wagner GP. 2018. Pervasive Correlated Evolution in Gene Expression Shapes Cell and Tissue Type Transcriptomes. *Genome Biology and Evolution* 10: 538–552.
Lin S, Lin Y, Nery JR, Urich MA, Breschi A, Davis CA, Dobin A, Zaleski C, Beer MA, Chapman WC, Gingeras TR, Ecker JR, Snyder MP. 2014. Comparison of the transcriptional landscapes between human and mouse tissues. *Proceedings of the National Academy of Sciences of the United States of America* 111: 17224–9.
Livezey BC. 2011. Grebes and flamingos: standards of evidence, adjudication of disputes, and societal politics in avian systematics. *Cladistics* 27: 391–401.
Livezey BC, Zusi RL. 2007. Higher-order phylogeny of modern birds (Theropoda, Aves: Neornithes) based on comparative anatomy. II. Analysis and discussion. *Zoological Journal of the Linnean Society*.
Losos JB. 1998. Contingency and determinism in replicated adaptive radiations of island lizards. *Nature* 279: 2115–2118.
Losos JB. 2011. Convergence, adaptation and constraint. *Evolution* 65: 1827–1840.
Maddison WP. 1990. A method for testing the correlated evolution of two binary characters: Are gains or losses Ccncentrated on certain branches of a phylogenetic tree? *Evolution* 44: 539.
Mayr G. 2004. Morphological evidence for sister group relationship between flamingos (Aves: Phoenicopteridae) and grebes (Podicipedidae). *Zoological Journal of the Linnean Society* 140: 157–169.

McInerney JO, O'Connell MJ, Pisani D. 2014. The hybrid nature of the Eukaryota and a consilient view of life on Earth. *Nature Reviews Microbiology* 12: 449–455.
Mitteroecker P, Huttegger SM. 2009. The Concept of Morphospaces in Evolutionary and Developmental Biology: Mathematics and Metaphors. *Biological Theory* 4: 54–67.
Miyamoto MM, Fitch WM. 1995. Testing species phylogenies and phylogenetic methods with congruence. *Systematic Biology*.
Moen DS, Morlon H, Wiens JJ. 2016. Testing convergence versus history: convergence dominates phenotypic evolution for over 150 million years in frogs. *Systematic Biology* 65: 146–160.
Moen DS, Wiens JJ. 2017. Microhabitat and Climatic Niche Change Explain Patterns of Diversification among Frog Families. *The American Naturalist* 190.
Moody EK, Lozano-Vilano ML. 2018. Predation drives morphological convergence in the *Gambusia panuco* species group among lotic and lentic habitats. *Journal of Evolutionary Biology* 31: 491–501.
Mounce RCP, Sansom R, Wills MA. 2016. Sampling diverse characters improves phylogenies: craniodental and postcranial characters of vertebrates often imply different trees. *Evolution* 70: 666–686.
Mugleston JD, Naegle M, Song H, Whiting MF. 2018. A Comprehensive Phylogeny of Tettigoniidae (Orthoptera: Ensifera) Reveals Extensive Ecomorph Convergence and Widespread Taxonomic Incongruence. *Insect Systematics and Diversity* 2.
Musser JM, Wagner GP. 2015. Character trees from transcriptome data: Origin and individuation of morphological characters and the so-called 'species signal'. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*.
Naser-Khdour S, Minh BQ, Zhang W, Stone E, Lanfear R. 2019. The Prevalence and Impact of Model Violations in Phylogenetics Analysis. *bioRxiv*: 460121.
O'Keefe FR, Wagner PJ. 2001. Inferring and Testing Hypotheses of Cladistic Character Dependence by Using Character Compatibility (J Wiens, Ed.).

Systematic Biology 50: 657–675.
Omland KE, Lanyon SM. 2000. Reconstructing plumage evolution in orioles (Icterus): Repeated convergence and reversal in patterns. *Evolution* 54: 2119–2133.
Orgogozo V, Morizot B, Martin A. 2015. The differential view of genotype–phenotype relationships. *Frontiers in Genetics* 6.
Pagel M. 1994. Detecting Correlated Evolution on Phylogenies: A General Method for the Comparative Analysis of Discrete Characters. *Proceedings of the Royal Society B: Biological Sciences* 255: 37–45.
Pankey MS, Minin VN, Imholte GC, Suchard MA, Oakley TH. 2014. Predictable transcriptome evolution in the convergent and complex bioluminescent organs of squid. *Proceedings of the National Academy of Sciences of the United States of America* 111: E4736-42.
Parker J, Tsagkogeorga G, Cotton JA, Liu Y, Provero P, Stupka E, Rossiter SJ. 2013. Genome-wide signatures of convergent evolution in echolocating mammals.
Patterson TB, Givnish TJ. 2002. Phylogeny, concerted convergence, and phylogenetic niche conservatism in the core Liliales: insights from rbcL and ndhF sequence data. *Evolution* 56: 233–252.
Peters RS, Meusemann K, Petersen M, Mayer C, Wilbrandt J, Ziesmann T, Donath A, Kjer KM, Aspöck U, Aspöck H, Aberer A, Stamatakis A, Friedrich F, Hünefeld F, Niehuis O, Beutel RG, Misof B. 2014. The evolutionary history of holometabolous insects inferred from transcriptome-based phylogeny and comprehensive morphological data. *BMC Evolutionary Biology* 14: 52.
Peters RS, Niehuis O, Gunkel S, Bläser M, Mayer C, Podsiadlowski L, Kozlov A, Donath A, van Noort S, Liu S, Zhou X, Misof B, Heraty J, Krogmann L. 2018. Transcriptome sequence-based phylogeny of chalcidoid wasps (Hymenoptera: Chalcidoidea) reveals a history of rapid radiations, convergence, and evolutionary success. *Molecular Phylogenetics and Evolution* 120: 286–296.
Pisani D, Benton MJ, Wilkinson M. 2007. Congruence of morphological and molecular phylogenies. *Acta Biotheoretica*.
Pisani D, Pett W, Dohrmann M, Feuda R, Rota-Stabelli O, Philippe H, Lartillot

N, Wörheide G. 2015. Genomic data do not support comb jellies as the sister group to all other animals. *Proceedings of the National Academy of Sciences* 112.
Price JJ, Friedman NR, Omland KE. 2007. Song and plumage evolution in the New World orioles (Icterus) show similar lability and convergence in patterns. *Evolution* 61: 850– 863.
Pyron R, Burbrink FT, Wiens JJ, Anisimova M, Hordijk W, Gascuel O, Valentin F, Wallace I, Wilm A, Lopez R. 2013. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evolutionary Biology* 13: 93.
de Queiroz K. 1996. Including the Characters of Interest During Tree Reconstruction and the Problems of Circularity and Bias in Studies of Character Evolution. *The American Naturalist* 148: 700–708.
Ran JH, Shen TT, Wang MM, Wang XQ. 2018. Phylogenomics resolves the deep phylogeny of seed plants and indicates partial convergent or homoplastic evolution between Gnetales and angiosperms. *Proceedings of the Royal Society B: Biological Sciences* 285: 20181012.
Read AF, Nee S. 1995. Inference from binary comparative data. *Journal of Theoretical Biology* 173: 99–108.
Reddy S, Kimball RT, Pandey A, Hosner PA, Braun MJ, Hackett SJ, Han KL, Harshman J, Huddleston CJ, Kingston S, Marks BD, Miglia KJ, Moore WS, Sheldon FH, Witt CC, Yuri T, Braun EL. 2017. Why do phylogenomic data sets yield conflicting trees? Data type influences the avian tree of life more than taxon sampling. *Systematic Biology* 66: 857–879.
Reeder TW, Townsend TM, Mulcahy DG, Noonan BP, Wood PL, Sites JW, Wiens JJ. 2015. Integrated analyses resolve conflicts over squamate reptile phylogeny and reveal unexpected placements for fossil taxa. *PLoS ONE* 10: e0118199.
dos Reis M, Donoghue PCJ, Yang Z. 2015. Bayesian molecular clock dating of species divergences in the genomics era. *Nature Reviews Genetics* 17: 71–80.
Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3: 217–223.
Rivera MC, Lake JA. 2004. The ring of life provides evidence for a genome

fusion origin of eukaryotes. *Nature* 431: 152–5.
Robinson DF, Foulds LR. 1981. Comparison of phylogenetic trees. *Mathematical Biosciences* 53: 131–147.
Rodríguez-Ezpeleta N, Brinkmann H, Roure B, Lartillot N, Lang BF, Philippe H. 2007. Detecting and Overcoming Systematic Errors in Genome-Scale Phylogenies. *Syst. Biol* 56: 389–399.
Sadleir RW, Makovicky PJ. 2008. Cranial shape and correlated characters in crocodilian evolution. *Journal of Evolutionary Biology* 21: 1578–1596.
Sanger TJ, Mahler DL, Abzhanov A, Losos JB. 2012. Roles for modularity and constraint in the evolution of cranial diversity among anolis lizards. *Evolution*.
Sansom RS, Wills MA. 2017. Differences between hard and soft phylogenetic data. *Proceedings of the Royal Society B: Biological Sciences* 284: 20172150.
Sansom RS, Wills MA, Williams T. 2016. Dental data perform relatively poorly in reconstructing mammal phylogenies: Morphological partitions evaluated with molecular benchmarks. *Systematic Biology* 66: syw116.
Sansom RS, Wills MA, Williams T. 2017. Dental data perform relatively poorly in reconstructing mammal phylogenies: Morphological partitions evaluated with molecular benchmarks. *Systematic Biology* 66: 813–822.
Sauquet H, von Balthazar M, Magallón S, Doyle JA, Endress PK, Bailes EJ, Barroso de Morais E, Bull-Hereñu K, Carrive L, Chartier M, Chomicki G, Coiro M, Cornette R, El Ottra JHL, Epicoco C, Foster CSP, Jabbour F, Haevermans A, Haevermans T, Hernández R, Little SA, Löfstrand S, Luna JA, Massoni J, Nadot S, Pamperl S, Prieu C, Reyes E, dos Santos P, Schoonderwoerd KM, Sontag S, Soulebeau A, Staedler Y, Tschan GF, Wing-Sze Leung A, Schönenberger J. 2017. The ancestral flower of angiosperms and its early diversification. *Nature Communications* 8: 16047.
Schluter D. 2000. *The ecology of adaptive radiation*. Oxford University Press.
Scotland RW, Olmstead RG, Bennett JR. 2003. Phylogeny Reconstruction: The Role of Morphology. *Systematic Biology* 52: 539–548.
Seiffert ER. 2007. A new estimate of afrotherian phylogeny based on simultaneous analysis of genomic, morphological, and fossil evidence. *BMC evolutionary biology* 7: 224.

Sharma PP, Kaluziak ST, Pérez-Porro AR, González VL, Hormiga G, Wheeler WC, Giribet G. 2014. Phylogenomic Interrogation of Arachnida Reveals Systemic Conflicts in Phylogenetic Signal. *Molecular Biology and Evolution* 31: 2963–2984.

Sibley CG, Ahlquist JE, Monroe BL. 1988. A Classification of the Living Birds of the World Based on Dna-Dna Hybridization Studies. *The Auk* 105: 409–423.

Simion P, Philippe H, Baurain D, Jager M, Richter DJ, Di Franco A, Roure B, Satoh N, Quéinnec É, Ereskovsky A, Lapébie P, Corre E, Delsuc F, King N, Wörheide G, Manuel M. 2017. A Large and Consistent Phylogenomic Dataset Supports Sponges as the Sister Group to All Other Animals. *Current Biology* 27.

Springer MS, Stanhope MJ, Madsen O, de Jong WW. 2004. Molecules consolidate the placental mammal tree. *Trends in Ecology & Evolution* 19: 430–438.

Springer MS, Teeling EC, Madsen O, Stanhope MJ, de Jong WW. 2001. Integrated fossil and molecular data reconstruct bat echolocation. *Proceedings of the National Academy of Sciences of the United States of America* 98: 6241–6246.

Stanhope MJ, Waddell VG, Madsen O, de Jong W, Hedges SB, Cleven GC, Kao D, Springer MS. 1998. Molecular evidence for multiple origins of Insectivora and for a new order of endemic African insectivore mammals. *Proceedings of the National Academy of Sciences of the United States of America* 95: 9967–72.

Suh A. 2016. The phylogenomic forest of bird trees contains a hard polytomy at the root of Neoaves. *Zoologica Scripta* 45: 50–62.

Thacker CE, Gkenas C. 2019. Morphometric convergence among European sand gobies in freshwater (Gobiiformes: Gobionellidae). *Ecology and Evolution* 9: 8087–8103.

Torres-Montúfar A, Borsch T, Ochoterena H. 2018. When Homoplasy Is Not Homoplasy: Dissecting Trait Evolution by Contrasting Composite and Reductive Coding (N MacLeod, Ed.). *Systematic Biology* 67: 543–551.

Torres C, van Tuinen M. 2013. Relationships of birds - molecules versus

morphology. *eLS*.
Tschopp P, Sherratt E, Sanger TJ, Groner AC, Aspiras AC, Hu JK, Pourquié O, Gros J, Tabin CJ. 2014. A relative shift in cloacal location repositions external genitalia in amniote evolution. *Nature* 516: 391–394.
Van Tuinen M, Butvill DB, Kirsch JA, Hedges SB. 2001. Convergence and divergence in the evolution of aquatic birds. *Proceedings. Biological sciences* 268: 1345–50.
Upham N, Esselstyn J, Jetz W. 2019. Inferring the mammal tree: species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLoS Biology*.
Varón A, Vinh LS, Wheeler WC. 2010. POY version 4: Phylogenetic analysis using dynamic homologies. *Cladistics* 26: 72–85.
Vickaryous MK, Hall BK. 2006. Human cell type diversity, evolution, development, and classification with special reference to cells derived from the neural crest. *Biological Reviews* 81: 425–455.
Wagner GP, Altenberg L. 1996. Perspective: Complex Adaptations and the Evolution of Evolvability. *Evolution* 50: 967.
Wagner GP, Gauthier JA. 1999. 1,2,3 = 2,3,4: a solution to the problem of the homology of the digits in the avian hand. *Proceedings of the National Academy of Sciences of the United States of America* 96: 5111–5116.
Wanninger A. 2015. Morphology is dead - long live morphology! Integrating MorphoEvoDevo into molecular EvoDevo and phylogenomics. *Frontiers in Ecology and Evolution* 3: 54.
Whelan N V., Kocot KM, Moroz LL, Halanych KM. 2015. Error, signal, and the placement of Ctenophora sister to all other animals. *Proceedings of the National Academy of Sciences* 112: 5773–5778.
Whelan N V, Kocot KM, Moroz TP, Mukherjee K, Williams P, Paulay G, Moroz LL, Halanych KM. 2017. Ctenophore relationships and their placement as the sister group to all other animals. *Nature ecology & evolution* 1: 1737–1746.
Whitfield JB, Kjer KM. 2008. Ancient Rapid Radiations of Insects: Challenges for Phylogenetic Analysis. *Annual Review of Entomology* 53: 449–472.
Wiens JJ. 2004. The Role of Morphological Data in Phylogeny Reconstruction. *Systematic Biology* 53: 653–661.

Wiley EO, Lieberman BS. 2011. *Phylogenetics : theory and practice of phylogenetic systematics*. Wiley-Blackwell.
Wolsan M, Suzuki S, Asahara M, Motokawa M. 2019. Dental integration and modularity in pinnipeds. *Scientific Reports* 9: 4184.
Wright AM, Lloyd GT, Hillis DM. 2016. Modeling Character Change Heterogeneity in Phylogenetic Analyses of Morphology through the Use of Priors. *Systematic Biology* 65: 602–611.
Zheng Y, Wiens JJ. 2016. Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Molecular Phylogenetics and Evolution* 94.

1. Osteological Characters Show Greater Congruence with Molecular Phylogenies than Soft Characters in Avian and Reptilian Morphological Datasets

Abstract

Despite increased use of genomic data in phylogenetics, morphological information remains vital for resolving evolutionary relationships, particularly for fossil taxa. The properties and models of evolution of molecular sequence data are well characterized and mature relative to those of morphological data. In particular, heterogeneity, integration and relative homoplasy of empirical morphological data could prove problematic for phylogenetic reconstruction if unaccounted for. Here we compare osteological and non-osteological characters of 28 morphological datasets of extant saurians in terms of their homoplasy relative to molecular trees. Analysis of individual avian datasets finds osteological characters to be significantly more consistent with molecular data than soft characters are. No significant differences were observed in 9 squamate datasets. Significant differences between morphological partitions were also observed in the age at which characters resolved on molecular trees; osteological character changes occur relatively earlier in deep branches whilst soft-tissue character transitions are more recent in shallow branches. The combined results demonstrate differences in evolutionary dynamics between morphological partitions. This may reflect evolutionary constraints acting on osteological characters, compared with the relative lability of soft characters. Furthermore, it provides some support to phylogenetic interpretations of fossil data, including dinosaurs, which are predominately osteological. Recent advances in amphibian and mammal phylogenetics may make these patterns possible to test for all tetrapods.

1. Introduction

Phylogenetics, the reconstruction of the evolutionary relationships between taxa, underpins evolutionary biology. It is necessary for understanding morphological and molecular evolution, since well-resolved, topologically accurate phylogenies allow us to track evolutionary changes through time. In the age of phylogenomics, morphology remains a vital source of phylogenetic information (Wiens, 2004; Lee & Palci, 2015; Wanninger, 2015) despite challenges such as the convergence and non-independence of characters. This is principally because morphological information is usually the only kind available from fossil taxa, and morphological data are therefore essential for their phylogenetic placement. An additional consideration is that morphological data provide an independent source of information with which to corroborate molecular data, with such consilience potentially strengthening hypotheses of evolutionary history, although simultaneous analyses of molecular and morphological data make the future role of morphology for consilience uncertain.

Underlying these considerations is the widespread distribution of homoplasy in morphological data i.e. convergence of phenotypic characters in distantly related species, often owing to the action of similar selective regimes. Morphological homoplasy has been demonstrated to confound phylogenetic reconstruction (for examples see Torres & van Tuinen, 2013; Reeder *et al.*, 2015). In order to best use morphological data, we need to understand and estimate the distribution of homoplasy as this could potentially aid in the building of more effective models for probabilistic inference methods. While convergence has also been observed in molecular data (Castoe *et al.*, 2009; Foote *et al.*, 2015; Zou & Zhang, 2016), it is nevertheless necessary to address these questions with regard to morphological data using molecular sequence data as an independent benchmark.

1.1. Morphological Partitioning and Congruence

While there has been a focus in recent literature on the use of different models for morphological phylogenetic data (Wright & Hillis, 2014; O'Reilly *et al.*, 2016; Puttick *et al.*, 2017), somewhat less attention has been paid to the characters themselves (but see Song & Bucheli, 2010; Mounce *et al.,* 2016; Sansom *et al.,* 2017). This is an important problem because there are several reasons to expect that different partitions of morphological data may give rise to different estimates of evolutionary history, such as selective regimes giving rise to convergence (e.g. Kivell *et al.,* 2013), evolutionary lability leading to rapid and superimposed changes (Omland & Lanyon, 2000; Wiens, 2009), modularity and mosaicism (Clarke & Middleton, 2008) and stochastic noise (e.g. Gaubert *et al.,* 2005). Differences have been shown in the phylogenetic signal conveyed by dental and osteological character partitions in mammals (Sansom *et al.,* 2017), hard and soft character partitions in various animals (Sansom & Wills, 2017) and craniodental and postcranial partitions in vertebrates (Mounce *et al.,* 2016).

The results of these and other studies (e.g. Sadleir & Makovicky, 2008; Goswami & Polly, 2010b) demonstrate that different groups of morphological characters evolve with different dynamics, rates and trends. Some characters evolve under strong confounding selection, while others retain phylogenetic signal which is recoverable when compared to a good estimate of the true underlying phylogeny. Characters evolving under selection can often be convergent (e.g. Sadleir & Makovicky, 2008), leading to suites of characters which, while compatible with each other, are incompatible with the underlying phylogeny (Holland *et al.*, 2010). In this case, genomic data can in theory help reveal which of these characters convey signal consistent with underlying evolutionary history, while acknowledging that the this is unknowable. This approach, i.e. evaluating the performance of subsets of morphological characters, has an important role to play in phylogenetics, as it demonstrates that the indiscriminate use of morphological characters without regard to convergence may be harmful to morphological phylogenetics, as is the same approach to molecular phylogenetics (e.g. Reddy *et al.,* 2017).

Assessing the consistency of morphological data with independent molecular trees and vice versa is useful because it reveals where data types are most consilient, and such consilience has the potential to strengthen or call into question existing hypotheses of phylogenetic history (Field *et al.*, 2014; McInerney, O'Connell, & Pisani, 2014; Reddy *et al.*, 2017). It can also provide potential insight into the modes of evolution of those characters which are not congruent with the molecular trees, for example convergence and modularity. Examining consistency and tracing morphological evolution on molecular trees provides insight into the basis of evolutionary change and allows us to interpret phenotypic traits in the context of evolutionary history. Although congruence has been tested for and discussed in previous literature, this study builds upon previous research by testing a wider group of datasets spanning birds and reptiles.

1.2. Rationale and Approach

Here, we use morphological data and molecular trees from birds and squamates, the two largest groups within the Sauropsida, to test hypotheses of the correspondence of morphological characters with molecular trees. Specifically, we test whether osteological characters differ from soft-tissue characters in their consistency with molecular phylogenies. Previous studies have shown differences in the phylogenetic signal conveyed between hard and soft characters (Sansom & Wills, 2017). Within the osteological partition, we test whether the postcranial and cranial subpartitions differ, and within the soft-tissue partition we test whether plumage and integument and myology subpartitions differ. We may expect these partitions and subpartitions of morphological data to contain different levels of homoplasy, considering differences in the action of sexual selection, directional selection, and ecological and developmental constraint on different character subsets. Furthermore, characters might be expected to be informative at different taxonomic ranks reflecting varying levels of evolutionary constraint and lability. There are *a priori* reasons to expect

different amounts of homoplasy in these different morphological regions; specifically, we hypothesize that soft characters such as plumage may be expected to contain more apparent homoplasy than osteological characters owing to higher evolutionary lability and convergence, and their role in sexual selection (Omland & Lanyon, 2000; Price, Friedman, & Omland, 2007). Additionally, different topologies are inferred from cranial and postcranial characters (Mounce *et al.*, 2016), and retention of only dental data in particular results in loss of phylogenetic signal (Sansom *et al.*, 2017), and we therefore hypothesize that there may be greater homoplasy in dental data.

It is particularly important to apply these analyses to osteological data, as this is usually the only data available when assessing the phylogenetic position of fossil species. We additionally indirectly test whether soft characters are more evolutionarily labile and therefore likely to resolve at different taxonomic levels by performing a simple, parsimony-based ancestral state reconstruction for each character, and then assigning each character state change an age based on the node ages of the molecular tree. We expect more labile traits to have more recent ages in view of higher frequency of change. We additionally directly test the relationship between retention index and character transition age, to determine whether any difference in transition age between partitions can be primarily explained by homoplasy.

2. Methods

2.1. Data Collection

Morphological data matrices for extant birds and squamates were compiled from three sources; a) Google Scholar, b) Graeme Lloyd's collection of matrices (Lloyd, 2009), specifically the Cenozoic bird collection, and c) the reference list of the supertree study of Davis & Page (2014).

Matrices were categorized and scored according to their taxonomic coverage, taxonomic level of terminals, number of taxa, number of characters, types of characters and source of data. We excluded datasets with fewer than twenty characters and 10 taxa. Characters were split into osteological and non-osteological partitions, and then further split into subpartitions (cranial, postcranial, integumentary and myological). Non-osteological morphological data in these datasets include mainly plumage, integument and myological data. It was necessary to exclude datasets for which the original data were either unavailable or unreadable. In many cases, matrices have been modified by various authors over time (e.g. Strauch, 1978), in which case the most recent version (e.g. Chu, 1995) was used to minimize pseudoreplication. We also identified non-independence between 3 matrices owing to taxonomic overlap, and excluded these from analyses accordingly. Additionally, some matrices were removed from analyses because of the unavailability of character descriptions or of the matrices themselves. A total of 18 avian and 8 squamate matrices were ultimately included, in addition to two large datasets with broad taxonomic coverage, one avian (Livezey & Zusi, 2007) and one squamate (Reeder *et al.,* 2015). Analysed datasets are available in the supplementary materials.
Molecular trees of birds were extracted from Jetz *et al.*'s (2012) companion website (birdtree.org) for comparison with avian morphological data. Trees based on either of the two phylogenies used as backbone constraints by Jetz *et al.,* (2012) are available to download, with the option to include all species of interest, or just those species with sequence data available.
Of the two separate backbone phylogenies used by Jetz *et al.,* (2012), the one using the Hackett *et al.,* (2008) topology as a constraint was used since a larger amount of data used to build this phylogeny. It should be noted that the Hackett *et al.,* (2008) tree uses β-fibrinogen, a gene which has been flagged as potentially problematic in recovering higher-level branching patterns (Mayr, 2011; Jetz *et al.*, 2012). However, since our analysis primarily uses small datasets (apart from two datasets) with narrow taxonomic spread, this should not affect this study. Although overall levels of morphological homoplasy will likely differ between molecular trees, it is

less clear that relative homoplasy between partitions will differ between molecular trees, unless there is agreement between morphological partitions and molecular partitions. We additionally performed analyses on an alternative Galloanserae supertree (Eo, Bininda-Emonds, & Carroll, 2009) for one of our datasets (Livezey, 1991), to test whether these results are robust to different estimates of evolutionary history.

Molecular trees were only extracted with taxa for which sequence data are available. The morphological data matrices were modified accordingly, by removing any taxa lacking molecular data. Additionally, some avian taxon names were altered in the morphological matrices to reflect alternative nomenclature used by Jetz *et al.,* (2012). Avibase (Lepage, 2003), a database of bird taxonomy, was used to establish synonymy. Invalid taxa of any other kind (i.e. extinct taxa or taxa for which no synonymy in Jetz *et al.,* 2012 could be established) were also deleted from the final morphological matrices. For each avian morphological matrix, 1,000 trees were sampled from the Jetz *et al.,* (2012) posterior distribution.

For squamates, the molecular supermatrix tree of Zheng & Wiens (2016) was used. The maximum likelihood tree includes around half of all squamates (4162 species) and 52 total mitochondrial and nuclear genes (12 genes for 4161 species and 44 genes for 161 species).

Some changes were made to matrices as necessary before analysis. For example, as the molecular data of Jetz *et al.,* (2012) includes taxa only at species level, any taxa in morphological matrices at the level of subspecies were changed to reflect this. For example, character coding for some subspecies were combined, some taxa were removed that overlapped between datasets to ensure independence, and taxa without corresponding molecular data were removed. In all cases, original character ordering and outgroup taxa were retained.

2.2. Homoplasy and Transition Ages

The fit of morphological characters relative to molecular trees was derived by calculating the retention indices of individual characters and, in the case of birds, averaging them over the 1000 trees for each dataset in TNT (Goloboff, Farris, & Nixon, 2008) (Figure 1). As these trees are from the posterior distribution it is possible that some may have low posterior probability. However, since the results are averaged over many trees and the MCMC algorithm spends more time sampling trees of higher posterior probability, it is likely that this will not affect results. The retention index of a cladistic character is a measure of its homoplasy, and is defined as the difference between the maximum number of steps on the tree and the number of state changes on that tree, divided by the difference between maximum number of steps on the tree and the number of state changes in the data. Values range from 0-1, with 1 indicating no homoplasy. Additionally, the ensemble retention index, the retention index for a group of characters, was calculated for partitions (osteological and non-osteological data) and subpartitions (cranial and postcranial osteological data and integumentary and myological data).

The difference in homoplasy between partitions was assessed by comparing mixed linear effects models in the R package nlme, with character retention index as the response variable and dataset treated as a random effect. Models compared were a null model, partition as a fixed effect and partition + clade as fixed effects. Heteroskedasticity was first accounted for by fitting different models of heteroskdasticity to the data as using the preferred model in all further analyses.

To test whether morphological partitions resolve on a phylogeny at different times and are informative at different taxonomic levels, the average ages of transition for osteological and non-osteological characters were compared. For birds, each informative, binary morphological character was applied to 100 trees from the posterior distribution of Jetz *et al.,* (2012) to estimate unambiguous character transitions in a parsimony framework (ACCTRAN) using the ancPropStateMat function in the R package Paleotree (Bapst, 2012). For squamates, the single maximum likelihood tree was used. This gives an estimate of the node or nodes at which a character transitions. To

estimate the age of transition, a random age between the age of the node of transition and its immediate ancestor was derived 100 times for each tree. The average of these ages was then taken as the transition age of a character on that tree. For characters transitioning multiple times on a tree, an average was taken of the ages for each of the branches along which the character transitions. The transition ages for each tree were then averaged to obtain the final transition age for each character (R script available in supplementary information). This approach pulls transition ages towards the middle of branches on which transitions occur, while adding a stochastic element. This allows for the comparison of multiple runs, in order to test how robust the results are to uncertainty in exact transition age. Note that this approach is likely to overestimate transition ages owing to the use of ACCTRAN and averaging ages along the branch preceding the transition. However, this is unlikely to significantly affect the results since the absolute age of transition is less important than the comparison of ages between partitions. The use of ACCTRAN as opposed to another algorithm was essentially arbitrary.

The average age of transition was compared between partitions for all datasets using a linear mixed model using the R package nlme, accounting for heteroskedasticity. A model treating partition (osteology and non-osteology) as fixed effect, and a model treating partition and clade (bird and reptile) as fixed effects were compared against a null. In all analyses, dataset was treated as a random effect. Additionally, to test for the effect of homoplasy on apparent transition age, a null model with age as response variable was compared with a model treating individual character retention indices as a fixed effect and models treating both retention index and partition as fixed effects, both with and without interaction terms.

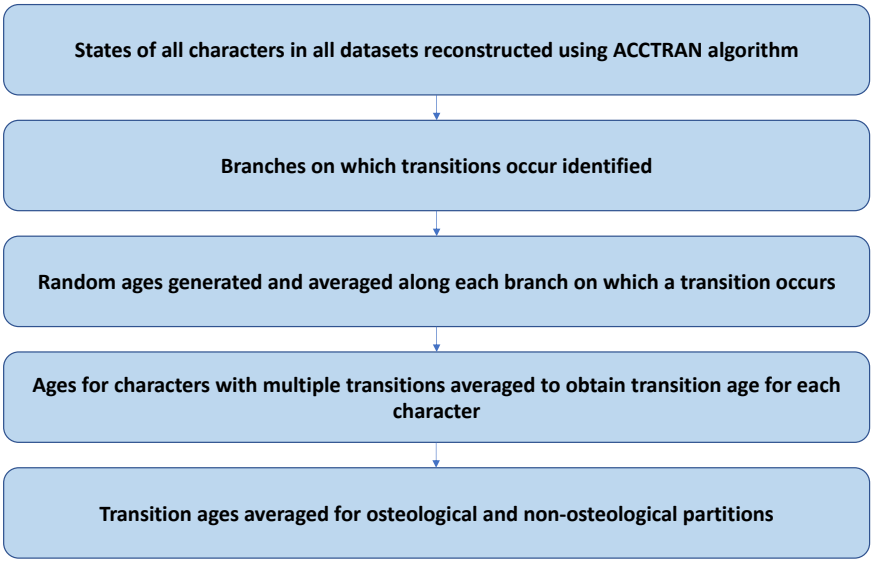


Fig 1. *Workflow for estimating transition ages of osteological vs non-osteological characters. Character state reconstructions were performed on all characters. One thousand random ages along branches on which transitions occur were simulated and averaged. For any characters transitioning multiple times, these ages were also averaged. These final transition ages for each character were then averaged for each partition to obtain relative transition ages for osteological and non-osteological data. Linear mixed effects models were also applied to the individual character transition ages.*

3. Results

3.1. Homoplasy

Osteological characters in morphological data are significantly more consistent with the molecular trees than soft tissue characters for avian and squamate datasets (paired t of ensemble retention indices, n = 28, p = 0.0004 (Figure 2), linear mixed effect model, likelihood ratio = 98.18, p = <.0001, least square mean for osteology = 0.666, non-osteology = 0.555). However, no significant difference was found between cranial and postcranial partitions (Figure 3), possibly because of the relative paucity of adequate data when splitting the osteological partition (paired t, n = 19, p = 0.5174). The relative homology between myology and integument could not

be compared between all datasets because there was a small number of characters in each subpartition. No significant difference was found between myological and integumentary or plumage characters for any individual datasets out of the 3 tested (Mann-Whitney *U*, Table 1). The addition of subpartition as a fixed effect did not significantly improve the explanatory power of linear models, although postcranial data had the highest least square mean retention index (0.682), compared with cranial (0.631), integument (0.525) and myology (0.673).

The single large morphological dataset covering all birds (Livezey & Zusi, 2007) was analysed separately because of taxonomic overlap and possible differences in properties between it and smaller datasets, such as character correlations, and found a significantly higher retention indices of soft characters with the underlying molecular phylogeny, as well as significantly higher consilience of cranial characters versus postcranial characters (Table 1). For squamates, the single large morphological dataset (Reeder *et al.,* 2015) shows significantly higher consilience of osteological and cranial characters (Table 1, p = 6.82e-05).

Table 1 at End of Document

Table 1. *Showing number of taxa, number of informative characters per partition, ensemble retention indices for partitions for each dataset and results of Mann-Whitney* U *comparing the mean retention index between partitions, where the null hypothesis is that the two sets of retention indices are drawn from the same distribution. Because of differences in taxonomic level, avian datasets are categorised by superorder or order, while squamate datasets are categorised by sub- or infraorder or superfamily.*

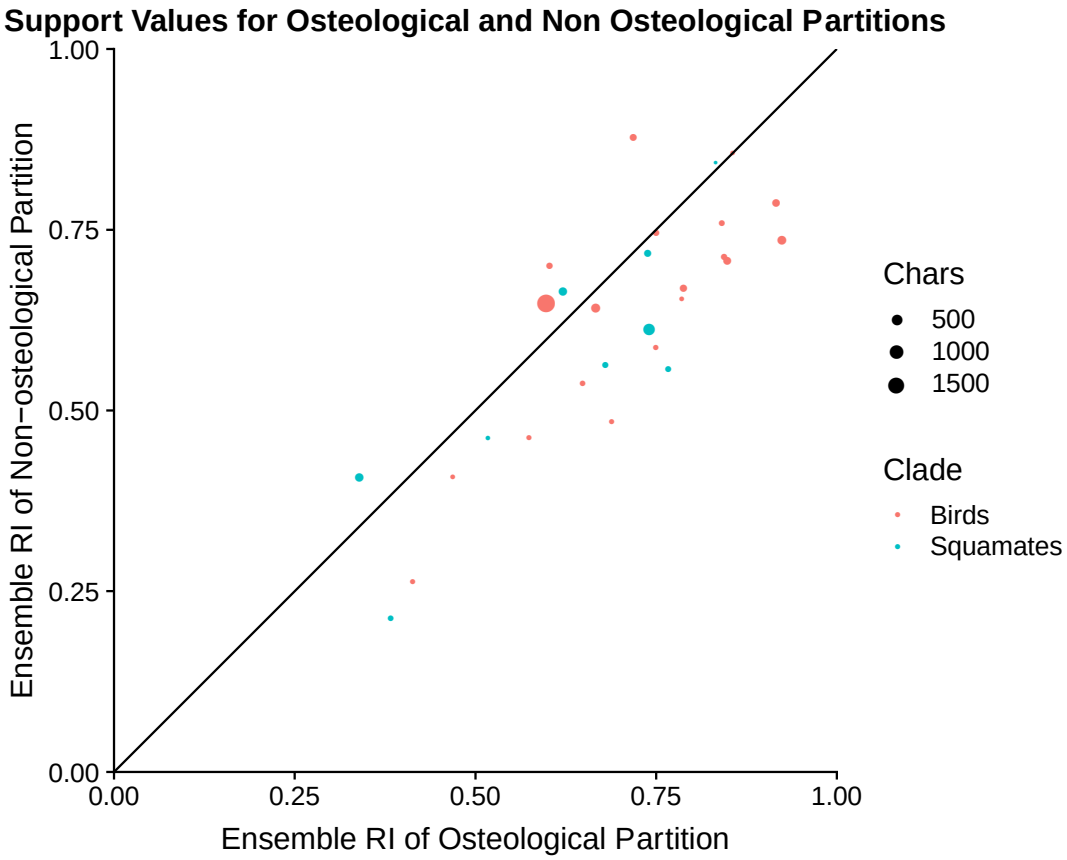


Fig 2. *Ensemble retention index between osteological and non-osteological partitions for birds and squamates. Osteological data has higher consistency in datasets falling below the x=y line. Colour corresponds to clade and dot size is proportional to dataset size (in number of total characters).*

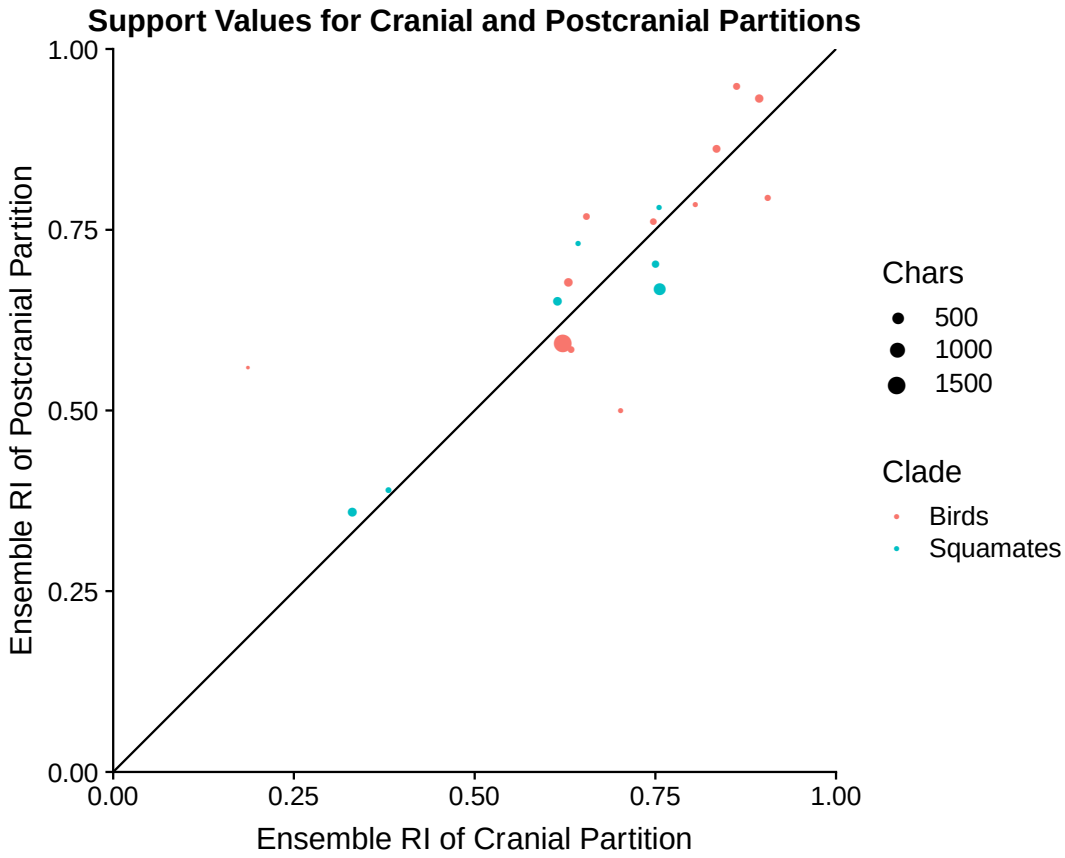


Fig 3. *Ensemble retention index between cranial and postcranial partitions for birds and squamates. Cranial data has higher consistency in datasets falling below the x=y line. Colour corresponds to clade and dot size is proportional to dataset size (in number of total osteological characters).*

A highly significant correlation (Pearson correlation coefficient, rho = 0.83, n = 27, p = 6.541e-08) between the ensemble retention indices between osteological versus soft character partitions across all datasets suggests that properties of the individual study or taxa are an important determinant of how well both sets of characters fit the molecular phylogeny, perhaps indicating that the age of datasets, coding practices of particular authors or underlying properties of the morphology are important factors determining character consistency.

There are more characters in osteological compared with non-osteological partitions (paired t test, t = 2.081, df = 26, p = 0.04705). However, there is no correlation between the number of characters and the ensemble retention index within either osteological (linear regression, F = 0.2282, df = 26, p = 0.637) or soft tissue partitions (linear regression, F = 0.3596, df = 26, p = 0.5539) suggesting that the marginally significant difference in the number of characters between partitions likely does not explain the observed relationship between retention index and data type. There is also no correlation between number of taxa and the ensemble retention index of either the osteological (linear regression, F = 1.555, df = 26, p = 0.224) or non-osteological (F = 0.909, df = 26, p = 0.349) partition, suggesting that dataset size does not affect the results.
Additionally, we find similar results on the alternative topology for Livezey, (1991), with an ensemble retention index of 0.82 for osteological data and 0.70 for non-osteological data, suggesting that our overall results are robust to different topologies.

3.2. Relative Transition Ages

Average character transition ages are significantly older for osteological characters relative to non-osteological characters (Figure 4, supplementary Figure 2). The best linear mixed effect model was that including both partition and clade as fixed effects (one-way anova, Table 2), indicating older ages for transitions in osteological data, and older ages for transitions in squamate data. Transition age data was highly heteroskedastic, with variability around the mean increasing with age. In all analyses, dataset was treated as a random effect.

Table 2. *Table comparing fit of three different linear mixed effect models to examine the effect of partition (osteological or non-osteological) on character transition age.*

Fixed Effects	**df**	**AIC**	**Log Likelihood**	**Likelihood Ratio**	**P value**
None	23	11355.15	-5654.575		
Partition	24	11326.53	-5639.264	30.62207	<.0001
Partition + Clade	25	11314.8	-5632.398	13.73356	<.0001

Models compared include a null model, a model including partition as a fixed effect, and a model including both partition and clade (birds or reptiles) as fixed effects. Dataset is treated as a random effect in all models. AICs, likelihood ratios and p values all prefer the most complex model, suggesting that partition and clade both have strong explanatory power. Transitions in osteological characters are significantly older, as are transitions in squamate characters.

For individual datasets, 5 out of 16 avian datasets and 3 out of 6 squamate datasets show significant differences in age of transition between individual osteological and non-osteological characters (Mann-Whitney *U* tests, Table 3).

Subpartitions (cranial vs postcranial and myology versus plumage) for all datasets were harder to assess statistically owing to paucity of data. There is no significant difference in cranial versus postcranial character transition ages between datasets for birds and squamates together (paired t, t = -0.36538, df = 14, p = 0.7203). Out of 15 assessed datasets, only 3 are significantly different between cranial and postcranial partitions, and 1 out of 2 datasets show a significant difference between myology and integument (Mann-Whitney *U* tests, Table 3).

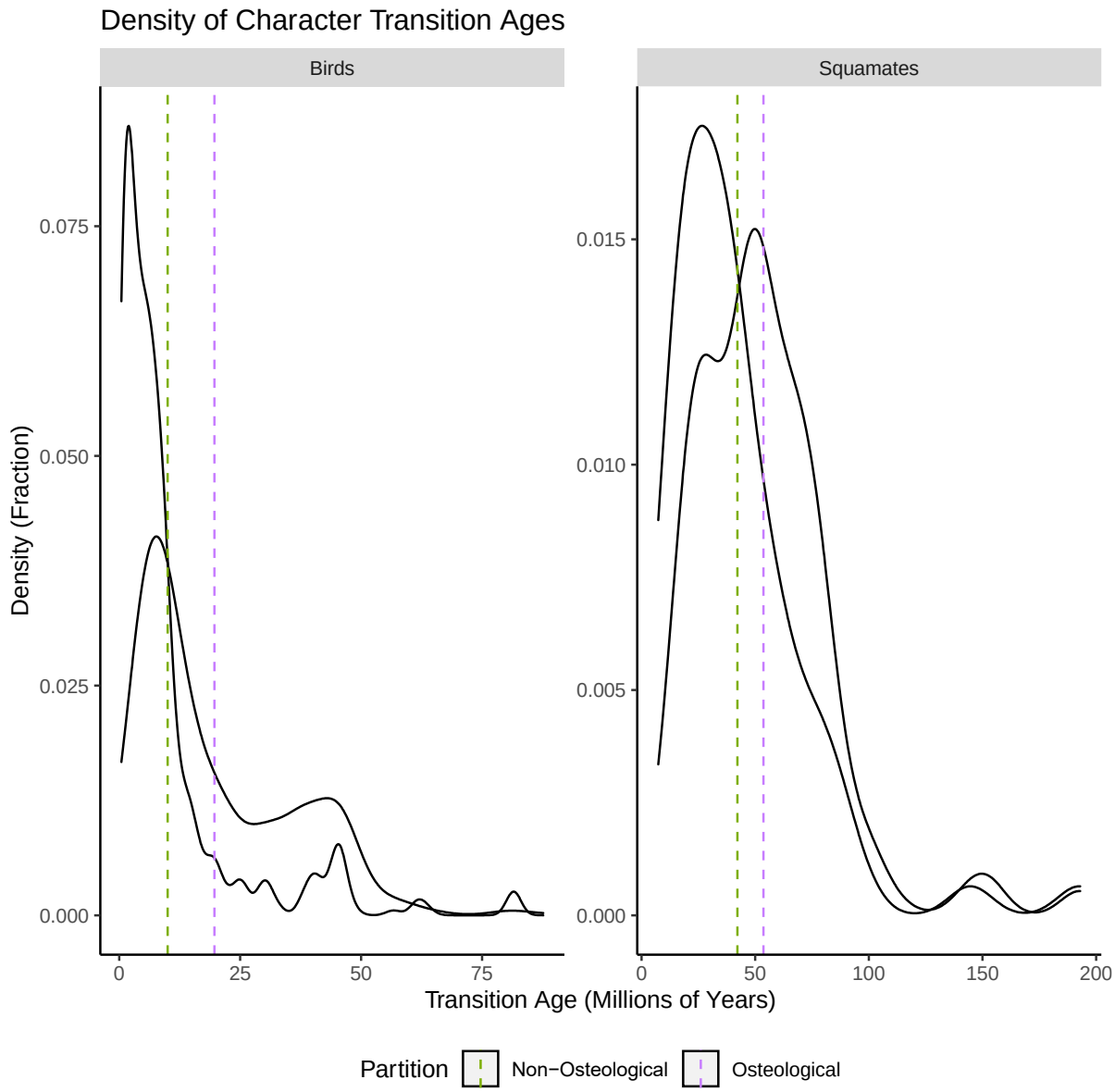


Fig 4. *Showing differences in pooled character transition ages in millions of years between partitions in all bird and squamate datasets. Dashed lines indicate mean ages of osteological (older) and non-osteological (younger) characters. Data is coloured by partition, with osteological data in purple and non-osteological data in green.*

There is a significant correlation between transition age and retention index (linear regression, F = 15.38, df = 955, p = 9.40E-05) in the Livezey and Zusi (2006) dataset, showing a direct association between the age of a character change and homoplasy. Additionally, there is a negative correlation between difference in the ensemble retention index between osteological characters and non-osteological characters, and the root age of the dataset in birds

(linear regression, F = 6.038, df = 18, p = 0.025, Supplementary Figure 3), indicating that osteological characters become relatively less congruent with molecular trees compared to soft characters further back in time. Given that both retention index and transition age differ between partitions, and retention index and transition ages are correlated (supplementary Figure 1), further linear mixed effect models were compared in order to further examine the relationship between homoplasy, transition age and partition. Retention index had significant explanatory power for transition age. Explanatory power was further improved with the addition of partition as a fixed effect, with an interaction term between retention index and partition, indicating that the difference in homoplasy between partitions explains the difference in transition ages between partitions (Table 3).

Table 3. *Table comparing fit of four different linear mixed effect models to examine the effect of homoplasy (as measured by the retention index) on apparent transition ages.*

Fixed Effects	**df**	**AIC**	**Log Likelihood**	**Likelihood Ratio**	**P value**
None	23	11355.15	-5654.575		
RI	24	11310.02	-5631.01	47.13005	<.0001
RI + Partition	25	11315.48	-5632.739	3.45623	0.063
RI * Partition	26	11297.53	-5622.767	19.94243	<.0001

Models compared include a null model, a model including retention index as a fixed effect, and models including both retention index and partition (osteology and non-osteology) as fixed effects. Character transition age is the response variable and dataset is treated as a random effect in all models. AICs, likelihood ratios and p values all prefer the most complex model, suggesting that retention index has strong explanatory power on transition age. Homoplasy is significantly stronger in younger characters, explaining the difference in transition age between osteological and soft characters.

4. Discussion

The results of this analysis of saurian datasets demonstrate that non-osteological characters exhibit greater levels of homoplasy relative to

molecular trees than osteological characters do, and that this homoplasy explains a difference in average character transition ages between these morphological regions. This is consistent with previous studies showing differences in phylogenetic properties between hard and soft morphological data (Sansom & Wills, 2017). Interestingly, however, soft characters have a higher consistency with molecular data than osteological characters in the single, larger (Livezey & Zusi, 2007) dataset. This may be owing to the coding style of the individual authors rather than a reflection of the properties of the morphological data themselves. Additionally, as this is a very large dataset, it may be affected by issues such as the non-independence of characters to a greater extent than the smaller datasets analyzed here. Alternatively, different classes of characters may be phylogenetically informative at different taxonomic levels, and this discrepancy is therefore explained by the difference in taxonomic breadth between the single broad dataset spanning all birds, and the datasets of smaller clades. This is supported by the fact that between individual avian datasets, the difference in congruence between osteological and soft characters decreases and appears to even reverse deeper in time. It is also possible that spurious higher-level relationships in the molecular topology have affected the results. Future analyses would benefit from comparisons between several molecular phylogenies, however this was not possible here as currently available molecular phylogenies for birds are limited in taxonomic coverage.

4.1. Evolutionary Constraint and Selection

In view of evolutionary and developmental constraints on the vertebrate body plan (Hu *et al.*, 2017), it seems sensible to suggest that osteological characters may tend to evolve more slowly than some soft characters. In birds, these constraints may particularly pertain to flight, which requires a specific bauplan, largely conserved across flying birds (Sullivan *et al.*, 2017). While also extremely important for flight, such constraints may not be

shared by some types of plumage characters which, by contrast, are considered relatively evolutionarily labile and therefore highly convergent owing to sexual selection in many avian taxa (Omland & Lanyon, 2000; Price *et al.*, 2007; Hofmann, Cronin, & Omland, 2007). In this case, the relative paucity of integumentary characters in the large Livezey & Zusi (2007) dataset (98 compared with 237 myological characters) may itself be a reflection of the difficulty in finding homologous plumage characters across all birds owing to the faster evolutionary rates of these characters.
It is therefore possible that superimposed changes in some plumage traits such as colouration, where characters may change multiple times on a tree, could result in saturation and convergence of these characters, while only through deep time and across a broad taxonomic range do osteological characters become convergent. Labile and ecologically important traits may be expected to poorly reflect evolutionary history (Price *et al.*, 2007; Kamilar & Muldoon, 2010), possibly because of this overwritten phylogenetic signal in characters accruing multiple changes. This cannot however explain why the same pattern is not observed in larger squamate datasets. Although reproductive success in squamates is often determined by physical capability and behaviour (Shine, Langkilde, & Mason, 2004; Swierk, Ridgway, & Langkilde, 2012), colouration is also important in many species (Chen *et al.*, 2012).

4.2. Character Transition Ages

To test whether different classes of characters are informative at different taxonomic levels, we compared the ages at which these different character types transition. Our results show that soft characters generally change on a phylogeny at younger ages than osteological characters for all datasets together, and that levels of homoplasy are directly related to age of transition for a character. This supports the hypothesis that different morphological regions may be phylogenetically informative at different taxonomic levels. Evolutionarily labile traits may contain the phylogenetic

signal to resolve the relationships between extremely young taxa, but may be subsequently overwritten over time, while slower-evolving traits become overwritten only in deeper time. For example, Benson & Choiniere (2013) suggest that although musculoskeletal evolution during the dinosaur-bird transition may have driven early rapid speciation, some of this potentially strong phylogenetic signal has been overwritten by subsequent diversification. Plumage characters, by contrast, may be phylogenetically useful within a genus, but are more likely to be homoplasious at the level of family or order. These changes may become saturated over time, a situation almost analogous to the long-branch attraction observed in molecular systematics. By contrast, traits that evolve slowly may be phylogenetically uninformative at lower taxonomic levels but may help to resolve relationships where they vary between higher taxa. This idea is supported by the fact that soft characters change more recently on phylogenies in both birds and squamates. Since ancestral state reconstructions were performed under a parsimony framework, which minimises transitions, evolutionary rates are not modelled and evolutionary lability and overwritten signal are expected to manifest as younger ages on trees.

The observed higher consistency of osteological characters in birds is consistent with previous studies suggesting that plumage contains weak phylogenetic signal as a result of homoplasy (Price *et al.*, 2007). However, while the non-osteological characters used in the meta-analysis have been split into the sub-partitions of myology and plumage or integument, there was not enough data from any of these partitions alone to compare with osteology. Given more data, splitting the non-osteological characters into more subpartitions may help to further resolve the distribution of support among character types by reducing the potentially conflicting signal between the morphologically disparate characters within the partition. This is because the non-osteological data are derived from plumage, myology, integument and in some cases other categories such as nervous tissue. As these all have different developmental and evolutionary origins, and evolutionary pressures, there is potential for conflicting signal between these data types.

An alternative explanation for the results presented here is that differences in the consilience of different morphological partitions with a molecular phylogeny may reflect differences in coding strategies in different subsets of morphological data rather than any inherent difference in the ability of these areas of morphology to elucidate evolutionary history per se. However, there is no obvious a priori reason to suspect that plumage or other soft-tissue characters are more difficult to accurately code than osteological characters. It may be the case that since osteological characters are more conserved across vertebrate species they are better understood and therefore easier to code, which is consistent with the idea that osteological characters contain more phylogenetic signal than other areas of morphology.

5. Conclusion

Examining the consistency of different classes of morphological characters against a molecular phylogeny provides a way of identifying consilience between these different data types. Here, we have identified greater consistency of osteological characters than soft tissue characters with molecular phylogenies, implying that hard characters may be more reliable in reconstructing evolutionary history, while relatively more evolutionarily labile soft characters such as plumage, which are affected to a greater extent by sexual and directional selection, and less affected by evolutionary constraint. This is supported by their more recent apparent transitions on phylogenetic trees.

The higher consistency of osteological characters with an underlying molecular tree is potentially good news for palaeontologists when placing fossil taxa, since osteological data are often the only phylogenetic information available when working with fossil species. However, caution must still be exercised since as the Livezey & Zusi (2007) data show, this is not always the case, and osteological data may become increasingly unreliable deeper in time.

These findings may have implications for the use of morphological data in phylogeny reconstruction. Future analyses, for example those using models of morphological evolution in probabilistic frameworks, may benefit from incorporating morphological partition schemes that reflect the shared evolutionary properties of characters within partitions, which may be identified by testing for differences in homoplasy, evolutionary rates or transition ages, or correlations. This may involve splitting hard and soft characters, or cranial and postcranial characters.

Whether these patterns apply more broadly across vertebrate species can be investigated with recent well-sampled molecular trees. This will be especially useful for teasing out any difference between the different types of soft tissue characters found in different vertebrates.

References

Arnold EN. 1997. Interrelationships and evolution of the east Asian grass lizards, Takydromus (Squamata: Lacertidae). *Zoological Journal of the Linnean Society* 119: 267–296.

Bapst DW. 2012. paleotree : an R package for paleontological and phylogenetic analyses of evolution. *Methods in Ecology and Evolution* 3: 803–807.

Bertelli S, Chiappe LM, Mayr G. 2014. Phylogenetic interrelationships of living and extinct Tinamidae, volant palaeognathous birds from the New World. *Zoological Journal of the Linnean Society* 172: 145–184.

Birdsley JS. 2002. Phylogeny of the Tyrant Flycatchers (Tyrannidae) Based on Morphology and Behavior. *The Auk* 119: 715–734.

Castoe TA, de Koning APJ, Kim HM, Gu W, Noonan BP, Naylor G, Jiang ZJ, Parkinson CL, Pollock DD. 2009. Evidence for an ancient adaptive episode of convergent molecular evolution. *Proceedings of the National Academy of Sciences* 106: 8986–8991.

Chen IP, Stuart-Fox D, Hugall AF, Symonds MRE. 2012. Sexual selection and the evolution of complex colour patterns in dragon lizards. *Evolution* 66:

3605–3614.
Chu PC. 1995. Phylogenetic reanalysis of Strauch's osteological data set for the Charadriiformes. *The Condor* 97: 174–196.
Clarke JA, Ksepka DT, Stucchi M, Urbina M, Giannini N, Bertelli S, Narváez Y, Boyd CA. 2007. Paleogene equatorial penguins challenge the proposed relationship between biogeography, diversity, and Cenozoic climate change. *Proceedings of the National Academy of Sciences of the United States of America* 104: 11545–11550.
Clarke J, Middleton K. 2008. Mosaicism, modules, and the evolution of birds: results from a Bayesian approach to the study of morphological evolution using discrete character data. *Systematic Biology* 57: 185–201.
Davis KE, Page RDM. 2014. Reweaving the tapestry: a supertree of birds. *PLoS Currents*.
Daza JD, Bauer AM. 2012. A new amber-embedded sphaerodactyl gecko from Hispaniola, with comments on morphological synapomorphies of the Sphaerodactylidae. *Breviora* 529: 1–28.
Eo SH, Bininda-Emonds ORP, Carroll JP. 2009. A phylogenetic supertree of the fowls (Galloanserae, Aves). *Zoologica Scripta* 38: 465–481.
Field DJ, Gauthier JA, King BL, Pisani D, Lyson TR, Peterson KJ. 2014. Toward consilience in reptile phylogeny: microRNAs support an archosaur, not a lepidosaur affinity for turtles. *Evolution & Development* 16: 189–196.
Foote AD, Liu Y, C Thomas GW, Vinař T, Alföldi J, Deng J, Dugan S, van Elk CE, Hunter ME, Joshi V, Khan Z, Kovar C, Lee SL, Lindblad-Toh K, Mancia A, Nielsen R, Qin X, Qu J, Raney BJ, Vijay N, W Wolf JB, Hahn MW, Muzny DM, Worley KC, Thomas Gilbert MP, Gibbs RA. 2015. Convergent evolution of the genomes of marine mammals. *Nature Genetics* 47: 272–275.
Frank-Hoeflich K, Silveira LF, Estudillo-López J, García-Koch AM, Ongay-Larios L, Piñero D. 2007. Increased taxon and character sampling reveals novel intergeneric relationships in the Cracidae (Aves: Galliformes). *Journal of Zoological Systematics and Evolutionary Research* 45: 242–254.
Frost DR, Etheridge R, Janies D, Titus TA. 2001a. Total evidence, sequence alignment, evolution of Polychrotid lizards, and a reclassification of the Iguania (Squamata: Iguania). *American Museum Novitates* 3343: 1–39.

Frost DR, Rodrigues MT, Grant T, Titus TA. 2001b. Phylogenetics of the lizard genus Tropidurus (Squamata: Tropiduridae: Tropidurinae): direct optimization, descriptive efficiency, and sensitivity analysis of congruence between molecular data and morphology. *Molecular Phylogenetics and Evolution* 21: 352–371.
Gaubert P, Wozencraft W, Cordeiro-Estrela P, Veron G. 2005. Mosaics of convergences and noise in morphological phylogenies: what's in a viverrid-like carnivoran? *Systematic Biology* 54: 865–894.
Goloboff PA, Farris JS, Nixon KC. 2008. TNT, a free program for phylogenetic analysis. *Cladistics* 24: 774–786.
Goswami A, Polly PD. 2010. The influence of modularity on cranial morphological disparity in Carnivora and Primates (Mammalia). *PLoS ONE* 5: e9517.
Hackett SJ, Kimball RT, Reddy S, Bowie RCK, Braun EL, Braun MJ, Chojnowski JL, Cox WA, Han KL, Harshman J, Huddleston CJ, Marks BD, Miglia KJ, Moore WS, Sheldon FH, Steadman DW, Witt CC, Yuri T. 2008. A phylogenomic study of birds reveals their evolutionary history. *Science* 320: 1763–1768.
Hitchmough R. 1997. A systematic revisions of the New Zealand Gekkonidae. Unpublished thesis, Victoria University of Wellington.
Hofmann CM, Cronin TW, Omland KE. 2007. Melanin coloration in New World orioles II: ancestral state reconstruction reveals lability in the use of carotenoids and phaeomelanins. *Journal of Avian Biology* 38: 172–181.
Holland BR, Spencer HG, Worthy TH, Kennedy M. 2010. Identifying cliques of convergent characters: concerted evolution in the cormorants and shags. *Systematic Biology* 59: 433–445.
Hu H, Uesaka M, Guo S, Shimai K, Lu TM, Li F, Fujimoto S, Ishikawa M, Liu S, Sasagawa Y, Zhang G, Kuratani S, Yu JK, Kusakabe TG, Khaitovich P, Irie N. 2017. Constrained vertebrate evolution by pleiotropic genes. *Nature Ecology & Evolution* 1: 1722–1730.
Irestedt M, Fjeldså J, Ericson PGP. 2004. Phylogenetic relationships of woodcreepers (Aves: Dendrocolaptinae) - incongruence between molecular and morphological data. *Journal of Avian Biology* 35: 280–288.

Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO. 2012. The global diversity of birds in space and time. *Nature* 491: 444–448.
Kamilar JM, Muldoon KM. 2010. The climatic niche diversity of Malagasy primates: a phylogenetic perspective. *PLoS ONE* 5: e11073.
Kearney M. 2003. Systematics of the Amphisbaenia (Lepidosauria:Squamata) Based on Morphological Evidence From Recent and Fossil Forms. *Herpetological Monographs* 17: 1–74.
Kivell TL, Barros AP, Smaers JB, Boesch C, Crockford C, Fawcett K, Inoue E, Inoue-Muruyama M, Morwood M, Jungers W. 2013. Different evolutionary pathways underlie the morphology of wrist bones in hominoids. *BMC Evolutionary Biology* 13: 229.
Lee MSY, Palci A. 2015. Morphological phylogenetics in the genomic age. *Current Biology* 25: R922–R929.
Lepage D. 2003. Avibase - The World Bird Database. Available at http://avibase.bsc-eoc.org.
Livezey BC. 1991. A Phyogenetic analysis and classification of recent Dabbling Ducks (Tribe Anatini) based on comparative morphology. *The Auk*: 471–507.
Livezey BC. 1996a. A phylogenetic analysis of modern pochards (Anatidae: Aythyini). *The Auk* 113: 74–93.
Livezey BC. 1996b. A phylogenetic analysis of geese and swans (Anseriformes: Anserinae), including selected fossil species. *Systematic Biology* 45: 415–450.
Livezey BC. 1998. A phylogenetic analysis of the Gruiformes (Aves) based on morphological characters, with an emphasis on the rails (Rallidae). *Philosophical Transactions of the Royal Society B: Biological Sciences* 353: 2077–2151.
Livezey BC, Zusi RL. 2007. Higher-order phylogeny of modern birds (Theropoda, Aves: Neornithes) based on comparative anatomy. II. Analysis and discussion. *Zoological Journal of the Linnean Society*.
Lloyd GT. 2009. Graeme T. Lloyd matrices. Available at http://www.graemetlloyd.com/matr.html.
Maurício GN, Areta JI, Bornschein MR, Reis RE. 2012. Morphology-based

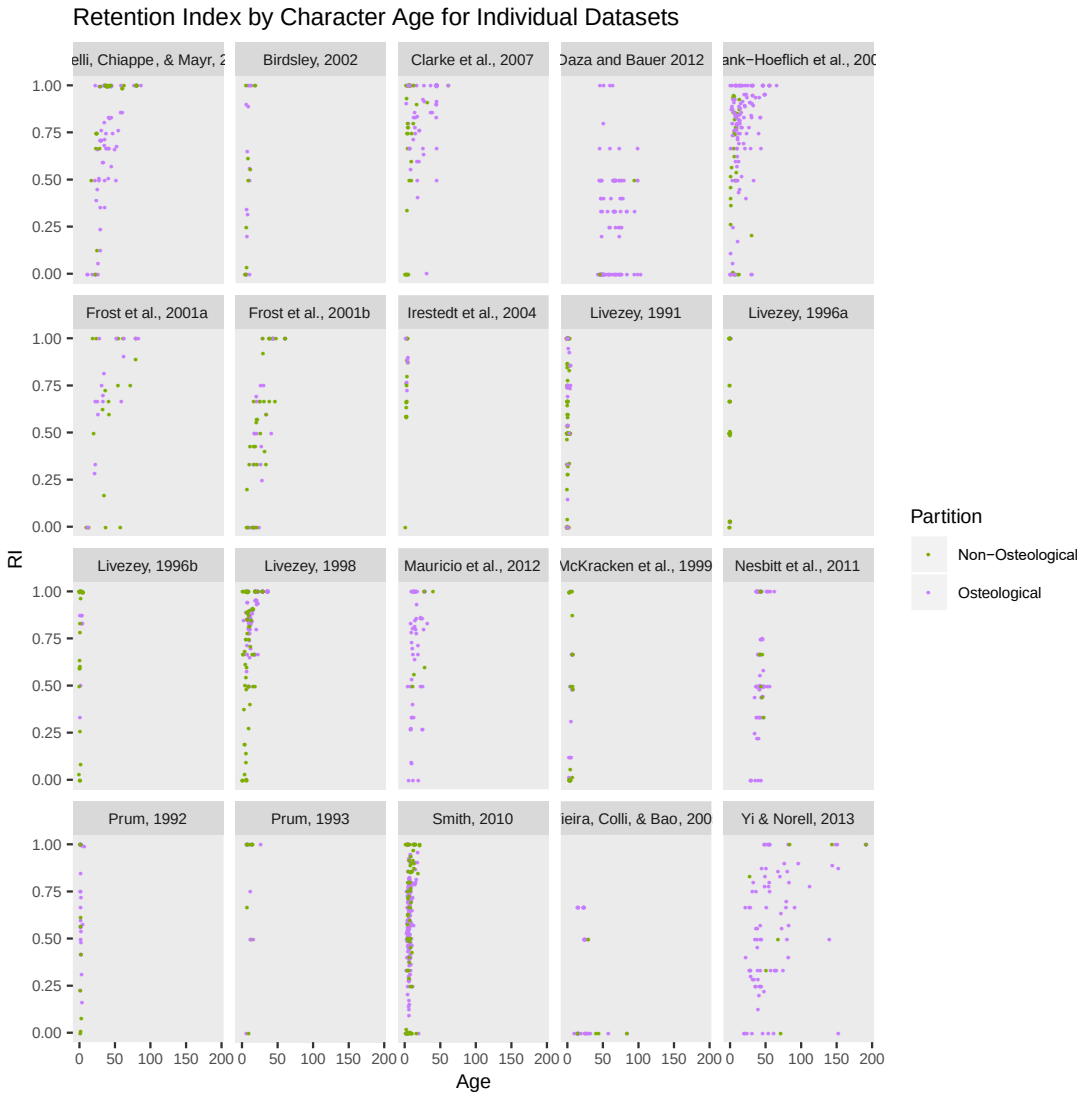
phylogenetic analysis and classification of the family Rhinocryptidae (Aves: Passeriformes). *Zoological Journal of the Linnean Society* 166: 377–432.
Mayr G. 2011. Metaves, Mirandornithes, Strisores and other novelties - a critical review of the higher-level phylogeny of neornithine birds. *Journal of Zoological Systematics and Evolutionary Research* 49: 58–76.
McCracken KG, Harshman J, McClellan DA, Afton AD. 1999. Data set incongruence and correlated character evolution: an example of functional convergence in the hind-limbs of stifftail diving ducks. *Systematic biology* 48: 683–714.
McInerney JO, O'Connell MJ, Pisani D. 2014. The hybrid nature of the Eukaryota and a consilient view of life on Earth. *Nature Reviews Microbiology* 12: 449–455.
Mounce RCP, Sansom R, Wills MA. 2016. Sampling diverse characters improves phylogenies: craniodental and postcranial characters of vertebrates often imply different trees. *Evolution* 70: 666–686.
Nesbitt SJ, Ksepka DT, Clarke JA. 2011. Podargiform affinities of the enigmatic Fluvioviridavis platyrhamphus and the early diversification of Strisores ("Caprimulgiformes" + Apodiformes). *PLoS ONE* 6: e26350.
O'Reilly JE, Puttick MN, Parry L, Tanner AR, Tarver JE, Fleming J, Pisani D, Donoghue PCJ. 2016. Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. *Biology letters* 12.
Omland KE, Lanyon SM. 2000. Reconstructing plumage evolution in orioles (Icterus): Repeated convergence and reversal in patterns. *Evolution* 54: 2119–2133.
Patten MA, Fugate M. 1998. Systematic relationships among the Emberizid sparrows. *The Auk* 115: 412–424.
Price JJ, Friedman NR, Omland KE. 2007. Song and plumage evolution in the New World orioles (Icterus) show similar lability and convergence in patterns. *Evolution* 61: 850– 863.
Prum R. 1992. Syringeal morphology, phylogeny, and evolution of the neotropical manakins (Aves, Pipridae). *American Museum Novitates* 3043: 1–65.

Prum RO. 1993. Phylogeny, biogeography, and evolution of the broadbills (Eurylaimidae) and asities (Philepittidae) based on morphology. *The Auk* 110: 304–324.
Puttick MN, O'Reilly JE, Tanner AR, Fleming JF, Clark J, Holloway L, Lozano-Fernandez J, Parry LA, Tarver JE, Pisani D, Donoghue PCJ. 2017. Uncertain-tree: discriminating among competing approaches to the phylogenetic analysis of phenotype data. *Proceedings of the Royal Society of London B: Biological Sciences* 284.
Reddy S, Kimball RT, Pandey A, Hosner PA, Braun MJ, Hackett SJ, Han KL, Harshman J, Huddleston CJ, Kingston S, Marks BD, Miglia KJ, Moore WS, Sheldon FH, Witt CC, Yuri T, Braun EL. 2017. Why do phylogenomic data sets yield conflicting trees? Data type influences the avian tree of life more than taxon sampling. *Systematic Biology* 66: 857–879.
Reeder TW, Townsend TM, Mulcahy DG, Noonan BP, Wood PL, Sites JW, Wiens JJ. 2015. Integrated analyses resolve conflicts over squamate reptile phylogeny and reveal unexpected placements for fossil taxa. *PLoS ONE* 10: e0118199.
Sadleir RW, Makovicky PJ. 2008. Cranial shape and correlated characters in crocodilian evolution. *Journal of Evolutionary Biology* 21: 1578–1596.
Sansom RS, Wills MA. 2017. Differences between hard and soft phylogenetic data. *Proceedings of the Royal Society B: Biological Sciences* 284: 20172150.
Sansom RS, Wills MA, Williams T. 2017. Dental data perform relatively poorly in reconstructing mammal phylogenies: morphological partitions evaluated with molecular benchmarks. *Systematic Biology* 66: 813–822.
Shine R, Langkilde T, Mason RT. 2004. Courtship tactics in garter snakes: how do a male's morphology and behaviour influence his mating success? *Animal Behaviour* 67: 477–483.
Smith NA. 2011. Taxonomic revision and phylogenetic analysis of the flightless Mancallinae (Aves, Pan-Alcidae). *ZooKeys* 91: 1–116.
Song H, Bucheli SR. 2010. Comparison of phylogenetic signal between male genitalia and non-genital characters in insect systematics. *Cladistics* 26: 23–35.
Strauch JG. 1978. The phylogeny of the Charadriiformes (Aves): a new

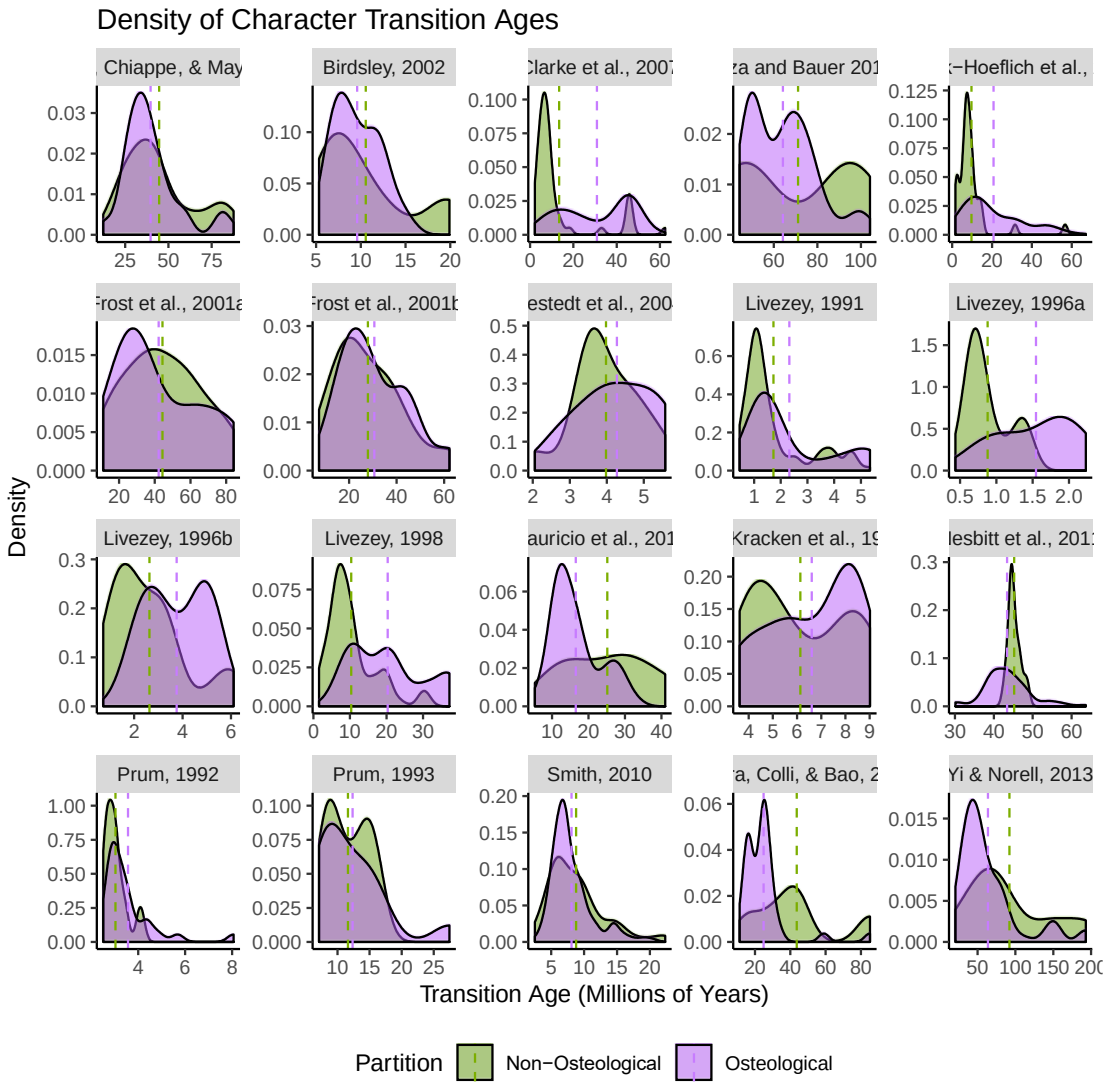
estimate using the method of character compatibility analysis. *The Transactions of the Zoological Society of London* 34: 263–345.
Sullivan TN, Wang B, Espinosa HD, Meyers MA. 2017. Extreme lightweight structures: avian feathers and bones. *Materials Today* 20: 377–391.
Swierk L, Ridgway M, Langkilde T. 2012. Female lizards discriminate between potential reproductive partners using multiple male traits when territory cues are absent. *Behavioral Ecology and Sociobiology* 66: 1033–1043.
Torres C, van Tuinen M. 2013. Relationships of birds - molecules versus morphology. *eLS*.
Vieira GHC, Colli GR, Bao SN. 2005. Phylogenetic relationships of corytophanid lizards (Iguania, Squamata, Reptilia) based on partitioned and total evidence analyses of sperm morphology, gross morphology, and DNA data. *Zoologica Scripta* 34: 605–625.
Wanninger A. 2015. Morphology is dead - long live morphology! Integrating MorphoEvoDevo into molecular EvoDevo and phylogenomics. *Frontiers in Ecology and Evolution* 3: 54.
Wiens JJ. 2004. The Role of Morphological Data in Phylogeny Reconstruction. *Systematic Biology* 53.
Wiens JJ. 2009. Estimating rates and patterns of morphological evolution from phylogenies: lessons in limb lability from Australian Lerista lizards. *Journal of Biology* 8: 19.
Winker K, Pruett CL. 2006. Seasonal migration, speciation, and morphological convergence in the genus Catharus (Turdidae). *The Auk* 123: 1052–1068.
Wright AM, Hillis DM. 2014. Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data (AFY Poon, Ed.). *PLoS ONE* 9: e109210.
Yi HY, Norell MA. 2013. New Materials of Estesia mongoliensis (Squamata: Anguimorpha) and the evolution of venom grooves in lizards. *American Museum Novitates* 3767: 1–31.
Zheng Y, Wiens JJ. 2016. Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards

and snakes) based on 52 genes and 4162 species. *Molecular Phylogenetics and Evolution* 94: 537–547.
Zou Z, Zhang J. 2016. Morphological and molecular convergences in mammalian phylogenetics. *Nature Communications* 7: 12758.

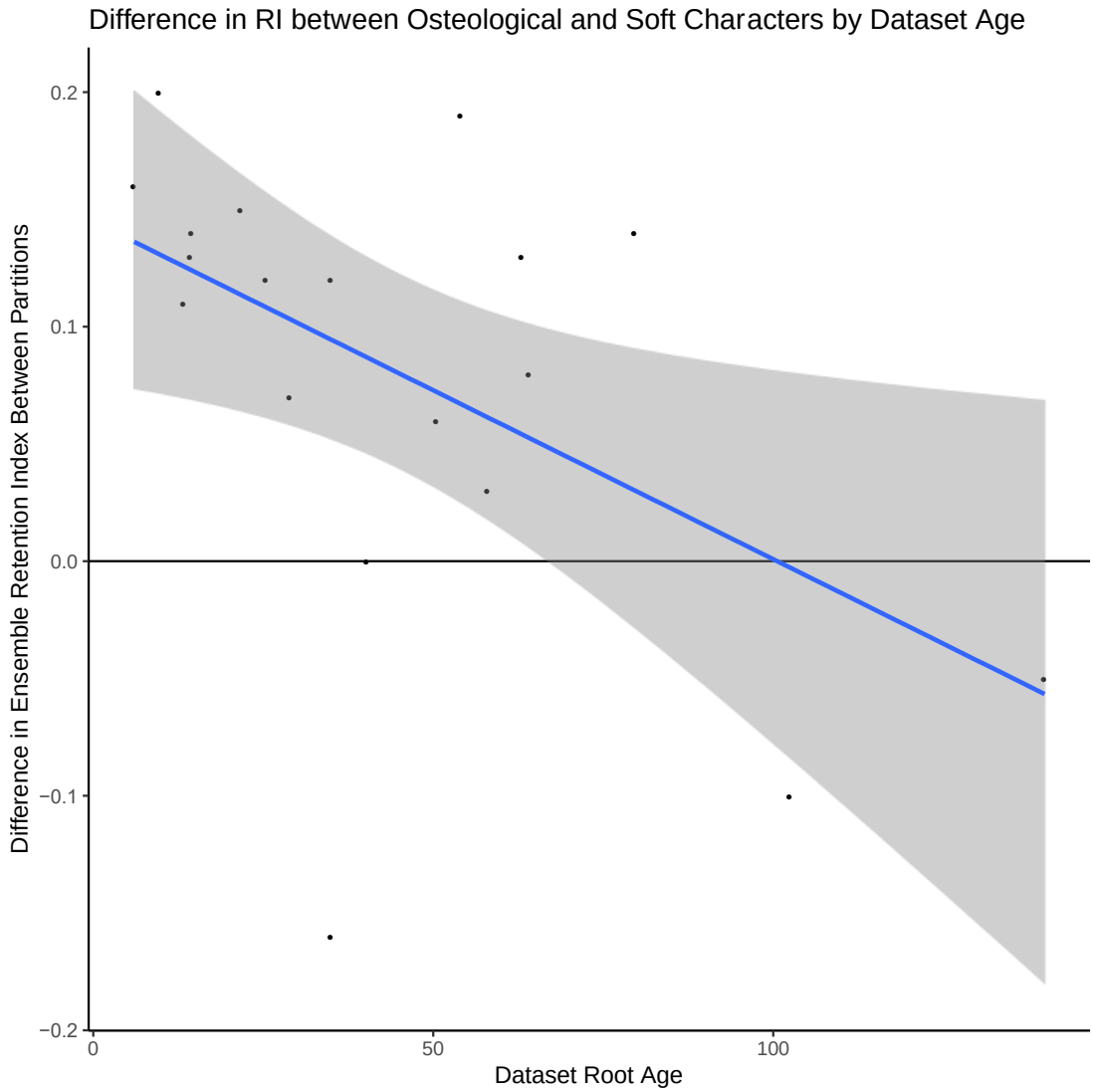
Supplementary Materials



Supplementary Figure 1. *Showing the retention index of individual characters on molecular trees by transition age in millions of years for each squamate and avian dataset. In most datasets, retention index tends to increase with age. Characters are coloured by partition, with osteological characters in purple and non-osteological characters in green.*



Supplementary Figure 2. *Showing the density of character transition ages in Millions of years on molecular trees individually for each squamate and avian dataset. In most datasets, average transition ages (dashed lines) are older for osteological than non-osteological. Plots are coloured by partition, with osteolgical data in purple and non-osteological data in green.*



Supplementary Figure 3. *Showing the difference between the ensemble retention indices of osteological and non-osteological partitions (osteological RI – non-osteological RI) on molecular trees within datasets, by root age of the dataset in millions of years. Regression line (p = 0.025) is in blue and standard error is indicated in grey. The difference between osteological and non-osteological fit to the molecular trees decreases and reverses in time, indicating a relatively poorer fit of osteological data, and relatievely better fit of non-osteological data, to molecular trees further back in time.*

Table 1. *Dataset dimensions and ensemble retention indices for partitions for each dataset*

				Osteological data		**Non-Osteological Data**		**P (Wilcoxon)**
Dataset	**Superorder/Order**	**Focal Clade**	**Taxa**	**Ensemble character retention index**	**Informative Characters**	**Ensemble character retention index**	**Informative Characters**	
Birds								
Chu, 1995	Charadriiformes	Various	70	0.84	62	0.76	5	0.1482
Smith, 2011	Charadriiformes	Pan-Alcidae	52	0.67	197	0.64	102	0.2762
Nesbitt, Ksepka, & Clarke, 2011	Cypselomorphae	Various	11	0.6	82	0.7	9	0.6089
Frank-Hoeflich *et al.*, 2007	Galloanserae	Cracidae	65	0.85	149	0.71	43	5.614e-05**
Livezey, 1991	Galloanserae	Anatini	49	0.79	28	0.67	115	0.1449
Livezey, 1996b	Galloanserae	Anserinae	25	0.85	31	0.71	51	0.0009672**
Livezey, 1996a	Galloanserae	Aythyini	12	0.75	7	0.59	38	0.4868
McCracken *et al.*, 1999	Galloanserae	Oxyurinae	11	0.65	25	0.54	28	0.1237
Livezey, 1998	Gruiformes	Rallidae	74	0.93	180	0.74	100	1.545e-11**
Birdsley, 2002	Passeriformes	Tyrannidae	31	0.58	27	0.46	12	0.9741

Irestedt, Fjeldså, & Ericson, 2004	Passeriformes	Dendrocolaptinae	38	0.79	9	0.66	21	0.08464
Maurício *et al.*, 2012	Passeriformes	Rhynocryptidae	34	0.76	80	0.7	8	0.6788
Patten & Fugate, 1998	Passeriformes	Emberizidae	27	0.47	6	0.4	24	0.8549
Prum, 1992	Passeriformes	Pipridae	30	0.69	29	0.49	10	0.04598*
Prum, 1993	Passeriformes	Eurylaimidae and Philepittidae	12	0.86	16	0.86	11	0.898
Winker & Pruett, 2006	Passeriformes	Catharus	17	0.41	11	0.26	29	0.1931
Clarke *et al.*, 2007	Sphenisciformes	Various	30	0.92	96	0.79	81	0.02412*
Bertelli, Chiappe, & Mayr, 2014	Tinamiformes	Tinamidae	17	0.72	85	0.88	39	3.86e-05**
Livezey & Zusi, 2007	Various	Various	139	0.6	1528	0.65	436	0.002469**
Squamates								
Daza & Bauer, 2012	Gekotta	Sphaerodactylidae	12	0.34	227	0.41	17	0.4753
Hitchmough, 1997	Gekotta	Diplodactylinae	12	0.83	10	0.84	9	1
Frost *et al.*, 2001a	Iguania	Polychrotidae	26	0.77	31	0.56	40	0.1222

Frost *et al.*, 2001b	Iguania	Tropiduridae	28	0.68	32	0.56	46	0.4259
Vieira, Colli, & Bao, 2005	Iguania	Corytophanidae	10	0.38	49	0.21	8	0.1132
Arnold, 1997	Lacertoidea	Takydromus	15	0.52	9	0.46	17	0.7854
Kearney, 2003	Lacertoidea	Amphisbaenia	36	0.74	118	0.72	16	0.3713
Yi & Norell, 2013	Lacertoidea	Various	25	0.62	209	0.67	25	0.09413
Reeder *et al.*, 2015	Various	Various	135	0.74	533	0.61	93	6.82e-05**

Showing number of taxa, number of informative characters per partition, ensemble retention indices for partitions for each dataset and results of Mann-Whitney U *comparing the mean retention index between partitions, where the null hypothesis is that the two sets of retention indices are drawn from the same distribution. Because of the differences in taxonomic level, avian datasets are categorised by superorder or order, while squamate datasets are categorised by sub- or infraorder or superfamily.*

Table 1 cont. *Dataset dimensions and ensemble retention indices for partitions for each dataset*

Cranial		**Postcranial**		**P (Wilcoxon)**	**Plumage**		**Myology**		**P (Wilcoxon)**
Ensemble character retention index	**Informative Characters**	**Ensemble character retention index**	**Informative Characters**		**Ensemble character retention index**	**Informative Characters**	**Ensemble character retention index**	**Informative Characters**	
0.91	23	0.8	39	0.02452*					
0.63	51	0.68	146	0.5143	0.572505	30	0.675355	72	0.493
0.64	35	0.59	47	0.5251					
0.84	79	0.86	70	0.173					
0.81	5	0.79	23	0.1494					
	4		27	NA					
	0		7	NA					
	0		25	NA					
0.9	36	0.93	144	0.01465*					

0.7	7	0.5	20	0.2419					
	9		0	NA					
0.75	39	0.76	41	0.1701					
	5		1	NA					
	0		29	NA					
	0		16	NA					
0.18	5	0.56	6	0.4015					
0.86	37	0.95	59	0.003018**					
0.66	37	0.77	48	0.08348					
0.63	357	0.59	1171	0.0005366**	0.651811	98	0.623933	237	0.09542
0.33	152	0.36	75	0.7183					
	2		8	NA					

0.76	17	0.78	14	0.8555					
0.64	21	0.73	11	0.9351					
0.38	37	0.39	12	0.8625					
				NA					
0.75	88	0.7	30	0.6918					
0.62	169	0.65	40	0.2796	0.711538	10	0.608696	11	0.6851
0.76	411	0.67	122	0.000001994*					

Showing number of taxa, number of informative characters per partition, ensemble retention indices for partitions for each dataset and results of Mann-Whitney U *comparing the mean retention index between partitions, where the null hypothesis is that the two sets of retention indices are drawn from the same distribution. Because of the differences in taxonomic level, avian datasets are categorised by superorder or order, while squamate datasets are categorised by sub- or infraorder or superfamily.*

2. The Prevalence of Correlated Morphological Characters and their Effect on Phylogenetic Reconstruction

Abstract

The independence of characters is an important principle for phylogenetic analysis. The presence of semi-autonomous, correlated pairs or suites of morphological characters as a result of selection or functional and ontogenetic constraints on body plan has the potential to confound phylogenetic reconstruction, but the presence of such traits in morphological character matrices is rarely tested or accounted for. It is therefore important to assess the prevalence and impact of these characters on phylogenetic data. Here, a meta-analysis approach is taken to identify correlated character pairs in 12 avian and squamate reptile morphological datasets, comparing osteological and soft tissue partitions, under a maximum likelihood framework. We find that correlated character pairs are present in all datasets. Consideration of the distribution of correlated pairs within and between osteological and non-osteological partitions finds that levels of correlation differ between partitions within datasets. Additionally, characters are more correlated within than between these partitions. Furthermore, simulation of datasets and trees finds that methods are sensitive to the number of taxa in individual datasets and to tree shape. This has implications for phylogenetic analysis, as the presence of correlation and modularity is known to mislead tree topology.

1. Introduction

Correlation of morphological characters is an increasingly recognized problem in phylogenetics (Sadleir & Makovicky, 2008; Goswami & Polly, 2010; Goswami *et al.*, 2014; Guillerme & Brazeau, 2018; Billet & Bardin, 2019). Correlation can arise as a result of developmental, functional and ecological integration (Cheverud, 1996; Harmon *et al.*, 2005; Klingenberg,

2008). Distantly related species may evolve independently to be morphologically similar as a result of a combination of common evolutionary pressures and functional constraints resulting in a limited number of possible evolutionary outcomes (Felice & Goswami, 2018). These shared pressures may lead not only to evolutionary convergence, but also to the presence of pairs or suites of correlated morphological characters as a result of parallel adaptation to multi-dimensional niche space (Harmon *et al.*, 2005), a process known as concerted convergence (Patterson & Givnish, 2002; Holland *et al.*, 2010). These correlations are often associated with adaptive peaks suitable for the occupation of this shared niche space. In addition, some morphological traits are necessarily correlated as a result of shared developmental origins or functional relatedness e.g. dentition (Labonne *et al.*, 2014; Billet & Bardin, 2019; Wolsan *et al.*, 2019), limbs (Hallgrímsson, Willmore, & Hall, 2002; Fabre *et al.*, 2014) and cranium (Goswami, 2006a, 2007). Functional and developmental integration have been most often examined within a morphometric context, where associations between traits can be found by analyzing the covariation of landmarks across several species using shape analysis.

In a phylogenetic context, morphological information may be coded into characters exhibiting correlation owing to ecological similarity and convergence, possibly inflating the estimated phylogenetic relatedness between these species if common morphological characteristics are assumed to be the result of common ancestry. For example, the convergence of multiple morphological traits associated with several ecomorphotypes in Caribbean *Anolis* lizards is substantial enough that cluster analysis using morphological data places species occupying the same niche on different islands together. However, these lizards often are more closely related to different ecomorphotypes on the same island than to those occupying the same niche space on a neighboring island (Losos, 1998). Additionally, integration within the cranium, an area often subject to strong selection relating to diet and habitat, has been shown to affect phylogenetic analysis in both carnivores (Goswami & Polly, 2010) and crocodilians (Sadleir & Makovicky, 2008).

These examples in isolated clades serve as a warning of how both convergence and concerted evolution of morphological characters in distantly related species in response to shared niche space can affect tree topology when these processes are ignored. It is therefore essential to evaluate the strength and prevalence of correlated evolution of morphological traits across larger clades, using partitions of cladistic morphological data (Billet & Bardin, 2019), to more closely examine which areas of morphology are more subject to selection and which, by contrast, may contain greater phylogenetic signal.

1.1. Detecting Correlations in Phylogenetic Data

Morphometric and cladistic data may capture broadly similar information about organismal form (Hetherington *et al.*, 2015). Indeed, this is necessarily true, for example when looking at relative lengths or distances between traits (i.e., these can be and are both coded in cladistic characters, and captured in multivariate analysis), meaning that such associations, though most often explicitly captured by morphometric analyses, are likely to feature prominently in cladistic data.

In phylogenetic analysis using morphological characters, characters are usually assumed to be uncorrelated (Felsenstein, 1973; Emerson & Hastings, 1998). This is essentially a problem of pseudoreplication, where correlated suites of characters are treated as independent data points, potentially inflating relatedness between morphologically similar species. This leaves these analyses open to being misled (Sadleir & Makovicky, 2008; Goswami & Polly, 2010; Goswami *et al.*, 2014).

Over the last few decades, several methods have been developed to detect correlations in morphological characters both with (Sadleir & Makovicky, 2008; Beaulieu, O'Meara, & Donoghue, 2013) and without (O'Keefe & Wagner, 2001; Sadleir & Makovicky, 2008; Holland *et al.*, 2010) an underlying phylogeny. Most recently, model-based approaches implemented under a Bayesian or likelihood framework have been used to

detect pairs of correlated characters along a phylogeny. In these models, a pair of characters is correlated if the rate of change of a character on a phylogeny is dependent on the state of the second. For example, this is the case if the change of a character from state 0 to state 1 occurs more frequently when the second character is in state 1. Thus, a model in which characters change independently of the state of a second character is an uncorrelated model, and models in which transition rates of characters differ according the state of the second are correlated models. This approach requires two or more morphological characters, and a tree on which to map these characters. Model comparison methods can then be used to discriminate between correlated and uncorrelated models of character pair evolution. This can be implemented in either a maximum likelihood or Bayesian framework (Pagel, 1994; Beaulieu *et al.*, 2013). This is not often done on large or multiple datasets, however, possibly because it is computationally expensive.

Here, a maximum likelihood approach is taken to identify correlated character pairs in 13 avian and squamate datasets, including a large dataset spanning crown birds (Livezey and Zusi 2006). Previously, this method has been implemented with individual character pairs or small groups of characters to test specific hypotheses of converted evolution between characters (Leslie *et al.*, 2015; Caruso, Eisen, & Case, 2016; Sauquet *et al.*, 2017; Billet & Bardin, 2019). We use this method to assess patterns of correlation in all pairwise character pairs across entire morphological datasets spanning the Sauropsida. This approach gives us an overall view of the prevalence and strength of correlation in datasets used for phylogenetic analysis across a wide range of vertebrate species, giving insight into what is likely to be a significant issue in morphological phylogenetic analysis. Additionally, datasets were split into osteological and soft character partitions in order to compare levels of correlation within and between these data types.

2. Materials and Methods

2.1. Datasets

Datasets collected and compiled for previous work (Chapter 1) were used for the following analyses. Briefly, the literature was systematically searched for avian and squamate morphological character matrices. These were chosen on the basis of taxonomy, number of characters and number of taxa, with minimal or no taxonomic overlap between datasets. Studies with fewer than five characters per partition after character removal, and ten total taxa, were excluded from analysis.

To enable correlation analysis, certain classes of characters needed to be removed: 1) multistate characters, as current methods do not allow for the assessment of correlation of non-binary characters, and 2) characters with missing data. Phylogenetically uninformative characters were also removed. Although retaining taxa was prioritized over retaining characters. Several datasets were removed from analysis for having too few characters after character removal. The final data sample comprised 12 datasets spanning Sauropsida, including 8 avian and 4 squamate datasets (Table 1). In addition, a much larger dataset spanning Aves (Livezey & Zusi, 2007) was separately analysed. This is due to taxonomic overlap and potential differences in properties between it and larger datasets. A comparable dataset for squamates was rejected on account of missing data.

As this method requires a dated topology to estimate rates of morphological evolution, dated molecular trees for birds and reptiles (Jetz *et al.*, 2012; Zheng & Wiens, 2016) were sourced from the literature. A tree from the post-burn in posterior distribution of the Hackett et al. (2008) backbone from Jetz et al. (2012) was chosen for birds as this tree contains molecular information from 6,663 bird species (around two thirds of all bird species), making this the largest molecular phylogeny of birds in terms of number of species sequenced to date. Similarly, the Zheng and Wiens (2016) squamate topology was chosen for taxonomic spread and high species coverage.

Note that while this method takes into account phylogeny, it cannot distinguish between other kinds of correlated characters. For instance, some characters may be correlated due to logical contingency, pseudoreplication or because they are ordered. They may also be correlated due to pleiotropy or integration of traits that do not fit into these categories. This method identifies all such correlations, without implying that all characters identified as correlated should be removed from phylogenetic analysis.

2.2. Assessing Correlations in Empirical Data

Three models of character pair evolution were fitted to all binary character pairs in all datasets using the corDISC function in the package CorHMM (Beaulieu *et al.*, 2013) implemented in R (R Core Team, 2018), with default settings, namely equal weighting of all states at the root, no ascertainment bias correction, and ancestral states were estimated by giving the optimal state at each node for the entire tree as opposed to integrating over the tree. Note that uncorrected models may overestimate the amount of change along branches, however this is unlikely to affect overall results, which focus on the differences in correlation between partitions. These models apply either 1, 2, or 8 rates to the matrix of all possible ways that a character in each pair could transition, excluding simultaneous transitions (e.g. 0,0 – 1,1). Of the 3 models, the Equal Rates model (ER) is an uncorrelated model, in which all rates in the rate matrix are equal, indicating that state changes in one character are unrelated to the state of the second character. The Symmetrical (SYM) model, in which half of the rates (0,0 – 0,1, 0,0 – 1,0, 0,1 – 1,1 and 1,0-1,1) in the rate matrix are the different to the other half (0,1 – 0,0, 1,0 – 0,0, 1,1 – 0,1 and 1,1 – 1,0), and the All Rates Different (ARD) model in which all rates are different, were also applied to the datasets. These are both correlated models, in which the rate of change in one of the characters in the pair is dependent on the state of the other. The exception to this was the significantly larger Livezy and Zusi (2007) dataset, for which

only the ER and SYM models could be applied because of the computational limitations of interrogating a dataset of this size.

Models were compared using AICc weights. The AICc weights for the two correlated models were added to obtain the cumulative AICc weights for both correlated models. In case of Livezey and Zusi (2007), the AICc weights of only the SYM model was used to assess correlation levels. Character pairs were judged to be correlated if the cumulative AICc weight was equal to or higher than 0.95 after Sauquet *et al.*, (2017), to reflect common criteria for statistical significance. Lowering this threshold will result on more correlated character pairs.

All characters were classified as belonging to either osteological or non-osteological partitions, and percentages of correlated characters both within and between partitions was then calculated for each dataset, and a paired t test was performed on these values. Additionally, mean AICc values within datasets were calculated and compared using the Kruskall-Wallis test, and the between-dataset results were compared using a paired t test. We additionally performed Dunn tests on these values to determine significant differences between the cumulative AICc weights for correlated models between all 3 character pair classes (osteological, non-osteological and between partition character pairs). Datasets were then combined in order to test linear mixed effects models using AICc weights of all character pairs as the response variable. A null model was tested against models treating partition and clade as fixed effects, with dataset treated as a random effect in all models. Heteroskedasticity was tested for prior to analysis by comparing linear mixed effect models accounting for different patterns of heteroskedasticity, and the best model used in the final analysis.

In addition to the above tests, the Colless Index, which calculates the total difference in the number of leaves on each branch descended from each internal node, was computed to assess tree symmetry (Blum, François, & Janson, 2006). Therefore, the higher the Colless Index, the higher the asymmetry of the tree.

2.3. Sensitivity Analysis

Morphological datasets and trees containing 10, 15, 20 and 30 taxa and each containing 10 characters were artifactually constructed. More realistic datasets might contain more characters. However, since the number of characters does not affect the detection of pairwise correlations, smaller datasets were constructed as they are less computationally expensive and the results simpler to interpret. Two separate datasets each were manually constructed of 20 and 30 taxa for a total of 6 constructed datasets. These datasets sizes are roughly in line with the smaller empirical datasets used here, and thus reflect minimum realistic dataset sizes. Beginning with 10 taxa, characters with states 0 and 1 were first constructed that transition together 100% of the time (i.e. identical taxa). Characters were then added that change together 90%, then 80% of the time. Using the 10-taxon dataset as a base, 0s were added to the dataset for the additional taxa in the 15-taxon dataset and one each of the 20- and 30- taxon datasets (see supplementary Tables 2 and 3 for the full 30 taxon dataset). In the second 20- and 30- taxon datasets characters were added with varying degrees of correlation, including alternating characters, unchanging characters, and characters transitioning every 3 taxa (see supplementary Table 3 for second full 30 taxon dataset). Thus, these data contained binary characters ranging from weakly to strongly correlated. This examines the effect of the underlying structure of the data on the detection of correlated character pairs.

To test the effect of tree topology, particularly tree symmetry, both highly asymmetrical trees were constructed and random trees were simulated for all datasets. A 30-taxon random tree was first simulated, with taxa successively removed from this same tree to fit the 20-, 15- and 10- taxon datasets. This was to ensure consistent trees for all datasets apart from the addition or removal of taxa (see supplementary Figure 14 for the full 30-taxon random tree). Additionally, a single 10 character and 30 taxon dataset with binary character states randomly distributed among taxa and

characters was simulated, by generating 30 vectors containing 0s and 1s of length 10 with replacement.
All three models of character evolution described above were fitted to this constructed and simulated data to investigate how well these models detect correlated characters, how sensitive these methods are to the number of taxa in a dataset and tree shape, and whether type ii errors are prevalent with these methods.

3. Results

3.1. Correlations in Empirical Datasets

Correlated character pairs were found in all datasets. The highest degree of overall correlation as measured by average AICc weights of correlated models was observed in the Chu, (1995) and Smith, (2011) datasets, both Charadriiformes (Figure 1, Table 1, supplementary Figures 2 and 11). Among the least correlated datasets were the Galloanserae datasets (Livezey, 1996a,b, Figure 1, Table 1, supplementary Figures 7 and 8) and some of the Passerine datasets (Prum, 1992; Winker & Pruett, 2006, Figure 1, Table 1, supplementary Figures 10 and 12). Few differences were found in the distribution of correlated character pairs within and between character partitions. Significant differences were found between the number of correlated characters within and between partitions for 2 out of 10 avian and 1 out of 3 squamate datasets (Fisher's exact test, Table 1, Figure 1). A significant difference was found between average AICc weights within and between partitions for 5 out of 10 avian and 0 out of 3 squamate datasets (Mann-Whitney *U* test, Table 1, Figure 2). A paired t-test of the percentages of correlated characters between datasets reveals no significant difference, though with slightly higher values for within partitions (n = 12, $p = 0.1411$), while a paired t test of the mean AICc weights reveals no significant difference. Mean AICc weights and percentage of correlated characters are

higher in non-osteological partitions for most datasets, but this was not significant.

When the character pairs are considered as within osteology, within non-osteology or between partitions (rather than just within partitions and between partitions), 5 of 10 avian and 2 of 3 squamate datasets show a significant difference in percentages of correlated characters (Fisher's exact test, Table 1, Figure 2), and 7 avian and 1 squamate datasets show significant differences in mean AICc weight (Kruskall-Wallis test, Table 1), with 5 of these significant differences being between osteological and non-osteological character pairs (Dunn tests, supplementary Table 1). Character pairs are significantly more correlated within than between partitions in the larger Livezey & Zusi (2007) dataset (Table 1, supplementary Figure 13). Models of cumulative AICc weights including partition (osteology, non-osteology and between partition) as a fixed effect had significantly greater explanatory power than a null model excluding partition, and models including only within and between partition character pairs. Of all linear mixed effect models tested, the model including all character pair classes (osteology, non-osteology and between), clade (avian versus squamate) and their interaction as fixed effects and dataset as random effect had the greatest explanatory power as measured by AIC weights. The preferred model according to the BIC splits the data only into partitions and excludes clade (Table 2). Overall, these results indicate a significant effect of partition on explanatory power, but there is ambiguity as to whether clade is important. Models splitting the character pairs into just 2 classes (within partition and between partition character pairs) also showed significantly higher correlation within partitions. This meta-analysis includes only the data from the smaller datasets. See supplementary figures 1-13 for visualization of the correlations between individual character pairs and datasets.

Table 1 at End of Document

Table 1. *Showing differences between proportion and strength of correlation between character pair classes. Character pairs are considered correlated if their cumulative AICc weight for correlated (SYM and ARD) models of character pair evolution are equal to or higher than 0.95. Strength of correlation is measured by average cumulative AICc weights of correlated models of character pair evolution. Character pair classes include a) within-osteological data character pairs, b) within-non-osteological data character pairs, and c) between-partition character pairs. Characters are generally more correlated (as measured by average AICc weights) within than between partitions. Differences in proportion and strength of correlation between osteological and non-osteological characters are mixed.*

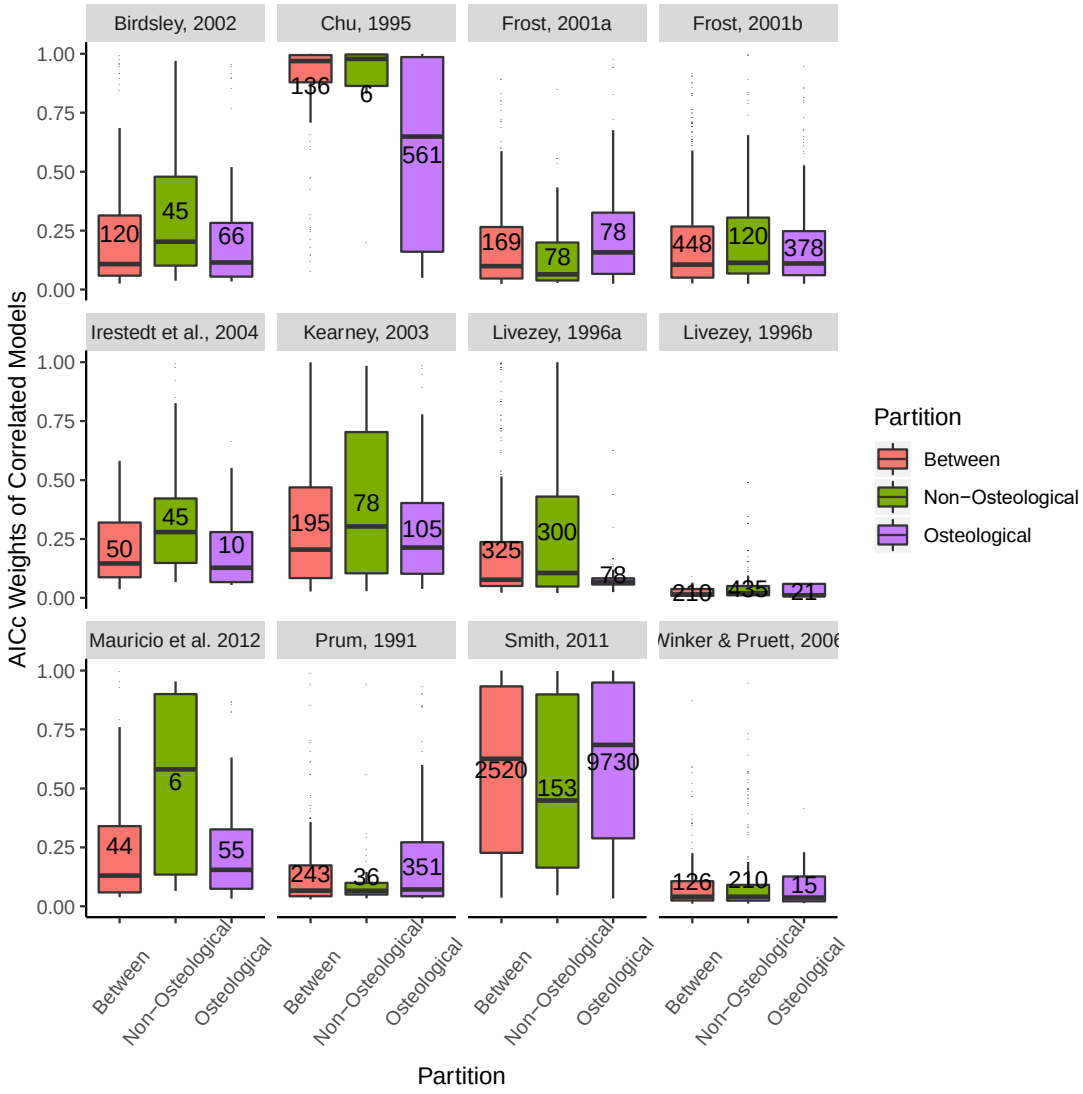


Figure 1. *Individual dataset boxplots of average cumulative AICc weights for correlated models of character pair evolution, showing all three classes of character pairs (within osteological, within non-osteological and between partition). Boxplots display number of characters. Lower means indicate lower correlation, higher means indicate higher correlation.*

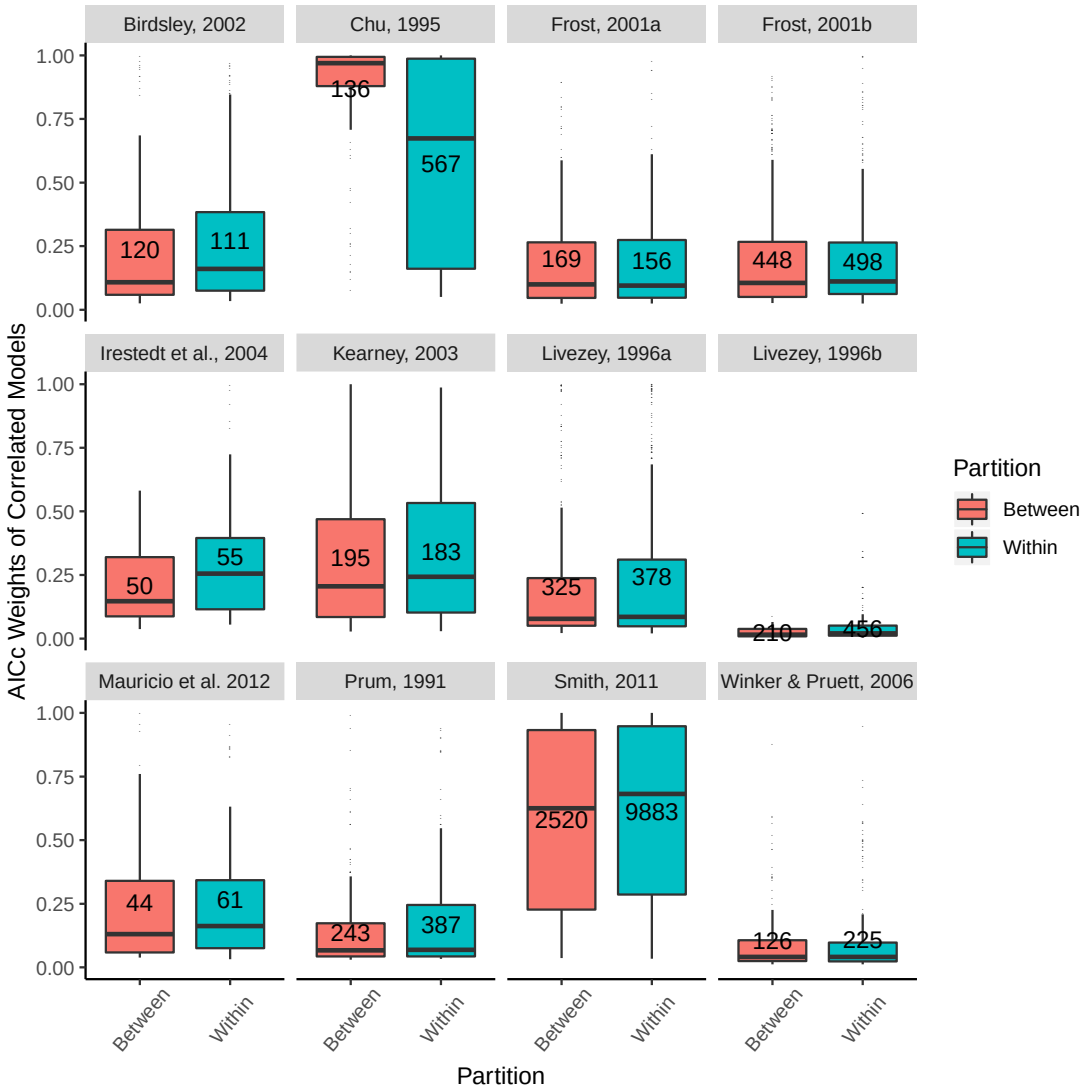


Figure 2. *Individual dataset boxplots of average cumulative AICc weights for correlated models of character pair evolution, showing within and between partition character pairs. Characters are more correlated within partitions in most datasets. Boxplots display number of characters. Lower means indicate lower correlation, higher means indicate higher correlation.*

Table 2. *The fit of five different linear mixed models comparing correlation levels within and between partitions*

Fixed Effects	df	AIC	BIC	Log L	Likelihood Ratio	P
None	15	6282	6399	-3126		
Character Pair Class WB (Within Partition Vs Between Partition)	16	6046	6171	-3007	238	<.0001
Character Pair Class All (Osteological Vs Non-Ostelogical Vs Between Partition)	17	5973	6105	-2969	76	<.0001
Character Pair Class All + Clade	18	5975	6115	-2970	0	0.7445
Character Pair Class All * Clade	20	5967	6123	-2964	12	0.0031

Table comparing the fit of five different linear mixed models comparing correlation levels within and between partitions. Models compared include a null model, a model including within-partition vs between partition character pairs as a fixed effect, a model including all three character pair classes as a fixed effect, and two models including all three character pair classes and clade as fixed effects. Dataset is included as a random effect in all models. AICs, likelihood ratios and P values (but not BIC) all prefer the most complex model, suggesting that partition and clade both have strong explanatory power. Models splitting character pairs into 3 rather than 2 classes have greater explanatory power, suggesting the importance of osteological vs non-osteological character pairs. The addition of clade as a fixed effect does not improve explanatory power except with an interaction term.

3.2. Sensitivity Analysis

In our simulated trees and datasets, artificially correlated morphological characters are not detected in either random or asymmetrical trees with as few as 10 or 15 taxa using the maximum likelihood method described above, and average AICc weights are lower in random trees (Table 3). Highly correlated characters are detected in trees with as few as 20 taxa. A higher number of correlated characters is detected in asymmetrical trees for the 10-taxon dataset, however average AICc weights are lower for random trees. In the second of the 20-taxon constructed datasets, the number of correlated characters is the same for both asymmetrical and random trees, while AICc weights are lower for random trees. More weakly correlated

characters are detected in trees with 30 taxa for both simulated datasets. Again, average AICc weights are lower in random than in asymmetrical trees. Overall, more correlated characters are detected in datasets with more taxa, and in asymmetrical trees. Characters that were constructed to be highly correlated have higher AICc weights (e.g. characters 1 and 2, and characters 5 and 6, supplementary Tables 1 and 2, supplementary Figure 13), while characters that were constructed to be less correlated have lower AICc weights (e.g. characters 2 and 4 in 30-taxon dataset 2, supplementary Table 2, supplementary Figure 13), indicating a low type i error rate. Correlated characters are not detected in the random simulated dataset on either tree, and average AICc weights are lower in the random dataset than in any of the constructed datasets except the 10- and 15- taxon matrices, indicating a low type ii error rate provided sufficient data.

Table 3. *number of correlated characters in constructed and simulated datasets*

Taxa	Asymmetrical tree corr %	Asymmetrical tree average weight	Asymmetrical tree sym score (Colless Index)	Random tree corr %	Random tree average weight	Random tree sym score (Colless Index)
10	0	0.013	36	0	0.006	18
15	0	0.179	91	0	0.029	42
20(1)	15.56	0.585	171	22.22	0.474	57
20(2)	8.89	0.405	171	8.89	0.163	57
30(1)	66.67	0.822	406	42.22	0.682	88
30(2)	42.22	0.709	406	24.44	0.549	88
30(Random)	0	0.103	406	0	0.106	88

Showing number of correlated characters (0.95 threshold) detected, average AICc weights for correlated models for constructed and simulated datasets trees containing 10, 15, 20 and 30 taxa, and tree symmetry scores.

4. Discussion

Overall, correlated character pairs in general are present in all datasets and the number of correlated character pairs is higher within than between partitions. The number of correlated characters also differs between partition (osteology and non-osteology). While there are significant differences in the number of correlated character pairs and strength of correlation between partitions within datasets, this varies between datasets. Thus, although non-osteological characters generally exhibit more within partition correlation than osteological characters, it is difficult to draw wider conclusions when analyzing individual datasets. Combining datasets finds that the distribution of character correlations differs between morphological partitions, with more correlation within non-osteological characters compared to osteological characters and more correlation within than between partition character pairs.

Previous work has shown that soft characters are more homoplastic than osteological characters (Chapter 1). We suggest that correlation and homoplasy in morphological characters are possibly related because of the concerted evolution of multiple morphological traits in response to selection. Concerted evolution occurs frequently in traits related strongly to ecology and environmental niche (Losos, 1998; Holland *et al.*, 2010; Leslie *et al.*, 2015; Caruso *et al.*, 2016; Billet & Bardin, 2019), and integration as detected by morphometric analysis is often strong in areas such as the cranium, which is often associated with strong ecological pressures resulting form dietary requirements (Goswami, 2006a, 2007; Sadleir & Makovicky, 2008). Ecologically important traits are likely to also be convergent in distantly related species that occupy a similar niche space (Losos, 1998). Indeed, entire body plans are sometimes repeated independently in several lineages (Van Valkenburgh, 2007). The results presented here and previously together are consistent with the idea that characters are correlated as a result of ecological and functional necessity, and that the evolutionary history of such characters is likely to reflect this, overwriting phylogenetic signal and becoming convergent.

While integration is clearly and consistently observed in traits relating to vertebrate body plan such as in the limbs and cranium (Hallgrímsson *et al.*, 2002; Goswami, 2006b; Sadleir & Makovicky, 2008), the results presented here suggest that soft tissues are at least equally and possibly more subject to concerted evolution. This may be especially true of evolutionarily labile characters, such as plumage, which are often subject to pressures such as sexual selection (Rubenstein & Lovette, 2009; Dunn, Armenta, & Whittingham, 2015) and less affected by constraint. Such characters may be more evolutionarily labile, and are therefore able to transition more frequently along lineages. Further, pseudoreplication may be an issue for plumage, since feathers tend to be structurally similar. There may be pleiotropic effects governing the development of feathers and colouration. Correlations between characters transitioning more frequently along branches may be detected at a higher rate than characters evolving more slowly on the same tree, since a greater number of transitions may offer greater statistical power. This therefore may alternatively explain the stronger apparent correlation between soft character pairs, since previous work has shown that soft tissue characters resolve on shallower branches on evolutionary trees, possibly implying faster rates (Chapter 1). However, a comparison of evolutionary rates between partitions would help to clarify this.

The presence of correlated characters in these datasets presents a problem for phylogenetic analysis that has been previously acknowledged but usually not accounted for. In particular, the stronger correlations detected in soft characters, whether because of inherently higher levels of correlation or greater detection owing to higher transition rates, may be prone to misleading phylogenetic analysis owing to pseudoreplication in these characters amplifying the signal.

Further to establishing the apparent prevalence and pattern of correlated characters in empirical datasets, it was necessary to test the sensitivity of the methods used. The results of the sensitivity analysis unsurprisingly show that correlations are not detected in trees with few taxa, potentially explaining the apparent paucity of correlated character pairs in the smaller

datasets analysed here. This is because the method used detects correlations by fitting models of character change over an evolutionary tree. Specifically, it detects whether a change in one character is dependent on the state of the second. The higher the number of taxa, therefore, the greater the statistical power and ability to detect significant concerted evolutionary change over a tree. This presents a problem for smaller morphological datasets, and suggests that a high number of taxa should be prioritized over a high number of characters when constructing or editing datasets with the view to detect correlated characters. A second factor in detecting correlations is tree shape. More correlations are detected in asymmetrical than random trees. Since the topology of empirical trees is likely to be closer to random trees than highly symmetrical trees, this may present less of a challenge when detecting correlations. However, tree shape should nevertheless be determined and accounted for in studies detecting correlations.

The methods described above allow us to identify pairs of correlated characters. Whilst this is a good way of identifying overall prevalence of correlation within datasets, it does not allow for the identification of entire suites, or modules, of correlated characters. This is a problem as correlated characters do not evolve together in pairs, but rather as entire groups of characters united by common function or ecology.

5. Conclusion

These results demonstrate that correlated character pairs are present in morphological datasets. Although soft characters generally exhibit more correlation, this differs between datasets. Correlated character pairs can be identified using existing maximum likelihood methods provided that there is a reliable independent (molecular) tree and a high enough number of taxa. However, even strongly correlated character pairs cannot be reliably identified in datasets with fewer than 20 taxa, with fewer correlated characters also detected in less symmetrical trees. These results show

empirically that correlated evolutionary change is frequently captured in morphological phylogenetic data, and thus has the potential to affect phylogenetic inference. This is particularly true of soft characters, which show a higher overall level of correlation than osteological characters. We therefore recommend that correlations are accounted for in models when building trees from morphological data using probabilistic methods, when molecular trees are available to use as a benchmark. In addition, when testing for correlated character pairs, it is advisable to use trees with at least 20-30 species.

These results are part of a broader emerging picture of the differences in the phylogenetic signal conveyed by different areas of morphology. For example, differences in phylogenetic signal have been tested for and found between dental and osteological characters (Sansom, Wills, & Williams, 2017) hard and soft characters (Sansom & Wills, 2017) and cranial and postcranial characters (Mounce, Sansom, & Wills, 2016). Differences in integration, as well as in levels of homoplasy and evolutionary rates, between morphological regions is likely to have an effect on phylogenetic inference, and thus be an important consideration in future work when developing tree-building methods using morphological data. In particular, models splitting morphology into partitions, and allowing these partitions to evolve at different rates may improve phylogenetic inference. When differences in evolutionary dynamics in pre-defined partitions are included in such models, this reduces the need for methods to find partitions, reducing computational effort and potentially reducing the risk of model overfitting.

The apparently lower levels of correlations in osteological characters may be reassuring to paleontologists when reconstructing evolutionary relationships between extinct species using data from fossils.

References

Beaulieu JM, O'Meara BC, Donoghue MJ. 2013. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. *Systematic Biology* 62: 725–737.

Billet G, Bardin J. 2019. Serial homology and correlated characters in morphological phylogenetics: modeling the evolution of dental crests in Placentals. *Systematic Biology* 68: 267–280.

Blum MGB, François O, Janson S. 2006. The mean, variance and limiting distribution of two statistics sensitive to phylogenetic tree imbalance. *The Annals of Applied Probability* 16: 2195–2214.

Caruso CM, Eisen K, Case AL. 2016. An angiosperm-wide analysis of the correlates of gynodioecy. *International Journal of Plant Sciences* 177: 115–121.

Cheverud JM. 1996. Developmental Integration and the Evolution of Pleiotropy. *American Zoologist* 36: 44–50.

Chu PC. 1995. Phylogenetic reanalysis of Strauch's osteological data set for the Charadriiformes. *The Condor* 97: 174–196.

Dunn PO, Armenta JK, Whittingham LA. 2015. Natural and sexual selection act on different axes of variation in avian plumage color. *Science Advances* 1: e1400155.

Emerson SB, Hastings PA. 1998. Morphological correlations in evolution: consequences for phylogenetic analysis. *Source: The Quarterly Review of Biology* 73: 141–162.

Fabre AC, Goswami A, Peigné S, Cornette R. 2014. Morphological integration in the forelimb of musteloid carnivorans. *Journal of anatomy* 225: 19–30.

Felice RN, Goswami A. 2018. Developmental origins of mosaic evolution in the avian cranium. *Proceedings of the National Academy of Sciences of the United States of America* 115: 555–560.

Felsenstein J. 1973. Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Systematic Biology* 22: 240–249.

Goswami A. 2006a. Morphological integration in the carnivoran skull.

Evolution 60: 169–183.
Goswami A. 2006b. Cranial modularity shifts during mammalian evolution. *The American naturalist* 168: 270–80.
Goswami A. 2007. Phylogeny, diet, and cranial integration in australodelphian marsupials. *PLoS ONE* 2: e995.
Goswami A, Polly DP. 2010. The influence of character correlations on phylogenetic analyses: a case study of the carnivoran cranium. In: Goswami A,, In: Friscia A, eds. *Carnivoran Evolution: New Views on Phylogeny, Form, and Function*. Cambridge: Cambridge University Press, 141–164.
Goswami A, Smaers JB, Soligo C, Polly PD. 2014. The macroevolutionary consequences of phenotypic integration: from development to deep time. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369: 20130254–20130254.
Guillerme T, Brazeau MD. 2018. Influence of different modes of morphological character correlation on phylogenetic tree inference. *bioRxiv*: 308742.
Hackett SJ, Kimball RT, Reddy S, Bowie RCK, Braun EL, Braun MJ, Chojnowski JL, Cox WA, Han KL, Harshman J, Huddleston CJ, Marks BD, Miglia KJ, Moore WS, Sheldon FH, Steadman DW, Witt CC, Yuri T. 2008. A Phylogenomic Study of Birds Reveals Their Evolutionary History. *Science* 320: 1763–1768.
Hallgrímsson B, Willmore K, Hall BK. 2002. Canalization, developmental stability, and morphological integration in primate limbs. *American journal of physical anthropology* 35: 131–58.
Harmon LJ, Kolbe JJ, Cheverud JM, Losos JB. 2005. Convergence and the multidimensional niche. *Evolution* 59: 409–421.
Hetherington AJ, Sherratt E, Ruta M, Wilkinson M, Deline B, Donoghue PCJ. 2015. Do cladistic and morphometric data capture common patterns of morphological disparity? *Palaeontology*.
Holland BR, Spencer HG, Worthy TH, Kennedy M. 2010. Identifying cliques of convergent characters: concerted evolution in the cormorants and shags. *Systematic Biology* 59: 433–445.
Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO. 2012. The global

diversity of birds in space and time. *Nature* 491: 444–8.
Klingenberg CP. 2008. Morphological integration and developmental modularity. *Annual Review of Ecology, Evolution, and Systematics* 39: 115–132.
Labonne G, Navarro N, Laffont R, Chateau-Smith C, Montuire S. 2014. Developmental integration in a functional unit: deciphering processes from adult dental morphology. *Evolution and Development* 16: 224–232.
Leslie AB, Beaulieu JM, Crane PR, Knopf P, Donoghue MJ. 2015. Integration and macroevolutionary patterns in the pollination biology of conifers. *Evolution* 69: 1573–1583.
Livezey BC. 1996a. A phylogenetic analysis of modern pochards (Anatidae: Aythyini). *The Auk* 113: 74–93.
Livezey BC. 1996b. A phylogenetic analysis of geese and swans (Anseriformes: Anserinae), including selected fossil species. *Systematic Biology* 45: 415–450.
Livezey BC, Zusi RL. 2007. Higher-order phylogeny of modern birds (Theropoda, Aves: Neornithes) based on comparative anatomy. II. Analysis and discussion. *Zoological Journal of the Linnean Society*.
Losos JB. 1998. Contingency and determinism in replicated adaptive radiations of island lizards. *Nature* 279: 2115–2118.
Mounce RCP, Sansom R, Wills MA. 2016. Sampling diverse characters improves phylogenies: craniodental and postcranial characters of vertebrates often imply different trees. *Evolution* 70: 666–686.
O'Keefe FR, Wagner PJ. 2001. Inferring and testing hypotheses of cladistic character dependence by using character compatibility. *Systematic Biology* 50: 657–675.
Pagel M. 1994. Detecting Correlated Evolution on Phylogenies: A General Method for the Comparative Analysis of Discrete Characters. *Proceedings of the Royal Society B: Biological Sciences* 255: 37–45.
Patterson TB, Givnish TJ. 2002. Phylogeny, concerted convergence, and phylogenetic niche conservatism in the core Liliales: insights from rbcL and ndhF sequence data. *Evolution* 56: 233–252.
Prum R. 1992. Syringeal morphology, phylogeny, and evolution of the

neotropical manakins (Aves, Pipridae). *American Museum Novitates* 3043: 1–65.
R Core Team. 2018. R: A language and environment for statistical computing.
Rubenstein DR, Lovette IJ. 2009. Reproductive skew and selection on female ornamentation in social species. *Nature* 462: 786–789.
Sadleir RW, Makovicky PJ. 2008. Cranial shape and correlated characters in crocodilian evolution. *Journal of Evolutionary Biology* 21: 1578–1596.
Sansom RS, Wills MA. 2017. Differences between hard and soft phylogenetic data. *Proceedings of the Royal Society B: Biological Sciences* 284: 20172150.
Sansom RS, Wills MA, Williams T. 2017. Dental data perform relatively poorly in reconstructing mammal phylogenies: morphological partitions evaluated with molecular benchmarks. *Systematic Biology* 66: 813–822.
Sauquet H, von Balthazar M, Magallón S, Doyle JA, Endress PK, Bailes EJ, Barroso de Morais E, Bull-Hereñu K, Carrive L, Chartier M, Chomicki G, Coiro M, Cornette R, El Ottra JHL, Epicoco C, Foster CSP, Jabbour F, Haevermans A, Haevermans T, Hernández R, Little SA, Löfstrand S, Luna JA, Massoni J, Nadot S, Pamperl S, Prieu C, Reyes E, dos Santos P, Schoonderwoerd KM, Sontag S, Soulebeau A, Staedler Y, Tschan GF, Wing-Sze Leung A, Schönenberger J. 2017. The ancestral flower of angiosperms and its early diversification. *Nature Communications* 8: 16047.
Smith NA. 2011. Taxonomic revision and phylogenetic analysis of the flightless Mancallinae (Aves, Pan-Alcidae). *ZooKeys* 91: 1–116.
Van Valkenburgh B. 2007. Deja vu: the evolution of feeding morphologies in the Carnivora. *Integrative and Comparative Biology* 47: 147–163.
Winker K, Pruett CL. 2006. Seasonal migration, speciation, and morphological convergence in the genus Catharus (Turdidae). *The Auk* 123: 1052–1068.
Wolsan M, Suzuki S, Asahara M, Motokawa M. 2019. Dental integration and modularity in pinnipeds. *Scientific Reports* 9: 4184.
Zheng Y, Wiens JJ. 2016. Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Molecular Phylogenetics*

and Evolution 94: 537–547.

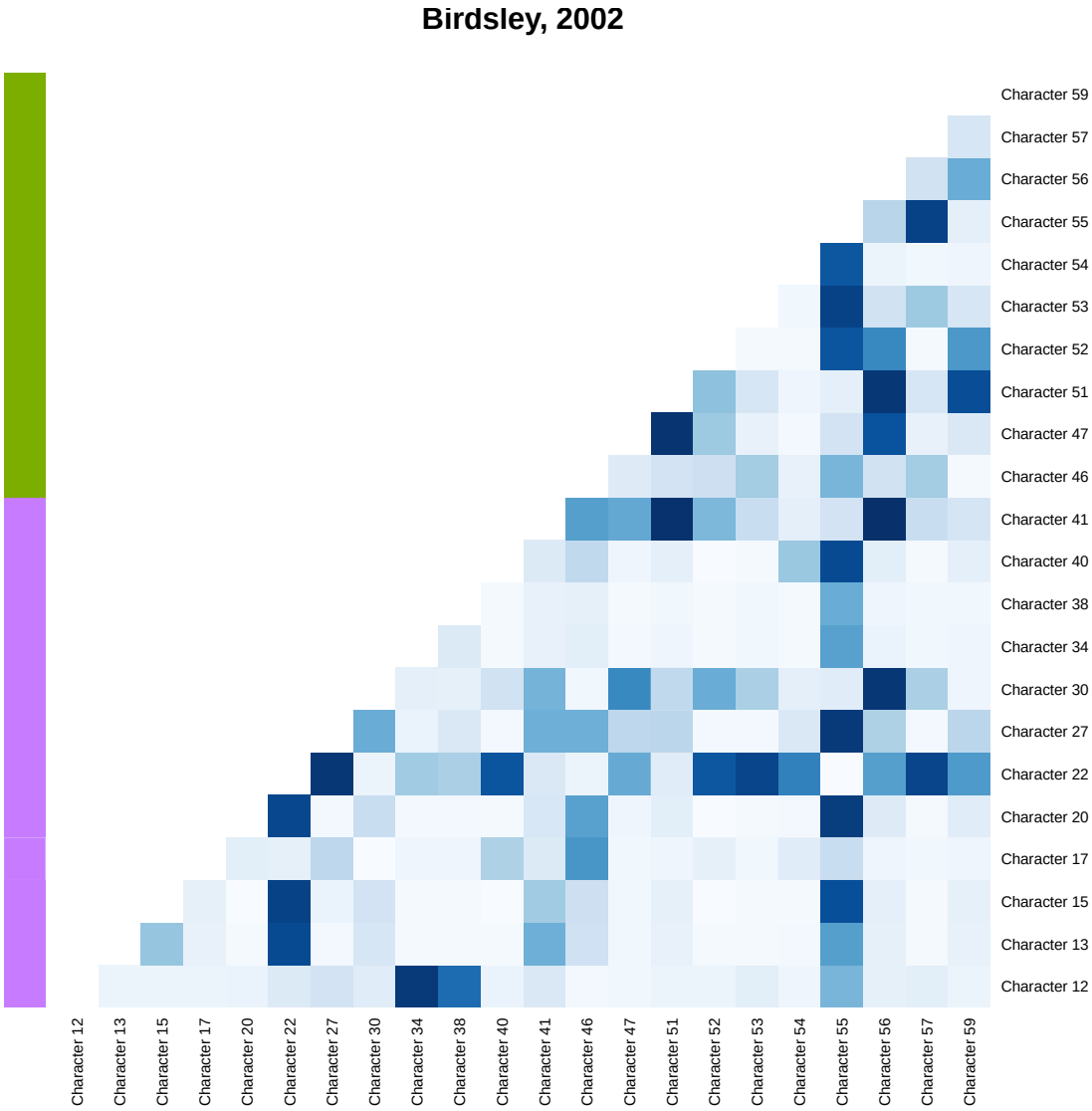
Supplementary Information

1. Empirical Data

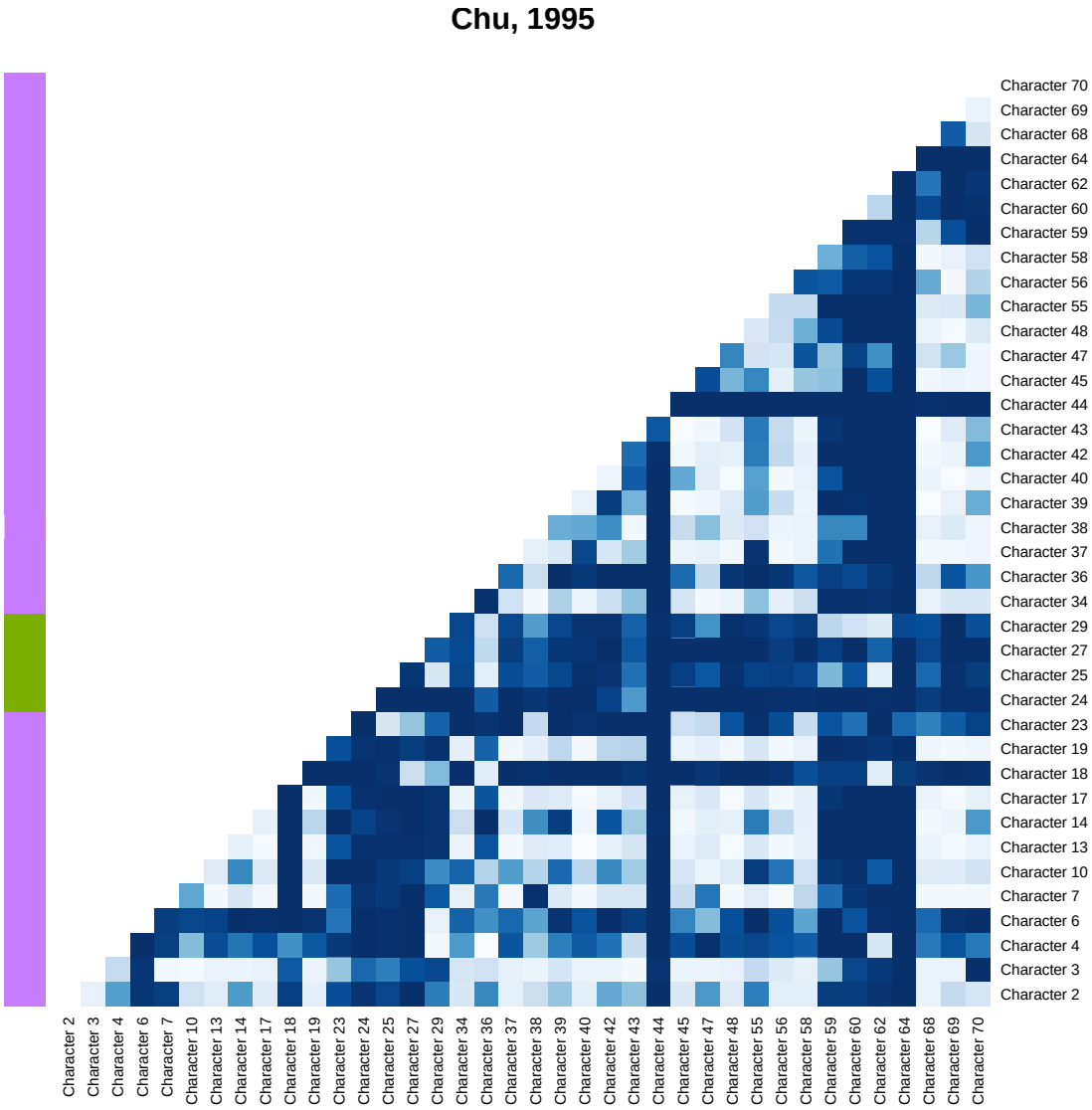
Supplementary Table 1. *results of Dunn tests of differences in correlation between character pair classes*

Dataset	Ost-NonOst	Ost-Between	NonOst-Between
Birds			
Birdsley, 2002	0.0117*	0.462	0.0048*
Chu, 1995	0.0447	0.0000*	0.4551
Irestedt, Fjeldså, & Ericson, 2004	0.0266	0.3765	0.0029*
Livezey, 1996a	0.0002*	0.0106*	0.0186*
Livezey, 1996b	0.0741	0.4851	0.0001*
Maurício *et al.*, 2012	0.0584	0.4788	0.0577
Prum, 1992	0.3503	0.1792	0.4791
Smith, 2011	0.0000*	0.0000*	0.0015*
Winker & Pruett, 2006	0.4403	0.4296	0.4706
Livezey & Zusi, 2007	0.0000*	0.0000*	0.0000*
Squamates			
Frost *et al.*, 2001b	0.0707	0.3514	0.0393
Frost *et al.*, 2001a	0.0000*	0.0214*	0.0028*
Kearney, 2003	0.0512	0.4454	0.0258

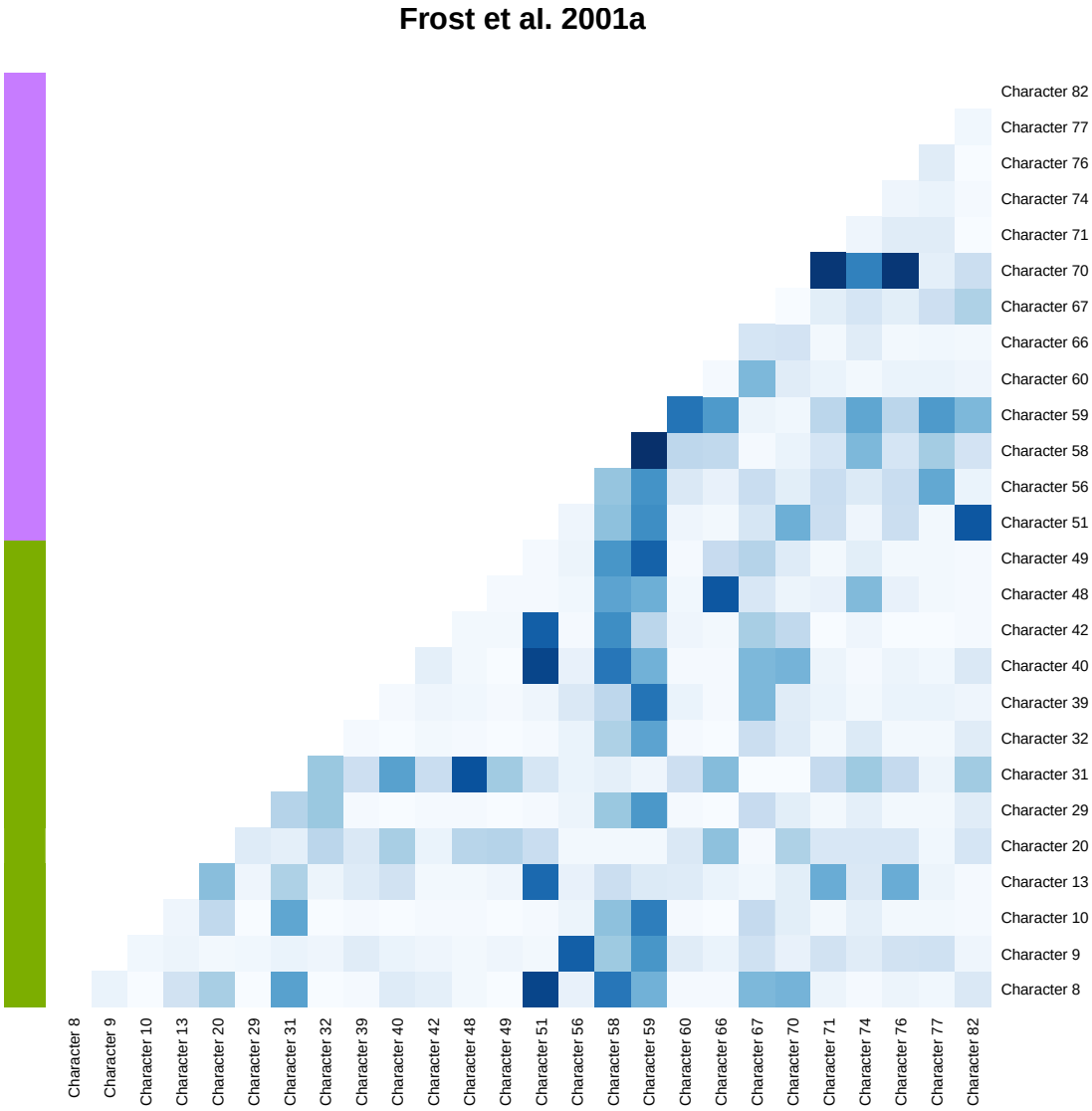
Showing results of the Dunn tests of differences in the cumulative AICc weights of correlated models of character pair evolution between three character pair classes (within-osteological, within-non-osteological, and between partition), finding significant differences in correlation between osteological and non-osteological partitions in only 5 datasets.



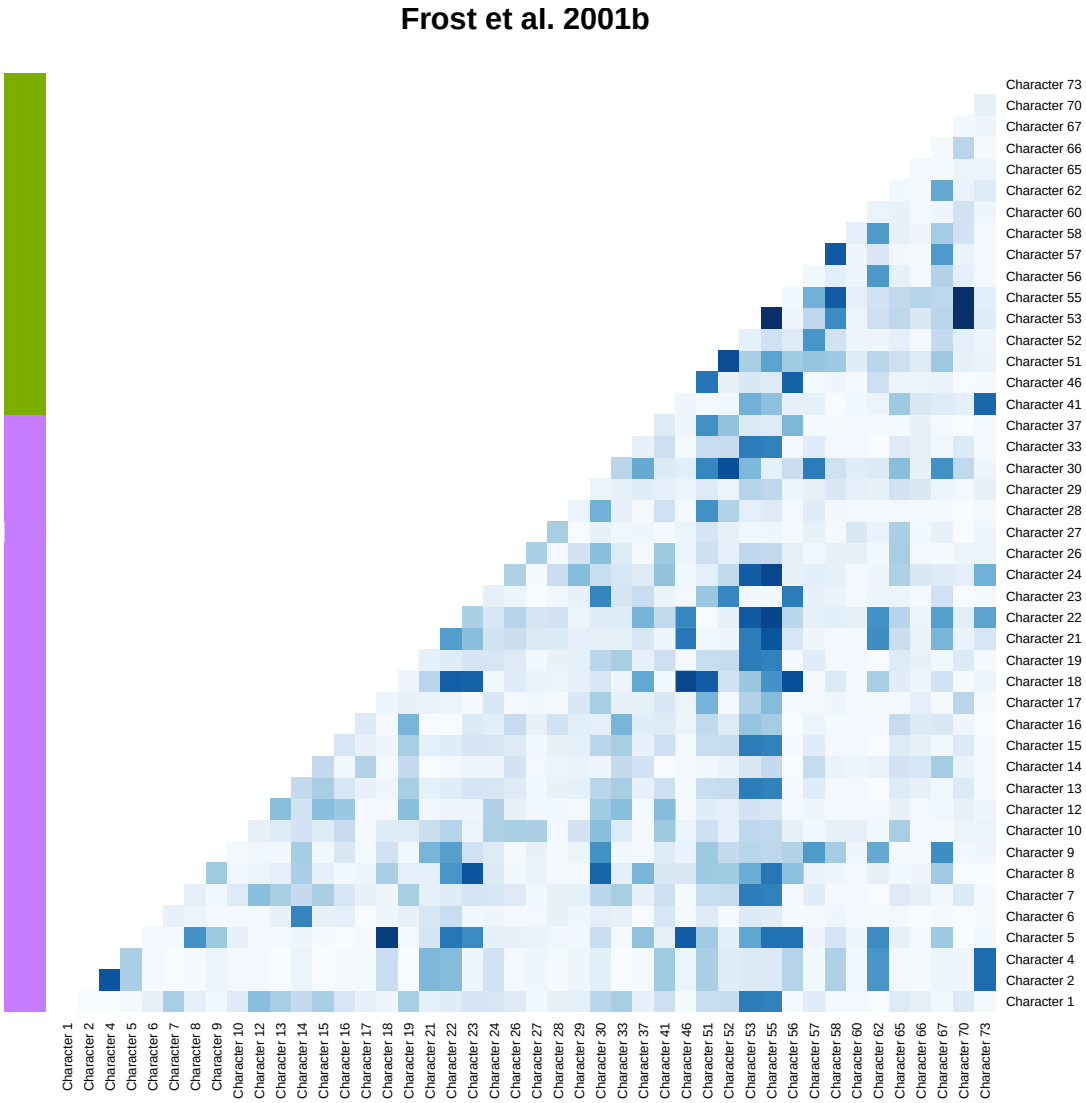
Supplementary Figure 1. *Heatmap showing the degree of correlation, as measured by AICc weight of fitted correlated models (SYM and ARD), between all character pairs tested in the Birdsley, (2002) dataset. Lighter cells show weaker correlation, while darker cells show stronger correlation. The sidebar delimits the partitions, with osteological characters in purple and non-osteological characters in green.*



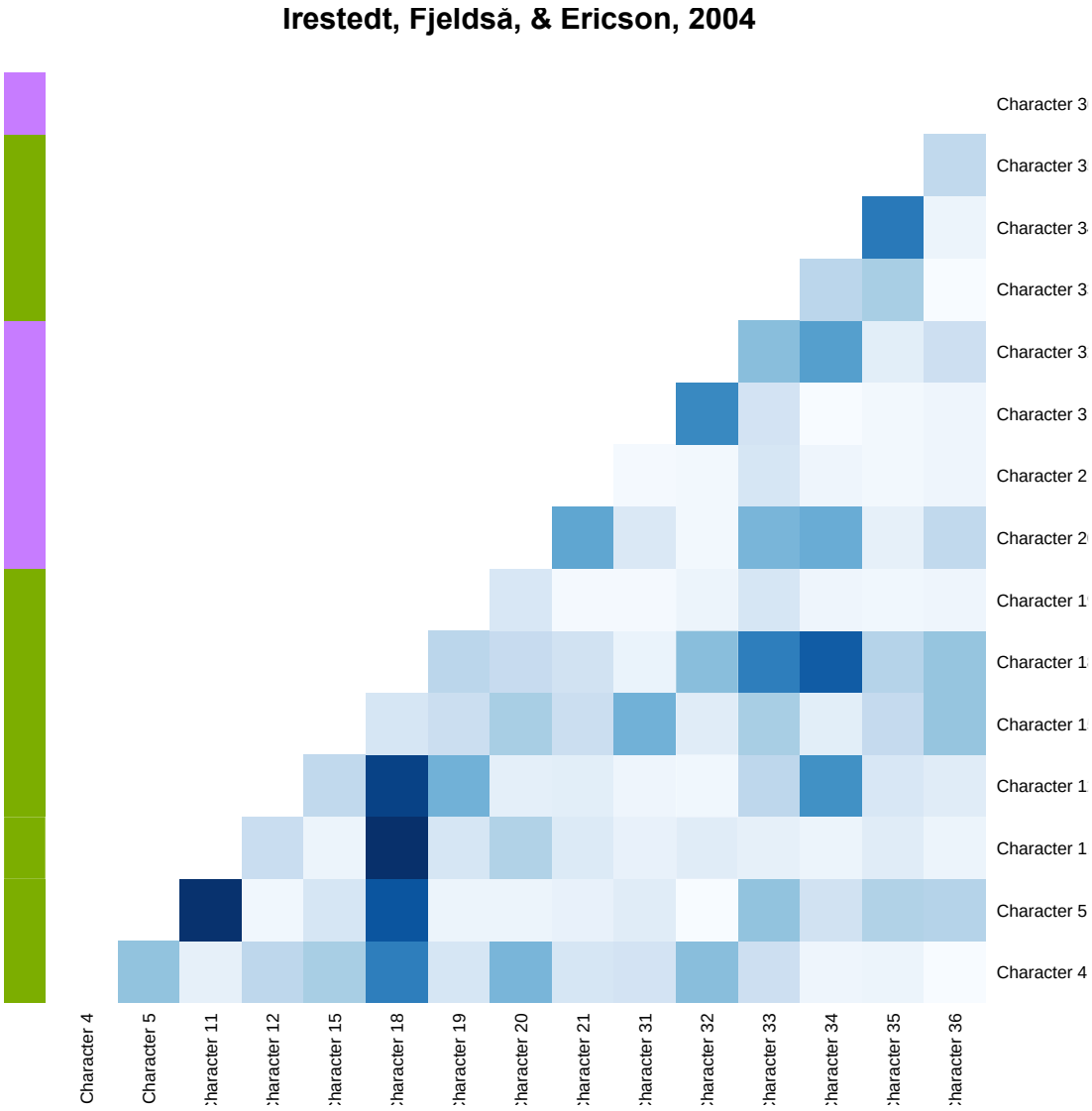
Supplementary Figure 2. *Heatmap showing the degree of correlation, as measured by AICc weight of fitted correlated models (SYM and ARD), between all character pairs tested in the Chu, (1995) dataset. Lighter cells show weaker correlation, while darker cells show stronger correlation. The sidebar delimits the partitions, with osteological characters in purple and non-osteological characters in green.*



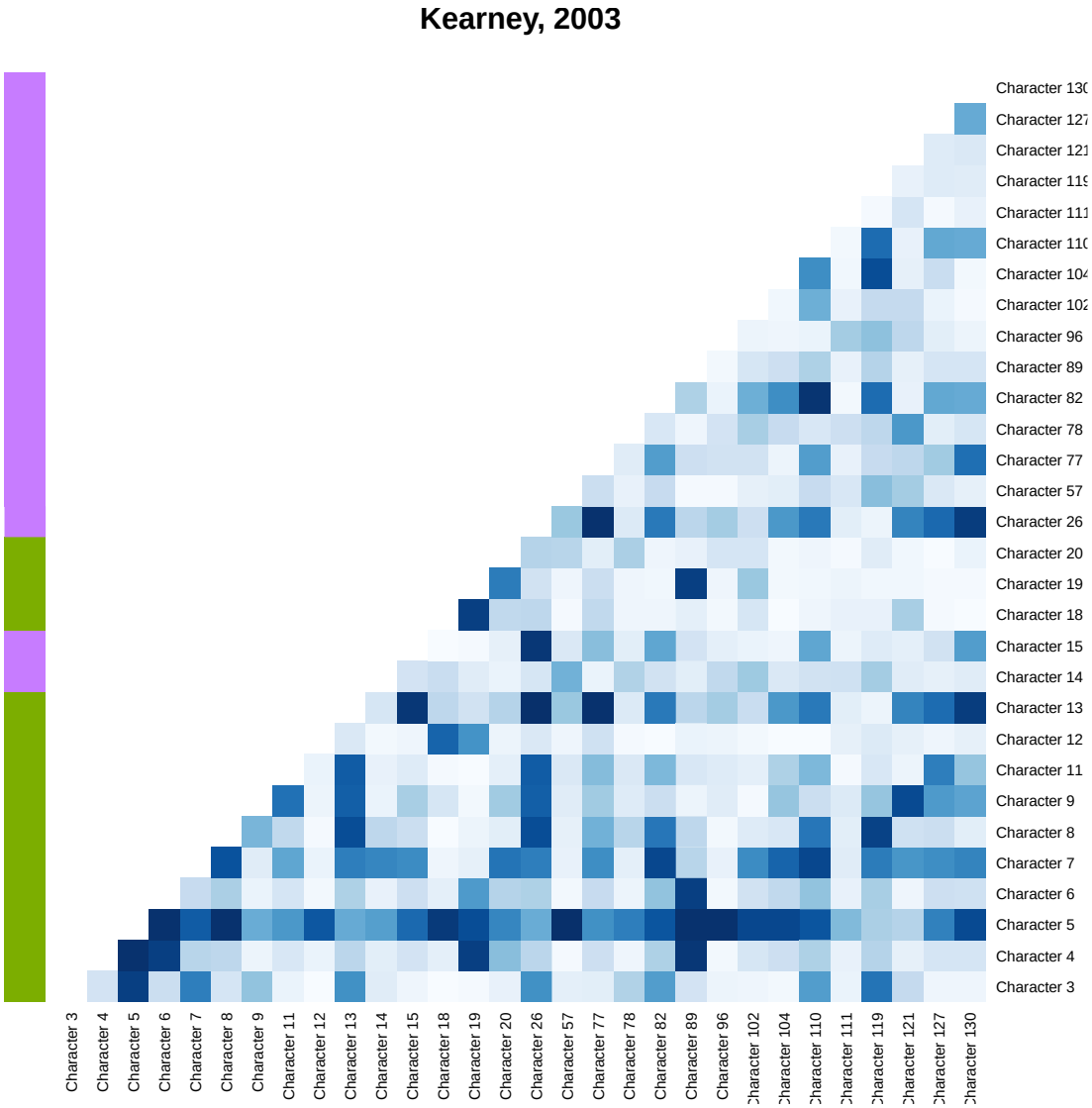
Supplementary Figure 3. *Heatmap showing the degree of correlation, as measured by AICc weight of fitted correlated models (SYM and ARD), between all character pairs tested in the Frost et al., (2001a) dataset. Lighter cells show weaker correlation, while darker cells show stronger correlation. The sidebar delimits the partitions, with osteological characters in purple and non-osteological characters in green.*



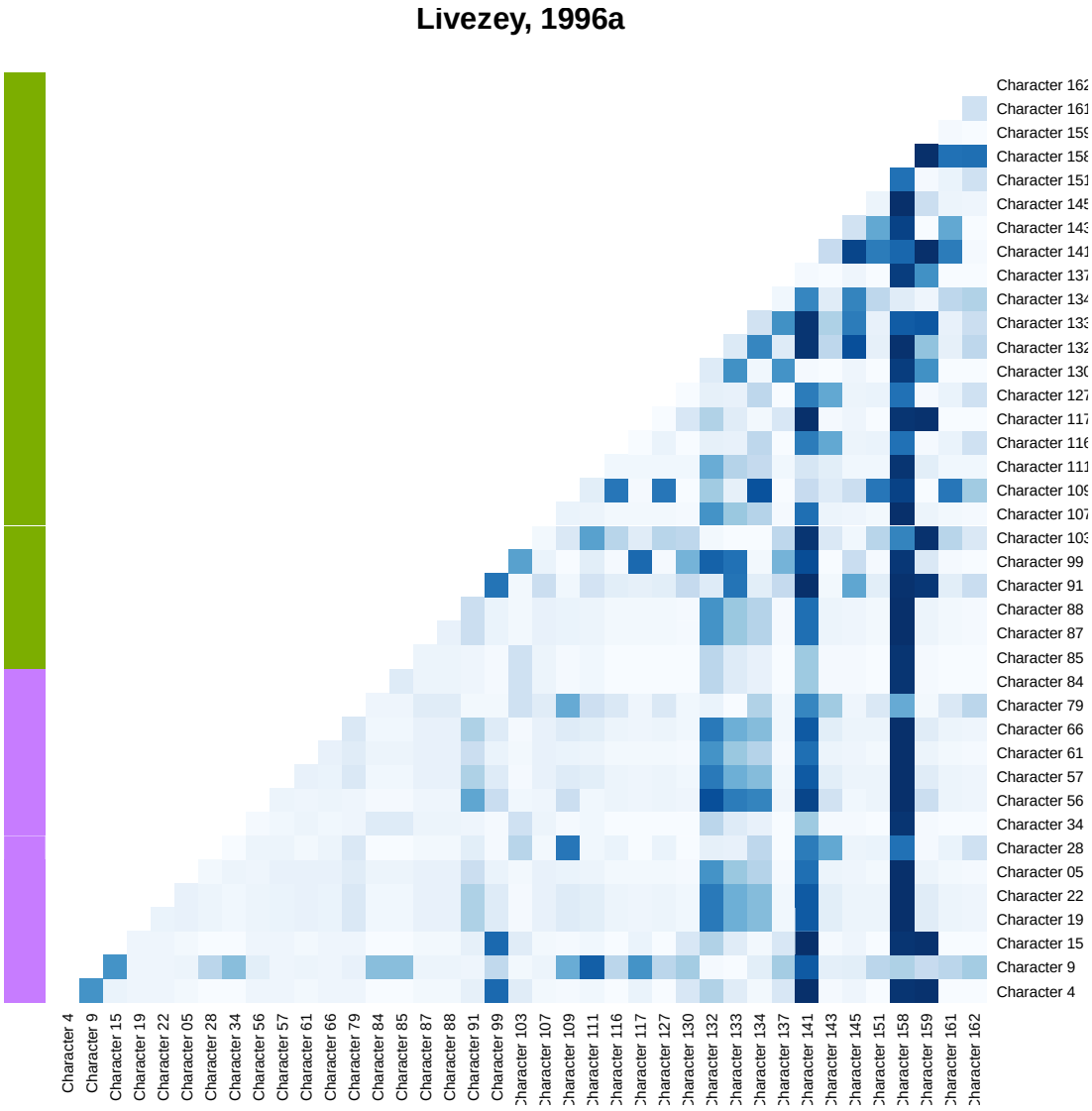
Supplementary Figure 4. *Heatmap showing the degree of correlation, as measured by AICc weight of fitted correlated models (SYM and ARD), between all character pairs tested in the Frost et al., (2001b) dataset. Lighter cells show weaker correlation, while darker cells show stronger correlation. The sidebar delimits the partitions, with osteological characters in purple and non-osteological characters in green.*



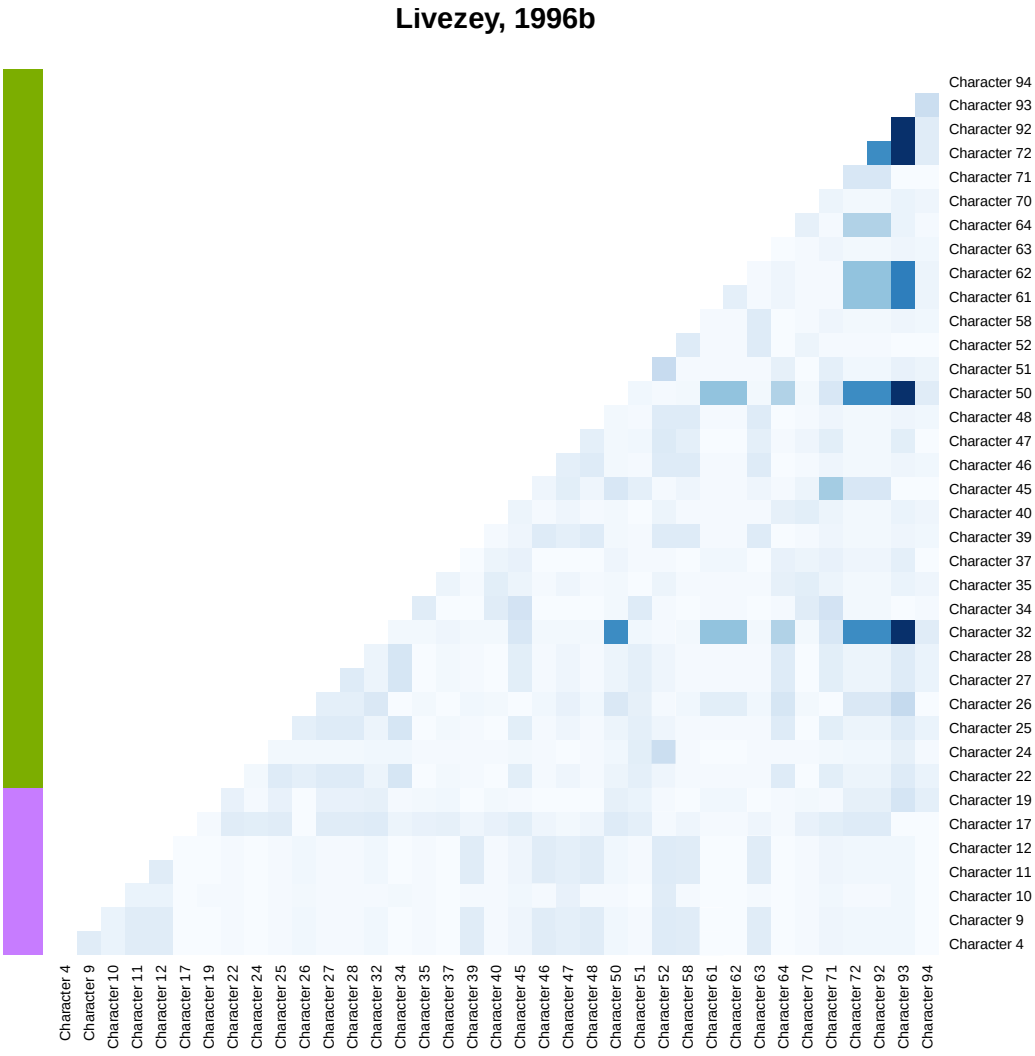
Supplementary Figure 5. *Heatmap showing the degree of correlation, as measured by AICc weight of fitted correlated models (SYM and ARD), between all character pairs tested in the Irestedt et al., (2004) dataset. Lighter cells show weaker correlation, while darker cells show stronger correlation. The sidebar delimits the partitions, with osteological characters in purple and non-osteological characters in green.*



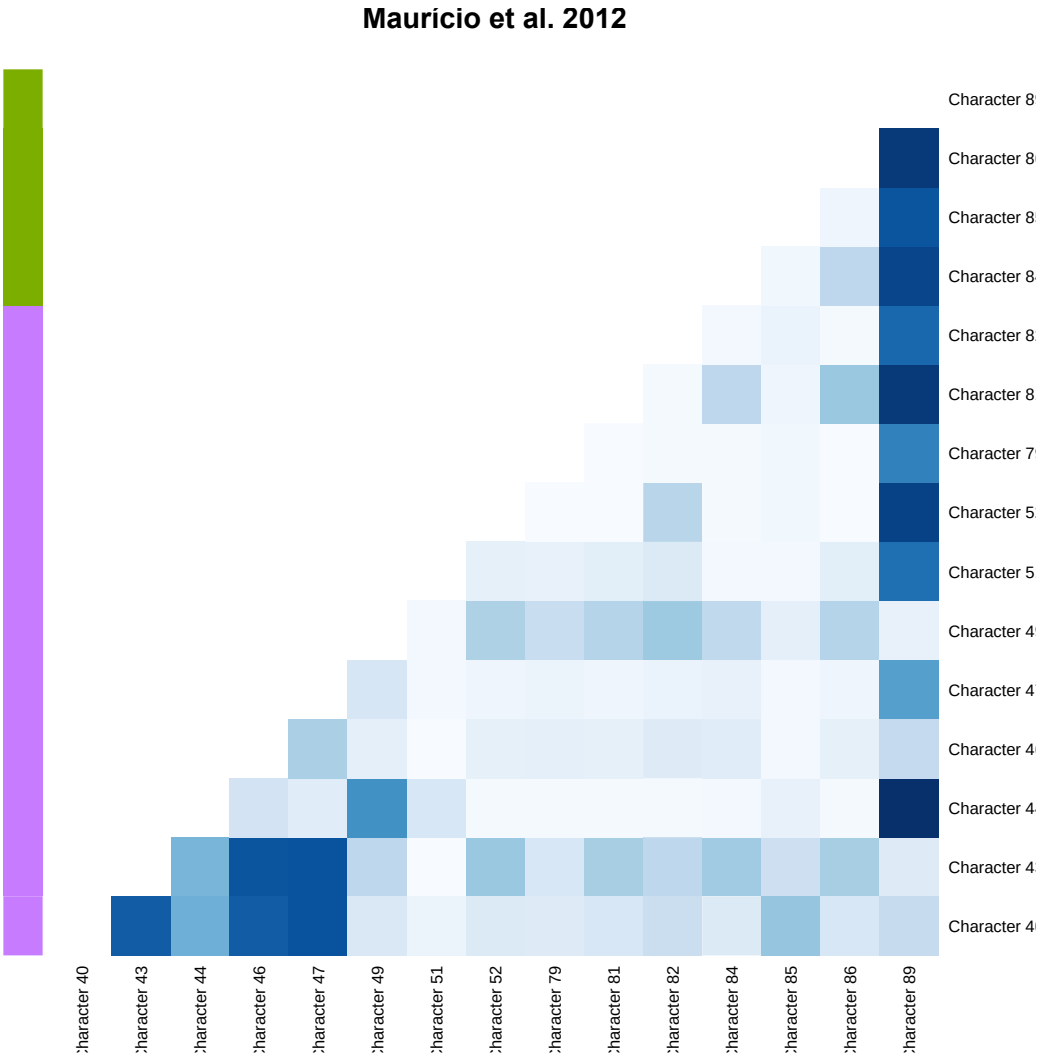
Supplementary Figure 6. *Heatmap showing the degree of correlation, as measured by AICc weight of fitted correlated models (SYM and ARD), between all character pairs tested in the Kearney, (2003) dataset. Lighter cells show weaker correlation, while darker cells show stronger correlation. The sidebar delimits the partitions, with osteological characters in purple and non-osteological characters in green.*



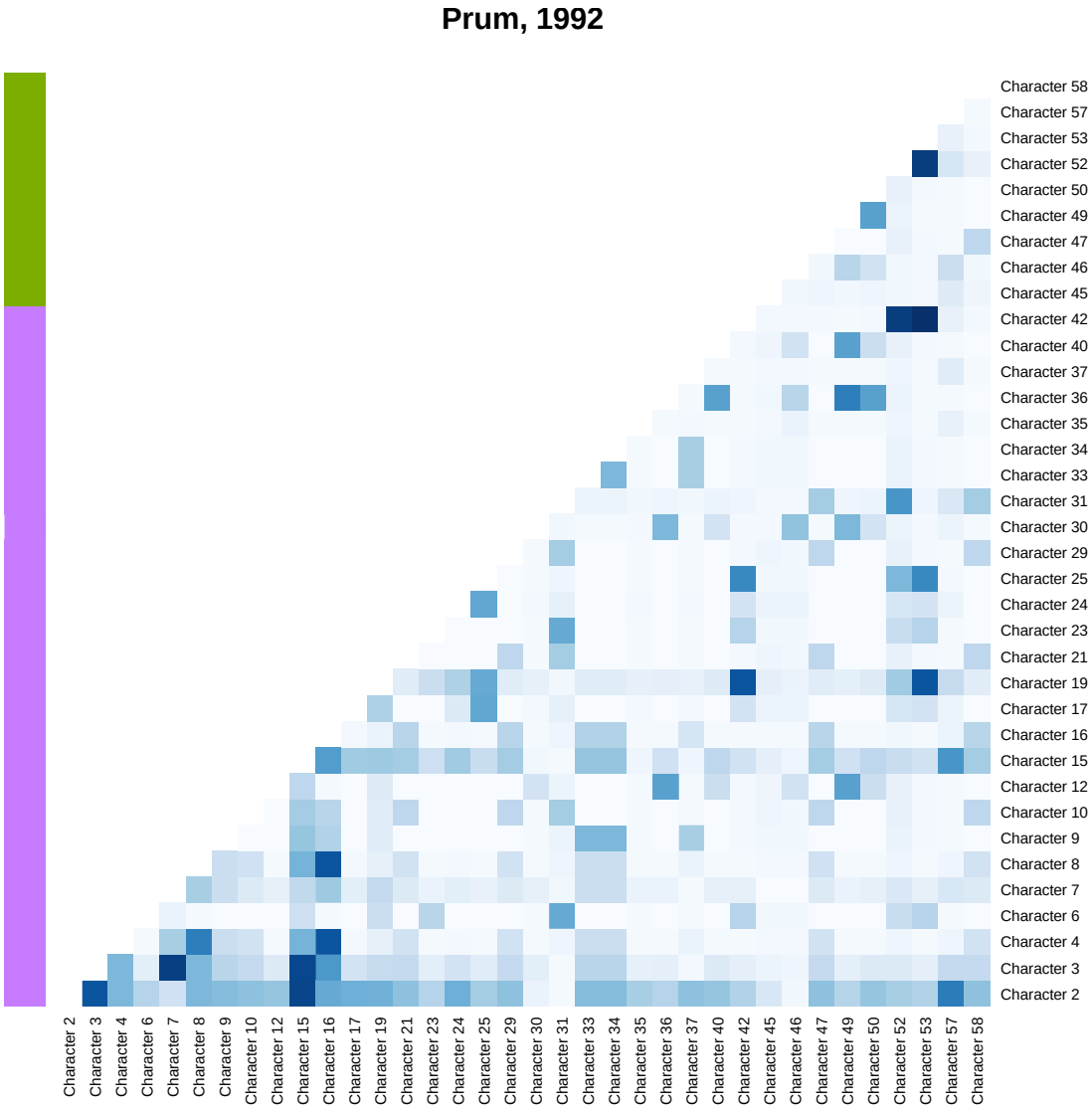
Supplementary Figure 7. *Heatmap showing the degree of correlation, as measured by AICc weight of fitted correlated models (SYM and ARD), between all character pairs tested in the Livezey, (1996a) dataset. Lighter cells show weaker correlation, while darker cells show stronger correlation. The sidebar delimits the partitions, with osteological characters in purple and non-osteological characters in green.*



Supplementary Figure 8. *Heatmap showing the degree of correlation, as measured by AICc weight of fitted correlated models (SYM and ARD), between all character pairs tested in the Livezey, (1996b) dataset. Lighter cells show weaker correlation, while darker cells show stronger correlation. The sidebar delimits the partitions, with osteological characters in purple and non-osteological characters in green.*

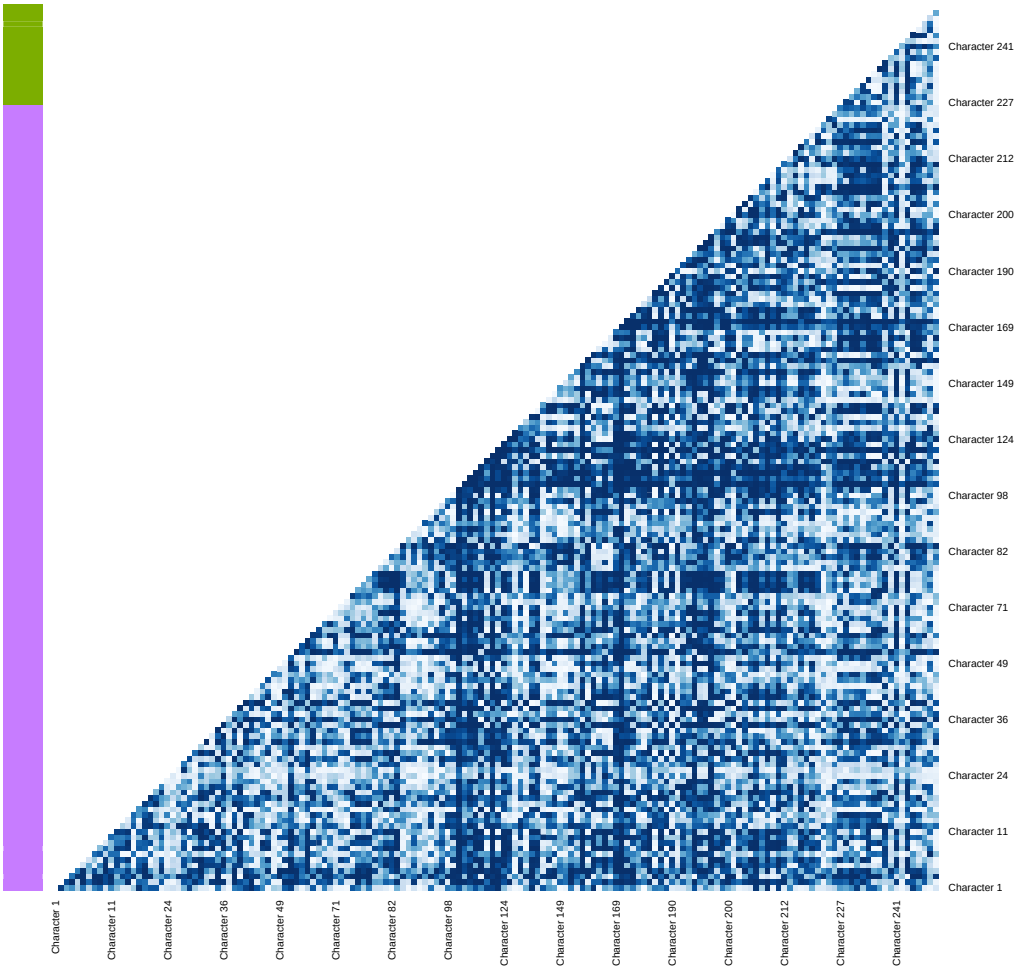


Supplementary Figure 9. *Heatmap showing the degree of correlation, as measured by AICc weight of fitted correlated models (SYM and ARD), between all character pairs tested in the Mauricio et al. (2012) dataset. Lighter cells show weaker correlation, while darker cells show stronger correlation. The sidebar delimits the partitions, with osteological characters in purple and non-osteological characters in green.*

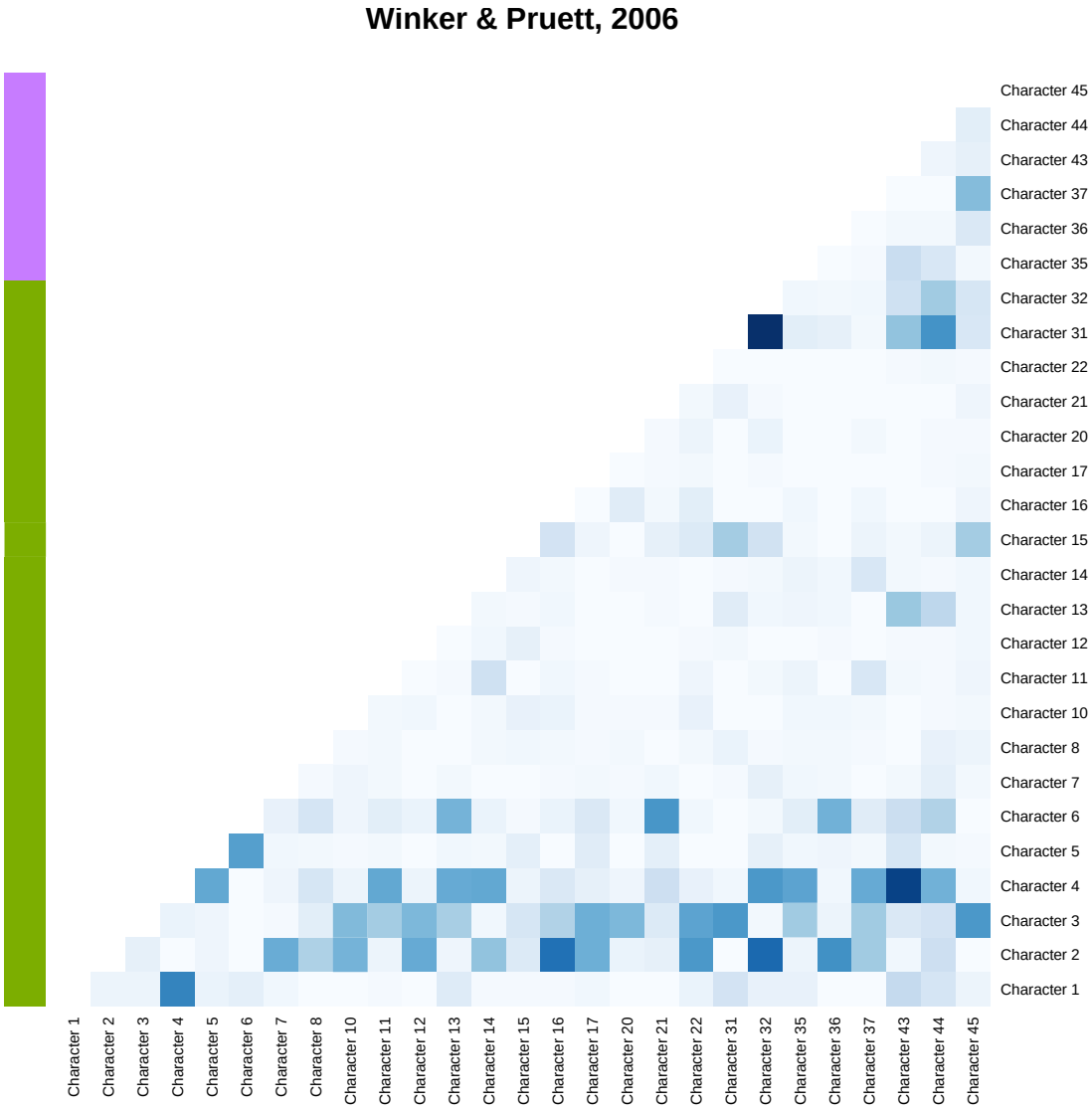


Supplementary Figure 10. *Heatmap showing the degree of correlation, as measured by AICc weight of fitted correlated models (SYM and ARD), between all character pairs tested in the Prum. (1992) dataset. Lighter cells show weaker correlation, while darker cells show stronger correlation. The sidebar delimits the partitions, with osteological characters in purple and non-osteological characters in green.*

Smith, 2011

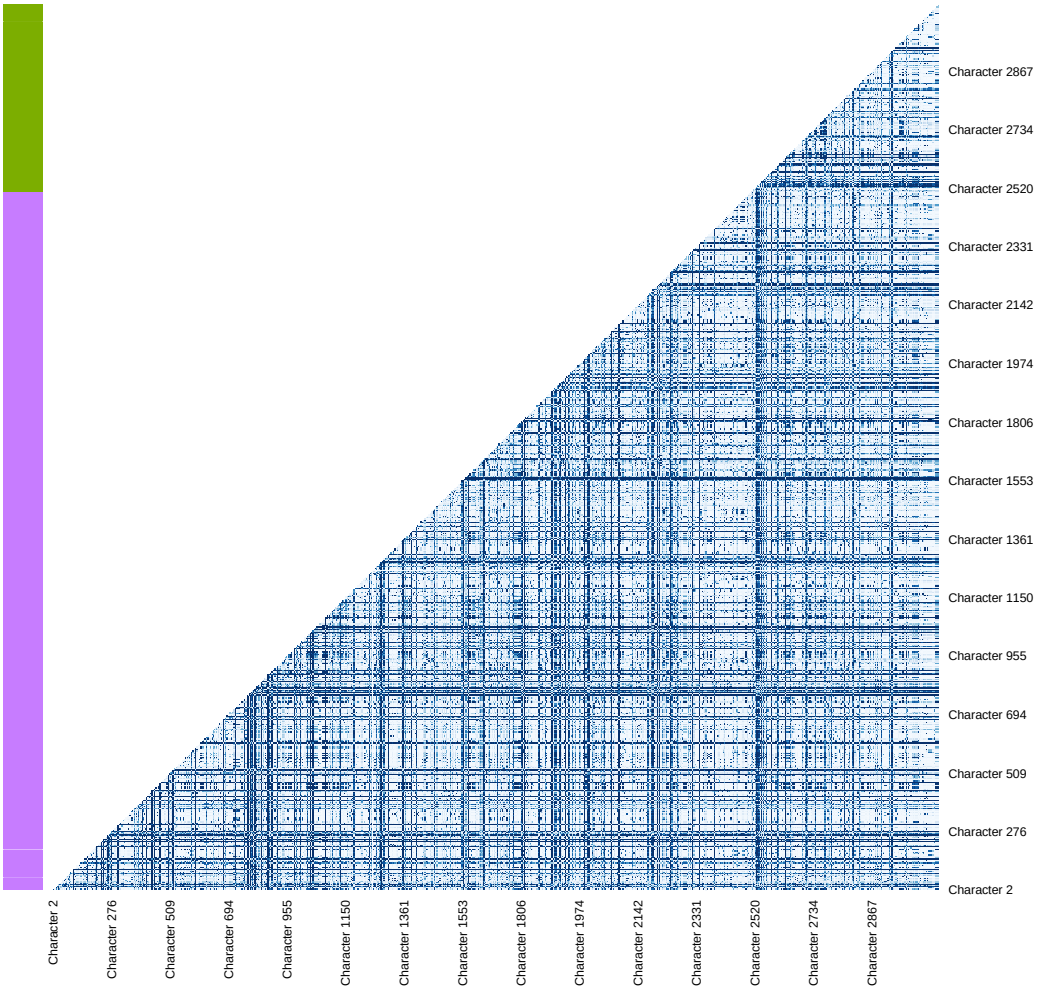


Supplementary Figure 11. *Heatmap showing the degree of correlation, as measured by AICc weight of fitted correlated models (SYM and ARD), between all character pairs tested in the Smith, (2011) dataset. Lighter cells show weaker correlation, while darker cells show stronger correlation. The sidebar delimits the partitions, with osteological characters in purple and non-osteological characters in green.*



Supplementary Figure 12. *Heatmap showing the degree of correlation, as measured by AICc weight of fitted correlated models (SYM and ARD), between all character pairs tested in the Winker & Pruett, (2006) dataset. Lighter cells show weaker correlation, while darker cells show stronger correlation. The sidebar delimits the partitions, with osteological characters in purple and non-osteological characters in green.*

Livezey & Zusi, 2006



Supplementary Figure 13. *Heatmap showing the degree of correlation, as measured by AICc weight of fitted correlated model (SYM), between all character pairs tested in the largest, Livezey & Zusi, (2006), dataset. Lighter cells show weaker correlation, while darker cells show stronger correlation. The sidebar delimits the partitions, with osteological characters in purple and non-osteological characters in green.*

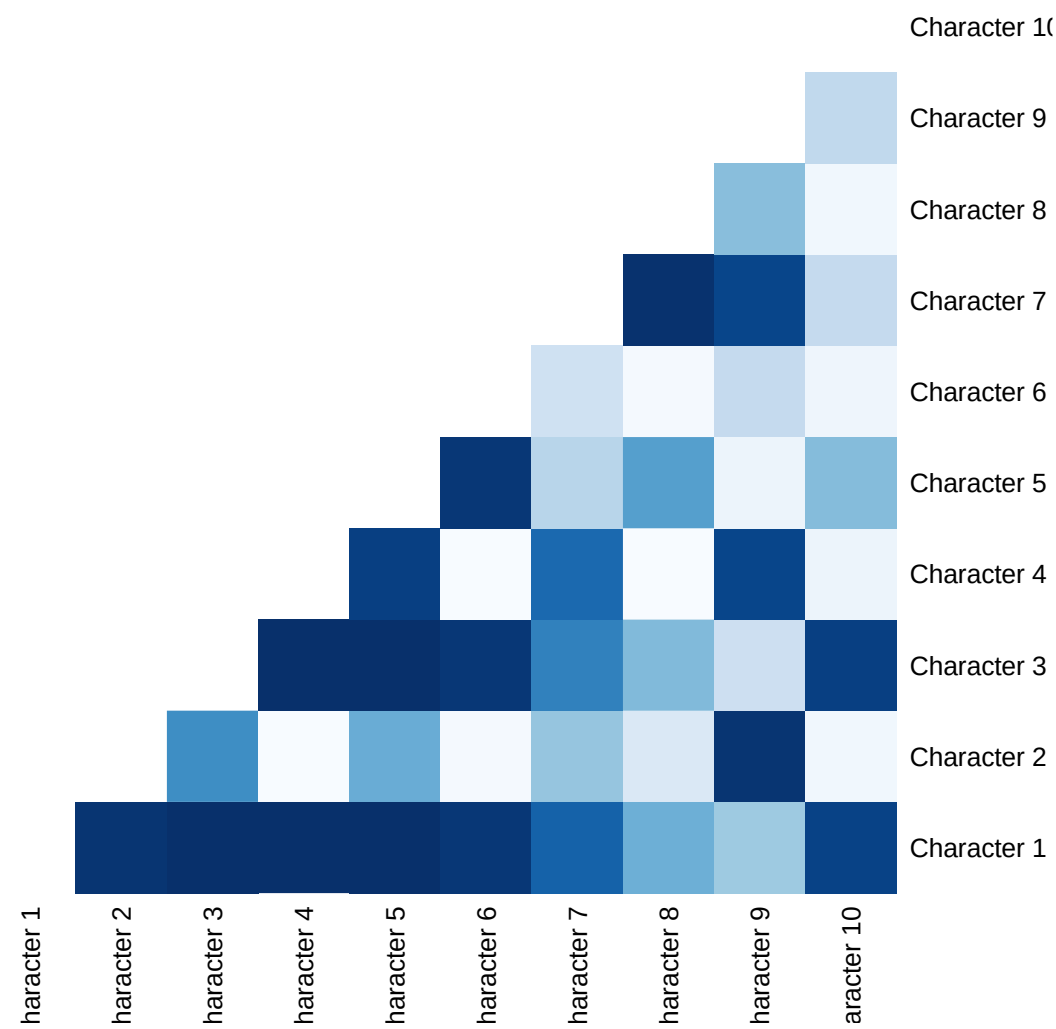
2. Constructed and Simulated Data

Taxa	Char1	Char2	Char3	Char4	Char5	Char6	Char7	Char8	Char9	Char10
t1	0	0	0	0	1	1	1	1	1	1
t2	0	0	1	1	1	1	0	0	1	0
t3	0	0	0	0	0	0	1	1	1	1
t4	0	0	0	0	0	0	0	0	1	0
t5	0	0	0	0	0	0	1	1	1	1
t6	1	1	1	1	1	1	0	0	0	0
t7	1	1	1	1	1	1	1	1	0	0
t8	1	1	1	1	1	1	0	0	0	0
t9	1	1	1	1	1	1	1	1	0	0
t10	1	1	1	1	1	1	0	0	0	0
t11	0	0	0	0	0	0	0	0	0	0
t12	0	0	0	0	0	0	0	0	0	0
t13	0	0	0	0	0	0	0	0	0	0
t14	0	0	0	0	0	0	0	0	0	0
t15	0	0	0	0	0	0	0	0	0	0
t16	0	0	0	0	0	0	0	0	0	0
t17	0	0	0	0	0	0	0	0	0	0
t18	0	0	0	0	0	0	0	0	0	0
t19	0	0	0	0	0	0	0	0	0	0
t20	0	0	0	0	0	0	0	0	0	0
t21	0	0	0	0	0	0	0	0	0	0
t22	0	0	0	0	0	0	0	0	0	0
t23	0	0	0	0	0	0	0	0	0	0
t24	0	0	0	0	0	0	0	0	0	0
t25	0	0	0	0	0	0	0	0	0	0
t26	0	0	0	0	0	0	0	0	0	0
t27	0	0	0	0	0	0	0	0	0	0
t28	0	0	0	0	0	0	0	0	0	0
t29	0	0	0	0	0	0	0	0	0	0
t30	0	0	0	0	0	0	0	0	0	0

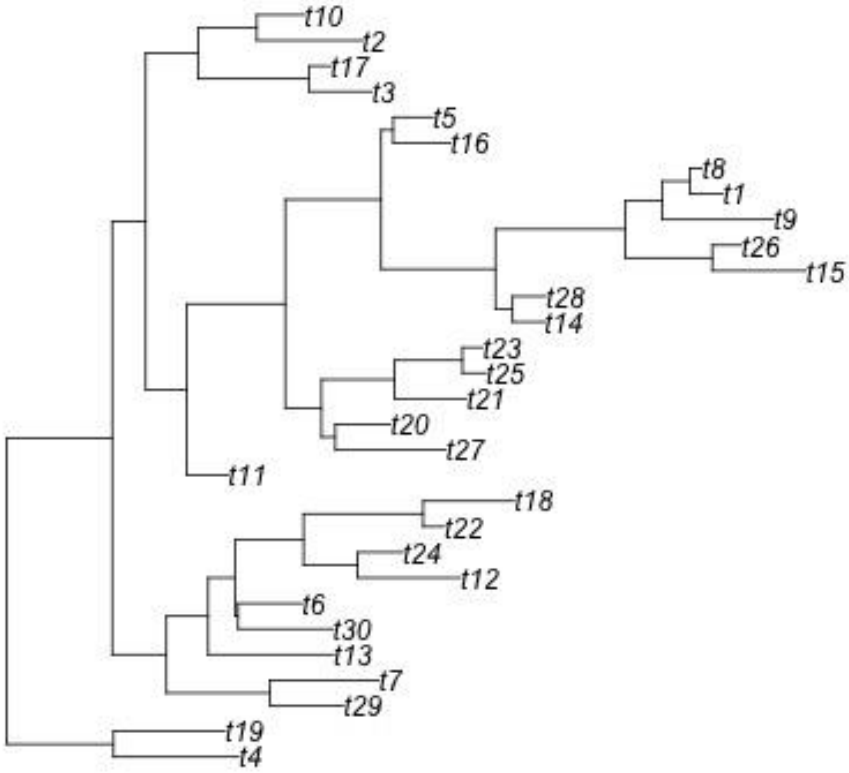
Supplementary Table 2. *Showing the first simulated 30 taxon dataset used for sensitivity analysis.*

Taxa	Char1	Char2	Char3	Char4	Char5	Char6	Char7	Char8	Char9	Char10
t1	0	0	0	0	1	1	1	1	1	1
t2	0	0	1	1	1	1	0	0	1	0
t3	0	0	0	0	0	0	1	1	1	1
t4	0	0	0	0	0	0	0	0	1	0
t5	0	0	0	0	0	0	1	1	1	1
t6	1	1	1	1	1	1	0	0	0	0
t7	1	1	1	1	1	1	1	1	0	0
t8	1	1	1	1	1	1	0	0	0	0
t9	1	1	1	1	1	1	1	1	0	0
t10	1	1	1	1	1	1	0	1	0	0
t11	0	1	0	1	0	0	0	0	0	1
t12	0	0	0	0	0	0	0	1	0	0
t13	0	1	0	0	0	0	0	0	0	1
t14	0	0	0	1	0	0	0	1	0	0
t15	0	1	0	0	0	0	0	0	0	1
t16	0	0	0	0	0	0	0	1	0	0
t17	0	1	0	1	0	0	0	0	0	1
t18	0	0	0	0	0	0	0	1	0	0
t19	0	1	0	0	0	0	0	0	0	1
t20	0	0	0	1	0	0	0	1	0	0
t21	0	1	0	0	0	1	0	0	0	1
t22	0	0	0	0	0	1	0	1	0	0
t23	0	1	0	1	0	1	0	0	0	1
t24	0	0	0	0	0	1	0	1	0	0
t25	0	1	0	0	0	1	0	0	0	1
t26	0	0	0	1	0	1	0	1	0	0
t27	0	1	0	0	0	1	0	0	0	1
t28	0	0	0	0	0	1	0	1	0	0
t29	0	1	0	1	0	1	0	0	0	1
t30	0	0	0	0	0	1	0	1	0	0

Supplementary Table 3. *Showing the second simulated 30 taxon dataset used for sensitivity analysis.*



Supplementary Figure 13. *Heatmap showing the degree of correlation, as measured by cumulative AICc weight of fitted correlated models (SYM and ARD), between all character pairs tested in the second 30-taxon simulated dataset on a random tree (dataset shown in supplementary Table 3 above). Lighter cells show weaker correlation, while darker cells show stronger correlation.*



Supplementary Figure 14. *The 30-taxon random tree used for sensitivity analysis.*

Table 1. *Differences between proportion and strength of correlation between character pair classes*

		Correlated Character Pairs (n of N)			Mean Correlated AICc Weight		
Dataset	**Taxa**	**Within Partition Character Pairs**	**Between Partition Character Pairs**	**P Fisher's**	**Within Partition Character Pairs**	**Between Partition Character Pairs**	**P Mann Whitney-*U***
Birds							
Birdsley, 2002	31	3/111	4/120	1	0.27	0.24	0.1445
Chu, 1995	173	181/567	30/136	0.02839	0.58	0.87	1.69E-11
Irestedt et al., 2004	38	2/55	0/50	0.4963	0.32	0.21	0.02332
Livezey, 1996a	20	20/378	14/325	0.5997	0.24	0.2	0.3381
Livezey,1996b	12	0/456	0/210	1	0.04	0.02	0.0003271
Mauricio et al., 2012	34	1/61	2/44	0.5759	0.27	0.26	0.6992
Prum, 1991	30	0/387	1/243	0.3857	0.17	0.14	0.3899
Smith, 2011	52	2434/9883	557/2520	0.008413	0.61	0.58	2.39E-05
Winker & Pruett, 2006	17	0/225	0/126	1	0.11	0.11	0.9218
Livezey and Zusi, 2007	139	53799/274498	23026/138188	2.20E-16	0.19	0.38	2.20E-16
Squamates							
Frost et al., 2001b	28	3/498	0/448	0.2513	0.19	0.2	0.3273
Frostet al., 2001a	26	1/156	0/169	0.48	0.19	0.2	0.6458
Kearney, 2003	35	6/183	0/195	0.01233	0.34	0.32	0.2413

Showing differences between proportion and strength of correlation between character pair classes. Character pairs are considered correlated if their cumulative AICc weight for correlated (SYM and ARD) models of character pair evolution are equal to or higher than 0.95. Strength of correlation is measured by average cumulative AICc weights of correlated models of character pair evolution. Character pair classes include a) within-osteological data character pairs, b) within-non-osteological data character pairs, and c) between-partition character pairs. Characters are generally more correlated (as measured by average AICc weights) within than between partitions. Differences in proportion and strength of correlation between osteological and non-osteological characters are mixed.

Table 1 cont. *Differences between proportion and strength of correlation between character pair classes*

Correlated Character Pairs (n of N)						
Osteological Character Pairs	**Non-Osteological Character Pairs**	**Between Partition Character Pairs**	**Osteological % Correlated Pairs**	**Non-Osteological % Correlated Pairs**	**Between Partition % Correlated Pairs**	**P Fisher's**
1/66	2/45	4/120	1.51	4.44	3.33	0.6012
177/561	4/6	30/136	31.55	66.67	22.06	0.01137
0/10	2/45	0/50	0	4.44	0	0.3636
0/78	20/300	14/325	0	6.667	4.31	0.02646
0/21	0/435	0/210	0	0	0	1
0/55	1/6	2/44	0	16.67	4.55	0.03831
0/351	0/36	1/243	0	0	0.41	1
2409/9730	25/153	557/2520	24.76	16.34	22.1	0.001101
0/15	0/210	0/126	0	0	0	1
51264/255970	2535/18528	23026/138188	20.03	13.68	16.66	2.20E-16
0/378	3/120	0/448	0	2.5	0	0.001997
1/78	0/78	0/169	1.28	0	0	0.48
2/105	4/78	0/195	1.9	5.13	0	0.006041

Showing differences between proportion and strength of correlation between character pair classes. Character pairs are considered correlated if their cumulative AICc weight for correlated (SYM and ARD) models of character pair evolution are equal to or higher than 0.95. Strength of correlation is measured by average cumulative AICc weights of correlated models of character pair evolution. Character pair classes include a) within-osteological data character pairs, b) within-non-osteological data character pairs, and c) between-partition character pairs. Characters are generally more correlated (as measured by average AICc weights) within than between partitions. Differences in proportion and strength of correlation between osteological and non-osteological characters are mixed.

Table 1 cont. *Differences between proportion and strength of correlation between character pair classes*

Mean AICc				
Osteological Character Pairs	**Non-Osteological Character Pairs**	**Between Partition Character Pairs**	**P Kruskal-Wallis**	**Tree Symmetry Score**
0.23	0.34	0.24	0.03	123
0.57	0.83	0.87	3.42E-11	835
0.23	0.35	0.21	0.01171	115
0.1	0.28	0.2	0.0009701	65
0.03	0.04	0.02	0.0005528	34
0.24	0.53	0.26	0.2709	166
0.17	0.13	0.14	0.64	72
0.61	0.5	0.58	1.58E-08	232
0.09	0.11	0.11	0.984	55
0.4	0.374	0.38	2.20E-16	856
0.18	0.23	0.2	0.2098	162
0.25	0.14	0.2	2.00E-04	106
0.3	0.4	0.32	0.1326	188

Showing differences between proportion and strength of correlation between character pair classes. Character pairs are considered correlated if their cumulative AICc weight for correlated (SYM and ARD) models of character pair evolution are equal to or higher than 0.95. Strength of correlation is measured by average cumulative AICc weights of correlated models of character pair evolution. Character pair classes include a) within-osteological data character pairs, b) within-non-osteological data character pairs, and c) between-partition character pairs. Characters are generally more correlated (as measured by average AICc weights) within than between partitions. Differences in proportion and strength of correlation between osteological and non-osteological characters are mixed.

3. Cliques of Morphological Characters in Avian and Squamate Datasets

Abstract

A key assumption in phylogenetic analysis is that characters are independent. Non-independent, correlated characters may impact phylogenetic reconstruction, but the general prevalence of internally-consistent cliques or clusters of characters in morphological datasets is unknown. Moreover it is known that different partitions, such as osteological and dental characters, support different estimates of evolutionary history, but not whether these partitions are more or less correlated. Here, we interrogate 19 avian and squamate morphological matrices identify the prevalence and size of internally-consistent groups of morphological characters. Through conducting UPGMA analysis, a hierarchical clustering method, on these datasets and mapping both character retention indices and non-osteological versus osteological partitions on the resulting UPGMA trees, we find that a) clique sizes are no larger then expected by chance in most datasets, b) most datasets contain few large cliques and many small cliques, c) that internally-consistent cliques are also more likely to be congruent with molecular trees, and d) that characters are more likely to be correlated within partitions. This demonstrates that subsets of morphology display different properties, which should be taken into account in phylogenetic analysis.

1. Introduction

Integration, modularity and correlation are central processes in the evolution of morphological traits (Goswami *et al.*, 2014), and a recognized problem in phylogenetic analysis (Sadleir & Makovicky, 2008; Goswami & Polly, 2010; Guillerme & Brazeau, 2018; Billet & Bardin, 2019). It is well established that many traits evolve non-independently in modules (Klingenberg, 2008) or cliques (Holland *et al.*, 2010; Blanke *et al.*, 2013) of

correlated characters, sometimes resulting from functional or developmental integration. Many of these correlated traits evolve in response to shared selective regimes related to occupying a similar ecological niche. Characters evolving in a correlated manner irrespective of phylogeny is known as concerted convergence (Patterson & Givnish, 2002), and has consequences for phylogenetic analysis as well as being of interest to researchers looking at ecological trait evolution.

While concerted convergence and correlation of characters have been observed in isolated clades at the generic and familial level (e.g. Patterson & Givnish, 2002; Holland *et al.*, 2010), it is yet to be established how widespread these phenomena are across morphological phylogenetic data in general. Further, many studies detecting character correlations in general have looked at only a subset of all available morphological data for tetrapods. For example, model-based methods have been used to detect correlated character pairs in floral traits (Sauquet *et al.*, 2017), in pollen morphology (Leslie *et al.*, 2015) and in dental morphology (Billet & Bardin, 2019), and clustering methods have been used to find suites of correlated insect head characters (Blanke *et al.*, 2013). Importantly, different subsets of characters in morphological datasets are likely to have different evolutionary properties and dynamics, meaning that we currently have an incomplete picture of the true prevalence of concerted morphological trait evolution in general.

1.1. Correlated Character Cliques in Phylogenetic Reconstruction

Given these limitations, it is important to interrogate patterns of correlated characters in entire morphological datasets spanning wide taxonomic levels. This is particularly important since the correlation and integration of characters poses a problem for phylogenetic reconstruction when morphological characters are used to infer evolutionary history (Sadleir & Makovicky, 2008; Goswami & Polly, 2010; Billet & Bardin, 2019). Correlation in morphological characters leads to a form of pseudoreplication

when these characters are treated as independent, potentially making species appear more closely related than they actually are. Where these characters display strong phylogenetic signal, support values are likely to be artificially inflated. Where these correlated characters do not agree with the underlying phylogeny, erroneously close relationships may be inferred between species that share a common morphology as a result of common selective regimes (for example, arising from occupying a shared ecological nice), as opposed to recent common ancestry. Identifying correlated character cliques may therefore be beneficial to phylogenetic analysis. In addition to improving model-based inference methods using morphological characters, identifying such groups of characters could potentially help resolve conflicts between molecular and morphological topologies.

1.2. Detecting Correlated Character Pairs and Cliques

To detect correlations in morphological data, tree-based approaches have been developed to determine whether characters are correlated given an underlying phylogeny. These include model-based methods used to detect character pairs under both Bayesian and maximum likelihood frameworks, and involve fitting correlated and uncorrelated models of state transition to pairs of characters on a given phylogeny (Pagel, 1994; Beaulieu, O'Meara, & Donoghue, 2013). These methods are useful in cases where there is high confidence in both the topology and ages of the trees used in these analyses. Molecular trees can be and often are used to offer a good estimate of evolutionary history, since molecular data are often considered more reliable than morphological data. Molecular phylogenies thus offer a useful way to characterize the relationships between morphological characters relative to an independent evolutionary framework.

However, it is nevertheless important to detect groups of correlated morphological characters by examining their internal consistency, independent of an underlying tree, given that there is no way to establish evolutionary history with absolute certainty. Further, finding groups of

characters that are internally consistent irrespective of whether they are congruent with underlying hypotheses of phylogeny will provide further insights into character evolution and phylogenetic reconstruction using these characters. Tree-independent methods of detecting correlated characters, such as those established by O'Keefe & Wagner, (2001), Sadleir & Makovicky, (2008) and more recently Holland *et al.*, (2010) have therefore been important to use alongside the tree-based methods described above. Furthermore, when the tree-independent correlation of characters has been detected, the internal consistency of characters can then be compared against the consistency of these characters on molecular trees.

1.3. Approach and Hypotheses

Here, we explore a number of questions relating to the nature of internally consistent cliques of characters in morphological data, firstly by clustering characters by similarity, largely following the methods of Holland *et al.*, (2010) and Blanke *et al.*, (2013). However, while these studies test specific hypotheses of character correlation, we interrogate 19 entire datasets spanning birds and squamates in order to determine the existence, prevalence, size and nature of cliques in morphological phylogenetic data in general. In addition to finding cliques in these datasets, we examine to what extent the internal consistency of morphological characters relates to their fit on previously published molecular trees, as measured by their retention index on these trees. In doing this, we are able to establish which cliques have low consistency with molecular trees, and therefore may be candidates for concerted convergence (Patterson & Givnish, 2002).

In general, in morphological datasets with good phylogenetic signal, we expect to see large cliques of characters that fit well on the molecular trees, and smaller cliques that fit the molecular trees poorly. We may expect that characters that are internally consistent will also share a similar consistency with molecular trees, since these characters will have similar cladistic

patterns regardless of any underlying phylogeny. We therefore also expect to find that some cliques of internally consistent characters share a poor fit on molecular trees. If we assume that these molecular trees are reliable, which is unknowable, we can take this as evidence for concerted convergence in these characters. Furthermore, the presence of cliques that poorly fit an underlying phylogeny could have potential ramifications for phylogenetic analysis if these characters are treated as independent in phylogenetic analysis. This is because, in the case that the molecular trees are in fact reliable, these characters can be interpreted as exhibiting poor phylogenetic signal.

Finally, we also determine whether characters are more or less internally consistent within partitions, namely osteological and soft character partitions. We expect that characters belonging to the same partition are more likely to be correlated and evolve in concert. This is because we expect that characters within these partitions are more likely to share selective regimes and developmental origins and/or genetic linkage. Furthermore, previous work has demonstrated that these partitions do display different properties with regards to convergence, transition times on molecular trees, and correlation on molecular trees using model-based likelihood approaches (Chapters 1 and 2).

In order to look for signal in both the fit of characters on molecular trees and the partition that these characters belong to, we employ methods normally used to detect phylogenetic signal, namely Pagel's lambda (Pagel, 1999), Blomberg's K (Blomberg, Garland, & Ives, 2003) and Fritz and Purvis's D (Fritz & Purvis, 2010). We treat the clustered characters as phylogenetic trees, and the retention index and partition of these characters as traits. We employ randomization procedures on a subset of the data to ensure that the methods we use perform as expected given our novel approach.

2. Methods

2.1. Datasets

To detect cliques of correlated characters, we took a meta-analysis approach using 19 avian and squamate datasets, including a large avian (Livezey & Zusi, 2007) and large squamate (Reeder *et al.*, 2015) dataset, spanning the entire clades. In brief, these datasets were compiled from Google Scholar, Graeme Lloyd's matrix collection (Lloyd, 2009), and the reference list of the supertree study of Davis & Page (2014). Datasets with few than 20 characters and 10 taxa were excluded from analysis. All uninformative characters were removed. Please see previous work (Chapter 1) for a more detailed discussion on dataset collection and treatment. In addition to the morphological datasets, two molecular trees were used in this analysis (Jetz *et al.*, 2012; Zheng & Wiens, 2016) for the calculation of homoplasy (the retention index) of characters on these trees. In almost all datasets, the number of characters is higher than the number of taxa although in some cases the character to taxon ratio is relatively low. It is possible that character to taxon ratios have an effect on analysis although this is not tested here.

2.2. Cluster Analysis

Following the methods of Holland *et al.,* (2010), we used a treeless approach to assess the internal consistency of morphological characters in these datasets. For each morphological dataset, we computed the pairwise excess indices (Holland *et al.*, 2010; Blanke *et al.*, 2013) between all character pairs in Python in order to produce character dissimilarity matrices. The excess, or dissimilarity, of a pair of characters is defined as the difference between its parsimony score on the most parsimonious tree that can be constructed with only those two characters, and the minimum possible parsimony scores for those characters.

To identify and visualise cliques, we then performed a UPGMA analysis on the resulting dissimilarity matrices (following Holland *et al.*, 2010) using the UPGMA function in the package Phangorn (Schliep, 2011) implemented in R (R Core Team 2018), in order to produce trees in which morphological characters are clustered according to their phylogenetic similarity. The UPGMA is a simple hierarchical clustering method that takes the pairwise dissimilarity matrix and constructs rooted tree where the tips, in this case morphological characters, are grouped according to their similarity. This method quickly constructs groups of characters with no homoplasy. Since methods to detect phylogenetic signal require that all branches have non-0 lengths, we then set negative branch lengths in the resultant UPGMA trees to 0 and added a small length (0.001) to all branches. UPGMA branch lengths indicate distance between cliques.

2.3. Clique Size Analysis

In avian datasets, we compared the maximum clique size of each dataset with the maximum clique size of shuffled datasets, to determine whether clique sizes in the empirical data are larger than expected by chance (Holland *et al.*, 2010). First, we calculated the size of the largest clique, i.e. the largest group of internally compatible characters (defined as groups of characters that can be mapped onto a tree without homoplasy), in each dataset. We then used a randomization procedure described in Holland *et al.*, (2010), which takes into account phylogenetic signal in the data, to create 100 shuffled datasets for each empirical dataset. We shuffled datasets on each of 100 trees taken from the posterior distribution of Jetz *et al.*, (2012), resulting in a null distribution of 100 shuffled datasets for each empirical dataset. Characters were shuffled such that their parsimony score on each input tree is the same as that of the empirical data. For each character in each dataset, all state changes were identified on a given input tree after character state reconstruction using the Fitch algorithm. The descendent taxa of each branch with a state change, and the new state, were

recorded and stored in an array along with the state at the root. Then a set of new characters matching dimensions of the original dataset was created. For each character, each taxon was randomly allocated a state from the root states array, and a taxon cluster and corresponding state were picked at random from the cluster and state arrays. The process of picking clusters with replacement was repeated until each character had the same parsimony score on the input tree as the original tree. This was repeated for each input tree (Holland *et al.*, 2010).

We then calculated the size of the largest internally consistent clique in each of these shuffled datasets. We determined whether there is a significant difference in clique size between the empirical and shuffled data by calculating the proportion of datasets in the null distribution which have maximum clique sizes the same as or smaller than in the unshuffled data, and treating this as the probability that the maximum clique size is significantly larger than expected by chance, following the procedure of Holland *et al.*, (2010). Statistical significance using this procedure indicates that the maximum clique size is significantly larger than expected by chance. This analysis could not be performed on the large avian dataset spanning all birds (Livezey & Zusi, 2007) owing to the size of the dataset. Additionally, this analysis could not be performed on squamate datasets, since this procedure requires several input trees, and here we use a single squamate molecular tree.

2.4. Comparison with Molecular Data

To determine whether characters that are internally consistent are also consistent with molecular trees, we used two metrics to estimate the relationship between clique distribution of characters and the molecular homoplasy of characters. Here, we used methods often used to detect phylogenetic signal - Pagel's lambda (Pagel, 1999) and Blomberg's K (Blomberg *et al.*, 2003), using the phylogsig function implemented in Phytools (Revell, 2012) in R. These methods test a null model of random

distribution of traits among tips against a Brownian motion model of trait clustering. We treated the UPGMA trees as phylogenetic trees (with morphological characters as tips) and the homoplasy, as measured by the retention index, of individual characters on molecular trees as the trait. Briefly, the retention index of character on a tree is a measure of fit of that character on that tree, with a retention index of 0 indicating an extremely poor fit on a tree, and a retention index of 1 indicating a perfect fit. For avian datasets, we calculated the homoplasy of each character on each of a set of 1,000 trees from the posterior distribution of Jetz *et al.*, (2012), and averaged these values to obtain a single value for each character. For squamate datasets, we calculated the homoplasy of each character on the topology of Zheng & Wiens, (2016).

Pagel's lambda determines whether the distribution of traits on the tips can be best described with a Brownian motion (random walk) model, indicating clustering, or a random model. The lambda metric describes to what extent the branches of a tree must be transformed in order for Brownian motion to best explain the observed data, where a value of 1 describes the original tree, and a value of 0 describes a completely unresolved tree. Thus, a value of 1 indicates strong phylogenetic signal, with traits evolving under strong Brownian motion, while a value of 0 indicates a completely random distribution of traits. The significance value indicates whether the distribution of traits is significantly different from random, but not whether it is significantly different to a Brownian motion model of significant clustering of traits. Therefore, to determine whether datasets exhibiting significant signal are more consistent with a Brownian motion model or a lambda model, where trait distribution is between random and Brownian motion, we also compared random, lambda and Brownian motion distribution models using AICc weights. For this, we used the fitContinuous function in the R package Geiger (Harmon *et al.*, 2008).

Blomberg's K compares the variance of Phylogenetically Independent Contrasts (PICs) in the observed data with the variance that would be expected under a Brownian motion model. A K value of 0 indicates that no phylogenetic signal is present in the data, while a K of 1 indicates that the

distribution of traits is best explained by Brownian motion. We conducted Blomberg's K with 1000 simulations.
These methods are not normally used on UPGMA trees. To check that these methods work correctly on our data, we shuffled the trait values on five of the datasets. We then computed both metrics on these random trait distributions, and compared the results against our empirical data.
To visualize the extent to which characters that are incongruent with the molecular tree form internally coherent cliques, the tips of each UPGMA tree are coloured on a scale according to their retention index (supplementary Figure 1).

2.5. Effect of Character Partition on Cliques

To determine whether characters belonging to the same partition (osteological versus non-osteological) share phylogenetic similarity in terms of clique composition, we use similar methods described above for homoplasy, instead treating partition as a trait. We use Fritz and Purvis's D (Fritz & Purvis, 2010) since this method is suitable for binary traits (i.e. osteological and non-osteological partitions), using the phylo.d function in the R package Caper (Orme *et al.*, 2018). This metric behaves similarly to Pagel's Lambda, fitting models of both random trait distribution (no phylogenetic signal) and Brownian motion (strong phylogenetic signal) to traits. The metric describes the 'clumpiness' of traits on phylogenetic trees, with a value of 1 indicating random distribution of traits, a value of 0 indicating evolution under Brownian motion, and values of less than 1 indicating more clustering than expected under Brownian motion.

3. Results

3.1. Clique Analysis

The cluster analysis resulted in highly asymmetrical UPGMA trees in most cases (e.g. Figure 1, Supplementary Figures 1, 2), where there are few large cliques and more small cliques and singletons.

In general the size of the largest cliques (in number of characters) was found to be within the range of the null distribution from clique sizes in shuffled datasets. Cliques were found to be larger than in shuffled data in 5 of 11 avian datasets, only two of which are significantly larger (Prum, 1992, 1993, Table 1, null distribution test). In just over half of cases, clique sizes are smaller than in shuffled datasets with the same parsimony score in 6 of 11 datasets. Overall, these results indicate that large cliques of characters are unusual in these datasets, and are often no larger than expected by chance.

Table 1. *Clique sizes in avian datasets*

Dataset	**Taxa**	**Characters**	**Largest Clique Size (no. characters)**	**Mean Largest Clique Size (Shuffled Data)**	***p* (null distribution test)**
Bertelli, Chiappe, & Mayr, 2014	17	124	34	52.51	0.98
Birdsley, 2002	31	39	12	10.69	0.33
Clarke *et al.*, 2007	30	174	72	80.49	0.71
Irestedt, Fjeldså, & Ericson, 2004	38	30	7	8.11	0.82
Livezey, 1996	25	82	30	36.32	0.86
Maurício *et al.*, 2012	34	88	30	23.62	0.14
Nesbitt, Ksepka, & Clarke, 2011	11	91	32	35.54	0.84
Patten & Fugate, 1998	27	30	5	4.56	0.44
Prum, 1993	12	27	23	18.71	0.01
Prum, 1992	30	39	24	18.03	0.02
Smith, 2011	52	299	19	36.49	1

Showing the total number of characters, the maximum clique size (size of the largest internally consistent clique of characters) in the unshuffled data, and the mean maximum clique sizes in 100 shuffled datasets. Values are compared by calculating proportions of datasets in the null distribution which have maximum clique sizes larger than or equal to that in the unshuffled data (p null distribution test). Clique sizes are significantly larger in unshuffled than shuffled data in 2 datasets, indicating that clique sizes are no larger than expected given their parsimony score in most datasets.

3.2. Phylogenetic Signal and Homoplasy

Characters in cliques shared similar levels of homoplasy relative to molecular trees. Treating retention index as a phylogenetic trait in UPGMA trees found evidence of signal in the homoplasy of morphological characters. For birds, 8 of 12 Pagel's lambda values and 8 of 12 Blomberg's K values are significantly different from 0, where 0 indicates random distribution of traits among tips. Of the 10 datasets displaying signal with at least one metric, 7 show significant signal with both. For squamates, 5 of 7 Pagel's lambda and 3 of 7 Blomberg's K values are significantly different from 0, with significant signal present in both metrics in 2 of these datasets. Overall, of the 14 total datasets in which there is significant phylogenetic signal with at least one metric, 9 datasets show significant signal with both, indicating reasonably high correspondence between these two methods (Table 2). However, many of these values are low, especially when using Blomberg's K, and of all datasets showing significant signal, a lambda model describes the trait distribution better than Brownian motion (Supplementary Table 1). This indicates that the signal is often weak even in cases where it is highly significant.

Table 2 at End of Document

Table 2. *Showing the dataset dimensions and presence of signal in homoplasy (retention index), and partition membership, of morphological characters on UPGMA trees clustering morphological characters in several avian and squamate datasets by their phylogenetic similarity using three metrics. We use Pagel's lambda and Blomberg's K to measure signal in homoplasy, and Fritz & Purvis's D to measure signal in partition membership. Here, UPGMA trees are treated as phylogenetic trees and the retention index of characters on molecular trees treated at the trait of interest. In both metrics used to measure signal in homoplasy, values close to 0 indicate close to random trait distribution, and values close to 1 indicate a distribution best explained by Brownian motion, indicating clumpiness of traits. Significant P values indicate that metrics are significantly different from 0, indicating the presence of signal*

in the trait distribution. In Purvis's D used to measure signal in partition membership, values close to 1 indicate close to random trait distribution, and values close to 0 indicate close to a distribution best explained by Brownian motion, indicating clumpiness of traits. Significant P values (random) indicate that metrics are significantly different from 1, indicating the presence of signal in the trait distribution. Significant P values (Brownian) indicate that the metrics are significantly different from 0, indicating no Brownian motion.

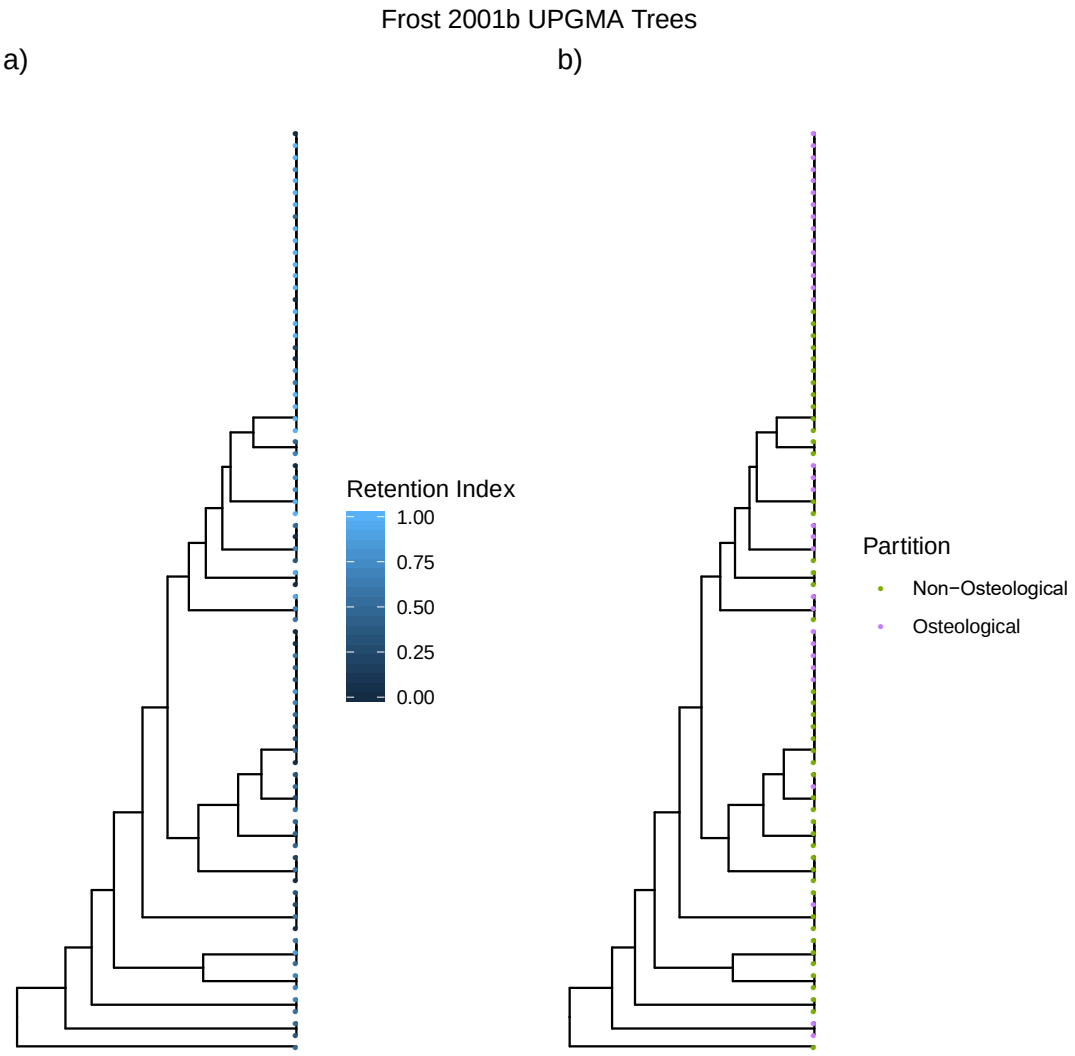


Figure 1. *Showing the distribution of a) the retention index and b) partition among the tips of an example UPGMA tree (Frost et al. 2001b), where tips are morphological characters. The trees show 3 characteristic patterns; firstly, the UPGMA trees are quite asymmetrical, demonstrating a non-normal distribution of clique sizes (i.e. there is a small number of large cliques and a larger number of small cliques). Secondly, the retention index is lower in smaller cliques, demonstrated by tips getting darker further down tree a. This suggests that larger cliques tend to be more congruent with molecular trees. Thirdly, smaller cliques are less likely to be composed of osteological characters, as demonstrated by an increase in 'greeness'*

further down tree b. Additionally, the presence of some medium-sized, darker (on tree a) cliques further down the tree suggest some 'concerted convergence', as these characters are internally consistent and have low retention indices.

When the retention index is randomized among UPGMA tips, both methods behave as expected for randomized data, with high p values, often close or equal to 1. This strongly suggests that the signal detected in the empirical data is likely real and not an artifact of our unusual use of non-phylogenetic trees for this analysis.

Visualisation of the retention indices (on molecular trees) on the tips of the UPGMA trees strongly suggests that larger character cliques are usually more consistent with molecular trees, with highly incongruent characters more likely to exist in smaller cliques or as singletons (Figure 1, Supplementary Figure 1). However, in several datasets, large cliques do exist in which congruence with molecular trees is low (e.g. Frost *et al.*, 2001b, Figure 1). These cliques possibly indicate instances of concerted convergence in these characters as described and discussed in, for example, Patterson & Givnish, (2002); Holland *et al.*, (2010); Blanke *et al.*, (2013).

3.3. Partitions

There is non-random distribution of the partitions of characters among UPGMA tips, indicating that characters are significantly more likely to cluster within partitions than expected by chance. This suggests higher correlation within than between partitions. In birds, 9 of 12 datasets show significant signal in partition (Table 2, Supplementary Figure 2). Three of these have trait distributions that are not significantly different from a distribution expected under Brownian motion. In squamates, 6 of 7 datasets show significant phylogenetic signal, with two of these not being significantly different from Brownian motion. Overall, D values in 15 of 19 datasets are significantly different from 1, 1 indicating random distribution of the binary trait among UPGMA tips. Of these, 4 are not significantly

different from 0, suggesting a Brownian motion distribution in these datasets.
Again, when traits are randomized among UPGMA tips, the D statistic is close to 1, not significantly different from random and significantly different from Brownian motion, suggesting that this method performs as expected even when using UPGMA trees.

4. Discussion

Morphological characters are more internally consistent within than between osteological and non-osteological partitions. This is consistent with our previous findings showing that character pairs are more correlated given a molecular tree within than between these partitions (Chapter 1). This result is strongly suggestive of integration within related morphological traits (Klingenberg, 2008), and is consistent with a view that traits are more likely to evolve in concert when they share genetic or developmental pathways. Further, the largest cliques tend to be composed of characters with low homoplasy. This suggests that in general, the bulk of characters used for phylogenetic analysis are not only internally consistent, but also likely to be more congruent with evolutionary history. Conversely, smaller cliques are generally made up of characters with lower retention indices, suggesting that homoplasious characters tend to contain signal that is disparate from other characters in general. This is consistent with stochastic noise in these characters, as opposed to concerted convergence (Patterson & Givnish, 2002) of traits. Nevertheless, in some datasets (e.g. Frost *et al.*, 2001), larger cliques of characters do exist which demonstrate high homoplasy, warranting further investigation into the makeup of these cliques, and the ecological, genetic, functional or developmental processes that unite their component characters. Additionally, cliques tend to be no larger than expected given the overall phylogenetic signal in the dataset. This study demonstrates new ways of looking at morphological characters,

detecting correlations and comparing character partitions. This potentially has wide utility in phylogenetics and studying morphological evolution.

4.1 Clique Sizes

Interestingly, maximum clique sizes are no larger than expected by chance in just over half of all analysed datasets. Overall, the clique size results imply that characters in these datasets do not cluster in internally consistent correlated cliques that are any larger or smaller than expected by chance. This may be because here we have analysed entire morphological datasets, presumably coded with the intent of constructing a phylogeny with as much data as possible. Further, the authors of these phylogenetic datasets may have avoided the use of characters that were thought to be obviously correlated. This contrasts with previous studies of this nature finding significantly larger cliques in their dataset than in shuffled data. This may be because these studies were using clique analysis to test pre-defined hypotheses of correlation, and therefore including only specific taxa or character types (Holland *et al.*, 2010; Blanke *et al.*, 2013).

We also used a randomization procedure that ensures that the shuffled datasets have the same level of congruence with molecular trees as the unshuffled data, since shuffling the data at random produces datasets that are more incongruent with molecular trees (Holland *et al.*, 2010). However, if the unshuffled and shuffled data share a parsimony score with respect to molecular trees this helps to ensure that any similarity between taxa, and ultimately the presence of larger maximum clique sizes in the unshuffled data compared with the shuffled data, are the result of convergence as opposed to phylogenetic similarity. However, this assumes that the underlying trees used for this procedure represent an accurate picture of evolutionary history.

4.2. Internal Consistency of Morphological Characters Versus Congruence With Molecular Data

Unsurprisingly, cliques of correlated characters that cluster together in UPGMA analysis are also generally more likely to share a similar level of consistency with molecular trees as measured by the retention index. This is demonstrated by the use of Pagel's lambda (Pagel, 1999) and Blomberg's K (Blomberg *et al.*, 2003) to search for signal in the retention index on the UPGMA trees, which were treated as phylogenetic trees. Additionally, cliques that are more congruent with molecular trees are often larger, suggesting that in general, morphological and molecular datasets contain broadly congruent signal. Characters in these cliques are likely correlated as a result of common ancestry, if we take consilience between morphological and molecular data to be suggestive of genuine phylogenetic signal in the morphological data. Conversely, cliques of characters that are homoplasious relative to molecular trees are internally consistent with each other. This is known in the literature as concerted convergence, and arises when taxa are evolving under shared ecological conditions (Patterson & Givnish, 2002). However, since we cannot distinguish the maximum clique sizes in most datasets from chance, we cannot conclude with certainty that these traits are evolving under concerted convergence.

Our use of UPGMA trees in this analysis is unusual, given that these trees are not phylogenetic in nature. The use of these clustered rather than bifurcating phylogenetic trees may have affected the results, for example the extremely low values of K even with very low p values, interpreted here as strong evidence of weak signal. Further, since common ancestry among UPGMA tips is not implied, concepts such as Brownian motion do not strictly apply to these data in a traditional sense. However, we argue that a Brownian motion model is relevant here, and in this context refers to the clustering of traits among tips owing to character similarity rather than common ancestry. This interpretation is strengthened by the fact that when we perform these same analyses on randomized trait data, the results are as we would expect for trait data randomly dispersed among the tips of

phylogenetic trees (i.e. low values of Pagel's lambda and Blomberg's K, p values close to 1), implying that the signal detected by these methods on our data is meaningful.

4.3. Internal Consistency of Subsets of Morphological Characters

Characters are significantly more correlated within osteological and non-osteological partitions, as measured by the significant signal detected when partition is treated as a trait on the UPGMA trees. This may reflect the different evolutionary trends and developmental origins of these character types. It has been shown previously that different partitions of morphological data infer different trees and have different phylogenetic properties, for example (Mounce, Sansom, & Wills, 2016; Sansom, Wills, & Williams, 2017; Sansom & Wills, 2017). Characters in these subsets may be more likely to share developmental origins or ecological properties, and therefore may be more likely to evolve in concert. It may also reflect the coding practices of the authors. For example, the oversplitting of related characters is a known phenomenon, and may result in non-independence. However, it is possible that these results are affected by data imbalance, in osteological data outweigh non-osteological data in most datasets. Purvis's D is affected by relative trait prevalence to some extent in smaller trees (Fritz & Purvis, 2010). Nevertheless, here both larger datasets and datasets with roughly equal prevalence of characters in either partition have similar results as smaller datasets and datasets with significant imbalance.
Both character partition and retention index on molecular trees display significant signal on UPGMA trees where characters are clustered by similarity. Osteological characters are also more congruent with molecular trees than non-osteological characters (Chapter 1). Therefore it is reasonable to speculate that cliques containing osteological characters are also cliques in which the characters have on average a higher retention index.

5. Conclusions

The size of the largest cliques in these datasets is no larger than in shuffled datasets in most cases. This suggests that, at least in whole datasets intended for phylogenetic reconstruction, morphological characters are no more internally correlated than expected by chance. Characters are also more internally correlated within osteological and non-osteological partitions, corroborating previous analyses suggesting higher levels of correlation within partitions when using an underlying molecular tree (Chapter 2). Further, these partitions themselves display a different level of congruence with molecular trees, with non-osteological characters apparently containing more homoplasy. In addition, cliques of characters displaying high congruence with molecular trees are often larger, with characters that are less congruent with molecular trees exhibiting less clustering. This is consistent with stochastic noise in many homoplasious characters as opposed to large-scale convergence. Further work exploring differences in levels of correlation between taxonomic rank would help to further elucidate macroevolutionary patterns of concerted evolution.

While these findings are reassuring when using morphological data for phylogenetic reconstruction, some internally consistent cliques do nevertheless exist which exhibit low congruence on molecular trees. These sets of characters are larger in some datasets than in others, and it therefore cannot be assumed when inferring phylogeny from morphological datasets that all characters are independent without explicitly testing for correlation *a priori*.

Overall, these results and others (Patterson & Givnish, 2002; Holland *et al.*, 2010; Blanke *et al.*, 2013) paint a picture of some concerted convergence in morphological datasets, although this is more apparent in some datasets than others. Patterns of convergence also differ between morphological subsets. We recommend therefore that correlations are tested for and taken into account when inferring phylogeny from morphological data, particularly when using probabilistic methods such as maximum likelihood

and Bayesian inference methods, paying particular attention to characters which may be prone to homoplasy. Correlated characters can be accounted for and their impact reduced by being reduced to fewer characters such as in composite coding (Billet & Bardin, 2019), or downweighted (Emerson & Hastings, 1998) in phylogenetic analysis.

References

Beaulieu JM, O'Meara BC, Donoghue MJ. 2013. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. *Systematic Biology* 62: 725–737.

Billet G, Bardin J. 2019. Serial homology and correlated characters in morphological phylogenetics: modeling the evolution of dental crests in Placentals. *Systematic Biology* 68: 267–280.

Blanke A, Greve C, Wipfler B, Beutel RG, Holland BR, Misof B, Forschungsmuseum Z, Koenig A. 2013. The identification of concerted convergence in insect heads corroborates Palaeoptera. *Systematic Biology* 13: 250–263.

Blomberg SP, Garland TJ, Ives AR. 2003. Testing for phylogenetic signal in comparative data: behavioural traits are more labile. *Evolution* 57: 717–745.

Davis KE, Page RDM. 2014. Reweaving the tapestry: a supertree of birds. *PLoS Currents*.

Emerson SB, Hastings PA. 1998. Morphological correlations in evolution: consequences for phylogenetic analysis. *The Quarterly Review of Biology* 73: 141–162.

Fritz SA, Purvis A. 2010. Selectivity in Mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conservation Biology* 24: 1042–1051.

Frost DR, Rodrigues MT, Grant T, Titus TA. 2001. Phylogenetics of the lizard genus Tropidurus (Squamata: Tropiduridae: Tropidurinae): direct optimization, descriptive efficiency, and sensitivity analysis of congruence between molecular data and morphology. *Molecular Phylogenetics and*

Evolution 21: 352–371.
Goswami A, Polly PD. 2010. The influence of character correlations on phylogenetic analyses: a case study of the carnivoran cranium. In: Goswami A,, In: Friscia A, eds. *Carnivoran Evolution*. Cambridge: Cambridge University Press, 141–164.
Goswami A, Smaers JB, Soligo C, Polly PD. 2014. The macroevolutionary consequences of phenotypic integration: from development to deep time. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369: 20130254–20130254.
Guillerme T, Brazeau MD. 2018. Influence of different modes of morphological character correlation on phylogenetic tree inference. *bioRxiv*: 308742.
Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W. 2008. GEIGER: investigating evolutionary radiations. *Bioinformatics* 24: 129–131.
Holland BR, Spencer HG, Worthy TH, Kennedy M. 2010. Identifying cliques of convergent characters: concerted evolution in the cormorants and shags. *Systematic Biology* 59: 433–445.
Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO. 2012. The global diversity of birds in space and time. *Nature* 491: 444–8.
Klingenberg CP. 2008. Morphological integration and developmental modularity. *Annual Review of Ecology, Evolution, and Systematics* 39: 115–132.
Leslie AB, Beaulieu JM, Crane PR, Knopf P, Donoghue MJ. 2015. Integration and macroevolutionary patterns in the pollination biology of conifers. *Evolution* 69: 1573–1583.
Livezey BC, Zusi RL. 2007. Higher-order phylogeny of modern birds (Theropoda, Aves: Neornithes) based on comparative anatomy. II. Analysis and discussion. *Zoological Journal of the Linnean Society*.
Lloyd GT. 2009. Graeme T. Lloyd matrices. Available at http://www.graemetlloyd.com/matr.html.
Mounce RCP, Sansom R, Wills MA. 2016. Sampling diverse characters improves phylogenies: craniodental and postcranial characters of vertebrates often imply different trees. *Evolution* 70: 666–686.

O'Keefe FR, Wagner PJ. 2001. Inferring and testing hypotheses of cladistic character dependence by using character compatibility. *Systematic Biology* 50: 657–675.
Orme D, Freckleton R, Thomas G, Petzoldt T, Fritz S, Isaac N, Pearse W. 2018. caper: Comparative Analyses of Phylogenetics and Evolution in R. R package version 1.0.1.
Pagel M. 1994. Detecting Correlated Evolution on Phylogenies: A General Method for the Comparative Analysis of Discrete Characters. *Proceedings of the Royal Society B: Biological Sciences* 255: 37–45.
Pagel M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401: 877–884.
Patterson TB, Givnish TJ. 2002. Phylogeny, concerted convergence, and phylogenetic niche conservatism in the core Liliales: insights from rbcL and ndhF sequence data. *Evolution* 56: 233–252.
Prum R. 1992. Syringeal morphology, phylogeny, and evolution of the neotropical manakins (Aves, Pipridae). *American Museum Novitates* 3043: 1–65.
Prum RO. 1993. Phylogeny, biogeography, and evolution of the broadbills (Eurylaimidae) and asities (Philepittidae) based on morphology. *The Auk* 110: 304–324.
Reeder TW, Townsend TM, Mulcahy DG, Noonan BP, Wood PL, Sites JW, Wiens JJ. 2015. Integrated analyses resolve conflicts over squamate reptile phylogeny and reveal unexpected placements for fossil taxa (P Wilf, Ed.). *PLOS ONE* 10: e0118199.
Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3: 217–223.
Sadleir RW, Makovicky PJ. 2008. Cranial shape and correlated characters in crocodilian evolution. *Journal of Evolutionary Biology* 21: 1578–1596.
Sansom RS, Wills MA. 2017. Differences between hard and soft phylogenetic data. *Proceedings of the Royal Society B: Biological Sciences* 284: 20172150.
Sansom RS, Wills MA, Williams T. 2017. Dental data perform relatively poorly in reconstructing mammal phylogenies: Morphological partitions evaluated with molecular benchmarks. *Systematic Biology* 66: 813–822.

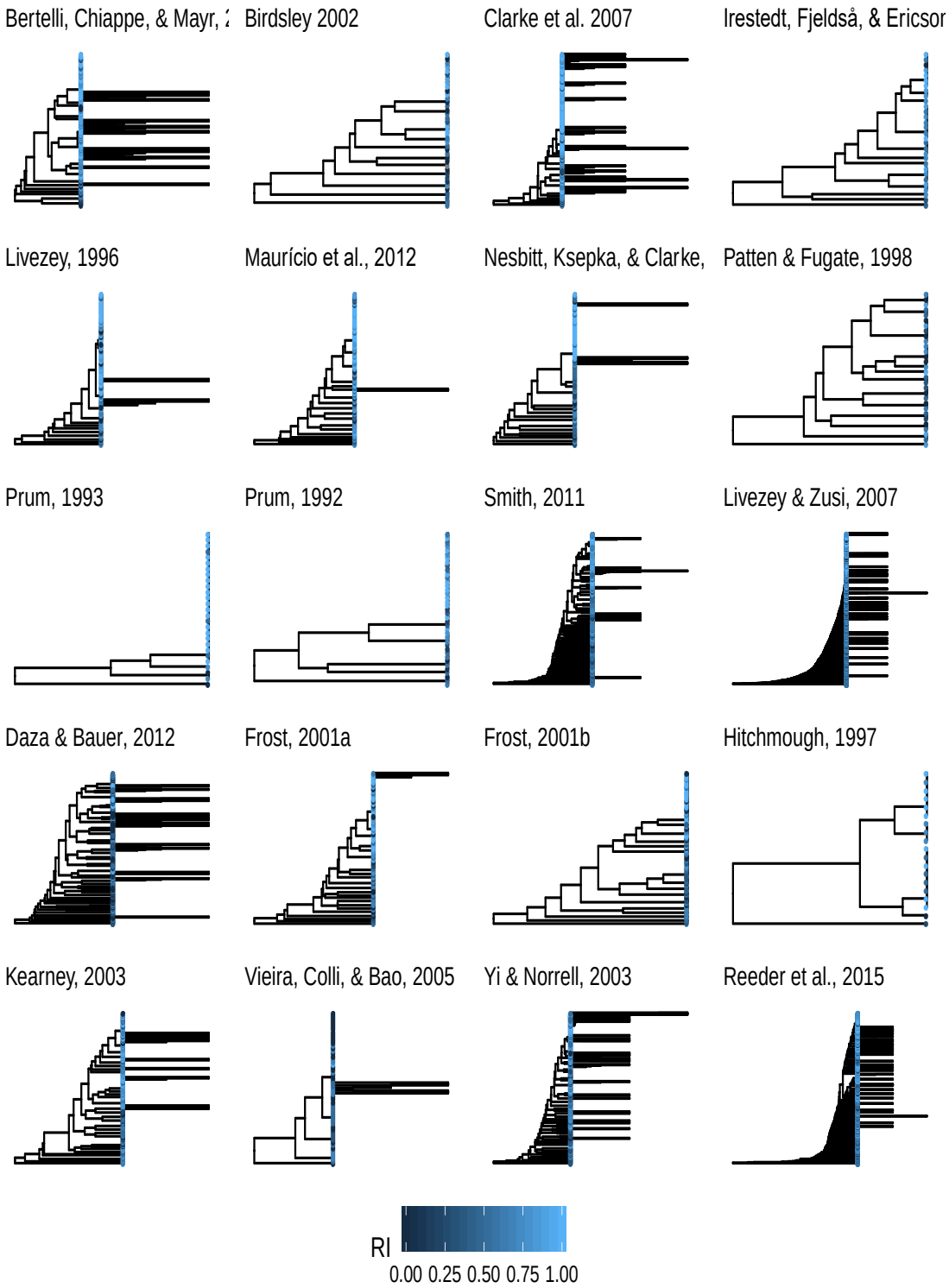
Sauquet H, von Balthazar M, Magallón S, Doyle JA, Endress PK, Bailes EJ, Barroso de Morais E, Bull-Hereñu K, Carrive L, Chartier M, Chomicki G, Coiro M, Cornette R, El Ottra JHL, Epicoco C, Foster CSP, Jabbour F, Haevermans A, Haevermans T, Hernández R, Little SA, Löfstrand S, Luna JA, Massoni J, Nadot S, Pamperl S, Prieu C, Reyes E, dos Santos P, Schoonderwoerd KM, Sontag S, Soulebeau A, Staedler Y, Tschan GF, Wing-Sze Leung A, Schönenberger J. 2017. The ancestral flower of angiosperms and its early diversification. *Nature Communications* 8: 16047.
Schliep KP. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics* 27: 592–593.
Zheng Y, Wiens JJ. 2016. Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Molecular Phylogenetics and Evolution* 94: 537–547.

Supplementary Information

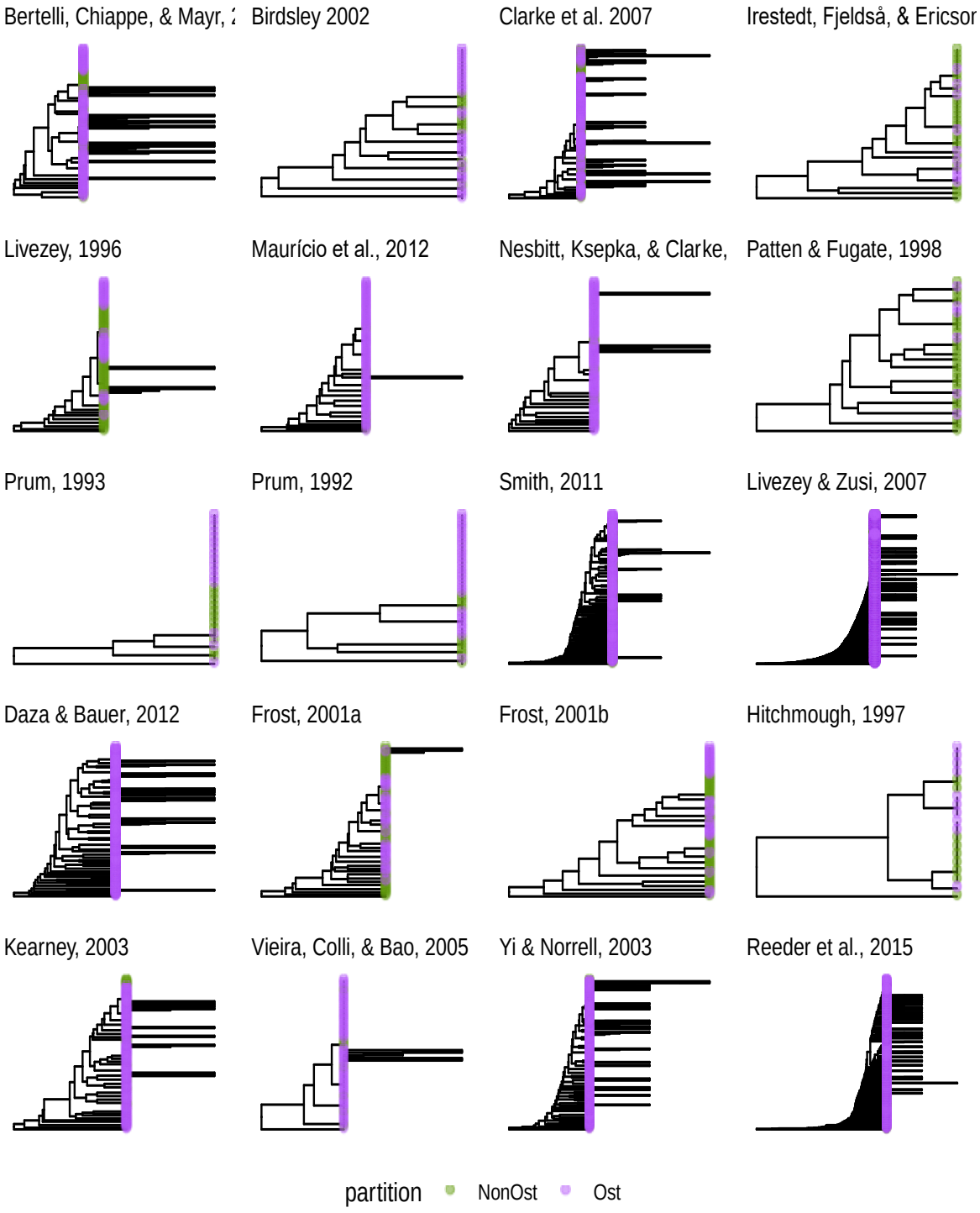
Supplementary Table 1. *AICc weights of 3 models of clustering of UPGMA tips by homoplasy*

Dataset			
Birds	**NoSig AICc Weight**	**Lambda AICc Weight**	**BM AICc Weight**
Bertelli, Chiappe, & Mayr, 2014	5.11E-08	1.00E+00	2.17E-44
Birdsley, 2002	7.02E-01	2.98E-01	2.36E-17
Clarke *et al.*, 2007	9.80E-07	1.00E+00	1.11E-81
Irestedt, Fjeldså, & Ericson, 2004	7.75E-01	2.25E-01	6.35E-21
Livezey, 1996	3.65E-01	6.35E-01	4.59E-25
Maurício *et al.*, 2012	6.61E-03	9.93E-01	2.22E-29
Nesbitt, Ksepka, & Clarke, 2011	4.47E-08	1.00E+00	1.19E-29
Patten & Fugate, 1998	7.75E-01	2.25E-01	2.96E-23
Prum, 1993	4.95E-03	9.95E-01	8.56E-07
Prum, 1992	1.39E-01	8.61E-01	1.53E-08
Smith, 2011	6.63E-08	1.00E+00	1.09E-187
Livezey & Zusi, 2007	5.30E-39	1.00E+00	0.00E+00
Squamates			
Daza & Bauer, 2012	1.25E-02	9.87E-01	6.93E-117
Frost *et al.*, 2001	2.62E-02	9.74E-01	1.21E-26
Frost *et al.*, 2001a	2.53E-01	7.47E-01	6.85E-41
Hitchmough, 1997	7.59E-01	2.41E-01	4.71E-06
Vieira, Colli, & Bao, 2005	6.39E-01	3.61E-01	4.17E-11
Yi & Norell, 2013	1.30E-05	1.00E+00	1.45E-122
Reeder *et al.*, 2015	1.51E-09	1.00E+00	0.00E+00

Showing AICc weights of 3 models of phylogenetic signal, fitted on the character retention index on molecular trees as tips on UPGMA trees. Models tested are a) no phylogenetic signal, where the retention index is randomly distributed among UPGMA tips, b) Lambda, where the retention index is non-randomly distributed but non-Brownian, and c) Brownian motion, where the retention index shows distinct clustering among UPGMA tips. The best model for each datasets is highlighted in green. Most datasets show a non-random distribution of retention index among tips, showing that internally consistent characters are also more likely to fit the underlying molecular phylogeny.



Supplementary Figure 1. *Showing character retention indices, where a high retention index indicates a good fit, on molecular trees mapped onto UPGMA trees of individual avian and squamate datasets, where characters are clustered by internal consistency. Characters with a low retention index are darker, while characters with a high retention index are lighter, Generally, characters are more likely to cluster together if they share similar consistency with molecular trees. In addition, many trees show a characteristic darkening further down the trees, showing that larger cliques are also likely to fir molecular trees better.*



Supplementary Figure 2. *Showing osteological (purple) and non-osteological (green) partitions mapped onto UPGMA trees of individual avian and squamate datasets, where characters are clustered by internal consistency. Generally, characters are more likely to cluster together if they belong to the same partition, demonstrating greater correlation within than between partitions.*

Table 2. *Signal in homoplasy and partition membership, of morphological characters on UPGMA trees*

				Distribution of Homoplasy in Character Clusters				**Distribution of Partition Membership in Character Clusters**		
				Pagel's Lambda		**Blomberg's K**		**Purvis's D**		
Dataset	**Taxa**	**Osteological Characters**	**Non-Osteological Characters**	**λ**	**P**	**K**	**P**	**D**	**P (Random)**	**P (Brownian)**
Birds										
Bertelli, Chiappe, & Mayr, 2014	17	85	39	0.312	0.00*	0.002	0.001*	0.329	0.00*	0.00*
Birdsley, 2002	31	27	11	0.192	0.42	0.001	0.193	0.449	0.01*	0.04*
Clarke *et al.*, 2007	30	96	78	0.373	0.00*	0.001	0.014*	0.368	0.00*	0.00*
Irestedt, Fjeldså, & Ericson, 2004	38	9	21	0	1	0.000	0.347	0.666	0.08	0.01*
Livezey, 1996	25	31	51	0.3000	0.07	0.001	0.043*	-0.059	0.00*	0.61
Maurício *et al.*, 2012	34	80	8	0.414	0.00*	0.001	0.039*	0.879	0.17	0.00*
Nesbitt, Ksepka, & Clarke, 2011	11	82	9	0.509	0.00*	0.002	0.001*	0.439	0.00*	0.04*
Patten & Fugate, 1998	27	6	24	0	1	0.000	0.493	0.991	0.43	0.00*
Prum, 1993	12	16	11	0.793	0.00*	0.008	0.008*	0.662	0.00*	0.85
Prum, 1992	30	29	10	0.294	0.01*	0.003	0.019*	0.032	0.00*	0.43
Smith, 2011	52	195	104	0.286	0.00*	0.000	0.057	0.534	0.00*	0.00*
Livezey & Zusi, 2007	139	1528	436	0.345	0.00*	0.000	0.023*	0.405	0.00*	0.00*
Squamates										
Daza & Bauer, 2012	12	227	17	0.300	0.00*	0.001	0.004*	0.672	0.00*	0.00*
Frost *et al.*, 2001b	28	32	46	0.271	0.00*	0.001	0.079	0.102	0.00*	0.27
Frost *et al.*, 2001a	26	31	40	0.249	0.04*	0.000	0.153	0.291	0.00*	0.05
Hitchmough, 1997	12	10	9	0.808	0.46	0.003	0.014*	0.381	0.01*	0.11

Vieira, Colli, & Bao, 2005	10	49	8	0.269	0.29	0.001	0.06	0.866	0.19	0.00*
Yi & Norell, 2013	25	209	25	0.352	0.00*	0.001	0.016*	0.783	0.01*	0.00*
Reeder *et al.*, 2015	135	533	93	0.295	0.00*	0.000	0.103	0.571	0.00*	0.00*

Showing the dataset dimensions and presence of signal in homoplasy (retention index), and partition membership, of morphological characters on UPGMA trees clustering morphological characters in several avian and squamate datasets by their phylogenetic similarity using three metrics. We use Pagel's lambda and Blomberg's K to measure signal in homoplasy, and Fritz & Purvis's D to measure signal in partition membership. Here, UPGMA trees are treated as phylogenetic trees and the retention index of characters on molecular trees treated at the trait of interest. In both metrics used to measure signal in homoplasy, values close to 0 indicate close to random trait distribution, and values close to 1 indicate a distribution best explained by Brownian motion, indicating clumpiness of traits. Significant P values indicate that metrics are significantly different from 0, indicating the presence of signal in the trait distribution. In Purvis's D used to measure signal in partition membership, values close to 1 indicate close to random trait distribution, and values close to 0 indicate close to a distribution best explained by Brownian motion, indicating clumpiness of traits. Significant P values (random) indicate that metrics are significantly different from 1, indicating the presence of signal in the trait distribution. Significant P values (Brownian) indicate that the metrics are significantly different from 0, indicating no Brownian motion.

4. Convergence, Correlations and Cliques: An Analysis of the Relative Performance of Morphological Character Partitions in Mammals

Abstract

Partitions of morphological data evolve under different selective regimes, resulting in different levels of homoplasy and integration. These differences raise questions about the ability of these data types to reconstruct evolutionary relationships. Dental and osteological data, for example, have been shown to imply different trees. Dental traits may be particularly prone to homoplasy and correlation as a result of trophic convergence and functional integration. Here, we take several approaches to compare osteological and dental partitions in 38 mammalian morphological datasets, testing for differences in homoplasy, transition ages and correlations using a newly published molecular tree of mammals as a benchmark. We also establish the internal consistency of characters and infer trees from these partitions. We find that osteological characters are less homoplastic (p = 0.001) and transition earlier ($p < 0.0001$) than dental characters relative to molecular trees. We find levels of correlation to be roughly equal between partitions, but characters are more correlated within than between partitions both with and without reference to the molecular tree. Further, we find that dental and osteological data imply different trees, with osteological trees being more similar to the molecular tree. These results converge on the conclusion that osteological and dental data have different properties. These differences should be taken into account when inferring evolutionary history using these data types.

1. Introduction

The convergence and non-independence of morphological traits have long been acknowledged to affect morphological data, and consequently phylogenetic reconstruction inferred from these data (Emerson & Hastings, 1998; Sadleir & Makovicky, 2008; Goswami & Polly, 2010; Billet & Bardin,

2019). These phenomena may arise as a result of selection on morphological traits – species occupying similar habitats or with similar diets, for example, often exhibit extraordinary morphological similarity despite being unrelated (e.g. McCurry *et al.*, 2017), sometimes to the extent of misleading phylogenetic analysis, necessitating the use of molecular data to infer relationships intractable to morphology alone (e.g. Stanhope *et al.*, 1998; Van Tuinen *et al.*, 2001). Selective regimes affecting characters which are ecologically, functionally or genetically linked can give rise to concerted convergence (Patterson & Givnish, 2002) potentially affecting entire suites of traits, leading to the presence of internally correlated cliques (Holland *et al.*, 2010; Blanke *et al.*, 2013). Despite their apparent prevalence, the extent of these phenomena in morphological datasets spanning wide taxonomic groups, as well as the relative prevalence of these issues between morphological subsets, remains largely unknown.

Despite these challenges, and the availability of molecular data (and its ability to produce well-resolved, reliable trees), morphology remains an indispensable source of biological information in phylogenetic analysis. This is in large part owing to the need to include fossil and rare species, for which molecular data are usually unavailable. Even when molecular data are available, however, morphological data remain vital since combined analyses allow the identification of consilience (Field *et al.*, 2014), provide greater statistical support to phylogenetic hypotheses (Lee & Camens, 2009), and help discriminate between hypotheses owing to the presence of hidden support (Gatesy & Arctander, 2000; Wahlberg *et al.*, 2005; Lee, 2009). However, questions remain about the properties of subsets of morphological characters. For example, do different subsets of morphological data display more convergence, faster evolutionary rates, or stronger internal correlation and integration?

Previous analyses of sauropsid datasets have demonstrated that there are differences in homoplasy, transition age and correlation relative to molecular trees (Chapter 1, 2) and internal consistency (Chapter 3) between osteological and non-osteological character partitions. Osteological characters were shown to be more consistent with molecular trees, and to

have older transition ages on average, and characters were more correlated within than between these partitions. Previous authors have additionally found heterogeneity in the phylogenetic signal conveyed between hard and soft characters (Sansom & Wills, 2017) and dental and osteological characters (Sansom, Wills, & Williams, 2017) using partition difference tests.

Differences in developmental and evolutionary origins, as well as selection acting on traits, may explain observed differences between these character subsets. For example, plumage in birds is driven by sexual selection and is relatively labile (Price, Friedman, & Omland, 2007). Additionally, cranial and dental evolution are driven by diet, which results in divergence, convergence and functional integration (Sakamoto & Ruta, 2012; Felice *et al.*, 2019; Sakamoto, Ruta, & Venditti, 2019; Godoy, 2019). This results in different levels of homoplasy and correlation between subsets of morphological data. Alternatively, observed differences may be driven by systematic differences in the way character subsets are coded by authors. Since these properties have been shown to impact phylogenetic inference (Sadleir & Makovicky, 2008; Goswami & Polly, 2010; Billet & Bardin, 2019), it is vital for evolutionary biologists to examine how general and widespread these phenomena are in subsets of morphological data, and whether they affect the tree topology resulting from phylogenetic analyses using these data.

1.1. Mammals

Mammals are a large and well studied, morphologically and ecologically disparate group of vertebrates, characterized by widespread convergence and parallelism between distantly related species, for example marine species, ungulate-like species and hedgehog-like species (Madsen *et al.*, 2001; Kelley & Motani, 2015; Gheerbrant, Filippo, & Schmitt, 2016; Mazel *et al.*, 2017).

As in many other groups, phylogenetic analyses of mammals have yielded contradictory results, not least as a result of convergence. Furthermore, the resolution of deep branching events has not been trivial (Foley, Springer, & Teeling, 2016). Areas of debate include the earliest divergence between the three major clades of placental mammals, namely Afrotheria, Xenarthra and Boreoeutheria. Each of the three possible branching orders between these clades has been supported at various times by molecular data (Scally *et al.*, 2001; Madsen *et al.*, 2001; Murphy *et al.*, 2001a,b; Delsuc *et al.*, 2002; Teeling & Hedges, 2013).

Recent advances in molecular mammalian phylogenetics (Upham, Esselstyn, & Jetz, 2019), as well as an abundance of morphological data, provide an opportunity to test some of the above questions on one of the most well studied and ecologically diverse vertebrate groups.

1.2. Approach and Hypotheses

Here, we use published morphological and molecular data to address questions relating to differences in the evolution and phylogenetic performance of osteological versus dental partitions in mammals. There are a priori reasons to examine the differences between these morphological regions, since dental and osteological data produce difference trees (Sansom *et al.*, 2017). Dental data are additionally often convergent owing to trophic similarity in unrelated animals (McCurry *et al.*, 2017), and correlated because of functional and developmental constraint (Billet & Bardin, 2019). We use the morphological datasets collected by Sansom *et al.*, (2017) for all analyses. As a benchmark, we use a newly published molecular tree (Upham *et al.*, 2019) for comparison with morphological data. We address the following topics:

1. The distribution of internal consistency of characters within and between partitions.
2. The distribution of homoplasy of morphological characters relative to molecular trees, and relative character transition ages.

3. The distribution of character correlations relative to molecular trees within and between partitions.
4. The phylogenetic hypothesis inferred from each partition.

We first examine the internal consistency of these partitions, using dissimilarity matrices to search for cliques of compatible characters (Holland *et al.*, 2010; Blanke *et al.*, 2013). As such, we directly test the hypothesis that there will be greater internal consistency within than between character partitions. We then compare the consistency and ages of these partitions on time-calibrated molecular trees using the retention index as a measure of homoplasy and using character state reconstructions to estimate average transition ages. Since dental data have been shown to display poor phylogenetic signal compared with osteological data (Sansom *et al.*, 2017), we hypothesize that dental data may exhibit poorer fit than osteological data on molecular trees. They may also be more evolutionarily labile than osteological characters, which may manifest as younger ages on molecular trees. These hypothesese are also consistent with previous work showing osteological data to be more consistent with molecular trees than soft characters in birds and reptiles (Chapter 1).

Further, we examine correlations between character pairs, again using molecular trees as a benchmark. We take a maximum-likelihood approach comparing models of trait evolution (Pagel, 1994). Here, we expect characters to exhibit greater correlation within than between partitions.

Finally, we generate new trees from the dental, osteological and combined data from the largest and most taxonomically diverse of our morphological datasets (O'Leary *et al.*, 2013) to look directly at the phylogenetic signal exhibited by each of the partitions. We chose the O'Leary dataset for this additional analysis in order to examine whether osteological and dental data support different hypotheses of early branching events in placental mammals. Some outstanding questions include the branching order of Boreoeutheria, Afrotheria and Xenarthra, and the Laurasiatheria polytomy. These questions are contentious, and an area of particular interest to mammal phylogeneticists. We expect these different partitions to support

different hypotheses of evolutionary history, although these may or may not be congruent with previous molecular or morphological analyses.

2. Methods

2.1. Data Collection

The morphological data sample comprised 38 published and edited data matrices containing osteological and dental character partitions (Sansom *et al.*, 2017). In brief, matrices were compiled from Google and Google Scholar, with minimum dimensions of 30 characters and 10 taxa, and minimum taxonomic overlap. See Sansom *et al.*, (2017) for details of data compilation. We retained fossil taxa in our analyses where possible, in order to gain a comprehensive look at trait evolution in crown mammals. Our datasets here include both a large dataset spanning all of Mammalia (O'Leary *et al.*, 2013), and smaller datasets covering narrower taxonomic levels..

For the molecular data, we used a new time-calibrated molecular tree of mammals (Upham *et al.*, 2019). This tree contains 5911 mammalian species, 4098 of which have molecular information available. The tree was constructed from 31 genes under a Bayesian framework, and dated using 17 fossil calibrations using node dating methods (Upham *et al.*, 2019). For all analyses involving these trees, we deleted any species from the character matrix that were lacking in molecular data. We additionally performed some of the following analyses (homoplasy and character transition ages) on the bat morphological dataset (Fracasso, De Oliveira Salles, & Perini, 2011) against an independent tree constructed from cytochrome b (Agnarsson *et al.*, 2011) for comparison. This is to test whether these methods are robust to different estimates of evolutionary history.

2.2. Cluster Analysis

To explore character compatibility in our morphological matrices, we first performed cluster analyses on all datasets following the methods of (Holland *et al.*, 2010; Blanke *et al.*, 2013) and our earlier work (Chapter 3). Since finding internally consistent cliques does not rely on a phylogeny, fossil species were retained for this analysis.

For each of the mammalian datasets, we initially constructed a dissimilarity matrix of pairwise excess indices in Python using scripts developed by previous authors (Holland *et al.*, 2010; Blanke *et al.*, 2013). Here, the excess of a pair of characters is defined as the number of extra steps required on the most parsimonious tree that can be constructed with those characters, over and above the minimum number of possible steps for those characters, where a pairwise parsimony score of 0 indicates perfect compatibility between 2 characters. We then constructed UPGMA trees from these dissimilarity matrices, using the package Phangorn (Schliep, 2011) in R (R Core Team, 2018), whereby characters cluster according to their phylogenetic similarity. We compared the size of the largest cliques in each dataset with a null distribution, following the procedure described in chapter 3 (Holland *et al.*, 2010; Blanke *et al.*, 2013)

We further used Purvis's D (Fritz & Purvis, 2010), implemented in the R package Caper (Orme *et al.*, 2018), to detect whether osteological and dental characters cluster together more than expected by chance on UPGMA trees. Purvis's D is a measure of phylogenetic signal of binary traits. On phylogenetic trees, Purvis's D tests a model of Brownian motion (indicating clustering) against a null model of random trait evolution. We randomized the trait values among tips on one dataset, to confirm that this method is not likely to spuriously detect internal consistency on UPGMA trees.

2.3. Homoplasy

For analysis of congruence of morphological partitions on molecular trees, we used 100 trees from the post-burn in posterior distribution of the phylogenetic analysis of Upham *et al.*, (2019), using the retention index as a measure of homoplasy. This approach takes some phylogenetic uncertainty into account. We only included species for which there was molecular data available, and we only used morphological datasets in which tips were species. The exception to this was a bat phylogeny (Fracasso *et al.*, 2011), to which we assigned species belonging to the taxa at the tips, in order to compare the congruence of this dataset on the Upham *et al.*, (2019) and alternative Agnarsson *et al.*, (2011) topologies. This is because this was a useful phylogeny to use for comparison because of the availability of alternative molecular data and a suitable number of taxa and characters.
We first calculated the individual retention indices for all informative characters in all datasets, and the ensemble retention indices for osteological and dental partitions for all datasets, averaging over the 100 trees, in TNT (Goloboff, Farris, & Nixon, 2008). The retention index (ri) of a character on a tree is defined as the difference between its maximum number of states changes on a tree (g) and its actual number of changes on that tree (s), divided by the maximum number of changes on the tree (g) and the minimum possible number of steps (m), so that ri = gs/gm. The retention index calculated for a group of characters is the ensemble retention index. A retention index of 0 indicates poor congruence with the tree, while a retention index of 1 indicates a perfect fit.
We compared the fit of these characters on the trees by comparing linear mixed effect models in the package nlme (Pinheiro *et al.*, 2019) implemented in R, accounting for heteroskedascticity and random dataset effects. We compared a null model, with no fixed effects, to a model where partition is treated as a fixed effect. In both models, the individual character retention index is the dependent variable and dataset is treated as a random effect.

We additionally tested whether there was any relationship between the morphological character clusters and the distribution of homoplasy of these characters relative to molecular data, i.e. whether characters cluster together on the UPGMA trees described above according to their retention index on molecular trees more than expected by chance, using Pagel's lambda (Pagel, 1999), implemented in the R package Phytools (Revell, 2012) to measure phylogenetic signal. Pagel's Lambda describes the extent to which branches on a tree must be transformed in order for a Brownian motion model to explain the trait distribution.

2.4. Relative Transition Ages

To compare the relative ages of partitions, we follow methods used previously (Chapter 1). We first performed character state reconstructions on each of 100 molecular trees for each dataset in R using the package Paleotree (Bapst, 2012), which estimates the node or nodes on the tree at which each character changes state. These reconstructions were conducted under a maximum parsimony framework using the ACCTRAN (accelerated transformation) algorithm, which minimizes distance from the root and favours secondary loss of traits as opposed to repeated gains. ACCTRAN is generally preferred over DELTRAN (delayed tranformation) for this reason. In our view, however, our choice of algorithm is arbitrary given we are interested in relative, not absolute transition ages. Nevertheless, we performed the same analysis using the MPR (most parsimonious reconstruction) algorithm on one of our datasets (Fracasso *et al.*, 2011), for comparison with our main transition age analysis.
For each tree, transition ages were then assigned to characters by generating and averaging 100 random ages between the nodes at which the characters transition and their ancestral node. Averages were also taken between the transition ages of characters transitioning multiple times on a tree to obtain an overall average transition age for each character. The final character transition ages were then compared between osteological and

dental partitions using linear mixed effect models accounting for heterskedasticity. Again, a null model was compared to a model with partition as a fixed effect. To test whether any difference in transition age can be explained by homoplasy, a further analysis was conducted with the character retention index treated as an additional fixed effect. Dataset was treated as a random effect in all analyses.

2.5. Correlations

In addition to the internal correlations described above, we derived morphological character correlations relative to molecular topologies in a maximum likelihood framework following the procedure of previous analyses (Chapter 2) using the corHMM package (Beaulieu, O'Meara, & Donoghue, 2013) implemented in R. For this analysis, multistate characters and characters with missing data were removed. We only included datasets for which there were 5 or more characters (=10 pairwise comparisons) in each partition after removal of characters.

We compared 3 models of character evolution on all character pairs in all datasets, namely an Equal Rates model (ER), a Symmetrical model (SYM) and an All Rates Different (ARD) model. In a rate matrix of a pair of characters, equal rates occur if the rate of change of one character in the pair is independent of the character state of the other. Rates in the rate matrix are symmetrical when the rate of change of one character in the pair depends on the state of the second, but not vise versa. When all rates in the rate matrix differ, the rate of change of one character in a pair is always dependent on the state of the other. Thus, ER is an uncorrelated model, while SYM and ARD are both correlated models. We compared these models for all character pairs using AICc weights, and summed the weights of the SYM and ARD models to obtain the cumulative weights for correlated models, as in Sauquet *et al.*, (2017). We used equal weights for ancestral states.

We then calculated the percentages of correlated character pairs in each partition, where character pairs are considered to be correlated if this cumulative AICc weight for the pair is more than or equal to 0.95. Since this choice of threshold is somewhat arbitrary, we also compared AICc weights between partitions by fitting linear mixed effects models similar to those described above. Namely, we compared a null model with a model with partition as a fixed effect.

2.6. Topologies Inferred from Osteological Versus Dental Data

To see whether partitions support different hypotheses of evolutionary history, we built new trees from the O'Leary *et al.,* (2013) dataset. The first phylogenetic analysis used the combined dental and osteological data, and we further analysed the two data types separately to create a total of 3 new sets of trees. All analyses were performed in TNT under a maximum parsimony framework using the tree bisection and reconnection algorithm with 1000 replications. We then compared these trees to each other and to the molecular tree using the Robinson-Foulds distance metric (Robinson & Foulds, 1981) as implemented in the R package Phangorn (Schliep, 2011).

3. Results

3.1. Cluster Analysis

Overall, results indicate that dental and osteological characters in these datasets cluster together more than expected by chance. UPGMA analyses of character dissimilarity matrices derived from mammalian morphological datasets resulted in 38 trees, on whose tips we then mapped the trait data (osteological and dental). Out of these 38 UPGMA trees, trait distribution was significantly different from random in 36, with strong evidence of significant clumping in 9 trees, as measured by Purvis's D metric for

measuring phylogenetic signal (Table 1, Figure 1). Here, significant clumping is indicated when trait distribution among UPGMA tips is not significantly different from that expected under a Brownian motion model of trait distribution. There is significant left-sided skew, i.e. towards lower values, in the distribution of Purvis's D values across datasets (Kolmogorov-Smirnov comparing left skew against a normal distribution, D = 0.507, $p < 0.0001$), as well as its being two-sided (Kolmogorov-Smirnov comparing a two-sided distribution against a normal distribution, $p < 0.0001$). In addition, clique sizes are no larger than expected by chance (supplementary Table 1).

Table 1. *The 'phylogenetic signal' of trait distribution among UPGMA tree tips*

				Relationship Between Character Clustering and Partition Membership		
Dataset	**Taxa**	**Osteological Characters**	**Dental Characters**	**D**	**P (random)**	**P (Brownian)**
Ahrens, (2012)	12	29	37	0.455	0	0
Asher *et al.*, (2005)	57	143	78	0.404	0	0.001
Asher *et al.*, (2010)	30	90	25	0.481	0	0.001
Boessenecker & Churchill, (2013)	17	55	30	0.554	0	0.001
Bi *et al.*, (2014)	37	270	132	0.656	0	0
Billet, (2011)	41	55	63	0.366	0	0
Boisserie, (2005)	14	24	12	0.223	0	0.264
Bryant, Russell, & Fitch, (1993)	24	20	16	0.572	0.014	0.001
Carleton, (1980)	71	27	15	0.33	0.001	0.038
Carstens, Lundrigan, & Myers, (2002)	38	19	43	0.423	0	0.005
Cerdeño, (1995)	41	46	25	0.679	0.031	0
Churchill, Boessenecker, & Clementz, (2014)	24	91	19	0.851	0.152	0
Domning, (1994)	31	45	14	0.225	0.001	0.163
Finarelli, (2008)	25	34	41	0.266	0	0.012
Fracasso *et al.*, (2011)	20	85	45	0.897	0.191	0
Froelich, (1999)	24	26	84	0.556	0	0
Gaubert *et al.*, (2005)	37	195	66	0.802	0.006	0
Gentry, (1992)	27	91	19	0.596	0.001	0

Gheerbrant *et al.*, (2005)	12	38	75	0.147	0	0.098
Giannini & Simmons, (2005)	49	129	36	0.025	0	0.494
He *et al.*, (2012)	11	40	33	0.017	0	0.449
Kielan-Jaworowska & Hurum, (2001)	11	23	28	0.127	0	0.2
Ladeveze, Pieter, & Smith, (2010)	12	37	20	0.7	0.018	0
LeCompte, Granjon, & Denys, (2002)	15	37	20	0.792	0.16	0
O'Leary *et al.*, (2013)	35	1434	765	0.867	0	0
Olivares & Verzi, (2015)	21	45	14	0.442	0.018	0
Oliviera *et al.*, (2011)	46	39	41	0.556	0.001	0
Prideaux & Warburton, (2010)	17	50	31	0.261	0	0.126
Sanchez-Villagra, Horovitz, & Motokawa, (2006)	14	105	47	0.055	0	0.24
Silcox *et al.*, (2010)	21	108	112	0.339	0	0
Springer, Kirsch, & Chase, (1997)	16	41	42	0.267	0	0.19
Steppan, (1993)	31	50	32	0.802	0.138	0
Strait & Grine, (2004)	14	35	52	0.869	0.104	0
Thorington, Pitassy, & Jansa, (2002)	20	33	44	0.342	0	0.005
Tomiya, (2011)	47	52	35	0.818	0.077	0
Weksler, (2006)	34	39	29	0.237	0.002	0.167
Wroe & Musser, (2001)	15	30	40	0.26	0	0.013
Zrzavý & Řičánková, (2004)	23	42	35	0.797	0.063	0

Showing the number of taxa, number of characters in each partition, and the 'phylogenetic signal' of trait distribution among UPGMA tree tips, where the traits are dental vs osteological data, as measured by Purvis's D for calculating phylogenetic signal in binary trait data. P (random) is the probability that the trait distribution is random, and p (Brownian) is the probability that the trait distribution is consistent with Brownian motion, or significant clumping. D values rage between 0 and 1, where 1 indicates random distribution and 0 indicates Brownian motion.

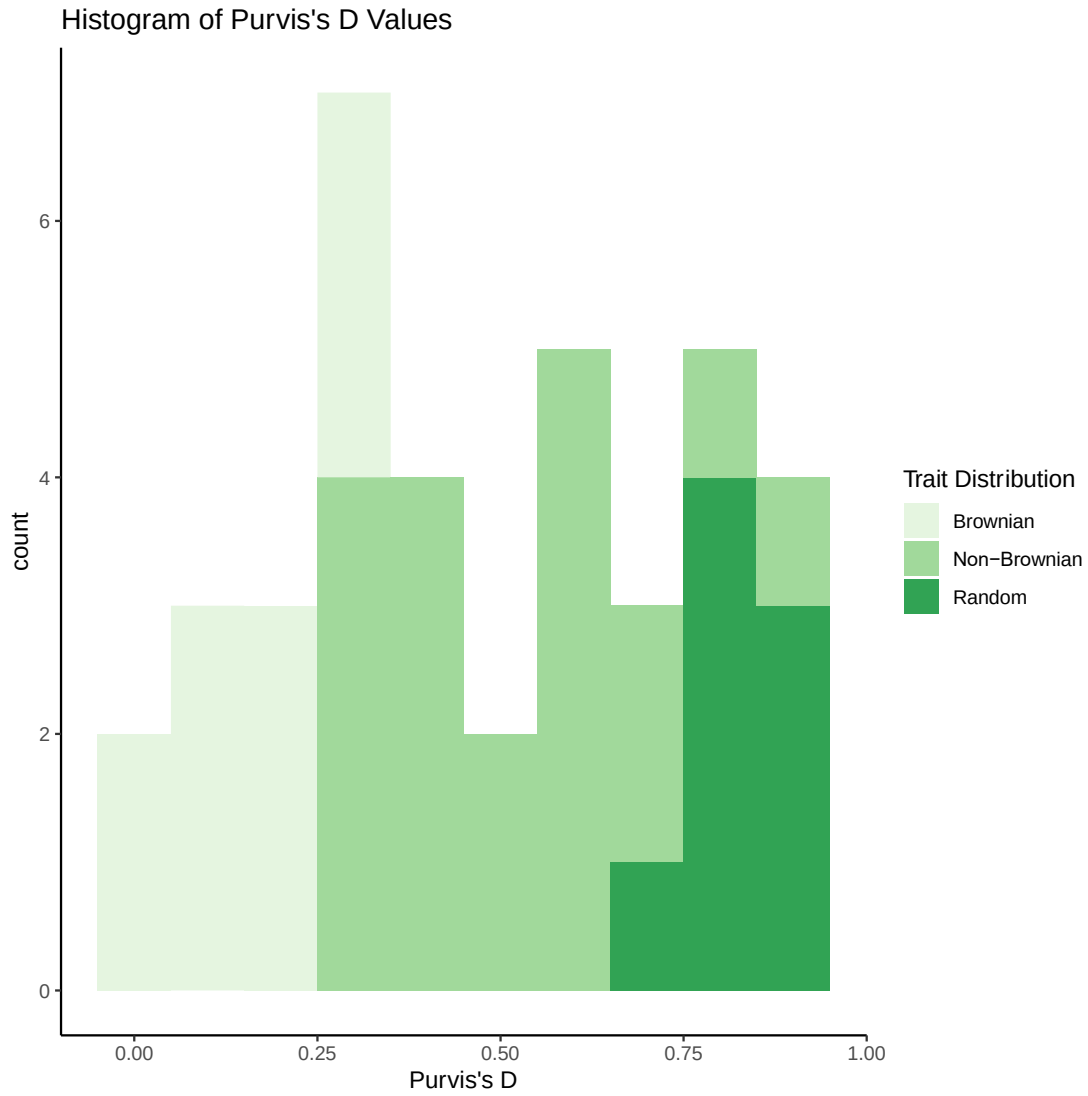


Figure 1. *Showing the distribution of Purvis's D values among the UPGMA trees, where characters are clustered by their internal consistency, showing the degree of clustering of characters by partition (osteological vs dental) in 39 morphological datasets. Colours relate to p values as given by Purvis's D, where pale green indicates a trait distribution consistent with a model of Brownian motion, indicating significant clustering of partitions. Dark green indicates a random trait distribution, while intermediate green indicates a non-random intermediate distribution that is also significantly different from Brownian ("Non-Brownian"). D values range between 0-1 and are binned in increments of 0.1, where 0 indicates strong clustering and 1 indicates a random trait distribution. Where Brownian motion explains the distribution within a dataset, characters cluster significantly by partition.*

3.2. Homoplasy

There was a significant difference in homoplasy as measured by the retention index on molecular trees between osteological and dental partitions. Homoplasy was lower in osteological partitions in (ensemble retention index in 11 out of 15 datasets), although this was marginally non-significant (paired t, t = 1.8, df = 14, p = 0.09, mean of differences = 0.068, Figure 2). The distribution of homoplasy in the Fracasso *et al.*, (2011) morphological dataset was similar relative to both molecular topologies (ensemble RI of 0.44 and 0.43 for osteological data, and 0.32 and 0.28 for dental data on the large mammal topology and the alternative topology respectively), suggesting that overall results are likely to be comparable even given different estimates of evolutionary history. Compared to a null model, the linear mixed effects model treating character partition as a fixed effect had significantly greater explanatory power, as measured by the AIC weights and likelihood ratios, taking into account dataset effects (Table 3, analysis 1).

In addition, there is an overall poor correspondence of homoplasy with internal consistency of characters. The distribution of character retention indices on molecular trees among UPGMA tips is significantly different from random in only 5 of 14 datasets as measured by Pagel's lambda (Table 2).

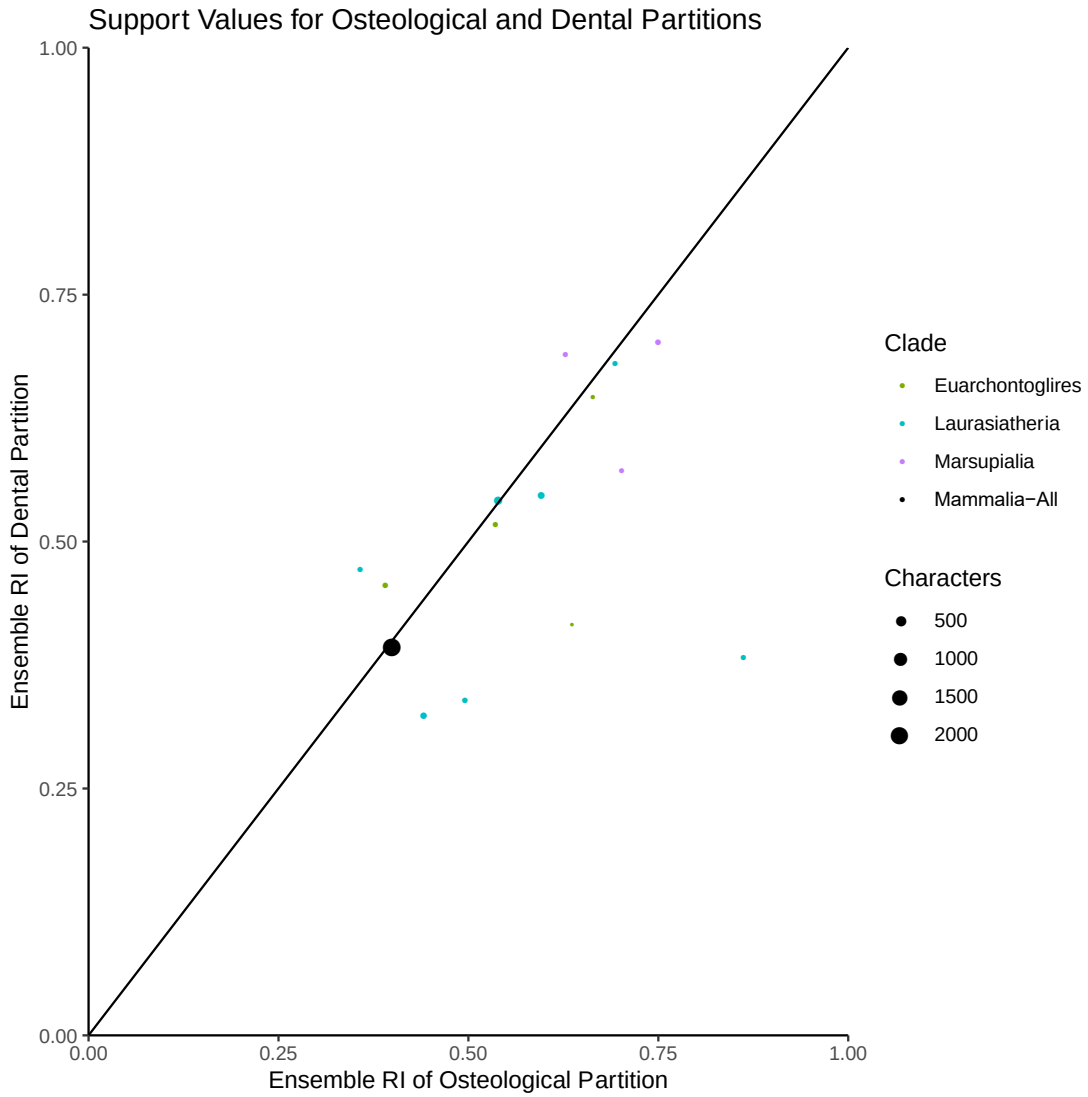


Figure 2. *Showing the ensemble retention index of osteological partitions by the ensemble retention index of dental partitions for each mammalian dataset. All datasets falling below the x=y line exhibit higher consistency of osteological than dental data with molecular trees. Datasets are coloured by clade and sized by number of characters.*

Table 2. *The distribution of homoplasy as measured by the character retention index among UPGMA tree tips*

	Relationship Between Character Clustering and Homoplasy Relative to Molecular Topology			
Dataset	**λ**	**LogL**	**LogL0**	***p***
Ahrens, (2012)	0	-21.293	-21.292	1
Carstens, Lundrigan, & Myers, (2002)	0	-4.661	-4.661	1
Fracasso *et al.*, (2011)	0	-39.731	-39.730	1
Gaubert *et al.*, (2005)	0	-40.357	-40.353	1
He *et al.*, (2012)	0.458	-26.748	-32.079	0.001*
LeCompte, Granjon, & Denys, (2002)	0.451	-7.631	-10.999	0.009*
O'Leary *et al.*, (2013)	0	-71.090	-71.054	1
Oliviera *et al.*, (2011)	0.524	-2.899	-5.553	0.021*
Prideaux & Warburton, (2010)	0.278	-7.278	-9.365	0.041*
Sanchez-Villagra, Horovitz, & Motokawa, (2006)	0.290	-35.284	-39.449	0.004*
Steppan, (1993)	0	5.982	5.983	1
Weksler, (2006)	0	-4.609	-4.609	1
Wroe & Musser, (2001)	0	-20.069	-20.068	1
Zrzavý & Řičánková, (2004)	0.115	-15.094	-14.981	1

Showing the distribution of homoplasy as measured by the character retention index among UPGMA tree tips using Pagel's Lambda, a metric for detecting phylogenetic signal in continuous traits. Significant p values indicate that the distribution of the retention index is different from random. Values of Lambda range between 0 and 1, where 0 indicates a random trait distribution and 1 indicates a trait distribution consistent with Brownian motion.

Table 3 at End of Document

Table 3. *Showing the results of all linear mixed effects models fitted to the data, testing whether osteological and dental partitions explain differences in homopasy, transition ages and correlations while taking into account dataset effects. There are 3 dependent variables in 4 total analyses including the effect of partition on homoplasy as measured by character retention indices on molecular trees, the effect of partition on transition ages as measured by estimating ages given a character state reconstruction on molecular trees, the effect of retention index versus partition on transition ages, and finally the effect of partition on the strength of correlation between character pairs. For the final analysis, within- and between-partition is tested as a fixed effect as well as partition itself. Dataset is treated as a random effect in all models, and partition treated as a fixed effect in all models except the null. Each model fitted for each analysis is given an AIC weight, where each weight is a value between 0-1, and the weights of each model add to 1 for each analysis. The highest AIC weight is given to the model which best describes the data. The best model in each analysis as measured by AIC weights is highlighted in green.*

3.3. Transition Ages

Average ages of osteological character transitions within datasets were older than dental characters in 10 out of 14 datasets (paired t, t = 2.244, df = 13, p = 0.043, mean of differences = 0.09 million years, supplementary Figure 2). The significant effect of partition on transition age remained when comparing linear mixed effect models (p = 0.0001, Table 3, analysis 2). However, when individual character retention indices are treated as a fixed effect, this also has greater explanatory power against a null (Table 3, analysis 3). Given that homoplasy and partition are related, it is possible that the retention index is driving the transition ages. To test this, we performed an additional analysis where models treating both the retention index and partition as fixed effects were compared against both a null and a model treating only the retention index as a fixed effect, and transition age as the response variable. In this analysis, adding partition as an additional fixed effect did not increase the explanatory power (Table 3, analysis 3),

demonstrating that the partition does not have an effect on the transition age independent of its correlation with homoplasy. Taken together, these results indicate a relationship between transition age and retention index (where more homoplasious characters transition later, supplementary Figure 3), but no difference in transition ages between partitions beyond that expected owing to differences in homoplasy between these partitions. Our additional analysis using the MPR algorithm to estimate relative transition ages for the Fracasso *et al.*, (2011) dataset yielded similar results to our main ACCTRAN analysis, albeit with a smaller age difference between partitions (osteological ages estimated by ACCTRAN versus MPR = 23.01 million years and 19.77 million years respectively, dental ages = 18.95 and 18.86 million years). This difference may suggest a slight sensitivity of transition age estimates to choice of character state reconstruction method, although the overall pattern of relative age difference between partitions remains the same for this dataset.

3.4. Correlations

Overall, characters are more correlated within than between partitions, with some datasets showing a difference in correlation between osteological and dental data. The latter result, however, differs between datasets. Correlated character pairs were identified in 8 out of 12 datasets (Table 3). Of these, a significant difference between the proportion of correlated between-partition character pairs and within-partition character pairs was observed in only 2 datasets, a higher proportion of correlated characters observed within partitions in both cases (Fisher's exact test, Table 3). Significant differences in the proportion of character pairs between all three character-pair classes (i.e. within osteological character pairs, within dental character pairs, and between partition character pairs) were observed in 4 datasets, with mixed results between dental and osteological data (Table 4). Given the high (and arbitrary) threshold, we also took into account the cumulative AICc weights of correlated (SYM and ARD) models of character

pair evolution. Average AICc weights are consistently (9 out of 12 datasets) higher within than between partitions (supplementary Figure 5). These differences were not significant within datasets (Mann-Whitney *U* tests, Table 4) but were marginally significant between datasets (paired t, t = 2.4179, df = 11, p = 0.03413). Kruskal-Wallis and Dunn tests showed significant differences in average AICc weights between all 3 character pair classes in 6 datasets (Table 3, supplementary Figure 4, supplementary Table 1).

We additionally compared linear mixed effects models treating pairwise cumulative AICc weights for correlated models as the response variable. A model splitting character pairs into all 3 classes was preferred over a null model and a model splitting character classes into only within- and between-character pairs (Table 3), suggesting an important, if inconsistent role of partition on correlation in these datasets.

Table 4 at end of Document

Table 4. *Showing differences between proportion and strength of correlation between character pair classes. Character pairs are considered correlated if their cumulative AICc weight for correlated (SYM and ARD) models of character pair evolution are equal to or higher than 0.95. Strength of correlation is measured by average cumulative AICc weights of correlated models of character pair evolution. Character pair classes include a) within-osteological data character pairs, b) within-dental data character pairs, and c) between-partition character pairs. Characters are generally more correlated (as measured by average AICc weights) within than between partitions. Differences in proportion and strength of correlation between osteological and dental characters are mixed.*

3.5. Evolutionary Relationships Inferred from Dental Versus Osteological Characters

Osteological and dental data each support different estimates of evolutionary relationships between mammals. Both the osteological and total morphological data in O'Leary *et al.,* (2013) dataset gave rise to two

most parsimonious trees each, while the dental data gave rise to a single most parsimonious tree.

The total morphological data produced trees more similar to the osteological data (average Robinson-Foulds distance = 35) than the dental data (RF distance = 53). The dental and osteological data produced trees more different to each other (RF distance = 60) than to trees produced by the combined data. The molecular topology was more similar to the topologies produced by the osteological data (RF distances = 40 and 42) and combined data (RF distance = 42) than the dental data (RF distance = 48). While Robinson-Foulds saturates quickly, since the inferred topologies are quite similar this should not affect these results.

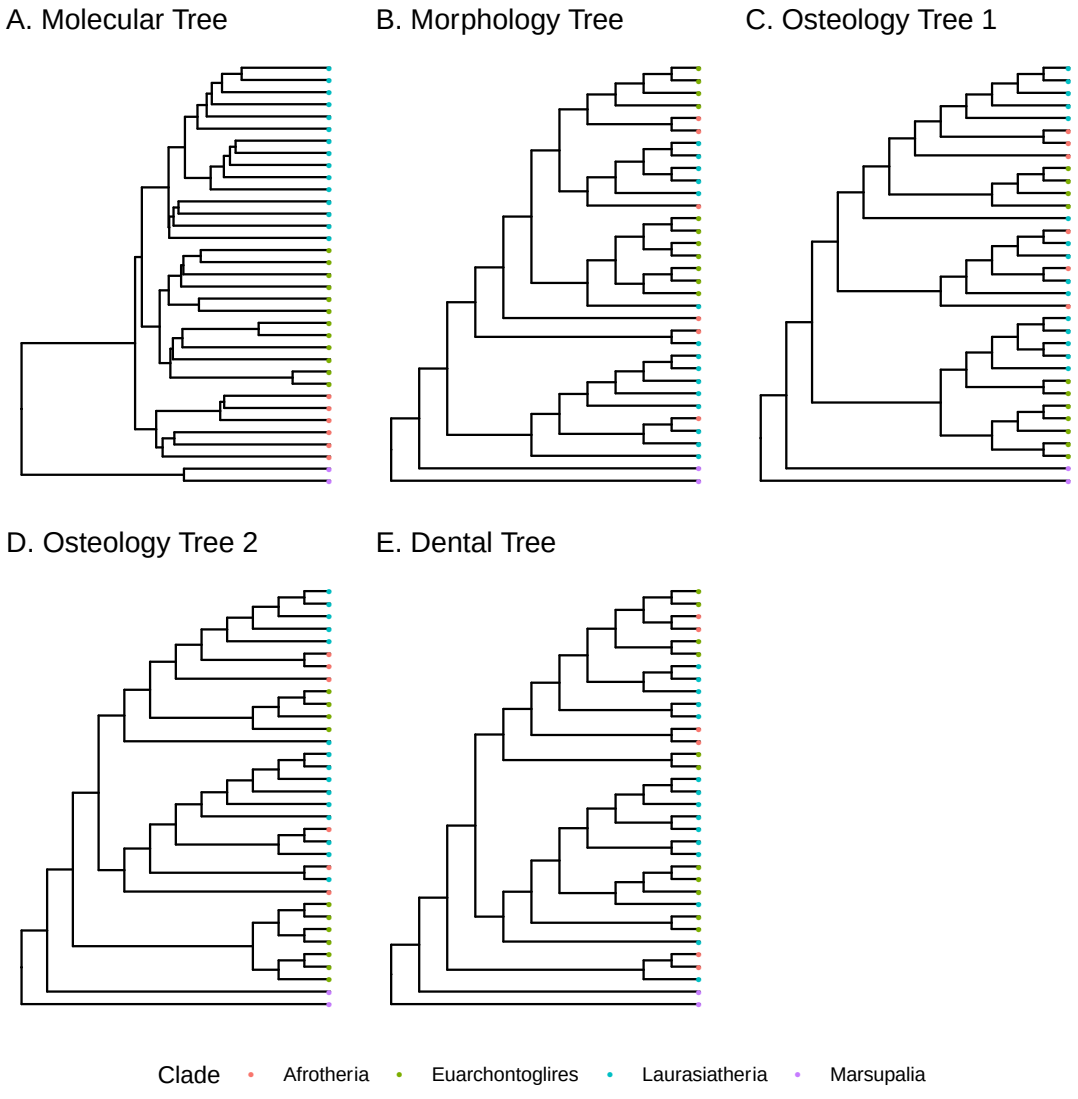


Figure 3. *Showing the trees containing O'Leary et al. (2013) taxa produced by a) Molecular data (Upham et al. 2019) b) Dental + Osteological data (consensus), c and d) Osteological data alone and e) Dental data alone. The topology produced by osteological characters alone is more similar to the molecular topology than either the dental or combined morphological data, demonstrating widespread convergence of dental characters in mammals. Both most parsimonious trees inferred from osteological data are presented because their consensus is highly unresolved owing to topological differences, whereas the two most parsimonious trees inferred from the combined morphological data are topologically similar.*

4. Discussion

The results presented here converge on a robust conclusion that dental and osteological characters display different trends and patterns of trait evolution with regards to homoplasy and correlation, resulting in different phylogenetic signal between these partitions. Osteological characters transition earlier and are less homoplastic than dental characters, and characters are more correlated within than between these partitions.

It can be argued that osteological are more consistent with independent data sources than dental data. This is because the molecular topology used here (Upham *et al.*, 2019) broadly corroborates widely accepted relationships between higher mammalian taxa, and the osteological data is more congruent with molecular data here as measured both by the retention index and Robinson-Foulds distances (Table 1, Figures 1 and 3). This result is in agreement with previous studies comparing dental and osteological data (Sansom *et al.*, 2017) and comparing osteological data with non-osteological data (Chapters 1, 2 3). It can thus be surmised that dental data display relatively elevated levels of homplasy, possibly the result of trophic convergence (Gingerich & Rose, 1979; Lazzari *et al.*, 2008; Goswami, Milne, & Wroe, 2011; Kelley & Motani, 2015).

In addition to levels of homoplasy, strength of correlation and integration differs between dental and osteological data. There is a consistent relationship between the internal consistency of morphological characters and *a priori* partitions, showing that characters are more similar within

these partitions than between them. This similarity in evolutionary dynamics between characters in the same partition, or at least in the way they are coded, is consistent across mammal datasets. Many of these datasets additionally display a Brownian motion patterning of trait distribution on these trees, constituting strong evidence of significant clumping, or clustering, of partitions among tree tips. This clustering of traits according to partition is indicative of stronger internal correlation within dental characters and within osteological characters compared with correlation between these partitions. Owing to the large number of datasets used here, covering much of Mammalia, and the strong agreement between datasets, we can have reasonable confidence in the generality of this result across mammals. This result is further supported by the stronger correlation of within than between partition character pairs on molecular trees as demonstrated by our comparison of models of trait evolution on molecular trees.

There are larger internally-consistent cliques mostly composed of osteological characters, which fit the molecular trees, and smaller cliques of dental characters which show a poorer fit to the tree in several datasets (e.g. Carstens *et al.*, 2002; He *et al.*, 2012, supplementary Figure 1). This is consistent with previous work showing that osteological characters are less homoplastic relative to molecular data than soft tissue characters (Chapter 1) and that osteological data show similar differences with dental data (Sansom *et al.*, 2017). Some datasets show the opposite pattern, with larger cliques of dental characters and smaller cliques of osteological characters, for example in thylacine marsupials (Wroe & Musser, 2001) and hominoids (Thorington *et al.*, 2002), consistent with widespread concerted convergence owing to ecological pressures (Patterson & Givnish, 2002).

Dental and Osteological data support different estimates of evolutionary history, both congruent an incongruent with previously proposed relationships. Osteological data correspond more closely than dental data to the total morphology data in the O'Leary *et al.*, (2013) dataset, which is unsurprising given that there are more osteological characters in this dataset. It is impossible to say with confidence which of the three

hypotheses of early branching is supported by these data, since there are no representatives of Xenarthra in this dataset. Further, neither the dental nor osteological data support accepted higher mammalian relationships, but instead produce trees inconsistent with molecular hypotheses of evolutionary history. For example, both the combined and osteological data place the lesser hedgehog tenrec *Echinops telfairi*, an Afrotherian, as sister to the European hedgehog *Erinacaes europaeus,* a Boreoeutherian. Interestingly, despite the general closer correspondence of osteology than dental data to molecular trees, only the dental data place the Afrotherian species *Echinops telfairi* and *Amblysomus hottentotus* together. Since this is inconsistent with molecular hypotheses and established relationships, this is reflective of convergence of osteological data in these species. Relationships supported by the osteological data include the placement of the Afrotherian paeungulate clade with the Laurasiatherian even-toed ungulates, a well-established case of morphological convergence (Gheerbrant *et al.*, 2016) and an intriguing placement of primates with bats, a spurious relationship previously proposed by Linneaus and other authors based on aspects of morphology (Pettigrew *et al.*, 1989). The results from this analysis are consistent with hidden support and convergence in morphological data, with different partitions supporting different relationships.

Taken together, all results indicate that dental and osteological data have evolved under different evolutionary regimes, and thus display different patterns of homoplasy and correlation. These results are consistent with the idea of concerted convergence (Patterson & Givnish, 2002) whereby some groups of characters displaying internal consistency are incongruent with underlying phylogenies owing to the action of evolutionary regimes shared by unrelated species. There is no evidence in the results presented here that groups of correlated homoplasious characters are larger or smaller than groups of correlated characters that support the underlying phylogeny. Nevertheless, such pairs or groups of characters can potentially mislead phylogenetic analysis when they are not identified *a priori*.

5. Conclusions

Overall, these results demonstrate the biological reality of dental and osteological partitions. They display different levels of homoplasy, external and internal correlation, transition ages, and produce different topologies. The treatment of these characters in phylogenetic reconstruction is thus an important consideration for future analyses involving morphological data, and the results presented here may necessitate careful consideration of partitioning schemes when using probabilistic methods. Further, there is reason to suspect that of these partitions, osteological data contains more phylogenetic signal. This is demonstrated by the better correspondence of osteological data than dental data with molecular trees and established hypotheses of evolutionary history.

The inconsistency between morphological trees and established phylogenetic hypotheses further highlights the danger of using only limited morphological data to infer evolution history. Further investigation into whether different subsets of morphological data support different hypotheses of mammalian evolutionary history may be warranted.

References

Agnarsson I, Zambrana-Torrelio CM, Flores-Saldana NP, May-Collado LJ. 2011. A time-calibrated species-level phylogeny of bats (Chiroptera, Mammalia). *PLoS Currents* 3: RRN1212.

Ahrens HE. 2012. Craniodental characters and the relationships of Procyonidae (Mammalia: Carnivora). *Zoological Journal of the Linnean Society* 164: 669–713.

Asher RJ, Maree S, Bronner G, Bennett NC, Bloomer P, Czechowski P, Meyer M, Hofreiter M. 2010. A phylogenetic estimate for golden moles (Mammalia, Afrotheria, Chrysochloridae). *BMC Evolutionary Biology* 10: 69.

Asher RJ, Meng J, Wible JR, McKenna MC, Rougier GW, Dashzeveg D, Novacek MJ. 2005. Stem Lagomorpha and the antiquity of Glires. *Science (New York,*

N.Y.) 307: 1091–4.
Bapst DW. 2012. paleotree : an R package for paleontological and phylogenetic analyses of evolution. *Methods in Ecology and Evolution* 3: 803–807.
Beaulieu JM, O'Meara BC, Donoghue MJ. 2013. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. *Systematic Biology* 62: 725–737.
Bi S, Wang Y, Guan J, Sheng X, Meng J. 2014. Three new Jurassic euharamiyidan species reinforce early divergence of mammals. *Nature* 514: 579–584.
Billet G. 2011. Phylogeny of the Notoungulata (Mammalia) based on cranial and dental characters. *Journal of Systematic Palaeontology* 9: 481–497.
Billet G, Bardin J. 2019. Serial homology and correlated characters in morphological phylogenetics: modeling the evolution of dental crests in Placentals. *Systematic Biology* 68: 267–280.
Blanke A, Greve C, Wipfler B, Beutel RG, Holland BR, Misof B, Forschungsmuseum Z, Koenig A. 2013. The identification of concerted convergence in insect heads corroborates Palaeoptera. *Systematic Biology* 13: 250–263.
Boessenecker RW, Churchill M. 2013. A Reevaluation of the Morphology, Paleoecology, and Phylogenetic Relationships of the Enigmatic Walrus Pelagiarctos (L VIRIOT, Ed.). *PLoS ONE* 8: e54311.
Boisserie JR. 2005. The phylogeny and taxonomy of Hippopotamidae (Mammalia: Artiodactyla): a review based on morphology and cladistic analysis. *Zoological Journal of the Linnean Society* 143: 1–26.
Bryant HN, Russell AP, Fitch WD. 1993. Phylogenetic relationships within the extant Mustelidae (Carnivora): appraisal of the cladistic status of the Simpsonian subfamilies. *Zoological Journal of the Linnean Society* 108: 301–334.
Carleton MD. 1980. Phylogenetic relationships in Neotomine-Peromyscine rodents (Muroidea) and a reappraisal of the dichotomy within New World Cricetinae. *Miscellaneous Publications Museum of Zoology* 157: 1–140.
Carstens BC, Lundrigan BL, Myers P. 2002. A Phylogeny of the Neotropical

Nectar-Feeding Bats (Chiroptera: Phyllostomidae) Based on Morphological and Molecular Data. *Journal of Mammalian Evolution* 9: 23–53.
Cerdeño E. 1995. Cladistic analysis of the family Rhinocerotidae (Perissodactyla). *American Museum Novitates* 3143: 1–25.
Churchill M, Boessenecker RW, Clementz MT. 2014. Colonization of the Southern Hemisphere by fur seals and sea lions (Carnivora: Otariidae) revealed by combined evidence phylogenetic and Bayesian biogeographical analysis. *Zoological Journal of the Linnean Society* 172: 200–225.
Delsuc F, Scally M, Madsen O, Stanhope MJ, de Jong WW, Catzeflis FM, Springer MS, Douzery EJP. 2002. Molecular Phylogeny of Living Xenarthrans and the Impact of Character and Taxon Sampling on the Placental Tree Rooting. *Molecular Biology and Evolution* 19: 1656–1671.
Domning DP. 1994. A phylogenetic analysis of the Sirenia. *Proceedings of the San Diego Society of Natural History* 29: 177–189.
Emerson SB, Hastings PA. 1998. Morphological correlations in evolution: consequences for phylogenetic analysis. *The Quarterly Review of Biology* 73: 141–162.
Felice RN, Tobias JA, Pigot AL, Goswami A. 2019. Dietary niche and the evolution of cranial morphology in birds. *Proceedings of the Royal Society B: Biological Sciences* 286.
Field DJ, Gauthier JA, King BL, Pisani D, Lyson TR, Peterson KJ. 2014. Toward consilience in reptile phylogeny: microRNAs support an archosaur, not a lepidosaur affinity for turtles. *Evolution & Development* 16: 189–196.
Finarelli JA. 2008. A Total Evidence Phylogeny of the Arctoidea (Carnivora: Mammalia): Relationships Among Basal Taxa. *Journal of Mammalian Evolution* 15: 231–259.
Foley NM, Springer MS, Teeling EC. 2016. Mammal madness: is the mammal tree of life not yet resolved? *Philosophical Transactions of the Royal Society B: Biological Sciences*.
Fracasso P, De Oliveira Salles L, Perini F. 2011. Upper molar morphology and relationships among higher taxa in bats. *Journal of Mammalogy* 92: 421–432.
Fritz SA, Purvis A. 2010. Selectivity in Mammalian extinction risk and threat

types: a new measure of phylogenetic signal strength in binary traits. *Conservation Biology* 24: 1042–1051.
Froelich DJ. 1999. Phylogenetic systematics of basal Perissodactyls. *Journal of Vertebrate Paleontology* 19: 140–159.
Gatesy J, Arctander P. 2000. Hidden Morphological Support for the Phylogenetic Placement of Pseudoryx nghetinhensis with Bovine Bovids: A Combined Analysis of Gross Anatomical Evidence and DNA Sequences from Five Genes (M Westneat, Ed.). *Systematic Biology* 49: 515–538.
Gaubert P, Wozencraft W, Cordeiro-Estrela P, Veron G. 2005. Mosaics of convergences and noise in morphological phylogenies: what's in a viverrid-like carnivoran? *Systematic Biology* 54: 865–894.
Gentry A. 1992. The subfamilies and tribes of the family Bovidae. *Mammal Review* 22: 1–32.
Gheerbrant E, Filippo A, Schmitt A. 2016. Convergence of Afrotherian and Laurasiatherian ungulate-like mammals: first morphological evidence from the Paleocene of Morocco. *PLOS ONE* 11: e0157556.
Gheerbrant E, Sudre J, Tassy P, Amaghzaz M, Bouya B, Iarochene M. 2005. Nouvelles données sur Phosphatherium escuilliei (Mammalia, Proboscidea) de l'Éocène inférieur du Maroc, apports à la phylogénie des Proboscidea et des ongulés lophodontes. *Geodiversitas* 27: 239–333.
Giannini NP, Simmons NB. 2005. Conflict and congruence in a combined DNA-morphology analysis of megachiropteran bat relationships (Mammalia: Chiroptera: Pteropodidae). *Cladistics* 21: 411–437.
Gingerich PD, Rose KD. 1979. Anterior Dentition of the Eocene Condylarth Thryptacodon: Convergence with the Tooth Comb of Lemurs. *Journal of Mammalogy* 60: 16–22.
Godoy PL. 2019. Crocodylomorph cranial shape evolution and its relationship with body size and ecology. *Journal of Evolutionary Biology*: jeb.13540.
Goloboff PA, Farris JS, Nixon KC. 2008. TNT, a free program for phylogenetic analysis. *Cladistics* 24: 774–786.
Goswami A, Milne N, Wroe S. 2011. Biting through constraints: cranial morphology, disparity and convergence across living and fossil carnivorous

mammals. *Proceedings of the Royal Society B: Biological Sciences* 278: 1831–1839.
Goswami A, Polly PD. 2010. The influence of character correlations on phylogenetic analyses: a case study of the carnivoran cranium. In: Goswami A,, In: Friscia A, eds. *Carnivoran Evolution*. Cambridge: Cambridge University Press, 141–164.
He K, Chen JH, Gould GC, Yamaguchi N, Ai HS, Wang YX, Zhang YP, Jiang XL. 2012. An Estimation of Erinaceidae Phylogeny: A Combined Analysis Approach (SO Kolokotronis, Ed.). *PLoS ONE* 7: e39304.
Holland BR, Spencer HG, Worthy TH, Kennedy M. 2010. Identifying cliques of convergent characters: Concerted evolution in the cormorants and shags. *Systematic Biology* 59: 433–445.
Kelley NP, Motani R. 2015. Trophic convergence drives morphological convergence in marine tetrapods. *Biology Letters* 11: 20140709.
Kielan-Jaworowska Z, Hurum JH. 2001. Phylogeny and Systematics of Multituberculate Mammals. *Palaeontology* 44: 389–429.
Ladeveze S, Pieter M, Smith T. 2010. First skull of Orthaspidotherium edwardsi (Mammalia, "Condylarthra") from the late Paleocene of Berru (France) and phylogenetic affinities of the enigmatic European family Pleuraspidotheriidae. *Journal of Vertebrate Paleontology* 30: 1559–1578.
Lazzari V, Charles C, Tafforeau P, Vianey-Liaud M, Aguilar JP, Jaeger JJ, Michaux J, Viriot L. 2008. Mosaic Convergence of Rodent Dentitions (JE Stajich, Ed.). *PLoS ONE* 3: e3607.
LeCompte E, Granjon L, Denys C. 2002. The phylogeny of the Praomys complex (Rodentia: Muridae) and its phylogeographic implications. *Journal of Zoological Systematics and Evolutionary Research* 40: 8–25.
Lee MSY. 2009. Hidden support from unpromising data sets strongly unites snakes with anguimorph 'lizards'. *Journal of Evolutionary Biology* 22: 1308–1316.
Lee MSY, Camens AB. 2009. Strong morphological support for the molecular evolutionary tree of placental mammals. *Journal of Evolutionary Biology* 22: 2243–2257.
Madsen O, Scally M, Douady CJ, Kao DJ, DeBry RW, Adkins R, Amrine HM,

Stanhope MJ, de Jong WW, Springer MS. 2001. Parallel adaptive radiations in two major clades of placental mammals. *Nature* 409: 610–614.
Mazel F, Wüest RO, Gueguen M, Renaud J, Ficetola GF, Lavergne S, Thuiller W. 2017. The Geography of Ecological Niche Evolution in Mammals. *Current Biology* 27: 1369–1374.
McCurry MR, Evans AR, Fitzgerald EMG, Adams JW, Clausen PD, McHenry CR. 2017. The remarkable convergence of skull shape in crocodilians and toothed whales. *Proceedings of the Royal Society B: Biological Sciences* 284: 20162348.
Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ. 2001a. Molecular phylogenetics and the origins of placental mammals. *Nature* 409: 614–618.
Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douady CJ, Teeling E, Ryder OA, Stanhope MJ, de Jong WW, Springer MS. 2001b. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science (New York, N.Y.)* 294: 2348–2351.
O'Leary MA, Bloch JI, Flynn JJ, Gaudin TJ, Giallombardo A, Giannini NP, Goldberg SL, Kraatz BP, Luo ZX, Meng J, Ni X, Novacek MJ, Perini FA, Randall ZS, Rougier GW, Sargis EJ, Silcox MT, Simmons NB, Spaulding M, Velazco PM, Weksler M, Wible JR, Cirranello AL. 2013. The placental mammal ancestor and the post-K-Pg radiation of placentals. *Science (New York, N.Y.)* 339: 662–7.
Olivares AI, Verzi DH. 2015. Systematics, phylogeny and evolutionary pattern of the hystricognath rodent *Eumysops* (Echimyidae) from the Plio–Pleistocene of southern South America. *Historical Biology* 27: 1042–1061.
OLiviera ÉV, Nova PV, Goin FJ, Avilla LDS. 2011. A new hyladelphine marsupial (Didelphimorphia, Didelphidae) from cave deposits of northern Brazil. *Zootaxa* 3041: 51.
Orme D, Freckleton R, Thomas G, Petzoldt T, Fritz S, Isaac N, Pearse W. 2018. caper: Comparative Analyses of Phylogenetics and Evolution in R. R package version 1.0.1.
Pagel M. 1994. Detecting Correlated Evolution on Phylogenies: A General Method for the Comparative Analysis of Discrete Characters. *Proceedings of*

the Royal Society B: Biological Sciences 255: 37–45.
Pagel M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401: 877–884.
Patterson TB, Givnish TJ. 2002. Phylogeny, concerted convergence, and phylogenetic niche conservatism in the core Liliales: insights from rbcL and ndhF sequence data. *Evolution* 56: 233–252.
Pettigrew JD, Jamieson BGM, Robson SK, Hall LS, McAnally KI, Cooper HM. 1989. Phylogenetic Relations Between Microbats, Megabats and Primates (Mammalia: Chiroptera and Primates). *Philosophical Transactions of the Royal Society B: Biological Sciences* 325: 489–559.
Pinheiro J, Bates D, DebRoy S, Sarkar D, Team RC. 2019. _nlme: Linear and Nonlinear Mixed Effects Models_. R package version 3.1-141.
Price JJ, Friedman NR, Omland KE. 2007. Song and plumage evolution in the New World orioles (Icterus) show similar lability and convergence in patterns. *Evolution* 61: 850– 863.
Prideaux GJ, Warburton NM. 2010. An osteology-based appraisal of the phylogeny and evolution of kangaroos and wallabies (Macropodidae: Marsupialia). *Zoological Journal of the Linnean Society* 159: 954–987.
R Core Team. 2018. R: A language and environment for statistical computing.
Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3: 217–223.
Robinson DF, Foulds LR. 1981. Comparison of phylogenetic trees. *Mathematical Biosciences* 53: 131–147.
Sadleir RW, Makovicky PJ. 2008. Cranial shape and correlated characters in crocodilian evolution. *Journal of Evolutionary Biology* 21: 1578–1596.
Sakamoto M, Ruta M. 2012. Convergence and Divergence in the Evolution of Cat Skulls: Temporal and Spatial Patterns of Morphological Diversity (A Goswami, Ed.). *PLoS ONE* 7: e39752.
Sakamoto M, Ruta M, Venditti C. 2019. Extreme and rapid bursts of functional adaptations shape bite force in amniotes. *Proceedings of the Royal Society B: Biological Sciences* 286: 20181932.
Sanchez-Villagra MR, Horovitz I, Motokawa M. 2006. A comprehensive

morphological analysis of talpid moles (Mammalia) phylogenetic relationships. *Cladistics* 22: 59–88.
Sansom RS, Wills MA. 2017. Differences between hard and soft phylogenetic data. *Proceedings of the Royal Society B: Biological Sciences* 284: 20172150.
Sansom RS, Wills MA, Williams T. 2017. Dental data perform relatively poorly in reconstructing mammal phylogenies: Morphological partitions evaluated with molecular benchmarks. *Systematic Biology* 66: 813–822.
Sauquet H, von Balthazar M, Magallón S, Doyle JA, Endress PK, Bailes EJ, Barroso de Morais E, Bull-Hereñu K, Carrive L, Chartier M, Chomicki G, Coiro M, Cornette R, El Ottra JHL, Epicoco C, Foster CSP, Jabbour F, Haevermans A, Haevermans T, Hernández R, Little SA, Löfstrand S, Luna JA, Massoni J, Nadot S, Pamperl S, Prieu C, Reyes E, dos Santos P, Schoonderwoerd KM, Sontag S, Soulebeau A, Staedler Y, Tschan GF, Wing-Sze Leung A, Schönenberger J. 2017. The ancestral flower of angiosperms and its early diversification. *Nature Communications* 8: 16047.
Scally M, Madsen O, Douady CJ, de Jong WW, Stanhope MJ, Springer MS. 2001. Molecular Evidence for the Major Clades of Placental Mammals. *Journal of Mammalian Evolution* 8: 239–277.
Schliep KP. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics* 27: 592–593.
Silcox MT, Bloch JI, Boyer DM, Houde P. 2010. Cranial anatomy of Paleocene and Eocene Labidolemur kayi (Mammalia: Apatotheria), and the relationships of the Apatemyidae to other mammals. *Zoological Journal of the Linnean Society* 160: 773–825.
Springer MS, Kirsch JAW, Chase JA. 1997. The chronicle of marsupial evolution. In: Givnish TJ,, In: Sytsma KJ, eds. *Molecular Evolution and Adaptive Radiation*. Cambridge: Cambridge University Press, 129–161.
Stanhope MJ, Waddell VG, Madsen O, de Jong W, Hedges SB, Cleven GC, Kao D, Springer MS. 1998. Molecular evidence for multiple origins of Insectivora and for a new order of endemic African insectivore mammals. *Proceedings of the National Academy of Sciences of the United States of America* 95: 9967–72.
Steppan S. 1993. Phylogenetic relationships among the phyllotini (rodentia:

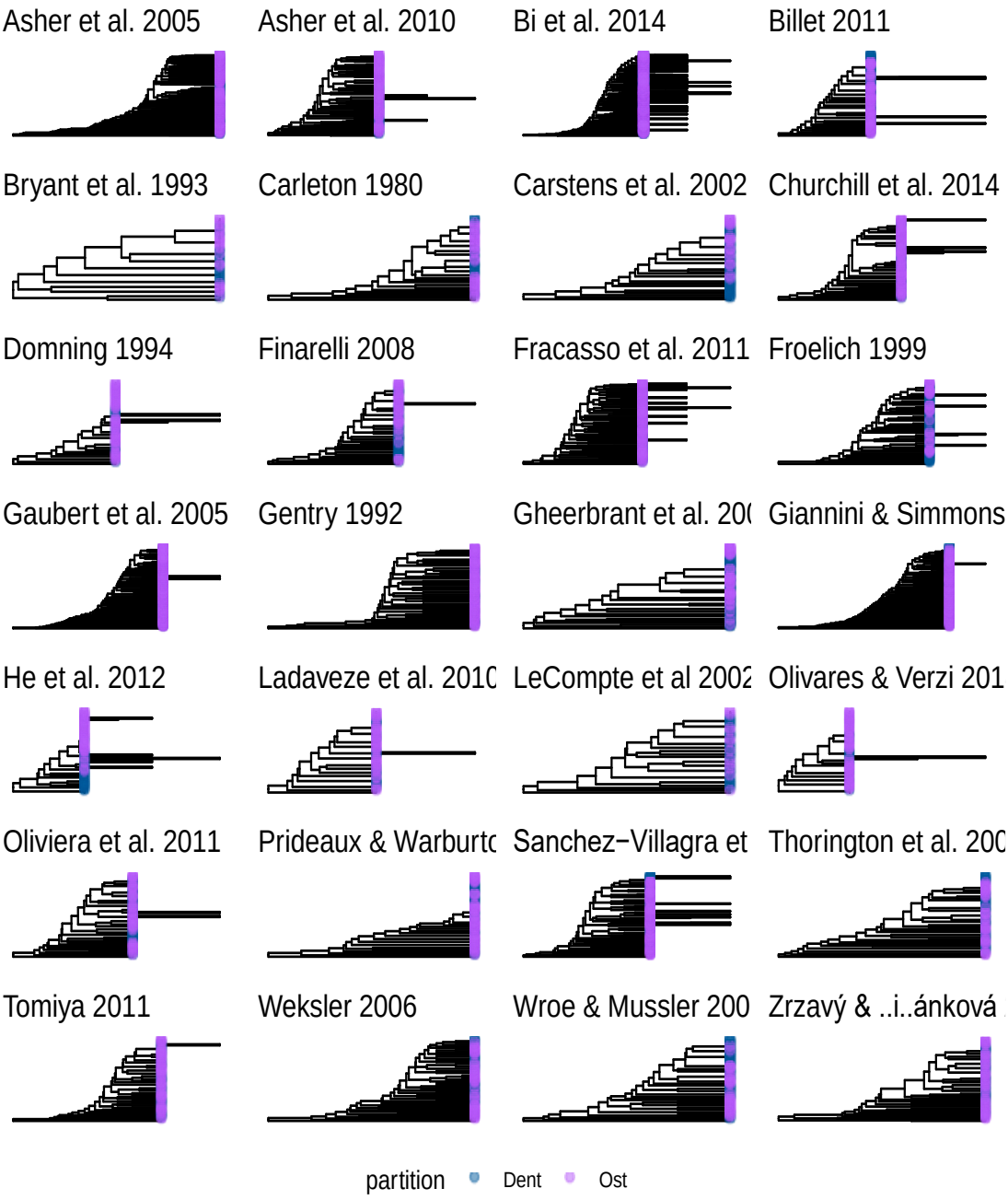
Sigmodontinae) using morphological characters. *Journal of Mammalian Evolution* 1: 187–213.
Strait DS, Grine FE. 2004. Inferring hominoid and early hominid phylogeny using craniodental characters: the role of fossil taxa. *Journal of Human Evolution* 47: 399–452.
Teeling EC, Hedges SB. 2013. Making the Impossible Possible: Rooting the Tree of Placental Mammals. *Molecular Biology and Evolution* 30: 1999–2000.
Thorington RW, Pitassy D, Jansa SA. 2002. Phylogenies of Flying Squirrels (Pteromyinae). *Journal of Mammalian Evolution* 9: 99–135.
Tomiya S. 2011. A New Basal Caniform (Mammalia: Carnivora) from the Middle Eocene of North America and Remarks on the Phylogeny of Early Carnivorans (A Goswami, Ed.). *PLoS ONE* 6: e24146.
Van Tuinen M, Butvill DB, Kirsch JA, Hedges SB. 2001. Convergence and divergence in the evolution of aquatic birds. *Proceedings. Biological sciences* 268: 1345–50.
Upham N, Esselstyn J, Jetz W. 2019. Inferring the mammal tree: species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLoS Biology*.
Wahlberg N, Braby MF, Brower AV., de Jong R, Lee MM, Nylin S, Pierce NE, Sperling FA., Vila R, Warren AD, Zakharov E. 2005. Synergistic effects of combining morphological and molecular data in resolving the phylogeny of butterflies and skippers. *Proceedings of the Royal Society B: Biological Sciences* 272: 1577–1586.
Weksler M. 2006. Phylogenetic Relationships of Oryzomine Rodents (Muroidea: Sigmodontinae): Separate and Combined Analyses of Morphological and Molecular Data. *Bulletin of the American Museum of Natural History* 2006: 1–149.
Wroe S, Musser A. 2001. The skull of Nimbacinus dicksoni (Thylacinidae : Marsupialia). *Australian Journal of Zoology* 49: 487.
Zrzavý J, Řičánková V. 2004. Phylogeny of Recent Canidae (Mammalia, Carnivora): relative reliability and utility of morphological and molecular datasets. *Zoologica Scripta* 33: 311–333.

Supplementary Information

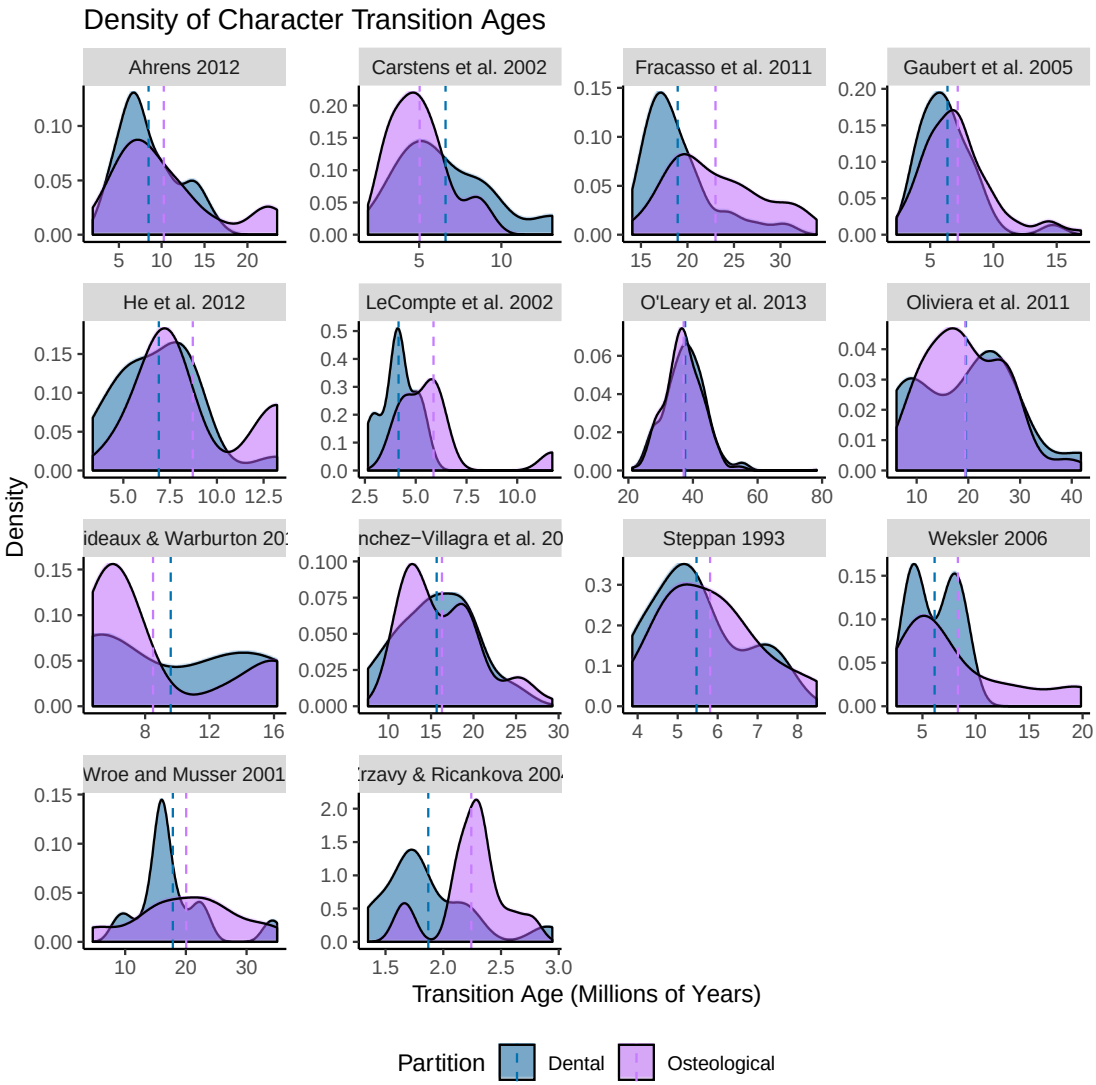
Supplementary Table 1*. Maximum clique sizes in mammalian morphological datasets*

Dataset	Characters	Max Clique Size Real Data	Mean Max Clique Size Shuffled Data	P
Ahrens, (2012)	66	9	11.91	0.95
Carstens, Lundrigan, & Myers, (2002)	62	16	17.19	0.74
Fracasso *et al.*, (2011)	130	4	3.86	0.68
Gaubert *et al.*, (2005)	261	15	20.2	0.95
He *et al.*, (2012)	73	31	42.82	0.97
LeCompte, Granjon, & Denys, (2002)	37	8	8.97	0.76
O'Leary *et al.*, (2013)	2199	7	6.89	0.59
Oliviera *et al.*, (2011)	80	8	12.47	1
Prideaux & Warburton, (2010)	81	47	42.34	0.18
Sanchez-Villagra, Horovitz, & Motokawa, (2006)	152	18	27.59	0.99
Steppan, (1993)	82	8	7.46	0.53
Weksler, (2006)	68	4	3.83	0.59
Wroe & Musser, (2001)	70	21	22.97	0.74
Zrzavý & Řičánková, (2004)	77	11	7.9	0.1

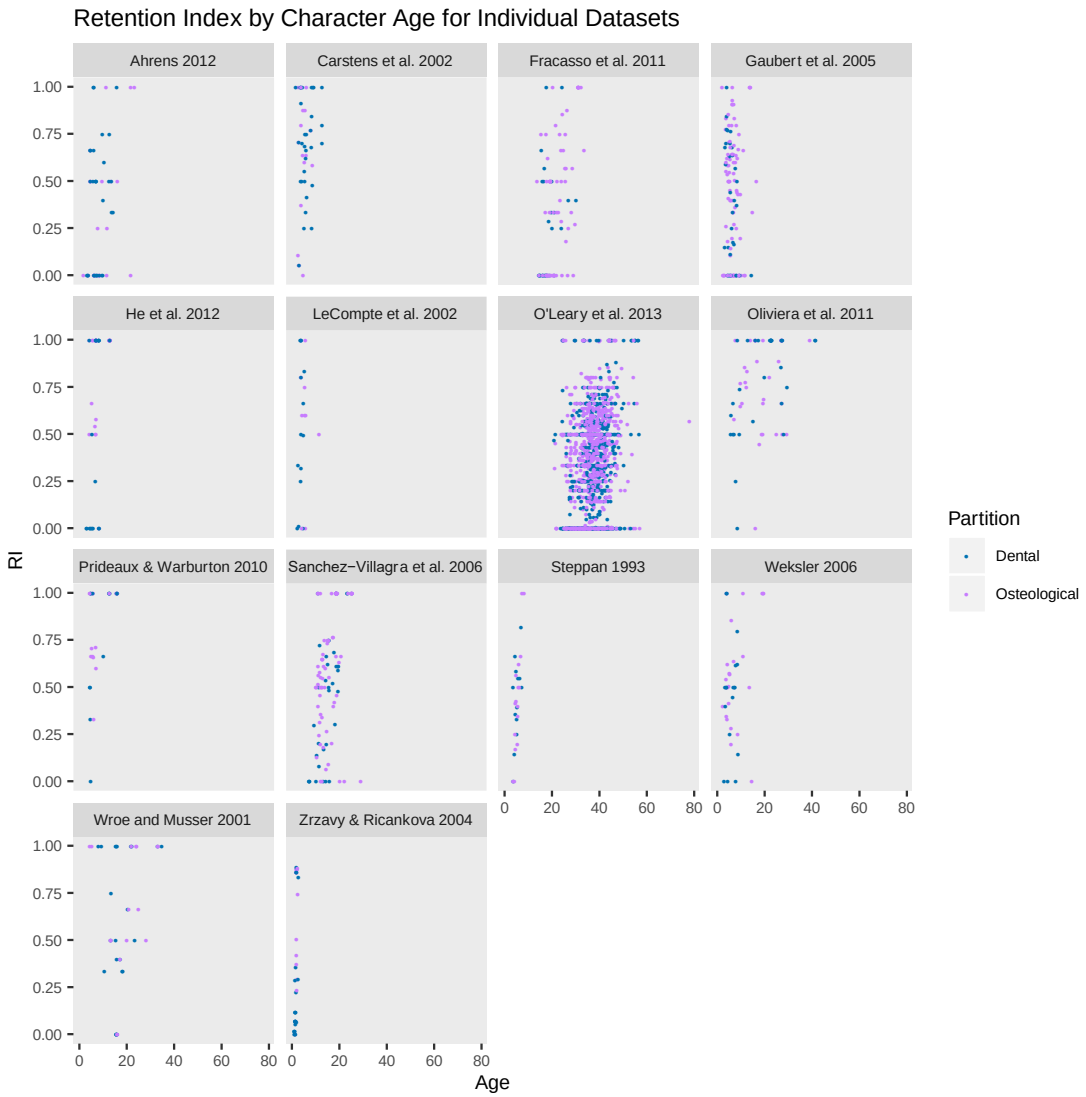
Showing the total number of characters, the maximum clique size (size of the largest internally consistent clique of characters) in the real data, and the mean maximum clique sizes in 100 shuffled datasets. Values are compared using by calculating proportions of datasets in the null distribution which have maximum clique sizes larger than or the same as that in the real data. Clique sizes are no larger than expected given their parsimony score in most datasets.



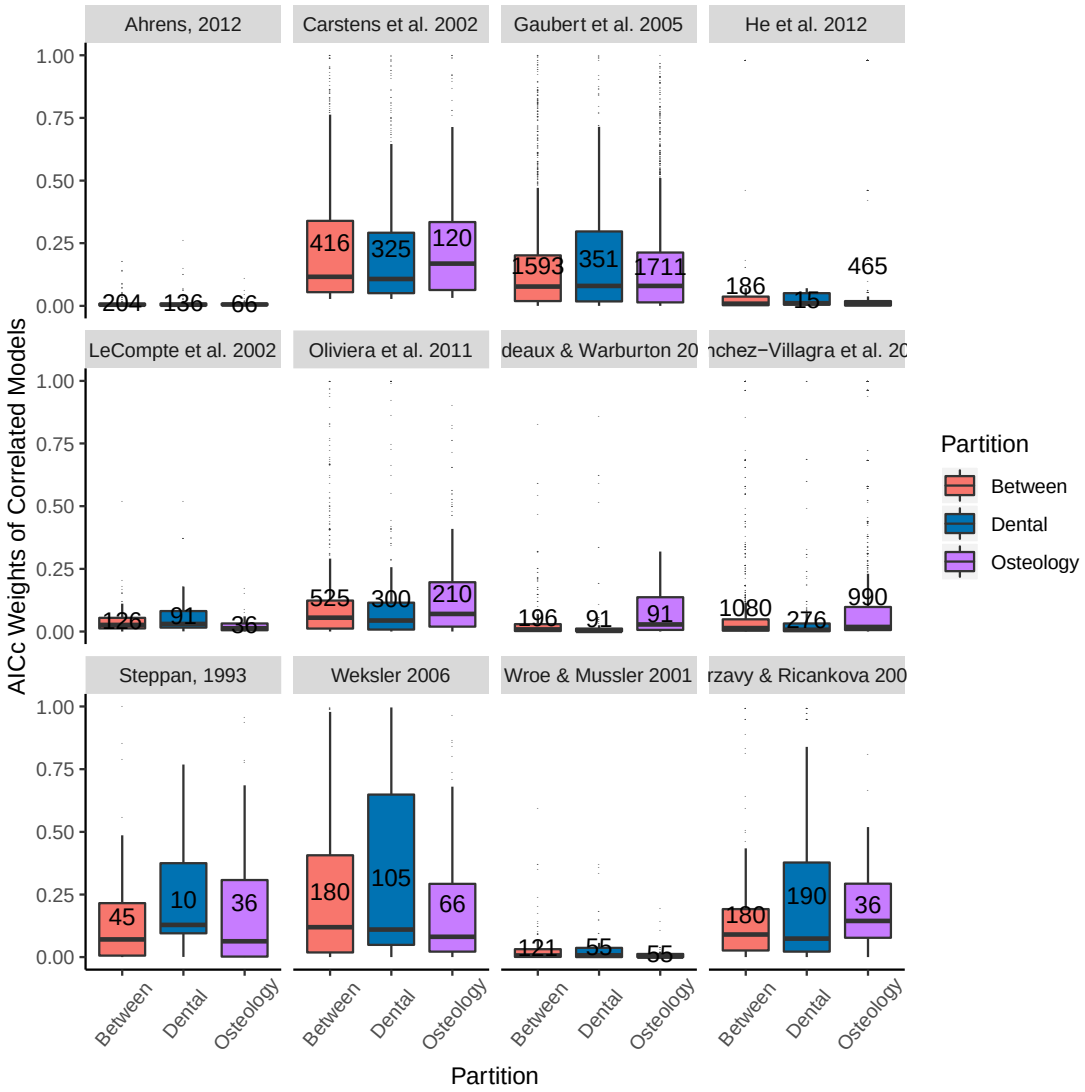
Supplementary Figure 1. *Showing the distribution of dental and osteological characters among tips of UPGMA trees where characters are clustered according to internal consistency. Dental characters are in blue while osteological characters are in purple. Almost all datasets display trait distribution significantly different from random. Some datasets (e.g. He et al., 2012, Carstens et al., 2002) show a distinct increase in 'blueness' further down the tree, i.e. in 'basal' characters. This is consistent with larger cliques of osteological characters (which are more congruent with molecular data), and smaller cliques of dental characters (which are less congruent with molecular data). Other datasets (e.g. Wroe & Mussler 2001, Sanchez-Villagra et al., 2006) show the opposite pattern, with large cliques of dental characters and smaller cliques of osteological characters, consistent with concerted convergence of dental characters in these datasets.*



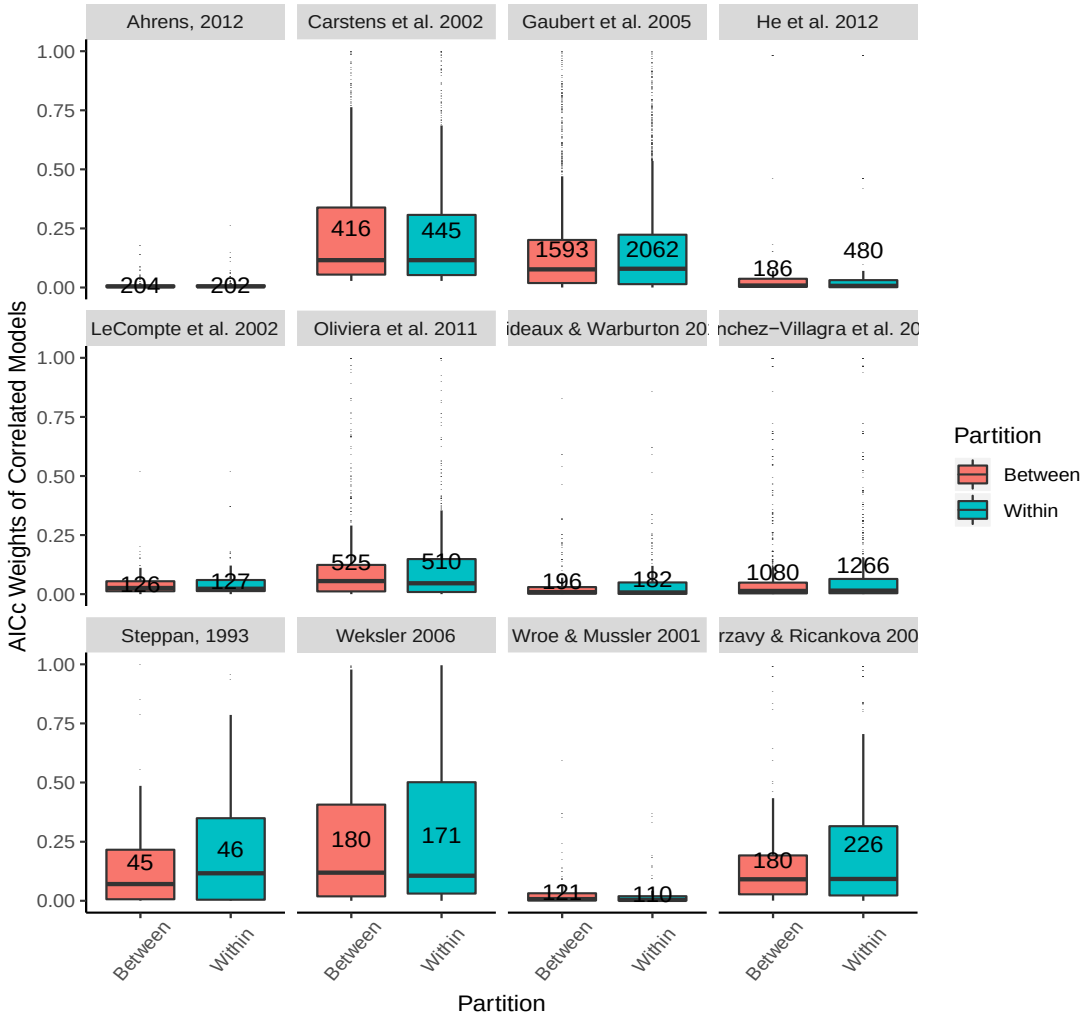
Supplementary Figure 2. *Showing the difference in density of transition ages between osteological (purple) and dental (blue) characters in millions of years for individual mammalian datasets. Dashed lines indicate average transition ages for each partition by dataset. Osteological characters transition earlier on average in most datasets.*



Supplementary Figure 3. *Showing the retention index of individual characters on molecular trees by transition age in millions of years for each mammalian dataset. In most datasets, retention index tends to increase with age. Characters are coloured by partition, with osteological characters in purple and dental characters in blue.*



Supplementary Figure 4. *Individual dataset boxplots of average cumulative AICc weights for correlated models of character pair evolution, showing all three classes of character pairs (within osteological, within dental and between partition) in mammalian morphological datasets. Boxplots display number of characters. Lower means indicate lower correlation, higher means indicate higher correlation.*



Supplementary Figure 5. *Individual dataset boxplots of average cumulative AICc weights for correlated models of character pair evolution, showing within and between partition character pairs. Characters are slightly more correlated within partitions in most datasets. Boxplots display number of characters. Lower means indicate lower correlation, higher means indicate higher correlation.*

Supplementary Table 1. *Showing results of the Dunn tests*

Dataset	Ost-Dent	Ost-Between	Dent-Between
Carstens et al. 2002	0.0023*	0.0341	0.0626
LeCompte et al. 2002	0.0002*	0.0052*	0.0563
Oliviera et al. 2011	0.0000*	0.0033*	0.029
Prideaux & Warburton 2010	0.0000*	0.0004*	0.0004*
Sanchez-Villagra et al. 2006	0.0000*	0.0000*	0.0000*
Wroe & Musser 2001	0.0273	0.0069*	0.4169

Showing results of the Dunn tests of differences in the cumulative AICc weights of correlated models of character pair evolution between three character pair classes (within-osteological, within-dental, and between partition) in all datasets with significant differences between these classes (Kruskal-Wallis tests, Table 4), finding significant differences in correlation between osteological and dental partitions in only 5 datasets.

Table 3. *Results of all linear mixed effects models comparing homoplasy, transition ages and correlations between partitions*

Analysis	Response Variable	Fixed Effects	df	AIC	AIC weight	BIC	Log L	Likelihood ratio	P
1. Homoplasy	Retention index	None	17	783.255	0.01245992	880.6456	-374.6275		
	Retention index	Partition	18	774.5096	0.98754008	877.629	-369.2548	10.74547	0.001
2. Transition Ages	Transition Age	None	17	13894.99	0.000720388	13992.38	-6930.494		
	Transition Age	Partition	18	13880.52	0.999279612	13983.64	-6922.26	16.46917	<.0001
3. Transition Ages with Homoplasy	Transition Age	None	17	13894.99	4.46E-36	13992.38	-6930.494		
	Transition Age	RI	18	13733.13	0.6268304	13836.25	-6848.564	163.85994	<.0001
	Transition Age	RI + Partition	19	13734.77	0.2760759	13843.62	-6848.384	0.36111	0.5479
	Transition Age	RI * Partition	20	13736.86	0.09709365	13851.44	-6848.432	0.09546	0.7573
4. Correlations	AICc Weights	None	15	-3606.686	0.02600639	-3497.545	1818.343		
	AICc Weights	Within/Between	16	-3610.341	0.16171606	-3493.924	1821.17	5.655086	0.0174
	AICc Weights	Partition/Between	17	-3613.569	0.81227755	-3489.876	1823.784	5.227939	0.0222

Showing the results of all linear mixed effects models fitted to the data, testing whether osteological and dental partitions explain differences in homopasy, transition ages and correlations while taking into account dataset effects. There are 3 dependent variables in 4 total analyses including the

effect of partition on homoplasy as measured by character retention indices on molecular trees, the effect of partition on transition ages as measured by estimating ages given a character state reconstruction on molecular trees, the effect of retention index versus partition on transition ages, and finally the effect of partition on the strength of correlation between character pairs. For the final analysis, within- and between-partition is tested as a fixed effect as well as partition itself. Dataset is treated as a random effect in all models, and partition treated as a fixed effect in all models except the null. Each model fitted for each analysis is given an AIC weight, where each weight is a value between 0-1, and the weights of each model add to 1 for each analysis. The highest AIC weight is given to the model which best describes the data. The best model in each analysis as measured by AIC weights is highlighted in green.

	Correlated Character Pairs (n of N)			Mean AICc Weight			Correlated Character Pairs		
Dataset	**Within Partition**	**Between Partitions**	**P Fisher's**	**Within Partitions**	**Between Partitions**	**P Mann-Whitney *U***	**Osteological Character Pairs**	**Dental Character Pairs**	**Between Partition Character Pairs**
Ahrens et al. 2012	0/202	0/204	1	0.012484 02	0.0129070 9	0.4457	0/66	0/136	0/204
Carstens et al. 2002	12/445	10/416	0.8318	0.242619 4	0.2542231	0.6397	5/120	7/325	10/416
Gaubert et al. 2005	11/2062	10/1593	0.8261	0.165384 4	0.1622268	0.6729	6/1711	5/351	10/1593
He et al. 2012	66/480	12/186	0.007264*	0.158388 3	0.0842370 6	0.246	66/456	0/15	12/186
LeCompte et al. 2002	0/127	0/126	1	0.054970 27	0.0451172 6	0.8798	0/36	0/91	0/126
Oliviera et al. 2011	16/510	11/525	0.3329	0.142303 1	0.1348383	0.8644	0/210	16/300	11/525
Prideaux & Warburton 2010	0/182	0/196	1	0.064170 61	0.0576123 8	0.9981	0/91	0/91	0/196
Sanchez-Villagra et al. 2006	69/1266	0/1080	0.000000000 00000022*	0.122255 7	0.0956309 4	0.08631	65/990	3/276	0/1080
Steppan 1993	1/46	0/45	1	0.221632 4	0.1624838	0.5829	1/36	0/10	0/45
Weksler 2006	11/171	9/180	0.6479	0.277985	0.2627441	0.7232	1/66	10/105	9/180
Wroe & Musser 2001	0/110	0/121	1	0.029386 78	0.0376734 6	0.09914	0/55	0/55	0/121
Zrzavý & Řičánková 2004	22/226	8/180	0.05523	0.241616 8	0.1722683	0.298	0/36	22/190	8/180

Table 4. *Showing differences between proportion and strength of correlation between character pair classes. Character pairs are considered correlated if their cumulative AICc weight for correlated (SYM and ARD) models of character pair evolution are equal to or higher than 0.95. Strength of correlation is measured by average cumulative AICc weights of correlated models of character pair evolution. Character pair classes include a) within-osteological data character pairs, b) within-dental data character pairs, and c) between-partition character pairs. Characters are generally more correlated (as measured by average AICc weights) within than between partitions. Differences in proportion and strength of correlation between osteological and dental characters are mixed.*

				Mean AICc Weight				
Osteological % Correlated Pairs	**Dental % Correlated Pairs**	**Between % Correlated Pairs**	**P Fisher's**	**Osteological**	**Dental**	**Between**	**P kruskal-Wallis**	**Tree Symmetry Score**
0	0	0	1	0.00974428	0.0138136	0.0129070	0.4736	40
4.166667	2.153846	2.403846	0.4404	0.2749716	0.2306739	0.2542231	0.01627*	175
0.3506721	1.424501	0.6277464	0.0537	0.1599507	0.1918721	0.1622268	0.4756	146
14.19355	0	6.451613	0.005309*	0.1626738	0.0255378	0.0842370	0.3822	12
0	0	0	1	0.0280967	0.0656015	0.0451172	0.001701*	43
0	5.333333	2.095238	0.0003564*	0.1537354	0.1343004	0.1348383	0.0003456*	247
0	0	0	1	0.0740624	0.0542949	0.0576123	0.00000007769*	67
6.565657	1.086957	0	0.00000000000022*	0.142353	0.0501676	0.0956309	0.00000000005842*	20
2.777778	0	0	0.5055	0.2186878	0.2322328	0.1624838	0.4854	126
1.515152	9.52381	5	0.08257	0.2143388	0.3179912	0.2627441	0.1826	191
0	0	0	1	0.01448972	0.0442838	0.0376734	0.04044*	39
0	11.57895	4.444444	0.00688*	0.2122556	0.24718	0.1722683	0.1002	85

Table 4 (cont). *Showing differences between proportion and strength of correlation between character pair classes. Character pairs are considered correlated if their cumulative AICc weight for correlated (SYM and ARD) models of character pair evolution are equal to or higher than 0.95. Strength of correlation is measured by average cumulative AICc weights of correlated models of character pair evolution. Character pair classes include a) within-osteological data character pairs, b) within-dental data character pairs, and c) between-partition character pairs. Characters are generally more correlated (as measured by average AICc weights) within than between partitions. Differences in proportion and strength of correlation between osteological and dental characters are mixed.*

General Discussion and Summary

1. Introduction

1.1. Summary of Results and Conclusions of Individual Chapters

The principal methods and findings of each of my results chapters are as follows.

Chapter 1: Osteological Characters Show Greater Congruence with Molecular Phylogenies than Soft Characters in Avian and Reptilian Morphological Datasets

Characters in avian and squamate morphological character matrices were classified into partitions and subpartitions. Homoplasy was calculated as the retention index of characters and ensemble index of partitions on independent molecular trees. Transition ages were calculated by performing ancestral state reconstructions under a parsimony framework, and averaging simulated ages between nodes at which characters transition and ancestral nodes.

Linear mixed effect models finds that osteological characters are more consistent than non-osteological (e.g. myological, integumentary) characters with molecular trees in avian and squamate datasets. Further, osteological characters are older, although their retention index on molecular trees explains their apparent age. There were no significant differences between smaller subpartitions (cranial, postcranial, integument, myology), partly owing to paucity of data. These findings are consistent with the view that soft characters are relatively labile. Osteological characters may reconstruct evolutionary history more faithfully than non-osteological characters. This finding is instructive for morphological phylogenetics, and particularly reassuring to palaeontologists working with extinct taxa for which only osteological characters are often available.

Chapter 2: The Prevalence of Correlated Morphological Characters and their Effect on Phylogenetic Reconstruction

Pairwise correlations between characters were tested for by fitting correlated and uncorrelated models of character evolution on pairs of binary characters in avian and squamate morphological datasets mapped onto independent molecular trees. Models were compared using AICc weights. Character pairs were split into two (within and between partition) and three (within osteological partition, within non-osteological partition, and between partition) classes. These models were also fitted to artificially constructed correlated data, and to simulated random data, to test for the effect of underlying data structure.

Characters are more correlated within than between osteological and non-osteological partitions. Patterns of correlation between partitions differ between datasets, although there is generally more correlation within non-osteological data. After rejecting a null model of no difference in correlation between character pair classes, linear mixed effect models find that splitting pairwise correlations into three character pair classes provides greater explanatory power than splitting the data into two classes. This implies the importance of partition. Overall, correlated characters are found in all datasets, but strength and pattern of correlations are clade and/or dataset dependent. Further, correlations are generally faithfully recovered in datasets with more than 20 taxa. No correlations are recovered in random data. This gives us confidence in the performance of these methods given sufficient data. Since correlated characters can affect phylogenetic reconstruction, this should be taken into account when using morphological data to infer evolutionary history.

Chapter 3: Cliques of Morphological Characters in Avian and Squamate Datasets

Cliques, or clusters, of internally consistent characters were found by constructing trees from character dissimilarity matrices, using the UPGMA clustering algorithm. The dissimilarity matrices describe the pairwise phylogenetic dissimilarity between characters. The largest cliques in each dataset were compared against shuffled data. Methods used to detect phylogenetic signal, i.e. whether traits cluster on tree tips more than expected by chance, were then applied to a) the retention indices of characters on independent molecular trees, and b) osteological versus non-osteological partitions, on UPGMA tips.

Cliques of characters are no larger than expected by chance in most datasets. Further, UPGMA trees are also asymmetrical, demonstrating the existence of few large cliques and many small cliques. Nevertheless, characters do cluster according to both their retention index and partition more than expected by chance, with larger cliques generally sharing high consistency with molecular trees, and often being composed of a higher proportion of osteological characters than smaller cliques. These results imply several things, that a) larger cliques are more likely to have high consistency with molecular data, implying phylogenetic signal, b) characters are more internally consistent, i.e. correlated, within partitions, and therefore that c) larger cliques of internally consistent characters, containing phylogenetic signal, are also more likely to be composed of osteological rather than non-osteological characters, although this differs between datasets. This is consistent with previous chapters showing that osteological characters are more consistent with molecular data, characters are more correlated within partitions, and that between-partition effects are inconsistent and dataset-dependent.

Chapter 4: Convergence, Correlations and Cliques: An Analysis of the Relative Performance of Morphological Character Partitions in Mammals

To test the generality of the above results, many of the methods previously described were applied to mammalian datasets, comparing osteological with dental data. I compared homoplasy, transition ages, character pair correlations on molecular trees and the presence of internally-consistent cliques between these partitions as described above, and additionally inferred and compared most parsimonious trees from these partitions. Osteological data are both less homoplasious and transition earlier than dental data, although the homoplasy explains the difference in transition ages. Correlations with reference to an independent molecular tree do not consistently differ between partitions. However, characters are more correlated within than between osteological and dental partitions both internally and when mapped onto a phylogeny. Trees inferred from osteological data rather than from dental data are closer in topology to the molecular tree. Overall, these results indicate different evolutionary dynamics between these partitions, and suggest that osteological data track evolutionary history more closely than dental data. This has implications for phylogenetic inference when using morphological data, since dental characters are often used alongside other morphological data to infer topology, particularly in fossil taxa where, besides osteological characters, dental characters are the only morphological data available.

2. Consistency, Convergence and Homoplasy

Some degree of homoplasy of morphological characters on molecular trees, as measured by both individual and ensemble retention indices measuring the fit of those data, is present in all partitions and subpartitions of morphological data tested. This is true in birds, squamates and mammals, regardless of the phylogeny used. However, clear differences exist in the levels of homoplasy between these partitions. Particularly, osteological data are more congruent with molecular trees than either soft characters or dental characters (Chapters 1 and 4). This finding has several possible interpretations and raises interesting questions about the nature of

osteological traits as compared with other subsets of morphology. One interpretation is that the relatively higher consilience between osteological than soft or dental with molecular data is evidence that osteological data convey more signal that is diagnostic of vertebrate clades. If true, this is likely to mean that trees produced from osteological data are likely to be more topologically accurate than trees inferred from other kinds of morphological data, as suggested by the closer similarity between osteological than dental trees with molecular trees (Chapter 4). Since osteological characters are preserved relatively well in the fossil record, this is a reassuring finding for palaeontologists working with fossil vertebrates, as well as being informative for phylogenetic inference using morphology in general.

However, this interpretation requires that we accept that molecular data faithfully recover evolutionary relationships, and therefore can be used as the standard against which to compare the quality of other phylogenetic data. This assumption is not without its problems; as previously discussed, it is increasingly recognised that molecular data are also subject to some degree of convergence in areas of the genome that are under selection (Parker *et al.*, 2013; Foote *et al.*, 2015; Brown *et al.*, 2019). It is possible that consilience between osteological and molecular data could therefore be in part the result of functional convergence. However, at least some molecular evolution is neutral, and owing to the objective nature and abundance of characters in molecular data, they are still widely considered more reliable than morphology in general for inferring evolutionary relationships. With this in mind, low consistency of individual characters or partitions on molecular trees built from multiple genomic loci can be regarded as evidence of homoplasy.

Differences in the selective regimes acting on morphological regions may be an explanation for the observed pattern. Dental (Peredo, Peredo, & Pyenson, 2018; Wysocki, 2019) and cranial (Kelley & Motani, 2015; McCurry *et al.*, 2017) traits are known to be influenced by convergent selection in many groups, likely meaning that phylogenetic signal in those characters is at least partially overwhelmed by directional phenotypic evolution related to

mastication. Some soft traits such as plumage, particularly with regards to colouration, may be relatively labile and convergent: limited character states owing to genetic and developmental constraint, as well as selection, have been suggested as explanations for this. For example, scale and bar patterns both evolve frequently in Anseriformes and Galloanserae plumage owing to relatively simple developmental pathways (Gluckman & Mundy, 2016). Sexual selection is also invoked as an explanation for lability in plumage traits (Price, Friedman, & Omland, 2007). Developmental constraints in addition to selection and lability might result in high levels of homoplasy as a result of rapid transitions between limited character states.

Lability of plumage, dental and other non-osteological traits is perhaps implied by their younger average transition ages on dated molecular trees (Chapter 1 and 4). The retention index and transition age of individual characters is correlated, and this model is not improved by splitting the data into partitions (Chapters 1 and 4), suggesting that the relationship between age and homoplasy is independent of the morphological region. If we take younger ages to be evidence of faster rates, this relationship is suggestive of a strong role of lability in generating convergence.

Questions surrounding the treatment of morphological data in phylogenetic analysis, such as in partitioning and model use, are still unresolved. However, homoplasy-based partitioning (Rosa, Melo, & Barbeitos, 2019), weighted parsimony (Goloboff, Torres, & Arias, 2018) and other methods have all recently been put forward as accurate methods of recovering relationships from morphological data. The relative homoplasy between morphological subsets is therefore a timely and relevant addition to current discussions around the use of morphological data for phylogenetic reconstruction in the age of genomics.

3. Correlation and Cliques

Two methods of looking for character correlations were used in this thesis. Firstly, a tree-dependent method was used to establish the correlations of characters taking into account a reference topology and branching times (Chapters 2 and 4). Secondly, the internal consistency of groups of characters was assessed using clustering methods (Chapters 3 and 4). Assessing correlations given an underlying phylogeny separates characters that appear interdependent owing to common ancestry from those are dependent owing to functional or other linkage. It does this by modelling character state changes of two characters over a topology, to determine if one of the characters transitions more frequently if the second is in state a as opposed to state b (Pagel, 1994). This method is useful when testing scenarios in which one trait requires or evolves more readily in the presence of another. For example, woody plants may be more likely to evolve from herbaceous plants that have a cambial layer than from those that do not, and thus over a tree there will be two distinct rates of transition from herbaceous to woody; a faster rate in the presence of a cambial layer, and a slower rate in its absence (Beaulieu, O'Meara, & Donoghue, 2013). When applying this method to many characters, an overall picture of integration and modularity, where character states are dependent on each other, can emerge. It is important to note here that both cladistic and morphometric data capture similar patterns of morphological disparity (Hetherington *et al.*, 2015). While continuous data are useful for morphometric studies of modularity, it is important to use methods like these to look for such patterns in discrete phylogenetic data.

Even with a very high threshold for correlation, correlated character pairs, given a molecular phylogeny, were found in many avian, reptilian and mammalian datasets (Chapters 2 and 4). Differences in degree of correlation between osteological, dental and soft characters were observed between datasets, due either to clade effects or character choice and coding of the authors. More in-depth work on individual clades would be needed to discriminate between these explanations. Interestingly, however, the datasets with the highest overall average correlation calculated by the

cumulative AICc weights of correlated models were both Charadriiformes, an order of waterbirds consisting of gulls, auks and waders.

Overall, soft characters are slightly more correlated given a phylogeny than osteological characters, as measured by least squared means, although the generality of this finding is unclear. Dental and osteological characters, on the other hand, display roughly equal correlation in mammals, although with a significant and slightly higher least square mean for dental data. This latter finding is slightly surprising given that dental data are known to be correlated as a result of functional, developmental and ecological linkage (Labonne *et al.*, 2014; Billet & Bardin, 2019). Further, characters are consistently more correlated within than between partitions. Biologically, this means that evolutionary change in osteological traits is likely to depend on the presence or absence of other osteological traits, more than it depends on the presence or absence of dental or soft traits. Likewise, the transition rates of soft traits depend on the state of other soft traits, and the same is true of dental traits. This is an unsurprising finding given that these morphological regions are likely to share evolutionary history, resulting in functional, developmental and genetic linkage, and thus be more tightly integrated.

The internal consistency of characters was determined by cluster analysis, where characters cluster on UPGMA trees according to their pairwise excess indices, or how many steps are required on the most parsimonious tree that can be constructed with two characters over and above the minimum possible steps for those characters. The absence of an independent tree in this method allows us to determine which characters are consistent with each other regardless of an underlying phylogeny. Similarly to tree-based correlated character pairs, analysis of clustering patterns using methods designed to detect phylogenetic signal in binary characters (Fritz & Purvis, 2010) finds that characters cluster by partition more than expected by chance in most datasets (Chapters 3 and 4). Additionally, larger cliques are more likely to be composed of osteological characters, and are more likely to be consistent with molecular trees. That osteological data form larger internally consistent cliques, in combination with their higher consistency

on molecular trees, reinforces the idea that osteological characters may be useful for phylogenetic reconstruction. This is because as well as supporting the molecular data, there is less conflicting signal within osteological data. On the other hand, soft and dental characters contain more conflicting signal, both in relation to molecular trees and within their respective partitions.

In addition, characters in phylogenetic datasets show no more internal correlation than expected by chance (Chapters 3 and 4), and characters are less correlated than in similar studies looking at specific characters (Sauquet *et al.*, 2017). Therefore, discovery of correlations seems to be contingent on testing specific hypotheses, involving characters and/or taxa where links between characters have been hypothesized *a priori* (Holland *et al.*, 2010; Leslie *et al.*, 2015; Billet & Bardin, 2019). Nevertheless, applying these methods without preconceived ideas about specific character correlations may be useful for uncovering surprising cases of concerted evolution, where functional or ecological links cannot necessarily be predicted *a priori*.

4. Concerted Convergence

Common evolutionary regimes link the homoplasy and correlation of morphological characters (Patterson & Givnish, 2002; Holland *et al.*, 2010; Blanke *et al.*, 2013). Functional, developmental and genetic linkage leading to integration of suites of traits serves to facilitate the evolution of specific morphotypes (Felice, Randau, & Goswami, 2018). Evolutionary regimes acting on traits guide suites of linked traits into adaptive peaks (Goswami *et al.*, 2014). These peaks in morphospace, also called ecomorphotypes, are convergent when similar pressures act on different species, resulting in common morphology that is not the result of shared ancestry. This is especially likely to occur when species already share a common genetic and biological framework. In practice, this means that concerted convergence is common between related species. On a narrow level, for example, striking

ecological convergence has been observed at the intrageneric level in *Anolis* lizards, where several ecomorphotypes have arisen multiple times independently in accordance with microhabitat occupation (Losos, 1998). At a broader scale, the same is true across frog species (Moen, Morlon, & Wiens, 2016). Parallel radiations involving the adaptation of many species to similar or shared environments have been observed in mammals (Madsen *et al.*, 2001; Gheerbrant, Filippo, & Schmitt, 2016) and birds (Fain & Houde, 2004). Although these examples are particularly striking, they may hint at a wider, more ubiquitous pattern of convergence in correlated or integrated traits. Ultimately, this may give evolution a degree of predictability, where similar environments and pressures can be expected to result in similar outcomes.

Results found here are consistent with the concept of concerted convergence, since both homoplasy and correlation are more widespread in dental and soft characters than in osteological characters. This implies the action of convergent selective regimes on integrated dental and soft characters. Although non-significant, homoplasy in cranial data was also higher than in postcranial data in most datasets, and cranial data also had a lower least square mean retention index.

Mastication is known to drive both convergence (Peredo *et al.*, 2018; Wysocki, 2019) and correlation (Billet & Bardin, 2019; Wolsan *et al.*, 2019) in dental characters. This is also true of the entire crania (Sadleir & Makovicky, 2008; McCurry *et al.*, 2017). Dental and cranial characters are additionally known to support different trees than their osteological and postcranial counterparts (Mounce, Sansom, & Wills, 2016; Sansom, Wills, & Williams, 2016). Trophic convergence requires coordinated responses of these traits, so the results presented here are consistent with the wider literature on the properties of dental and cranial morphology.

Soft traits are also ecologically important, and responsive to pressures specific to habitat and microhabitat, as well as being subject to sexual selection (Omland & Lanyon, 2000; Price *et al.*, 2007). In conjunction with limited developmental pathways (Gluckman & Mundy, 2016) this leads to repeated and coordinated convergence of form.

5. Conclusions and Implications

The results presented in this thesis all tend towards the conclusion that different subsets of morphological characters display different phylogenetic properties. This includes differences in the levels of homoplasy, transition ages, correlations with respect to a phylogeny and internal consistency of osteological, soft and dental data. Not only are differences in overall correlation observed between partitions, characters are also more likely to be correlated or internally consistent within partitions than between them. These are unsurprising findings, and may reflect different evolutionary trends in these character types; for example differences in rates, lability, constraint and directional or sexual selection. Further, this may also reflect differences in the evolutionary and developmental origins of these traits (Liang *et al.*, 2018). Previous work has highlighted differences in the properties of morphological subsets with regards to phylogenetic signal (Mounce *et al.*, 2016; Sansom *et al.*, 2016; Sansom & Wills, 2017). This thesis continues along this vein, shedding further light on the nature of phenotypic evolution in different morphological subsets. These results demonstrate again that morphological characters are not equivalent, and should not be treated as such in phylogenetic or other analyses involving morphological data.

Possibly the most important implications of these results involve the use of morphological characters in phylogenetic reconstruction. As well as reflecting the nature of morphological evolution, these results also stem from and shed light on the manner in which they are coded. For example, correlated traits are often coded as multiple characters, a form of pseudoreplication. One possibility for dealing with correlated characters is composite coding, reducing the number of characters coded for correlated traits (Torres-Montúfar, Borsch, & Ochoterena, 2018; Billet & Bardin, 2019). Additionally, the different levels of homoplasy in different partitions may need to be addressed when applying models to morphological characters in

phylogenetic analysis. Studies fitting models to several empirical datasets suggest that homoplasy-based partitioning, as a proxy for rates, outperforms other methods of phylogenetic inference using morphology (Rosa *et al.*, 2019). While not performing as well as directly partitioning morphology by homoplasy, partitioning morphology by region has been successful compared to unpartitioned approaches (Tarasov & Génier, 2015). Using the partitions tested here as a prior, or as a proxy for homoplasy and rates, could potentially help to circumvent issues with overparametrization and reduce computational power.

Despite the challenges and issues that the results presented here raise, there are also some reassuring findings. Generally, the largest cliques of characters are also those characters which best reflect evolutionary history, if we accept that the molecular data offers a good estimate of topology. This implies that generally, larger groups of internally consistent morphological characters tend to contain phylogenetic signal. Additionally, the most congruent overall signal comes from osteological data, the data type already most often used for the phylogenetic placement of fossils. This is a useful finding when constructing phylogenies containing extinct species, although the use of dental data should be considered carefully (Sansom *et al.*, 2016), as these are also preserved in the fossil record. The accurate placement of fossil species in turn is indispensible for dating phylogenies (Kimura *et al.*, 2015).

This thesis also in part addresses the conflict between molecular and morphological data. The results presented here imply the possibility that the hidden support often seen in morphological datasets (Gatesy *et al.*, 2003; Thompson, Bärmann, & Asher, 2012; Reeder *et al.*, 2015) may often come from agreement between osteological and molecular data. While the interpretation of hidden support and the synergistic effects of combining data is unclear (Thompson *et al.*, 2012), consilience between data is an important theme in evolutionary biology (Field *et al.*, 2014; McInerney, O'Connell, & Pisani, 2014). Ultimately, is not possible to infer evolutionary relationships with absolute certainty, and thus consilience between independent datasets, and uncovering hidden support in seemingly

disparate data types, is one way to address phylogenetic uncertainty and add support to existing hypotheses.

5. Directions for Future Work

The explicit measurement of how the homoplasy and correlation uncovered here affects phylogenetic analysis, and how to correct this, would be beneficial. Correlated characters have previously been shown to affect phylogenetic reconstruction (Sadleir & Makovicky, 2008; Goswami & Polly, 2010; Guillerme & Brazeau, 2018), and homoplasy is known to mislead relationships. While trees were inferred here from mammal datasets under a parsimony framework, comparing signal in dental and osteological characters, a more comprehensive approach involving different statistical approaches, different partitions, and different clades would be a logical further step. Specifically, applying these methods to invertebrates, partitioning into analogous structures (hard and soft tissue), would be of particular interest since invertebrates make up the majority of animal diversity. Furthermore, inferring trees from all data but using partitioning schemes to differentiate between regions, as mentioned above, would be beneficial for comparison with previous studies examining methods of modelling morphological data in phylogenetic inference. Data could be partitioned further in larger datasets, although as shown here, this reduces statistical power as the number of traits in each partition is reduced. Relating the signal conveyed by homoplastic and correlated traits to ecology and habitat would also help to uncover information about modes of phenotypic evolution, differentiating stochastic noise from real biological processes. Clustering species by morphology using multivariate techniques could also do this. While morphometric techniques are often used with osteological data, similar multivariate analyses using soft characters would be useful to compare how species cluster when using soft versus osteological data.

References

Beaulieu JM, O'Meara BC, Donoghue MJ. 2013. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. *Systematic Biology* 62: 725–737.

Billet G, Bardin J. 2019. Serial homology and correlated characters in morphological phylogenetics: modeling the evolution of dental crests in Placentals. *Systematic Biology* 68: 267–280.

Blanke A, Greve C, Wipfler B, Beutel RG, Holland BR, Misof B, Forschungsmuseum Z, Koenig A. 2013. The identification of concerted convergence in insect heads corroborates Palaeoptera. *Systematic Biology* 13: 250–263.

Brown AP, McGowan KL, Schwarzkopf EJ, Greenway R, Rodriguez LA, Tobler M, Kelley JL. 2019. Local ancestry analysis reveals genomic convergence in extremophile fishes. *Philosophical Transactions of the Royal Society B: Biological Sciences* 374: 20180240.

Fain MG, Houde P. 2004. Parallel radiations in the primary clades of birds. *Evolution* 58: 2558–2573.

Felice RN, Randau M, Goswami A. 2018. A fly in a tube: Macroevolutionary expectations for integrated phenotypes. *Evolution* 72: 2580–2594.

Field DJ, Gauthier JA, King BL, Pisani D, Lyson TR, Peterson KJ. 2014. Toward consilience in reptile phylogeny: microRNAs support an archosaur, not a lepidosaur affinity for turtles. *Evolution & Development* 16: 189–196.

Foote AD, Liu Y, C Thomas GW, Vinař T, Alföldi J, Deng J, Dugan S, van Elk CE, Hunter ME, Joshi V, Khan Z, Kovar C, Lee SL, Lindblad-Toh K, Mancia A, Nielsen R, Qin X, Qu J, Raney BJ, Vijay N, W Wolf JB, Hahn MW, Muzny DM, Worley KC, Thomas Gilbert MP, Gibbs RA. 2015. Convergent evolution of the genomes of marine mammals.

Fritz SA, Purvis A. 2010. Selectivity in Mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conservation Biology* 24: 1042–1051.

Gatesy J, Amato G, Norell M, Desalle R, Hayashi C. 2003. Combined Support

for Wholesale Taxic Atavism in Gavialine Crocodylians. *Syst. Biol* 52: 403–422.
Gheerbrant E, Filippo A, Schmitt A. 2016. Convergence of Afrotherian and Laurasiatherian ungulate-like mammals: first morphological evidence from the Paleocene of Morocco. *PLOS ONE* 11: e0157556.
Gluckman TL, Mundy NI. 2016. Evolutionary pathways to convergence in plumage patterns. *BMC Evolutionary Biology* 16: 172.
Goloboff PA, Torres A, Arias JS. 2018. Weighted parsimony outperforms other methods of phylogenetic inference under models appropriate for morphology. *Cladistics* 34: 407–437.
Goswami A, Polly PD. 2010. The influence of character correlations on phylogenetic analyses: a case study of the carnivoran cranium. In: Goswami A,, In: Friscia A, eds. *Carnivoran Evolution*. Cambridge: Cambridge University Press, 141–164.
Goswami A, Smaers JB, Soligo C, Polly PD. 2014. The macroevolutionary consequences of phenotypic integration: from development to deep time. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369: 20130254–20130254.
Guillerme T, Brazeau MD. 2018. Influence of different modes of morphological character correlation on phylogenetic tree inference. *bioRxiv*: 308742.
Hetherington AJ, Sherratt E, Ruta M, Wilkinson M, Deline B, Donoghue PCJ. 2015. Do cladistic and morphometric data capture common patterns of morphological disparity? *Palaeontology*.
Holland BR, Spencer HG, Worthy TH, Kennedy M. 2010. Identifying cliques of convergent characters: Concerted evolution in the cormorants and shags. *Systematic Biology* 59: 433–445.
Kelley NP, Motani R. 2015. Trophic convergence drives morphological convergence in marine tetrapods. *Biology Letters* 11: 20140709.
Kimura Y, Hawkins MTR, McDonough MM, Jacobs LL, Flynn LJ. 2015. Corrected placement of Mus-Rattus fossil calibration forces precision in the molecular tree of rodents. *Scientific Reports* 5: 14444.
Labonne G, Navarro N, Laffont R, Chateau-Smith C, Montuire S. 2014.

Developmental integration in a functional unit: deciphering processes from adult dental morphology. *Evolution and Development* 16: 224–232.
Leslie AB, Beaulieu JM, Crane PR, Knopf P, Donoghue MJ. 2015. Integration and macroevolutionary patterns in the pollination biology of conifers. *Evolution* 69: 1573–1583.
Liang C, Musser JM, Cloutier A, Prum RO, Wagner GP. 2018. Pervasive Correlated Evolution in Gene Expression Shapes Cell and Tissue Type Transcriptomes. *Genome Biology and Evolution* 10: 538–552.
Losos JB. 1998. Contingency and determinism in replicated adaptive radiations of island lizards. *Nature* 279: 2115–2118.
Madsen O, Scally M, Douady CJ, Kao DJ, DeBry RW, Adkins R, Amrine HM, Stanhope MJ, de Jong WW, Springer MS. 2001. Parallel adaptive radiations in two major clades of placental mammals. *Nature* 409: 610–614.
McCurry MR, Evans AR, Fitzgerald EMG, Adams JW, Clausen PD, McHenry CR. 2017. The remarkable convergence of skull shape in crocodilians and toothed whales. *Proceedings of the Royal Society B: Biological Sciences* 284: 20162348.
McInerney JO, O'Connell MJ, Pisani D. 2014. The hybrid nature of the Eukaryota and a consilient view of life on Earth. *Nature Reviews Microbiology* 12: 449–455.
Moen DS, Morlon H, Wiens JJ. 2016. Testing convergence versus history: convergence dominates phenotypic evolution for over 150 million years in frogs. *Systematic Biology* 65: 146–160.
Mounce RCP, Sansom R, Wills MA. 2016. Sampling diverse characters improves phylogenies: craniodental and postcranial characters of vertebrates often imply different trees. *Evolution* 70: 666–686.
Omland KE, Lanyon SM. 2000. Reconstructing plumage evolution in orioles (Icterus): Repeated convergence and reversal in patterns. *Evolution* 54: 2119–2133.
Pagel M. 1994. Detecting Correlated Evolution on Phylogenies: A General Method for the Comparative Analysis of Discrete Characters. *Proceedings of the Royal Society B: Biological Sciences* 255: 37–45.
Parker J, Tsagkogeorga G, Cotton JA, Liu Y, Provero P, Stupka E, Rossiter SJ.

2013. Genome-wide signatures of convergent evolution in echolocating mammals.
Patterson TB, Givnish TJ. 2002. Phylogeny, concerted convergence, and phylogenetic niche conservatism in the core Liliales: insights from rbcL and ndhF sequence data. *Evolution* 56: 233–252.
Peredo CM, Peredo JS, Pyenson ND. 2018. Convergence on dental simplification in the evolution of whales. *Paleobiology* 44: 434–443.
Price JJ, Friedman NR, Omland KE. 2007. Song and plumage evolution in the New World orioles (Icterus) show similar lability and convergence in patterns. *Evolution* 61: 850– 863.
Reeder TW, Townsend TM, Mulcahy DG, Noonan BP, Wood PL, Sites JW, Wiens JJ. 2015. Integrated analyses resolve conflicts over squamate reptile phylogeny and reveal unexpected placements for fossil taxa. *PLoS ONE* 10: e0118199.
Rosa BB, Melo GAR, Barbeitos MS. 2019. Homoplasy-Based Partitioning Outperforms Alternatives in Bayesian Analysis of Discrete Morphological Data (L Harmon, Ed.). *Systematic Biology* 68: 657–671.
Sadleir RW, Makovicky PJ. 2008. Cranial shape and correlated characters in crocodilian evolution. *Journal of Evolutionary Biology* 21: 1578–1596.
Sansom RS, Wills MA. 2017. Differences between hard and soft phylogenetic data. *Proceedings of the Royal Society B: Biological Sciences* 284: 20172150.
Sansom RS, Wills MA, Williams T. 2016. Dental data perform relatively poorly in reconstructing mammal phylogenies: Morphological partitions evaluated with molecular benchmarks. *Systematic Biology* 66: syw116.
Sauquet H, von Balthazar M, Magallón S, Doyle JA, Endress PK, Bailes EJ, Barroso de Morais E, Bull-Hereñu K, Carrive L, Chartier M, Chomicki G, Coiro M, Cornette R, El Ottra JHL, Epicoco C, Foster CSP, Jabbour F, Haevermans A, Haevermans T, Hernández R, Little SA, Löfstrand S, Luna JA, Massoni J, Nadot S, Pamperl S, Prieu C, Reyes E, dos Santos P, Schoonderwoerd KM, Sontag S, Soulebeau A, Staedler Y, Tschan GF, Wing-Sze Leung A, Schönenberger J. 2017. The ancestral flower of angiosperms and its early diversification. *Nature Communications* 8: 16047.
Tarasov S, Génier F. 2015. Innovative Bayesian and Parsimony Phylogeny of

Dung Beetles (Coleoptera, Scarabaeidae, Scarabaeinae) Enhanced by Ontology-Based Partitioning of Morphological Characters (BJ Mans, Ed.). *PLOS ONE* 10: e0116671.
Thompson RS, Bärmann E V., Asher RJ. 2012. The interpretation of hidden support in combined data phylogenetics. *Journal of Zoological Systematics and Evolutionary Research* 50: 251–263.
Torres-Montúfar A, Borsch T, Ochoterena H. 2018. When Homoplasy Is Not Homoplasy: Dissecting Trait Evolution by Contrasting Composite and Reductive Coding (N MacLeod, Ed.). *Systematic Biology* 67: 543–551.
Wolsan M, Suzuki S, Asahara M, Motokawa M. 2019. Dental integration and modularity in pinnipeds. *Scientific Reports* 9: 4184.
Wysocki MA. 2019. Fossil evidence of evolutionary convergence in juvenile dental morphology and upper canine replacement in sabertooth carnivores. *Ecology and Evolution* 9: 12649–12657.

Appendices

Appendix 1. Example TNT Matrix with First 100 Molecular Trees

1.1. Bertelli, Chiappe & Mayr, 2014

```
nstates nogaps;
xread

157 17

Megapodius_freycinet
        00001010?00000[01]0000?0?1?0000000??000?00130?0110031??0?10020111100000101
00000110101000000000011000?000?00000011003?10020?0?????1??2?????????????????????000?0
??0??
Rhea_americana
        00111102101001010 20?000100011010000111011110000120??1????0000???11?100????
??011??0?0?00?20??0??10?010000?110?111?10200?0?????0?0?????????????0????????000?0???
??
Tinamus_major
        10000101010120001110000100001000010101111201000011001100011011000000021100
00100011011100101010021001210000110002110?11210121010121100100000001000101121121 1
01
Crypturellus_tataupa
        10000101010121001110110100001000?01111112001001010011111021010100000021111
012?012101110000103002100113000011110210021121111001012011010000000100110112112 11
01
Crypturellus_parvirostris
        10000101010121001110110100001000?01111112001001010011111021010100000021111
012?012101110000103002100113000011110210021121111001012011010000000100110????????
??
Crypturellus_obsoletus
        10000101010121001110010100001000010101111200100101001111021010100000021111
011001210111000010300210011300001111021002112111100101201101000000010011 0????????
??
Crypturellus_variegatus
        10000101010121001110000100001000010101111 20?000101001111021010100000021111
011001210111000010200210012300001111021002 1????????????????????????????????????????
??
Crypturellus_undulatus
        10000101010121001110?00100001000010101111 20?000001001111021010100000021111
011001210111000010200210012300001111021002 1????????????????????????????????????????
??
Crypturellus_soui
        10000101010121001110000100001000010101111200000001001111021010100000?21111
011?012101110000102002100113000011110210021121111001012011010000000100110112112 11
11
Nothura_maculosa
        100000[01]10100011111111101000011011[01]011111121010020100112 01[23]2011100
000120000012100211111010011311211021300101111111100[23]110101301000011001000010100
0?011110110000
Nothura_darwinii
        10000011010001111111111010000110111011111121010020100112012201 11[01]0010120
000012100211111010011311211021300101111111003110101301000011001000010100 0?0111101
10000
Nothoprocta_cinerascens
        1[01]00101201001101111100010010111 01[01]011111121020020011112012[12]010100
0[01]0[01]20100012100211111010011311211021301101111111003111100?02000100000011110
10000021110110000
Nothoprocta_perdicaria
        11001012?10001?111111001001011101011221222102102 0?????201310111000101200001
012000211111010011311211021301101111111100[23]111100?020001100000111101000?0211101
10000
Rhynchotus_rufescens
        11111112110000011110100100101000101 0?111220100201101120 12?010100010120 00[
01]012[01]0021[01]111010011311211 02?201001111111003111101002000110000011110000?0
1??????????
Eudromia_elegans
        10000012011010010112010000001001110111110200001221000110023011200000[01]21
```

```
1110120002101100010102002100123100111111000020101013120110000101110110111000011111
10001
Tinamotis_pentlandii
       101000010110?00101120100000111002101111112010001100000100230112000000020111
012000210010101110200210112311111121111011010100?020110000100110110111?00111010000
01
Tinamotis_ingoufi
       100000110110010101120100000111002101111112010001100000100230112000000020111
01200021001[01]10111020021011231011111111110110??????????????????????????????????
?????
;

ccode   +  148;

xgroup=0 (osteological) 0.115 ;
xgroup=1 (nonosteological) 116.156 ;



tread

((((2 ((7 8) ((3 5) (4 6)))) ((16 (14 15)) ((12 (13 11)) (10 9)))) 1) 0)*
((1 ((((10 9) ((12 13) 11)) ((15 14) 16)) (2 ((4 (6 (3 5))) (7 8))))) 0)*
((1 (((((13 11) 12) (10 9)) ((14 15) 16)) (2 ((4 (5 3)) (7 (6 8)))))) 0)*
(((((14 (15 16)) ((13 (11 12)) (9 10))) (2 ((7 8) (6 (5 (3 4)))))) 1) 0)*
((1 ((2 ((((5 3) 4) 6) (8 7))) (((15 14) 16) ((10 9) (12 (11 13)))))) 0)*
((((2 (6 (((5 3) 4) (7 8)))) (((14 16) 15) ((9 10) (11 (12 13))))) 1) 0)*
((1 (((((3 4) 5) ((7 6) 8)) 2) (((10 9) (13 (12 11))) ((15 16) 14)))) 0)*
((1 (((14 (16 15)) ((10 9) ((13 12) 11))) (((5 (3 4)) (7 (8 6))) 2))) 0)*
((((((15 16) 14) ((10 9) ((12 13) 11))) (2 ((6 (4 (3 5))) (7 8)))) 1) 0)*
((1 ((((10 9) ((12 13) 11)) ((15 14) 16)) (2 ((4 (6 (3 5))) (7 8))))) 0)*
((((((15 16) 14) ((10 9) (13 (11 12)))) (2 (((6 8) 7) (5 (4 3))))) 1) 0)*
((1 (((2 (((7 6) 8) ((3 5) 4))) (((11 13) 12) (10 9))) (14 (15 16)))) 0)*
((1 ((2 ((5 (3 4)) (8 (7 6)))) ((14 (15 16)) (((13 12) 11) (9 10))))) 0)*
((1 ((((16 15) 14) (((5 (3 4)) ((6 8) 7)) 2)) ((10 9) (12 (13 11))))) 0)*
((((((16 15) 14) (((12 13) 11) (9 10))) (2 ((6 (4 (3 5))) (7 8)))) 1) 0)*
((1 (((((12 11) 13) (10 9)) ((15 16) 14)) (2 ((8 7) (6 ((5 3) 4)))))) 0)*
((((2 ((7 (6 8)) ((3 4) 5))) ((14 (16 15)) ((12 (13 11)) (9 10)))) 1) 0)*
(((((14 (16 15)) ((10 9) ((12 13) 11))) (2 ((7 (6 8)) ((3 5) 4)))) 1) 0)*
((1 ((((10 9) (11 (13 12))) (((((5 4) 3) 6) (8 7)) 2)) (14 (16 15)))) 0)*
((1 ((((14 15) 16) (2 (((3 5) (4 6)) (8 7)))) ((13 (12 11)) (10 9)))) 0)*
((1 (((14 (16 15)) ((12 (13 11)) (10 9))) (2 (((6 8) 7) (4 (5 3)))))) 0)*
((1 ((((10 9) ((11 12) 13)) ((15 16) 14)) (((7 8) (5 (6 (3 4)))) 2))) 0)*
((((((16 15) 14) ((13 (11 12)) (9 10))) (2 ((4 (5 3)) (8 (7 6))))) 1) 0)*
(((((14 (15 16)) ((10 9) ((13 11) 12))) (2 (((8 6) 7) (4 (3 5))))) 1) 0)*
((((2 (((3 5) 4) ((8 6) 7))) (((15 16) 14) ((11 (13 12)) (9 10)))) 1) 0)*
((((2 ((7 8) ((3 5) (4 6)))) ((((13 11) 12) (9 10)) (14 (16 15)))) 1) 0)*
((1 ((2 (((5 3) 4) (8 (6 7)))) (((12 (11 13)) (10 9)) (14 (16 15))))) 0)*
((((2 ((5 (3 4)) (8 (6 7)))) ((14 (15 16)) ((9 10) ((13 11) 12)))) 1) 0)*
((1 ((2 (((5 3) 4) (8 (7 6)))) (((15 16) 14) ((10 9) (13 (12 11)))))) 0)*
((1 ((((9 10) ((12 11) 13)) ((14 15) 16)) (2 ((3 (5 4)) ((7 6) 8))))) 0)*
((1 ((2 ((4 (5 3)) ((6 7) 8))) (((11 (13 12)) (10 9)) (14 (15 16))))) 0)*
((1 ((((8 (6 7)) ((5 3) 4)) 2) ((16 (15 14)) ((10 9) ((11 12) 13))))) 0)*
((1 (((16 15) 14) ((((11 12) 13) (10 9)) ((((7 6) 8) ((3 5) 4)) 2)))) 0)*
((1 (((((7 6) 8) (4 (5 3))) 2) (((10 9) (13 (11 12))) ((14 16) 15)))) 0)*
((1 ((2 (((8 7) 6) (4 (3 5)))) ((((11 13) 12) (10 9)) (14 (16 15))))) 0)*
((1 (((((7 8) ((4 6) (5 3))) 2) ((10 9) ((11 13) 12))) (14 (16 15)))) 0)*
((1 (((14 (15 16)) ((10 9) ((13 12) 11))) (2 (((3 5) 4) ((7 8) 6))))) 0)*
((1 (((((12 11) 13) (10 9)) ((15 14) 16)) ((((5 3) 4) (7 (8 6))) 2))) 0)*
((1 ((2 ((4 (3 5)) ((6 8) 7))) (((16 15) 14) ((12 (11 13)) (10 9))))) 0)*
((((((12 (13 11)) (10 9)) (14 (16 15))) (2 ((7 8) ((3 5) (6 4))))) 1) 0)*
((1 ((((14 16) 15) ((10 9) (11 (13 12)))) ((((6 (5 3)) 4) (7 8)) 2))) 0)*
((((((16 15) 14) ((10 9) ((13 12) 11))) (2 (((7 6) 8) ((3 5) 4)))) 1) 0)*
((1 ((2 (((6 8) 7) ((5 3) 4))) ((16 (15 14)) (((11 13) 12) (10 9))))) 0)*
((((((10 9) ((13 11) 12)) ((15 16) 14)) (2 ((8 7) ((4 6) (5 3))))) 1) 0)*
((1 ((((16 15) 14) (((12 13) 11) (10 9))) (((6 ((4 3) 5)) (7 8)) 2))) 0)*
((1 (((((5 3) (6 4)) (7 8)) 2) (((15 16) 14) (((11 13) 12) (10 9))))) 0)*
((((2 ((6 ((5 3) 4)) (8 7))) (((16 15) 14) (((12 11) 13) (10 9)))) 1) 0)*
((1 ((2 (((3 5) 4) (8 (6 7)))) (((16 15) 14) ((10 9) (11 (13 12)))))) 0)*
(((((14 (16 15)) ((9 10) (12 (13 11)))) (2 ((4 (6 (5 3))) (7 8)))) 1) 0)*
((1 (((((12 11) 13) (9 10)) ((16 15) 14)) (2 ((7 (6 8)) ((3 5) 4))))) 0)*
((1 ((((5 (4 3)) (8 (7 6))) 2) ((16 (14 15)) ((10 9) (13 (11 12)))))) 0)*
((((((11 (12 13)) (9 10)) ((15 16) 14)) (2 (((5 3) 4) ((6 7) 8)))) 1) 0)*
((((2 (((7 6) 8) ((3 5) 4))) (((16 15) 14) (((11 13) 12) (9 10)))) 1) 0)*
((1 ((2 ((8 (7 6)) (4 (5 3)))) ((((12 13) 11) (10 9)) (16 (15 14))))) 0)*
(((((14 (15 16)) (((12 13) 11) (10 9))) (2 (((3 5) 4) ((7 6) 8)))) 1) 0)*
((1 (((((12 13) 11) (10 9)) ((14 15) 16)) (2 ((5 (4 3)) (8 (7 6)))))) 0)*
```

```
((((2 (((5 3) (4 6)) (8 7))) ((14 (15 16)) (((13 12) 11) (10 9)))) 1) 0)*
((1 ((((14 15) 16) ((10 9) ((11 13) 12))) (((8 (7 6)) (4 (5 3))) 2))) 0)*
((1 ((2 (((6 7) 8) ((3 5) 4))) (((15 14) 16) (((13 11) 12) (10 9))))) 0)*
((1 ((2 (((5 3) (4 6)) (7 8))) (((15 16) 14) (((11 12) 13) (10 9))))) 0)*
((((2 (((4 3) 5) ((7 6) 8))) ((14 (16 15)) ((11 (12 13)) (9 10)))) 1) 0)*
((1 ((((15 14) 16) ((12 (11 13)) (10 9))) (2 ((8 7) ((6 4) (5 3)))))) 0)*
((1 (((14 (16 15)) ((13 (11 12)) (10 9))) (2 (((6 7) 8) (5 (4 3)))))) 0)*
((1 ((2 ((8 7) (((4 3) 5) 6))) ((((11 13) 12) (10 9)) (15 (16 14))))) 0)*
(((((16 (14 15)) (((13 11) 12) (10 9))) (2 (((6 7) 8) (5 (3 4))))) 1) 0)*
((((2 (((3 5) (6 4)) (8 7))) ((((11 12) 13) (9 10)) ((16 15) 14))) 1) 0)*
((1 ((((10 9) ((13 11) 12)) ((15 16) 14)) (((8 (6 7)) (5 (3 4))) 2))) 0)*
((1 ((2 ((8 7) ((6 4) (3 5)))) (((10 9) (11 (12 13))) (16 (14 15))))) 0)*
((((2 ((7 (8 6)) ((5 3) 4))) ((16 (14 15)) ((10 9) ((11 12) 13)))) 1) 0)*
((1 ((((10 9) (11 (12 13))) ((16 15) 14)) (2 (((8 6) 7) (4 (3 5)))))) 0)*
((1 ((((14 15) 16) ((13 (11 12)) (10 9))) (2 ((((4 3) 5) 6) (7 8))))) 0)*
((1 (((10 9) ((11 13) 12)) (((16 15) 14) (2 ((8 7) ((4 6) (5 3)))))) 0)*
((((2 ((8 (6 7)) ((3 5) 4))) ((((11 13) 12) (9 10)) (14 (15 16)))) 1) 0)*
((1 ((((10 9) (12 (11 13))) (14 (16 15))) (((4 (3 5)) (7 (6 8))) 2))) 0)*
((1 ((((10 9) (12 (13 11))) (14 (16 15))) (((((5 3) 4) 6) (7 8)) 2))) 0)*
((1 ((((11 (12 13)) (9 10)) ((15 16) 14)) (2 ((7 8) (((5 3) 4) 6))))) 0)*
((((((15 14) 16) ((11 (12 13)) (9 10))) (2 (((5 3) 4) ((6 7) 8)))) 1) 0)*
((((((10 9) (13 (11 12))) (16 (15 14))) (2 (((3 4) 5) (8 (7 6))))) 1) 0)*
((((2 ((7 8) ((6 4) (3 5)))) (((14 15) 16) ((10 9) ((12 13) 11)))) 1) 0)*
((1 ((2 ((7 8) ((3 5) (6 4)))) ((14 (16 15)) (((11 13) 12) (10 9))))) 0)*
((1 (((14 (15 16)) (((13 12) 11) (10 9))) (((4 (5 3)) (8 (7 6))) 2))) 0)*
((((((14 15) 16) ((13 (12 11)) (10 9))) (2 ((8 (7 6)) ((5 3) 4)))) 1) 0)*
((((2 ((8 (6 7)) ((3 4) 5))) (((16 15) 14) (((13 12) 11) (9 10)))) 1) 0)*
((1 (((14 (16 15)) ((10 9) (13 (11 12)))) (((8 7) ((3 5) (6 4))) 2))) 0)*
((1 ((((7 8) ((6 (5 3)) 4)) 2) (((15 16) 14) ((11 (13 12)) (10 9))))) 0)*
((1 (((((11 13) 12) (9 10)) (16 (15 14))) (2 (((3 5) 4) ((8 6) 7))))) 0)*
((1 ((2 (((5 3) 4) (7 (6 8)))) (((14 16) 15) ((13 (11 12)) (10 9))))) 0)*
((1 ((2 (((4 6) (5 3)) (8 7))) ((((11 12) 13) (10 9)) (14 (16 15))))) 0)*
((((2 ((6 ((3 5) 4)) (7 8))) ((((13 12) 11) (9 10)) ((16 15) 14))) 1) 0)*
((1 ((((15 16) 14) (((11 12) 13) (10 9))) (2 (((6 4) (5 3)) (8 7))))) 0)*
((1 ((((7 (6 8)) ((3 5) 4)) 2) ((16 (15 14)) ((10 9) ((12 13) 11))))) 0)*
((1 ((2 (((5 3) 4) ((7 6) 8))) (((13 (12 11)) (10 9)) (15 (14 16))))) 0)*
(((((14 (16 15)) ((10 9) (12 (13 11)))) (2 ((6 (4 (3 5))) (7 8)))) 1) 0)*
((1 ((2 (((7 6) 8) (4 (3 5)))) (((14 15) 16) ((10 9) ((13 11) 12))))) 0)*
((((2 ((8 7) ((6 4) (3 5)))) ((((11 12) 13) (9 10)) ((16 15) 14))) 1) 0)*
((((((10 9) (13 (12 11))) (14 (16 15))) (2 ((8 (7 6)) (4 (5 3))))) 1) 0)*
((((((14 15) 16) (2 ((8 7) (((5 3) 4) 6)))) ((9 10) ((11 13) 12))) 1) 0)*
((((2 (((5 3) 4) (8 (7 6)))) (((16 15) 14) (((11 13) 12) (10 9)))) 1) 0)*
((1 ((((14 15) 16) ((11 (13 12)) (10 9))) (2 (((3 5) (4 6)) (8 7))))) 0)*
((1 (((14 (15 16)) ((10 9) ((12 13) 11))) (((6 (4 (5 3))) (8 7)) 2))) 0)*

;
proc /;
comments 0
;
```

Appendix 2. Code

The following sections display the important code developed for this thesis.

2.1. CIRI (Consistency Index Retention Index)

Below is the TNT code, christened CIRI, used for computing the individual character retention indices and ensemble retention indices for partitions, used in Chapter 1.

```
macro-;
macro* 3 50000;
macro [ 2000000;
macro=;
var-;
log/;
```

```
if (ntrees==-1) errmsg no trees!; end

/** deactivate taxa not in tree and create matrix with trees excluding them**/
/** up to three character groups specified, xgroup 7 being excluded ***/

loop 0 ntax
       if (isintree[0 #1]) continue else taxcode -#1; end
stop
log $dataset<.ciri.tnt;
xread-*;
xgroup*;
log/;
tsave* $dataset<.ciri.tnt +;
save-.;
tsave/;


proc $dataset<.ciri.tnt;

log $dataset<.log;


/**calculate ensemble retention indices for dataset paritions and individual
character retention indices **/

var: minmax[(nchar+1) 2] ris[(nchar+1)] mina minb minc mind mine maxa maxb maxc
maxd maxe theactual RIa RIb RIc RId RIe acta actb actc actd acte ncha nchb nchc
nchd nche avria avrib avric avrid avrie tax ch;

set mina 0; set minb 0; set minc 0; set mind 0; set mine 0; set maxa 0; set maxb
0; set maxc 0; set maxd 0; set maxe 0; set RIa 0; set RIb 0; set RIc 0; set RId
0; set RIe 0; set ncha 0; set nchb 0; set nchc 0; set nchd 0; set nche 0; set
avria 0; set avrib 0; set avric 0; set avrid 0; set avrie 0;

quote plap;

loop 0 nchar

       set ris[#1] 0;
       if ((isinfo[#1]) && (minsteps[#1]>0))
              set minmax [#1 0] minsteps [#1];
              set minmax [#1 1] maxsteps [#1];
              if (isinxgroup[0 #1])
                     set mina ('mina' + 'minmax[#1 0]'); set maxa ('maxa' +
'minmax[#1 1]'); set ncha ++;
              end
              if (isinxgroup[1 #1])
                            set minb ('minb' + 'minmax[#1 0]'); set maxb
('maxb' + 'minmax[#1 1]'); set nchb ++;

              end
              if (isinxgroup[2 #1])
                     set minc ('minc' + 'minmax[#1 0]'); set maxc ('maxc' +
'minmax[#1 1]'); set nchc ++;
              end
              if (isinxgroup[3 #1])
                     set mind ('mind' + 'minmax[#1 0]'); set maxd ('maxd' +
'minmax[#1 1]'); set nchd ++;
              end
              if (isinxgroup[7 #1])
                     set ris[#1] 100000;           end
       else   set ris[#1] 100000;
       end
stop

quote plop;
loop 0 ntrees

       set acta 0; set actb 0; set actc 0; set actd 0; set acte 0;
       loop 0 nchar

              if ((isinfo[#2]) && (minsteps[#2]>0))
                     set theactual length [#1 #2];
                     if (isinxgroup[0 #2])
                            set acta ('acta' + 'theactual');

                     end
```

```
                        if (isinxgroup[1 #2] )
                        set actb ('actb' + 'theactual');
                        end

                        if (isinxgroup[2 #2] )

                                set actc ('actc' + 'theactual');
                        end

                        if (isinxgroup[3 #2] )
                                set actd ('actd' + 'theactual');
                        end
                        set ris[#2] ('ris[#2]' + (('minmax[#2 1]'-
'theactual')/('minmax[#2 1]'-'minmax[#2 0]')) );
                end
        stop
        if ('ncha'>0)
                set RIa ('RIa'+ (('maxa'-'acta')/('maxa'-'mina')) );
        end
        if ('nchb'>0)
                set RIb ('RIb'+ (('maxb'-'actb')/('maxb'-'minb')) );
        end
        if ('nchc'>0)
                set RIc ('RIc' + (('maxc'-'actc')/('maxc'-'minc')) );
        end
        if ('nchd'>0)
                set RId ('RId' + (('maxd'-'actd')/('maxd'-'mind')) );
        end
stop

quote plip;

if ('ncha'>0)
        set RIa ('RIa'/(ntrees+1));
        end
if ('nchb'>0)
        set RIb ('RIb'/(ntrees+1));
                end
if ('nchc'>0)
        set RIc ('RIc'/(ntrees+1));
                end
if ('nchd'>0)
        set RId ('RId'/(ntrees+1));
        end

var : output[('ncha'+'nchb'+'nchc'+'nchd') 3] count;

set count 0;

loop 0 nchar

        set ris[#1] ('ris[#1]'/(ntrees+1));

        if ('ris[#1]'<=1)
                set output['count' 0] (#1+1);

                set output['count' 2] 'ris[#1]';
                if (isinxgroup[0 #1])
                        set output['count' 1] 0;
                        set avria ('avria'+'ris[#1]');
                end
                if (isinxgroup[1 #1])
                        set output['count' 1] 1;
                        set avrib ('avrib'+'ris[#1]');
                end
                if (isinxgroup[2 #1])
                        set output['count' 1] 2;
                        set avric ('avric'+'ris[#1]');
                end
                if (isinxgroup[3 #1])
                        set output['count' 1] 3;
                set avrid ('avrid'+'ris[#1]');
                end
                set count++;
        end
```

```
stop
var ris*;
quote plup;
if ('ncha'>0)
       set avria ('avria'/'ncha');
end
if ('nchb'>0)
       set avrib ('avrib'/'nchb');
end
if ('nchc'>0)
       set avric ('avric'/'nchc');
end
if ('nchd'>0)
       set avrid ('avrid'/'nchd');
end

set tax (ntax+1); set ch ('ncha'+'nchb'+'nchc'+'nchd');


Quote ciri6-2;
Quote output for retention indices of individual characters (informative only);

var output*;

quote total of included characters and informative characters 'tax' 'ch';

quote partition 1 has ensemble RI 'RIa' 'ncha' characters;
quote partition 2 has ensemble RI 'RIb' 'nchb' characters;
quote partition 3 has ensemble RI 'RIc' 'nchc' characters;
quote partition 4 has ensemble RI 'RId' 'nchd' characters;


quote partition 1 has average ri 'avria';
quote partition 2 has average ri 'avrib'
quote partition 3 has average ri 'avric';
quote partition 4 has average ri 'avrid';

log/;
proc/;
```

2.2. Transition Ages

The following R code first identifies and removes uninformative and multistate characters, before estimating character transition ages under an ACCTRAN framework as described in Chapter 1.

```
Library(paleotree)

colallchanges<-matrix(,nrow=(ncol(Dataset)),ncol=(Ntip(DatasetT1)*2-1)) #nrow =
 number of characters +1 (i.e. ncol(dataset)), ncol = number of internal nodes +
 number of tips (i.e. nnodes)
row.names(colallchanges)<-names(Dataset)
nroot<-(Ntip(DatasetT1)+1)
nnodes<-(Ntip(DatasetT1)*2-1)

#Establish which characters to remove (multistate and uninformative), and remove
 them from colallchanges

Datasetnew_matrix <- Dataset[,-1]
row.names(Datasetnew_matrix) <- Dataset[,1]

Datasetrmchars <- for (j in 1:ncol(Datasetnew_matrix)){

ch<- as.vector(Datasetnew_matrix[,j])

nC=0 #number of scored characters

c0=0 #whether or not state 0 exists in this character (1 = yes)
c1=0
c2=0
```

```
c3=0
c4=0
c5=0
c6=0
c7=0
c8=0
c9=0

n0=0 #number of 0s in this character
n1=0
n2=0
n3=0
n4=0
n5=0
n6=0
n7=0
n8=0
n9=0

for (i in (ch)) {
 if (i == 0) {c0=1
      n0=n0+1
      nC=nC+1}
 if (i == 1) {c1=1
      n1=n1+1
      nC=nC+1}
 if (i == 2) {c2=1
      n2=n2+1
      nC=nC+1}
 if (i == 3) {c3=1
      n3=n3+1
      nC=nC+1}
 if (i == 4) {c4=1
      n4=n4+1
      nC=nC+1}
 if (i == 5) {c5=1
      n5=n5+1
      nC=nC+1}
 if (i == 6) {c6=1
      n6=n6+1
      nC=nC+1}
  if (i == 7) {c7=1
    n7=n7+1
    nC=nC+1}
  if (i == 8) {c8=1
    n8=n8+1
    nC=nC+1}
  if (i == 9) {c9=1
    n9=n9+1
    nC=nC+1}}


nstates=c0+c1+c2+c3+c4+c5+c6+c7+c8+c9

if    (nstates!=2) {print (j)} else
 {if(n0==(nC-1)) {print(j)} else
 {if(n1==(nC-1)) {print(j)} else
 {if(n2==(nC-1)) {print(j)} else
 {if(n3==(nC-1)) {print(j)} else
 {if(n4==(nC-1)) {print(j)} else
 {if(n5==(nC-1)) {print(j)} else
 {if(n6==(nC-1)) {print(j)} else
 {if(n7==(nC-1)) {print(j)} else
 {if(n8==(nC-1)) {print(j)} else
 {if(n9==(nC-1)) {print(j)}}}}}}}}}}

}



rmchars  <- read.csv("Datasetrmchars.csv") #make this file from returned
 characters
rmchars <- rmchars$Chars
rmchars <- as.vector(rmchars)

colallchanges<-matrix(,nrow=(ncol(Dataset)),ncol=(Ntip(DatasetT1)*2-1)) #nrow =
 number of characters +1 (i.e. ncol(dataset)), ncol = number of internal nodes +
```

```
 number of tips (i.e. nnodes)
row.names(colallchanges)<-names(Dataset)

Datasetrmrows <- paste("Col", rmchars, sep = "")
Datasetkeeprows<- setdiff(colnames(Dataset), Datasetrmrows)
Datasetcolallchanges <- colallchanges[Datasetkeeprows,]
colchanges<-Datasetcolallchanges
colchanges<-colchanges[-1,]
colallchanges<-colallchanges[-1,]

#the big loop - 100 trees[l], each character reconstruction[j], each anc-
 descendant[i], extract branch ages 100 times [k]
pb = txtProgressBar(min = 0, max = 100, initial = 0)
colavavagechanges<-matrix(,nrow=nrow(colchanges),ncol=100)
colavavavagechanges<-vector(length=nrow(colchanges))
for (l in 1:100) {
 DatasetT_this <- DatasetT[[l]]
 setTxtProgressBar(pb,l)



 for (j in row.names (colchanges)){
      Col_this<-Dataset[,j]
      Col_this<-as.vector(Col_this)
      Col_this <- noquote(Col_this)
      names(Col_this) <- DatasetTips
      char_recon<-ancPropStateMat(Col_this,DatasetT_this,type="ACCTRAN")

      for (i in nroot:nnodes){
            child_node<-Descendants(DatasetT1,i,type=("children"))
            vec1 <- c(char_recon[child_node[1]], char_recon[i])
            vec1<-sort(vec1)
            diff1 <- vec1[2] - vec1[1]
            if (diff1 <= 0.5) {(colchanges[j,child_node[1]]<-0)}
 else{colchanges[j,child_node[1]]<-1}
            vec2 <- c(char_recon[child_node[2]], char_recon[i])
            vec2<-sort(vec2)
            diff2 <- vec2[2] - vec2[1]
            if (diff2 <= 0.5) {(colchanges[j,child_node[2]]<-0)}
 else{colchanges[j,child_node[2]]<-1}
      }
 }


#Create table for branch start and end dates
 nodedepth <- node.depth.edgelength(DatasetT_this)
 nodeage <- nodedepth[1] - nodedepth
 nodeages<-matrix(,nrow=nnodes,ncol=2)
 for (i in 1:nnodes){
      nodeages[i,2]<-nodeage[i]
      anc_node<-Ancestors(DatasetT1,i,type=("parent"))
      if(i==nroot){nodeages[i,1]<-NA}else{nodeages[i,1]<-nodeage[anc_node]}
 }

#select random ages from each transition on each branch and find averages for 100
 random iterations
 colavagechanges<-matrix(,nrow=nrow(colchanges),ncol=100)
 for (k in 1:100){
      this_colagechanges<-matrix(,nrow=nrow(colchanges),ncol=ncol(colchanges))
      for (i in 1:nrow(colchanges)){
            for (j in 1:ncol(colchanges)){
                  if(j != nroot){
                        if(colchanges[i,j]==1){
                        this_colagechanges[i,j]<-
 runif(1,nodeages[j,2],nodeages[j,1])
                        }
                  }
            }
            colavagechanges[i,k]<-mean(this_colagechanges[i,],na.rm=TRUE)
      }
 }

 for (i in 1:nrow(colchanges)){
      colavavagechanges[i,l]<-mean(colavagechanges[i,])
 }
 for (i in 1:nrow(colavavagechanges)){
 colavavavagechanges[i] <- mean(colavavagechanges[i,], na.rm=TRUE)
```

```
}
}
row.names(colavavagechanges)<-row.names(colchanges)
names(colavavavagechanges)<-row.names(colchanges)
```

2.3. Fitting Uncorrelated and Correlated Models to Datasets

R code fitting models of trait evolution to pairs of binary characters, first described in Chapter 2. “ER” in the below code is changed to “SYM” or “ARD” as necessary.

```
library(corHMM)
library(doParallel)
library(foreach)

#set rownames
######################
samp2 <- morph_matrix[,-1]
rownames(samp2) <- morph_matrix[,1]
morph_matrix <- samp2

#make blank results matrix
######################
colnames <-  names(morph_matrix)
result_matrix<-
matrix(0,ncol(morph_matrix),ncol(morph_matrix),dimnames=list(colnames, colnames))

#for row i, for each column (i:j) do corDISC
######################
temp_result <- 0
for (i in 1:nrow(result_matrix)){
par_result <- foreach(j=i:ncol(morph_matrix)) %dopar% {
temp_result <- corDISC(morph_tree1, data.frame(rownames(morph_matrix),
morph_matrix[i], morph_matrix[j]), ntraits=2, rate.mat=NULL, model=("ER"),
node.states=("joint"), lewis.asc.bias=FALSE, p=NULL, root.p=NULL, ip=1, lb=0,
ub=1000, diagn=FALSE)
temp_result$AICc}
par_result <- (unlist(par_result))
for(k in i:ncol(morph_matrix)){
result_matrix[i,k] <- par_result[(k - (i-1))]}}
```