*Biological Journal of the Linnean Society*, 2020, **131**, 231–273. With 15 figures.

## INVITED REVIEW

# An effect size statistical framework for investigating sexual dimorphism in non-avian dinosaurs and other extinct taxa

EVAN T. SAITTA[1*], MAXIMILIAN T. STOCKDALE[2], NICHOLAS R. LONGRICH[3], VINCENT BONHOMME[4], MICHAEL J. BENTON[5], INNES C. CUTHILL[6], and PETER J. MAKOVICKY[7]

[1]*Life Sciences Section, Integrative Research Center, Field Museum of Natural History, Chicago, IL 60605, USA*
[2]*School of Geographical Sciences, University of Bristol, Bristol, BS8 1RL, UK*
[3]*Department of Biology and Biochemistry and Milner Centre for Evolution, University of Bath, Bath, BA2 7AY, UK*
[4]*Institut des sciences de l'évolution, Université de Montpellier, Montpellier 5, F-34095, France*
[5]*School of Earth Sciences, University of Bristol, Bristol, BS8 1RJ, UK*
[6]*School of Biological Sciences, University of Bristol, Bristol, BS8 1TH, UK*
[7]*Department of Earth and Environmental Sciences, University of Minnesota, Minneapolis, MN 55455, USA*

*Received 23 April 2020; revised 5 June 2020; accepted for publication 10 June 2020*

Despite reports of sexual dimorphism in extinct taxa, such claims in non-avian dinosaurs have been rare over the last decade and have often been criticized. Since dimorphism is widespread in sexually reproducing organisms today, under-reporting in the literature might suggest either methodological shortcomings or that this diverse group exhibited highly unusual reproductive biology. Univariate significance testing, especially for bimodality, is ineffective and prone to false negatives. Species recognition and mutual sexual selection hypotheses, therefore, may not be required to explain supposed absence of sexual dimorphism across the grade (a type II error). Instead, multiple lines of evidence support sexual selection and variation of structures consistent with secondary sexual characteristics, strongly suggesting sexual dimorphism in non-avian dinosaurs. We propose a framework for studying sexual dimorphism in fossils, focusing on likely secondary sexual traits and testing against all alternate hypotheses for variation in them using multiple lines of evidence. We use effect size statistics appropriate for low sample sizes, rather than significance testing, to analyse potential divergence of growth curves in traits and constrain estimates for dimorphism magnitude. In many cases, estimates of sexual variation can be reasonably accurate, and further developments in methods to improve sex assignments and account for intrasexual variation (e.g. mixture modelling) will improve accuracy. It is better to compare estimates for the magnitude of and support for dimorphism between datasets than to dichotomously reject or fail to reject monomorphism in a single species, enabling the study of sexual selection across phylogenies and time. We defend our approach with simulated and empirical data, including dinosaur data, showing that even simple approaches can yield fairly accurate estimates of sexual variation in many cases, allowing for comparison of species with high and low support for sexual variation.

ADDITIONAL KEYWORDS: dinosaurs – effect size statistics – fossils – sexual dimorphism – sexual selection.

---

*Corresponding author. E-mail: evansaitta@gmail.com

## INTRODUCTION

When Charles Darwin introduced the concept of sexual selection as a variant of natural selection (Darwin, 1871), it began a line of scientific investigation that has grown into one of the most important theories of biology (Gould & Gould, 1989). Sexual selection describes the drivers of adaptations for reproductive competition, through mate attraction or intrasexual competition, as opposed to an individual's survival in the ecological context of factors such as resource competition, predation or physical stress (Andersson, 1994). Often, these drivers appear to act against those involved in viability selection, producing novel anatomies, physiologies and behaviours that might increase reproductive success at the expense of the individual's survival (Endler, 1988). Features thought to evolve in response to sexual selection include ornamental display structures such as the long tail feathers of peacocks (*Pavo cristatus*) (Petrie *et al.*, 1991), the songs of humpback whales (*Megaptera novaeangliae*) (Smith *et al.*, 2008) and songbirds (Passeri) (Nowicki *et al.*, 1998), and weapons (i.e. armaments) such as the antlers of deer (Cervidae) (Vanpé *et al.*, 2007) and the spurs of roosters (*Gallus gallus*) (Rico-Guevara & Hurme, 2019). The traits produced by sexual selection are secondary sexual characteristics, as opposed to the primary sexual characteristics of the reproductive anatomy itself (i.e. sex organs). Sexual selection frequently consists of two main types: competition between individuals of one sex for mates and preferences in mate choice. Often, these types are expressed as male-male competition and female mate choice (Andersson & Simmons, 2006). However, some species show the reverse of these patterns, a condition commonly known as sex-role reversal (Barlow, 2005). Others engage in a more mutual form of courtship, such as tufted puffins (*Fratercula cirrhata*) (Blackburn, 2004), little blue penguins (*Eudyptula minor*) (Waas, 1988), and white-fronted Amazon parrots (*Amazona albifrons*) (Skeate, 1984), possibly related to intensive biparental care or monogamy (Szekely *et al.*, 2000).

In many species, both intrasexual competition and intersexual mate choice can occur and to varying degrees (Hunt *et al.*, 2009), which can make studying sexual selection in the fossil record all the more challenging. For example, male satin bowerbirds (*Ptilonorhynchus violaceus*) compete with each other by stealing feathers from rivals' bowers that they use to display to females, and the females then subsequently selectively choose males with which to mate based on the quality of their bowers (Borgia & Gore, 1986). Hidden mate preferences can even exist in species whose mating systems do not in practice allow for that preference to be expressed through mate choice; for example, female mosquitofish (*Gambusia holbrooki*), a species in which males 'scramble' for mates without female choice, show mate preferences for males with exaggerated traits under experimental conditions (Gould *et al.*, 1999). Therefore, for any given species, it is imperative to consider the selective pressure within and between each sex, rather than simplifying descriptions of mating systems to entirely male-male competition or female-mate choice (Clutton-Brock, 2007). It is now understood that sexual selection is not only an important evolutionary driver, but that its effect is widespread among organisms and can be complex and nuanced.

One of the most frequent manifestations of sexual selection is sexual dimorphism. Sexually dimorphic organisms exhibit differences between the sexes in the distributions of certain anatomical, physiological or behavioural traits (Lande, 1980). Dimorphic traits influenced by sexual selection can include some of the most elaborate products of evolution: highly complex behaviours such as bird songs (Catchpole, 1987), colours such as 'super black' light-absorbing feathers in some birds of paradise (Paradisaeidae) (McCoy *et al.*, 2018), and exaggerated anatomical structures such as the tusks of elephants (Elephantidae) (Chelliah & Sukumar, 2013) or enlarged mandibles of stag beetles (*Cyclommatus metallifer*) (Goyens *et al.*, 2015). One of the most common forms of sexual dimorphism is sexual size dimorphism, in which one sex grows to a larger size than the other on average.

Sexual dimorphism need not always be expressed as the presence vs. absence of a particular characteristic, such as externally protruding tusks of male narwhals (*Monodon monoceros*), which are normally absent on females (Gerson & Hickie, 1985). Instead, it can often be a difference in degree where one sex is underdeveloped in the trait, such as canine length in gorillas (*Gorilla*) (Schwartz & Dean, 2001). The magnitude of dimorphism can vary greatly between different species (i.e. the effect size between male and female distributions, most typically quantified as differences in measures of centrality between the distributions). For example, primates show interspecific variation in the magnitude of sexual dimorphism with respect to body mass and canine tooth length (Clutton-Brock *et al.*, 1977; Harvey *et al.*, 1978). Statistically, a truly monomorphic species (i.e. difference between the male and female distributions is precisely zero) is not expected in finite populations of empirical data (Nakagawa & Cuthill, 2007), irrespective of the strength of sexual selection acting on that population. For example, if the measured heights of a group of two people are 1.8 m and 1.6 m and the heights of a second group of two people are 1.8

m and 1.6 m, we would calculate the difference in their average height to be exactly zero. However, few would predict that this difference would remain precisely zero if this empirical data were recorded to the nearest nanometre. Therefore, it is important to remember that terms like dimorphic and monomorphic are often used subjectively to indicate whether a species shows relatively high or low sexual variation, respectively. It is better to think simply in terms of the magnitude of sexual variation, without forcing species into binary descriptive categories of monomorphic or dimorphic (and this informs our use of effect size statistics below).

Thanks to the potential conspicuousness, complexity and variability of many sexual dimorphisms, as well as the possibility of testing functional hypotheses and their relation to underlying selective pressures in extant organisms, sexual dimorphisms are a major topic of research and an important quantifiable proxy for sexual selection (Fairbairn *et al.*, 2007). Sexual variation can also appear in traits capable of fossilizing in some environments, including certain soft tissues (Parry *et al.*, 2018) (e.g. biomineralized or pigmented anatomy).

However, sexual selection can act in ways that do not always produce visible variation between the sexes, such as sperm competition (Parker, 1970; Birkhead & Møller, 1998). Furthermore, factors can work to counter sexual selection and reduce dimorphism, such as female bovids sporting horns when under predation pressure or intrasexual competition over resources (Packer, 1983; Caro *et al.*, 2003; Robinson & Kruuk, 2007), predation risk countering sexual selection in the coloration of male poeciliid fish (Endler, 1984), or male lions (*Panthera leo*) reducing their mane thickness in warmer climates (West & Packer, 2002). A good example of confounding between sexual functions and secondary functions/biological trade-offs of sexually selected traits is the dichromatism of the polygynandrous eclectus parrot (*Eclectus roratus*); male coloration is a trade-off between conspicuous sexual display and camouflage during foraging, whereas female coloration is driven by competition for nest hollows, without any opposing need for camouflage (Heinsohn *et al.*, 2005). Finally, some sexual dimorphisms are difficult to study because of human limitations. For example, some birds once considered monomorphic in colour based on trichromatic human vision are actually dimorphic when studied with spectroscopic techniques that reveal ultraviolet colour variation, which is detectable by tetrachromatic avian vision (Burkhardt, 1989; Hunt *et al.*, 1998; Santos *et al.*, 2006).

Here, we (1) highlight the under-reporting of sexual dimorphism in non-avian dinosaurs compared to other extinct taxa over the last decade, which reflects the current debate as to whether this group exhibited unusual social/sexual biology or if methodological shortcomings are at play. (2) We then discuss arguments against sexual selection or dimorphism in non-avian dinosaurs, showing that hypotheses explaining a supposed lack of dimorphism throughout the grade can be flawed based on current evolutionary theory. (3) Furthermore, these explanatory hypotheses are likely unnecessary, as we show in our summary of the abundant evidence for sexual selection and probable sexual variation in non-avian dinosaurs. These early sections provide evidence from evolutionary/game theory, modern dimorphisms/extant phylogenetic brackets, and fossils to show that some degree of sexual variation in anisogamic populations, including non-avian dinosaurs, is the expectation, not the exception. In this context, the most appropriate methodologies and statistical approaches can be selected. (4) We then show that significance testing methods used to argue that dimorphism in non-avian dinosaurs lacks evidence are highly prone to type II error. (5) Finally, we present our framework to study sexual dimorphism in extinct taxa that utilizes effect size statistics and controls for alternate hypotheses for observed variation.

## NON-AVIAN DINOSAURS: UNIQUE BIOLOGY OR METHODOLOGICAL SHORTCOMINGS?

Detecting sexual dimorphism in the fossil record is complicated by difficulty in distinguishing sexual variation from ontogenetic variation, interspecific variation, and relatively continuous intra-population variation or polymorphisms unrelated to sex (Brusatte, 2012). Furthermore, certain characteristics can also show intra-individual variation, such as contour vs. flight feathers (Lucas & Stettenheim, 1972) or anterior vs. posterior osteoderms (Gilmore, 1914; Carpenter, 1998). Therefore, when fossil specimens are incompletely preserved, within-body variation might be confused for sexual variation. When studying fossils, taphonomic effects must also be considered, such as plastic deformation or partial preservation as a result of scavenging, transport, decay, diagenesis, weathering or erosion (Parry *et al.*, 2018). Some characteristics that are often sexually selected or dimorphic have limited or no fossilization potential, such as various soft tissues or mating behaviours. Sample sizes of many fossil species are often small as well as geographically and stratigraphically dispersed. For most fossil specimens, it is impossible to assign a sex with certainty, except in exceptional cases such as specimens with eggs (Sato *et al.*, 2005) or embryos (Caldwell & Lee, 2001) preserved *in situ* in the body cavity, claspers in chondrichthyans (Maisey, 2009), bacula (Abella *et al.*, 2013) or reproductive medullary bone (Lee & Werning, 2008).

Without destructive sampling for bone histology, it can sometimes be challenging to infer developmental maturity among specimens exhibiting potentially dimorphic traits. Studying sexual selection in fossils is further limited because behavioural observations generally cannot be made, except for minor inferences from trace fossils or pathologies, for example, and behavioural experimentation is entirely precluded (Hone & Faulkes, 2014). Furthermore, it is difficult to hypothesize the function of a candidate secondary sexual characteristic or the behaviour of an extinct species without close extant relatives or without obvious modern analogues based on ecology, overall body plan or similar anatomical traits. Many fossils have unusual structures not quite like those of any extant species, such as stegosaur plates.

Despite these challenges, there have been many proposed sexual dimorphisms in extinct species, along with discussions of sexual selection in extinct organisms more generally (Knell *et al.*, 2013a). Some of these extinct species are recent with comparable extant relatives and analogues, or are known from many fossil specimens (e.g. invertebrates). Sexual dimorphism in fossil ostracods has even been used to test hypotheses regarding the relation between sexual selection and extinction risk (Martins *et al.*, 2018). However, some examples are relatively ancient, such as ammonoids (Neige *et al.*, 1997), and unique, such as trilobites (Cederström *et al.*, 2011). Vertebrate examples include fossil hominids (Reno *et al.*, 2003) and other primates (Krishtalka *et al.*, 1990), proboscidians (Smith & Fisher, 2011), perissodactyls (Gingerich, 1981), artiodactyls (Sánchez *et al.*, 2010), pinnipeds (Cullen *et al.*, 2014), felids (Meachen-Samuels & Binder, 2010), dicynodonts (Sullivan *et al.*, 2003), pterosaurs (Wang *et al.*, 2014), birds (Chinsamy *et al.*, 2013), phytosaurs (Zeigler *et al.*, 2002), basal archosauromorphs (Sengupta *et al.*, 2017), ichthyosauriforms (Motani *et al.*, 2018), pachypleurosaurs (Cheng *et al.*, 2009), and chondrichthyans (Lund, 1982), among others. The evolution and function of certain putative secondary sexual characteristics in extinct taxa, such as the antlers of male 'Irish elk' (*Megaloceros giganteus*), have historically been heavily discussed (Gould, 1974; Kitchener, 1987; Lemaître *et al.*, 2014). Reversed dimorphism in moa (*Dinornis*), by definition an extinct dinosaur, is thought to have been so extreme that the sexes were previously considered to be different species (Bunce *et al.*, 2003). Many published claims have not been challenged in the literature or broader media, presumably due to ubiquity of sexual variation in vertebrates.

Noticeable exceptions, however, are non-avian dinosaurs, for which claims of sexual dimorphism have recently been highly debated and criticized (Padian & Horner, 2011; Hone *et al.*, 2012; Hone & Mallon, 2017; Mallon, 2017). Hone *et al.* (2020) state, "To date, no dinosaur has been determined to exhibit sexual dimorphism under rigorous analysis" (p. 13). An examination of over a decade of recent abstracts from the annual meeting of the Society of Vertebrate Paleontology reveals that abstracts proposing or concluding sexual variation in non-avian dinosaurs are highly under-represented compared with those for all other fossil taxa, in relation to the prevalence of abstracts on non-avian dinosaurs at the meeting ($\chi^2$ goodness of fit test on summed counts over an 11-year period of non-avian dinosaur dimorphism abstracts compared with dimorphism abstracts of all other taxa using the online tool from vassarstats.net: degrees of freedom = 1; expected count = 34, observed count = 15, unadjusted $\chi^2 = 13.11$, $P$-value = 0.0004 for potential dimorphism; expected count = 24, observed count = 2, unadjusted $\chi^2 = 24.96$, $P$-value = < 0.0001 for concluded dimorphism) (Table 1). Does this under-reporting reflect a highly unusual social/sexual system in non-avian dinosaurs or differences in how dinosaur researchers interpret fossil data compared to other palaeontologists? Is the debate around non-avian dinosaur dimorphism a case of biology, or does it stem from methodological shortcomings and/or preconceived notions about a lack of dimorphism?

## FLAWED ALTERNATIVES TO SEXUAL SELECTION AND DIMORPHISM IN NON-AVIAN DINOSAURS

Claims of sexual dimorphism in non-avian dinosaurs have varied in sample size, methodology, and whether or not they were approached quantitatively, with some studies criticized for using relatively little data or analysis (Chapman *et al.*, 1997; Mallon, 2017). Published reports include proposals of dimorphism in *Tyrannosaurus rex* (Larson, 1994, 2008), *Coelophysis bauri* (Rinehart *et al.*, 2009), *Coelophysis* (= *Syntarsus*) *rhodesiensis* (Raath, 1990), *Kentrosaurus aethiopicus* (Barden & Maidment, 2011), *Plateosaurus* (Weishampel & Chapman, 1990), *Stegoceras validum* (Chapman *et al.*, 1981), *Protoceratops andrewsi* (Dodson, 1976), *Allosaurus fragilis* (Smith, 1998), *Citipati osmolskae* [notably with a sample size of only two (Persons *et al.*, 2015)], and *Hesperosaurus* (= *Stegosaurus*) *mjosi* (Saitta, 2015), among others [see Table 1 of Mallon (2017) for more examples]. Beyond morphological dimorphisms, behavioural dimorphisms that might indirectly result from sexual selection in troodontids and oviraptorids have also been hypothesized in the form of unequal parental care, specifically paternal care [Varricchio *et al.* (2008); although see Birchard

**Table 1.** Annual meeting of the Society of Vertebrate Paleontology abstracts from 2008 to 2018, excluding Preparators' Session, technical, education, outreach and related abstracts. Observed counts are abstracts at least proposing possible sexual variation of anatomical traits that are not primary sexual characteristics. Expected counts calculated based on the percentage of all abstracts focusing on non-avian dinosaurs among the total number of abstracts in that year's meeting, rounded to the nearest whole number. Percentage of conclusive abstracts is the percentage of observed abstracts that more conclusively propose sexual variation as the explanatory hypothesis. Total values over this 11-year period subjected to $\chi^2$ goodness of fit tests. Each abstract considered an independent observation regardless of repeated authorship

| Year | Abstracts focusing on non-avian dinosaurs | Total abstracts | Abstracts implying anatomical secondary sexual variation as at least a possibility in fossils | | | | Abstracts more conclusively implying anatomical secondary sexual variation in fossils | | | | % conclusive abstracts | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Expected count other taxa | Expected count non-avian dinosaurs | Observed other taxa | Observed non-avian dinosaurs | Expected count other taxa | Expected count non-avian dinosaurs | Observed other taxa | Observed non-avian dinosaurs | Other taxa | Non-avian dinosaurs |
| 2008 | 113 | 601 | 9 | 2 | 10 | 1 | 6 | 1 | 6 | 1 | 60 | 100 |
| 2009 | 133 | 750 | 12 | 2 | 14 | 0 | 7 | 2 | 9 | 0 | 64 | NA |
| 2010 | 126 | 695 | 13 | 3 | 13 | 3 | 8 | 2 | 9 | 1 | 69 | 33 |
| 2011 | 136 | 774 | 18 | 4 | 19 | 3 | 12 | 2 | 14 | 0 | 74 | 0 |
| 2012 | 127 | 628 | 10 | 2 | 11 | 1 | 6 | 2 | 8 | 0 | 73 | 0 |
| 2013 | 151 | 777 | 15 | 4 | 17 | 2 | 13 | 3 | 16 | 0 | 94 | 0 |
| 2014 | 161 | 889 | 16 | 4 | 20 | 0 | 11 | 2 | 13 | 0 | 65 | NA |
| 2015 | 145 | 774 | 7 | 1 | 8 | 0 | 3 | 1 | 4 | 0 | 50 | NA |
| 2016 | 156 | 779 | 24 | 6 | 28 | 2 | 18 | 5 | 23 | 0 | 82 | 0 |
| 2017 | 136 | 649 | 11 | 3 | 13 | 1 | 9 | 2 | 11 | 0 | 85 | 0 |
| 2018 | 161 | 744 | 10 | 3 | 11 | 2 | 8 | 2 | 10 | 0 | 91 | 0 |
| Total | 1545 | 8060 | 145 | 34 | 164 | 15 | 101 | 24 | 123 | 2 | 75 | 13 |

*et al*. (2013) for a counterarguement]. Recently, claims of sexual dimorphism in non-avian dinosaurs, or at least demonstrable evidence for it, have been rejected by some (e.g. Mallon, 2017). The postulated absence of sexual dimorphism (or absence of evidence for it) in non-avian dinosaurs has been explained in various ways: an artifact resulting from limited sample size, taphonomic information loss, methodological shortcomings, or, especially regarding 'exaggerated'/'bizarre' traits, as at least partly a result of one of two other biological phenomena termed the species recognition hypothesis and the mutual sexual selection hypothesis.

### *Species recognition hypothesis*

Signals can function to discriminate con- from heterospecific individuals in both sexual and non-sexual contexts, such as flocking/shoaling/herding to reduce predation risk (Krause & Ruxton, 2002). In the latter context, benefits are likely similar for both sexes, so no dimorphism evolves. Although mixed-species groups certainly form, similarity of morphology and behaviour may favour preferential association with one's own species, there being advantages when it comes to group cohesion and escape from predators that seek to separate a prey individual from a group (Croft *et al*., 2009). A signal evolved to facilitate same-species aggregation could reasonably be described as a trait for species recognition. However, there is no selection driving exaggeration of the trait beyond the minimum for successful detection, so the expectation is that such traits would be relatively low-cost 'road signs' rather than the costly 'advertisements' produced through sexual selection (Maynard Smith & Harper, 2003; Knell & Sampson, 2011).

The other context in which species recognition is invoked is in mate choice. Padian & Horner (2011, 2013, 2014) contrast this with sexual selection and have been criticized for doing so (Knell & Sampson, 2011; Hone & Naish, 2013; Knell *et al*., 2013b; Knapp *et al*., 2018). The two concepts are not readily separated (Paterson, 1980, 1985; Ryan & Rand, 1993). When considering pre-zygotic reproductive barriers, mating with the wrong species is simply an extreme form of sub-optimal mate choice and, because the marginal cost of sperm production is usually lower than that of eggs, selection for mating with the right species will often be higher in females than males. Thus, species recognition for mate choice predicts sexual dimorphism, or lack thereof, in a similar fashion to sexual selection. It is also possible that signals used for species/mate recognition might be exaggerated in order to increase an individual's appeal to the opposite sex [e.g. 'supernormal stimuli' or 'sensory exploitation' hypotheses (Tinbergen, 1948; Ryan & Keddy-Hector, 1992)]. For species recognition, as related to mate choice, to be separable from sexual selection, the fitness of consequences of mating with different individuals of one's own species would have to be identical. Given the abundance of examples of discriminating mate choice in extant birds, this scenario is unlikely in non-avian dinosaurs (Hone & Naish, 2013).

One reason why Padian & Horner (2011, 2013, 2014) feel that species recognition for mate choice is readily separable from sexual selection is that they adopt a non-standard definition of what constitutes a sexually selected character. For them, only discrete anatomical traits, such as horns, that are present in one sex and not the other count as sexually selected. This is inconsistent with current sexual selection theory, and indeed Darwin's own writings (Knell & Sampson, 2011; Hone & Naish, 2013; Knell *et al*., 2013b; Mendelson & Shaw, 2013; Borkovic & Russell, 2014). Padian & Horner (2014) claim, "The term "sexual selection" should only be used when one sex uses a feature not present in the other sex to attract mates or repel rivals for mates" (p. 709). If the identification of a sexually selected character requires an extreme dimorphism in the form of binary presence vs. absence, then numerous cases of sexual dimorphism seen in modern organisms, expressed as differences in degree, would be rejected. For example, one of the most commonly studied sexually selected traits is body size, which is continuously variable and thus would be excluded under this unreasonably stringent paradigm. Their argument for non-sexual species recognition, therefore, depends on negative evidence, namely that there be no clear examples of presence vs. absence sexual dimorphisms in non-avian dinosaurs.

There are plenty of examples of species recognition in the context of mate choice that do not require the sorts of exaggerated anatomical structures that are the focus of Padian & Horner (2011, 2013, 2014) hypothesis. Detailed observations of breeding pedigrees show that some organisms have little difficulty in identifying conspecifics or members of the same newly speciating hybrid lineage and that sufficient character displacement and reproductive isolation can occur rapidly, such as in Galapagos finches (*Geospiza*) which could be argued to lack many of the sorts of conspicuous morphological structures of other species (but do have songs acquired through imprinting) (Grant & Grant, 2008, 2010; Lamichhaney *et al*., 2018). Even domestic dogs (*Canis familiaris*), one of the most morphologically diverse species in external appearance, can identify conspecifics from sight alone (Autier-Dérian *et al*., 2013) despite being highly olfactory in their social signalling. These dog experiments reiterate that species recognition does not require unique exaggerated structures that sexual selection theory predicts and that Padian & Horner (2011, 2013, 2014) instead propose ought to strictly be markers of species recognition. This is especially true when potentially less energetically costly, physiological/behavioural alternatives for species recognition exist (rather than novel, exaggerated anatomical traits) or simply alternatives that might not be apparent in fossils (Hone & Naish, 2013), such as pheromones of moths (Lepidoptera) (Löfstedt, 1993), courtship displays of fireflies (*Photinus*) that differ in flash pattern between species (Lewis *et al*., 2004), or species-specific bird songs (Emlen, 1972; Nelson, 1989; Seddon, 2005).

Sexual dimorphism in anatomical traits is not even required for sexual selection to operate. For example, dimorphisms can be behavioural (Nottebohm & Arnold, 1976). Sexual selection can also operate under no dimorphism of secondary sexual characteristics at all, such as sperm competition (Parker, 1970; Birkhead & Møller, 1998).

### *Mutual sexual selection hypothesis*

Another hypothesis proposed for non-avian dinosaurs is far better founded in mechanisms of current sexual selection theory than the species recognition hypothesis. The mutual sexual selection hypothesis was originally proposed in light of an apparent lack of extinct archosaurian cranial crests exhibiting a presence vs. absence pattern of expression (Hone *et al*., 2012). Accordingly, the purported lack of dimorphism in non-avian dinosaurs could be due to mutual sexual selection whereby males and females show equal preference for the same trait when choosing mates (or traits used in intrasexual mating competition are equally important to both sexes), resulting in minimal to no difference in the trait distribution between the sexes.

We do not imply here that various degrees of mutual mate choice or intrasexual competition cannot occur in both sexes of a species. Instead, we discuss a scenario in which mutual sexual selection minimizes sexual variation in a particular trait such that the species might appear to be monomorphic—for example, going beyond a case where both sexes possess an ornament, but where it is expressed to a similar extent in both sexes. Furthermore, this is not to say that previous authors (e.g. Hone *et al*., 2012) were attempting to propose mutual sexual selection as a 'blanket hypothesis' to be applied to all non-avian dinosaurs, to the exclusion of any dimorphism. However, in order to justify our statistical approach below, we are required to show why mutual sexual selection resulting in minimal to no dimorphism could not be proposed as an alternative to traditional patterns of sexual variation widely across non-avian dinosaurs, as was attempted with the species recognition hypothesis.

Although the mutual sexual selection hypothesis could explain a supposed lack of sexual dimorphism in non-avian dinosaurs, assuming a lack of dimorphism throughout the grade may be flawed. Abandoning this assumption would then make the mutual sexual selection hypothesis, at least in many cases, unnecessary to invoke. Furthermore, the hypothesis assumes that the effect of any intrasexual competition is nullified by subsequent mate choice processes and that the combined influence of intra- and intersexual selection is equal between the two sexes—a big assumption to apply across all non-avian dinosaurs. A monomorphic equilibrium produced through this sort of mutual sexual selection is likely a rare social/sexual system in modern species (thoroughly demonstrated in a few species [e.g. Jones & Hunter, 1993; Kraaijeveld *et al*., 2004; Nolan *et al*., 2010]), further compounded by the fact that social/sexual systems can evolve rapidly (Liker *et al*., 2013). For example, although Kraaijeveld *et al*. (2007) summarize experiments on 14 bird species with these sorts of mutual ornaments that tested if the ornaments are involved in mate choice (see Table 3 therein), these examples may be far exceeded by the number of bird species that lack this precise type of mutual sexual system, given a modern bird diversity of 11 000–18 000 species (Barrowclough *et al*., 2016). Indeed, some birds are model organisms for sexual dimorphism [e.g. birds of paradise or peafowl, grouse, and pheasants (Phasianidae)] as well as reversed dimorphism [e.g. jacanas (Jacanidae) (Emlen & Wrege, 2004)], and it is common knowledge among birders that many species have dimorphic plumage (Sibley, 2014). Although this topic is understudied (Lihoreau *et al*., 2008), the prevalence of such well-balanced mutual sexual selection is unknown, with no indication that it might be as or more common than 'classic' sexual systems (Fig. 1). Given the plausible relative scarcity of well-balanced mutual sexual selection among living species, it is unlikely that a highly diverse and disparate grade of animals with at least ~165 million years of evolutionary history and global biogeographic range showed stasis in a social or sexual system, considering the varied and frequent forms of anatomical dimorphism in their living descendants.

Another challenge for this hypothesis is that it is strongly helped by *in vivo* experiments/observations of behaviour and mate preferences between the two sexes, which is not possible for extinct species. Such experiments can rule out alternative functions for female ornaments/armaments. Selection pressures for female ornaments/armaments other than females competing for mates or mate choice by the males can therefore be tested (vice versa under sex-role reversal). Alternative functions of these female structures could include defence against predators or competition with other females for resources other than mates (Stankowich & Caro, 2009; Hone *et al*., 2012; Tobias *et al*., 2012). Additionally, genetic correlation (Kraaijeveld *et al*., 2007) can result in females possessing alleles for a trait that is only actively selected for in males: females produce female offspring with the males whose trait they prefer. For example, females that prefer larger males as mates might produce large daughters when they mate with those males, even if males show no preference for larger females as mates. If expressing those traits is costly to females, expression can be sex-limited (Rice & Chippindale, 2001; Parker, 2006); but if the costs are low, a trait that is only adaptive for males can be expressed in females.

## EVIDENCE FOR SEXUAL SELECTION AND SEXUAL VARIATION IN NON-AVIAN DINOSAURS

Under a game theoretic framework, unequal initial investment into reproduction between the sexes is expected to yield different optimal strategies. Therefore, one of the underlying principles of sexual selection theory is that anisogamy tends to result in behavioural, physiological and anatomical sexual dimorphism (Schärer *et al.*, 2012). Furthermore, statistically, any finite population of males and females is expected to show some non-zero difference between the distributions of the sexes (Nakagawa & Cuthill, 2007), regardless of whether sexual selection is acting or not.

Non-avian dinosaurs should have exhibited sexual variation according to extant phylogenetic bracketing. Both birds and crocodilians are anisogamic and exhibit various types of sexual dimorphism, including in body size and behaviours (Owens & Hartley, 1998; Platt *et al.*, 2009). Even when only examining dimorphism in body masses of extant birds (Dunning, 2007), without taking into account the prevalent dimorphism in integumentary structures, coloration and behaviour, disruptive selection against monomorphic body mass at the macroevolutionary scale is evident (Fig. 1). The $\log_{10}$-transformed distribution of male:female mass is bimodal with peaks corresponding to dimorphism magnitudes approximately +/–10% off from parity between the sexes (i.e. a peak at males ~90% the size of females and a far larger peak at males ~110% the size of females). This serves as a reminder that many sexual dimorphisms are of subtle magnitude. Sexual selection affects rates and directions of phenotypic evolution at the macroevolutionary scale in some birds (Cooney *et al.*, 2019). Many crocodilians are known to exhibit size dimorphism, with males growing faster than females to attain larger adult sizes (Wilkinson & Rhodes, 1997; Hone & Mallon, 2017), in addition to behavioural dimorphism, such as male 'water dance' displays in American alligators (*Alligator mississippiensis*) (Vliet, 1989; Moriarty & Holt, 2011).

As an aside, while a non-trivial fraction of bird body mass dimorphism conforms to a reversed pattern with larger females (negative values on the histogram), most birds follow the widespread trend among amniotes toward larger males (Fig. 1). Further, a strong positive effect of sexual selection on male, but not female, rates of interspecific divergence of plumage colour has been

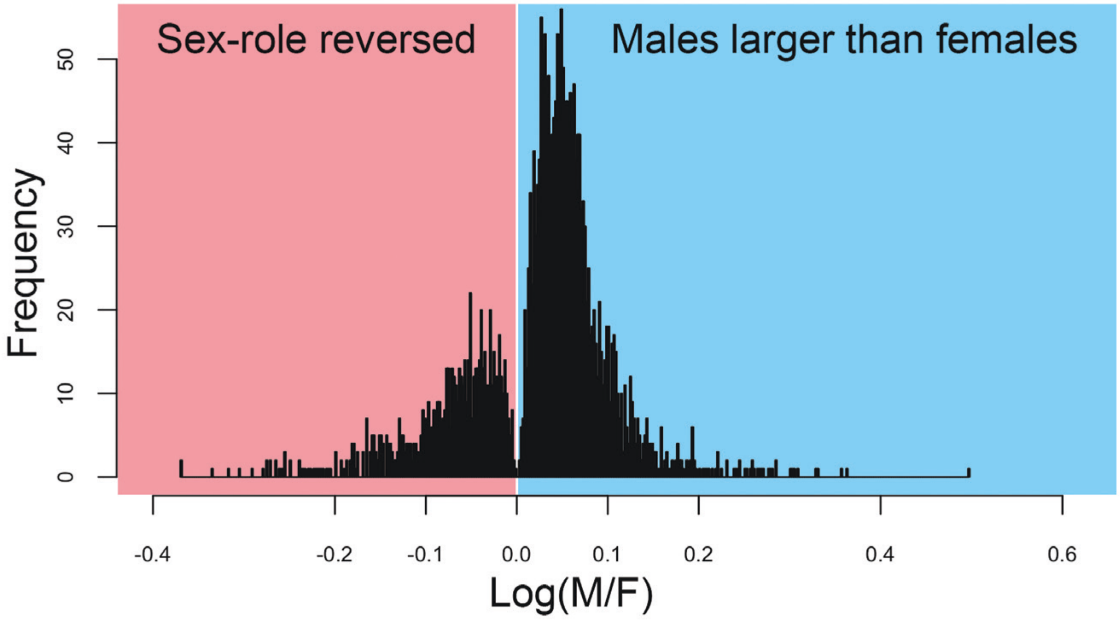

**Figure 1.** Extant bird body mass sexual dimorphism. Data from Dunning (2007) and presented as the $\log_{10}$-transformed average male body mass divided by average female body mass (M/F). Sample size = 2576 taxa. Sample sizes and data quality vary widely between taxa. All entries with both male and female values in Dunning (2007) were included.

demonstrated in Tyrannida (Cooney *et al*., 2019). These points weaken the hypothesis that larger, more robust specimens of certain non-avian theropods represent females (Carpenter, 1990; Raath, 1990; Larson, 1994; Chapman *et al*.,1997). Those hypotheses were based on patterns of body size dimorphism in extant birds of prey. Falconiformes were hypothesized to show reversed size dimorphism due to biomechanical compensation for increases in wing loading during gestation in a clade whose hunting is dependent on flight performance (Wheeler & Greenwood, 1983). Such an analogy is likely inappropriate for flightless theropods like *Coelophysis* or *Tyrannosaurus*.

Non-avian dinosaurs were diverse and morphologically disparate (Barrett *et al*., 2009; Brusatte *et al*., 2012) through their global geographic distribution and duration of at least ~165 million years. Many lineages of non-avian dinosaurs possessed elaborate or exaggerated structures (e.g. horns, frills in ceratopsians, plates, spikes, spur-like claws in *Iguanodon*, elongated tusk-like teeth in heterodontosaurs, clubs, casques, cranial domes, feathers, bristles, keratinous epidermal spines or sail-like hyper-elongated neural spines forming a dorsal crest). These resemble modern ornaments or armaments known to be, at least partly, under sexual selection or to exhibit sexual variation (Molnar, 1977), such as horns (Bro-Jørgensen, 2007), casques (Karsten *et al*., 2009; Taylor *et al*., 2017), cranial domes (Wilson, 2001), feathers/feather coloration (Møller & Höglund, 1991), bristles (Scott & Payne, 1934), spurs (Møller, 1992), tusks (Cabrera & Stankowich, 2018), keratinous epidermal spines (Ord & Stuart-Fox, 2006) or sail-like crests (Taylor *et al*., 2017). Many non-avian dinosaurs show high levels of intraspecific variation of these structures, and some structures developed under positive allometric growth with delayed onset (Hayashi *et al*., 2009; Hone *et al*., 2016a), which can be consistent with sexual selection (Bonduriansky, 2007). Extreme dimorphisms in non-avian dinosaurs, whereby a trait is present in one sex and absent in the other, might be at risk of being interpreted as entirely ontogenetic or interspecific, as well as obscured by incomplete fossil records. It is also worth noting that even if a structure primarily functions in sexual display/combat and whose evolution is primarily driven by sexual selection, it can still have secondary functions/effects.

Although some structures, such as ceratopsian frill epoccipitals, have been suggested to show fluctuating asymmetry (Longrich, 2010; Longrich *et al*., 2010) previously claimed to function as honest signalling of mate quality (Møller & Höglund, 1991; Møller, 1992; Grammer & Thornhill, 1994; Ditchkoff *et al*., 2001), the connection between fluctuating asymmetry and sexual selection has been doubted due to difficulty in replicating results (Balmford *et al*., 1993).

### *Likely armaments*

Beyond the commonly sexually variable trait of body size, there is good evidence (Farke, 2014) that non-avian dinosaurs had structures morphologically analogous to armaments of modern animals (i.e. they might have used the structures for combat), with biomechanical analyses suggesting the ability to use the structures as weapons (i.e. they could have used the structures for combat), and with pathologies consistent with intraspecific combat (i.e. they likely did use the structures for combat).

Ceratopsian horns developed late in ontogeny and have been hypothesized to be involved in mate competition (Sampson *et al*., 1997). *Triceratops* and *Centrosaurus* show variation in their horns that mirrors sexual variation in many modern bovids (Poissant *et al*., 2008). In these bovids, males have large, highly curved horns with wide bases whose tips point back towards the skull to allow for non-lethal sparring/head-butting and withstanding associated forces. Female bovids have smaller, thinner horns whose tips point away from the skull to allow for stabbing defence against predators (Packer, 1983; Caro *et al*., 2003). Similarly, while smaller and juvenile *Triceratops* and *Centrosaurus* specimens have thin horns that tend to point upward, larger specimens tend to have large, broad, downward-curving horns (Horner & Goodwin, 2006; Frederickson & Tumarkin-Deratzian, 2014) that might represent an anteriorly curving analogue to the posteriorly curving pattern in many bovids, despite the fact that researchers have attributed this horn variation entirely to ontogeny (or taxonomy) and not sex. Since these bony horn cores were likely covered and further extended by a keratin sheath [which can sometimes preserve as calcium phosphate (Brown *et al*., 2017; Saitta *et al*., 2018; Saitta & Vinther, 2019)], the potential for morphological variation *in vivo* is even greater than the observed skeletal variation. Morphologically complex sheathing can be present on simple horn corns of modern species, like pronghorn antelope (*Antilocapra americana*) whose single tipped horn core supports a two-pronged keratin sheath (Davis *et al*., 2011). Most extant species use their horns for intraspecific combat, even Jackson's chameleons (*Trioceros jacksonii*), which can exhibit dramatic sexual dimorphism in some subspecies (Waring, 1997). Even unusual horn morphologies are used in sparring, such as the spiral horns of greater kudus (*Tragelaphus strepsiceros*) that are at risk of becoming locked together during fights between males (Owen-Smith, 1993), the highly inwardly-curved horns of cape buffalo (*Syncerus caffer*) (Turner *et al*., 2005), or the giant horns of ankole-watusi cattle (*Bos taurus*) (Huber *et al*., 2008). Pathologies on *Triceratops* skulls are consistent with intraspecific sparring based on the geometry of where horn tips would contact a rival while sparring (Farke, 2004; Farke *et al.*, 2009). In a study of macroevolutionary and biogeographic trends in ceratopsians, Knapp *et al.* (2018) could not explain their exaggerated structures simply by non-sexual species recognition. Note that, as Knapp *et al.* (2018) are careful to qualify, treating all exaggerated structures (e.g. ceratopsian horns) as ornaments can be problematic given the importance of distinguishing armaments from ornaments in sexual selection (McCullough *et al.*, 2016).

Pachycephalosaur cranial domes are morphologically and histologically similar to extant species with domes or thickened skull roofs used for head butting, such as duiker (Cephalophinae), bighorn sheep (*Ovis canadensis*), cape buffalo (*Syncerus caffer*) or musk oxen (*Ovibos moschatus*), and the alignment between their skull and vertebral column are also similar (Galton, 1970; Sues, 1978; Alexander, 1989; Giffin, 1989; Carpenter, 1997; Snively & Theodor, 2011). They show intraspecific variation between flattened and domed crania, which has a strong ontogenetic signal (Schott *et al.*, 2011) but could also be influenced by sex—possibly an example of extreme presence vs. absence dimorphism (Galton, 1971). Biomechanical analyses of pachycephalosaur domes using finite element analysis and simple physical calculations show that they were capable of withstanding stress and strain from head-butting, especially when a plausible amount of keratin sheathing around the dome is included (Alexander, 2006; Snively & Cox, 2008; Snively & Theodor, 2011). Finally, pathologies on their cranial domes are consistent with injuries sustained from head-butting, as seen in extant head-butting species (Peterson & Vittore, 2012; Peterson *et al.*, 2013).

Although ankylosaur tail clubs lack obvious modern analogues, some extant species use tail whipping in intraspecific combat, such as the lizard *Agama agama* (Schall *et al.*, 1989; Arbour & Zanno, 2018). Ankylosaur tail clubs have been shown through finite element analysis to be capable of withstanding the stress and strain from use as an armament (Arbour & Snively, 2009). Although limited, some possible pathologies in anterior caudal vertebrae and tail clubs have been noted (Arbour & Currie, 2011). Stegosaur tail spikes have been suggested to function in defence against predators based on pathologies in *Allosaurus* bones (Carpenter *et al.*, 2005) and in *Stegosaurus* tail spikes themselves (McWhinney *et al.*, 2001). Like some ankylosaur specimens, anterior caudal vertebrae pathologies have been noted on several North American stegosaur specimens, including the wide-plated hypothesized male of *Hesperosaurus mjosi* (Saitta, 2014). Whether stegosaur tail spikes would have been excluded from intraspecific combat due to excessive damage and potential lethality (i.e. the 'total war' avoidance hypothesis [Maynard Smith & Price, 1973]) remains to be determined.

### *Likely ornaments*

Although ornamental function in extinct species is difficult to study given the lack of behavioural data, many non-avian dinosaur structures are consistent with display or inconsistent with mechanical usage as an armament (Hone *et al.*, 2012). Hadrosaur casques house elaborate nasal passages that could have produced low-frequency sounds (Weishampel, 1981) and exteriorly (although not in internal structure) resemble casques of modern animals that can exhibit sexual variation or be under sexual selection, such as hornbills (Bucerotidae) (Gamble, 2007), chameleons (Chamaeleonidae) (Karsten *et al.*, 2009) or casque-headed lizards (Corytophanidae) (Taylor *et al.*, 2017). A soft tissue caruncle, similar to the snoods, wattles and combs of some modern birds, has been described in *Edmontosaurus* based on preserved skin impressions (Bell *et al.*, 2014), which could represent a sexual ornament, assuming that this is not a taphonomic artefact from decay.

Sail-like, hyper-elongated, vertebral neural spines consistent with sexual ornaments (Isles, 2009) in spinosaurs (e.g. *Spinosaurus*), sauropods (e.g. *Amargasaurus*), ornithopods (e.g. *Ouranosaurus*), carcharodontosaurs (e.g. *Concavenator* or *Acrocanthosaurus*) and ceratopsians (e.g. *Leptoceratops* or *Koreaceratops*) resemble the sail-like crests and elongated neural spines of various lizards. These include casque-headed lizards like the plumed basilisk (*Basiliscus plumifrons*) (Taylor *et al.*, 2017), sailfin lizards like the Philippine sailfin lizard (*Hydrosaurus pustulatus*) (Ord & Stuart-Fox, 2006), or chamaeleons like the crested chameleon (*Trioceros cristatus*) (Klaver & Böhme, 1992), which can show sexual dimorphism in these sail-like crests. Midsagittal, dorsal, keratinous epidermal extensions along the back of hadrosaurs like *Brachylophosaurus* (Murphy *et al.*, 2006) and the spines of diplodocid sauropods (Czerkas, 1992) resemble spines of agamid lizards like the crowned forest dragon (*Lophosaurus dilophus*), which are sexually dimorphic in some agamids (Ord & Stuart-Fox, 2006).

*Psittacosaurus* tail bristles, which might not be present on all specimens, may be structurally and developmentally similar to display bristles in extant birds, like the beards of mature male turkeys (*Meleagris gallopavo*) or the bristles on the head of the Congo peafowl (*Afropavo congensis*) (Mayr *et al.*, 2016). Some feathers (e.g. head feathers, remiges or rectrices) and feather-bearing bones (e.g. pygostyles or ulnae with quill knobs) could be partly consistent with display, particularly in taxa incapable of flying or gliding (Barsbold *et al.*, 2000; Turner *et al.*, 2007; Zelenitsky *et al.*, 2012). 'Palaeocolour' reconstructions of some non-avian dinosaurs reveal iridescent feather colours consistent with social/sexual signalling (e.g. *Microraptor*), including a 'rainbow' iridescence in *Caihong* analogous to hummingbirds (Trochilidae) (Li *et al.*, 2012; Hu *et al.*, 2018). Colour and colour pattern reconstructions will likely represent a major area of sexual selection research on non-avian dinosaurs and other extinct taxa in the future (Vinther, 2015; Roy *et al.*, 2019), especially given the large sample sizes of some feathered dinosaur taxa from China.

Other traits that lack obvious modern analogues have also been hypothesized as display structures among other functions, such as stegosaur plates (Saitta, 2014, 2015), which show a delayed growth pattern compared to the rest of the skeleton (Hayashi *et al.*, 2009), or hyper-elongated ceratopsian frills, also under differing growth patterns to the rest of the skeleton (i.e. allometric growth or delayed onset) (Sampson *et al.*, 1997; Hone *et al.*, 2016a). It is possible that plates and frills share at least some functional similarities (e.g. sexual display) with sail-like dorsal crests or cranial casques, respectively.

Trace fossils have been used to suggest lekking display behaviour in theropod dinosaurs based on footprints and scratch marks (Lockley *et al.*, 2016), although scratch marks can be produced through non-lekking or non-sexual behaviours (e.g. antagonistic displays, territorial marking or searching for resources).

## PREVIOUS METHODS OF INVESTIGATION INTO SEXUAL DIMORPHISM IN FOSSILS

Methods for detecting sexual dimorphism in fossil taxa vary according to the evidence they invoke, the alternate hypotheses they test, their commitment to quantitative data and the statistical methods used. In the extreme, some authors (Padian & Horner, 2011, 2013, 2014) require binary differences (i.e. presence vs. absence) to accept dimorphism in fossils. Such a qualitative approach would preclude studying dimorphism in traits such as body size, which may be the most common type of dimorphism among animals. Here, we focus on two recent statistical investigations into sexual dimorphism. Both methods represent univariate significance tests that can be ineffective for detecting a signal of sexual dimorphism and suffer from low statistical power (i.e. a tendency to fail to detect an effect when present)—a serious concern when studying datasets with low sample sizes, as is the case with most vertebrate fossils. These approaches appear to be inconsistent with the American Statistical Association's recent statement about over-reliance on *P*-values that says, "Scientific conclusions and business or policy decisions should not be based only on whether a *P*-value passes a specific threshold" (Wasserstein & Lazar, 2016: p. 131).

### *The bimodality method*

Hartigans' dip test for unimodality has been used to test for sexual dimorphism in non-avian dinosaurs and to conclude that there is no evidence for it (Mallon, 2017). Mallon (2017) writes that "no evidence for sexual dimorphism was found in any of the examined taxa" (p. 495), although is careful to state, "This is not to say that dinosaurs were not sexually dimorphic (phylogenetic inference suggests they may well have been), only that the available evidence precludes its detection" (p. 495). Although Mallon (2017) only examined non-avian dinosaur datasets, it is unlikely that other fossil groups would have passed these tests either, because a key point regarding significance testing vs. effect size statistics remains. There is a subtle, but important, difference between a claim of 'no dimorphism' in non-avian dinosaurs and a claim of 'no evidence for dimorphism'. It is certainly true, and sometimes acknowledged, that failure to achieve a certain *P*-value is not evidence for the absence of an effect, but even to say 'no evidence' is potentially wrong. This is because such a statement can be the product of excellent data that estimates an effect to be near zero, with tight confidence intervals around that point; or it can be the result of poor data and broad confidence intervals. For the distinction to be clear, effect sizes should be estimated and uncertainty quantified. This is even more sensible when context is included. There are many lines of evidence from the theoretical to the empirical (both modern and fossil) for sexual variation in non-avian dinosaurs, regardless of the weight given to different arguments by different researchers. Our methodologies, particularly our statistical analyses, should reflect this evidence, which does not happen when a null hypothesis is set to monomorphism by convention.

Beyond the tempting conclusion of no evidence for dimorphism in non-avian dinosaurs, the bimodality method suffers from further issues. Mallon (2017) reports a method whereby the data are first tested for normality with Shapiro-Wilk and Anderson-Darling tests, followed by Hartigans' dip test for unimodality. The method cannot accommodate ontogenetic effects unless juveniles or sufficiently young individuals are excluded from the dataset [as Mallon (2017) did in some analyses]. Especially when lacking histological evidence of growth rate or other indicators of sexual maturity (e.g. medullary bone or *in situ* foetuses/eggs), excluding smaller individuals risks excluding the smaller sex of a potentially dimorphic species and also reduces sample size.

A key problem of the Mallon (2017) approach is that sexually variable traits do not always exhibit great enough effect size to produce a bimodal distribution in a single variable, irrespective of sample size. In other words, the magnitude of the dimorphism (i.e. the difference in measures of centrality between male and female distributions) relative to the intrasexual variation (i.e. the spread of the male and female distributions) might not be great enough to produce a negative curvature in the centre of the combined male and female distribution. This can be true even for the theoretical population (i.e. a hypothetically infinite sample size), let alone in a random sampling of that population. This statistical problem has been known for some time, as stated by Schilling *et al.* (2002: p. 233), "a mixture of equally weighted normal distributions with common standard deviation σ is bimodal if and only if the difference between the means of the distributions is greater than 2σ". Schilling *et al.* (2002) demonstrated this using human height as an example of an accepted sexual dimorphism whose effect size is too small to produce a bimodal distribution, even in theoretical population distributions (i.e. if one were able to measure an infinite number of men and women).

As acknowledged by Mallon (2017), the significance testing used (Hartigans' dip test) is highly susceptible to false negatives, or type II errors, and therefore low statistical power. Dip tests are thus an ineffective method for identifying sexual dimorphism in an extinct species. Just how large might the magnitude of dimorphism have to be in order to pass the Hartigans' dip test? Inspired by the illustrative example of Schilling *et al.* (2002), we have statistically modelled the heights of adult men and women and subjected them to Hartigans' dip test (Fig. 2), since human height is a commonly accepted and familiar example of sexual dimorphism. We randomly generated data for women using a normal distribution whose height was fixed at an average ($\mu$) of 162 cm with a standard deviation ($\sigma$) of 7 cm, a reasonable approximation of observed distributions of women's heights (Schilling *et al.*, 2002). We then randomly generated data for men's height using a normal distribution whose average ($\mu$) was allowed to vary, while keeping the standard deviation ($\sigma$) also fixed at 7 cm, in order to study the influence of effect size without confounding the impact of varying standard deviations. After randomly generating male and female data, the data were combined into a single distribution, analogous to not independently knowing the sex of any individual. We also allowed for the total sample sizes to vary from 20 to 20 000 in seven steps (under a log-scale), while keeping the ratio of men:women at 1:1 in each sample to avoid the impact of unequal sex ratios. At each combination of male average height and total sample size, we generated 1000 datasets, performed a Hartigan's dip test (10 000 Monte Carlo replicates) on each dataset, and then took the average *P*-value ($\alpha$ = 0.05) produced over those 1000 iterations. In order to produce a significant average *P*-value in

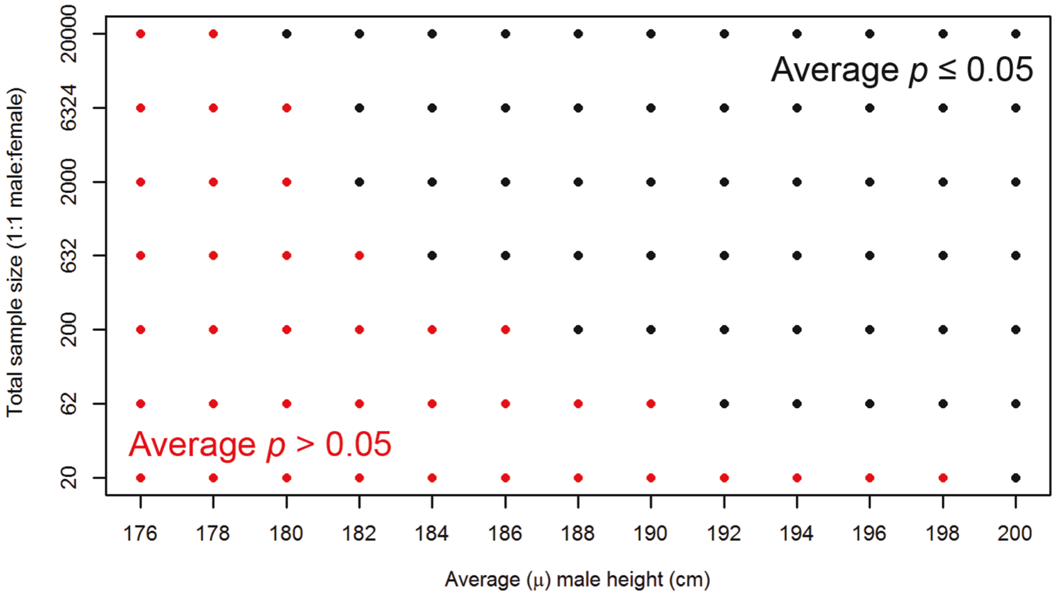

**Figure 2.** Hartigans' dip testing in simulated human height dimorphism. Each point at a given magnitude of dimorphism and sample size represents the average *P*-value of 1000 iterations of data simulation. Red indicates average *P*-values > 0.05. Black indicates average *P*-values ≤ 0.05.

support of a bimodal distribution at sample sizes consistent with those typical of fossil vertebrates (e.g. 20 and 62 on our log-scale), the average height of men has to be ~192–200 cm—an extreme magnitude of dimorphism relative to natural human populations. Given a more realistic average male height of 176 cm (Schilling *et al*., 2002), no sample size would yield a significant average *P*-value in support of bimodality. This is because the underlying theoretical population distribution at this magnitude of dimorphism is in fact unimodal, and so this result is not simply a matter of insufficient sample size. As described by Schilling *et al*. (2002), this is where the equally weighted (i.e. 1:1 men:women) normal distributions with a common standard deviation ($\sigma$ = 7 cm) have a difference in means (176 - 162 cm = 14 cm) that is not greater than $2\sigma$ (2 × 7 cm = 14 cm). Therefore, any observed *P*-value over 0.05 at this particular condition would actually represent a true negative, despite the fact that these theoretical parameters are a fairly realistic approximation of naturally occurring sexual dimorphism in the heights of men and women in many human populations.

A further problem for the use of bimodality tests to examine sexual dimorphism is that changing the proportion of males to females in the sample or the ratio of standard deviations between the male and female distributions can require even greater differences between the averages of the male and female distributions in order to yield a bimodal theoretical population of combined males and females (Fig. 3) (Schilling *et al*., 2002).

Mallon (2017) himself demonstrated that the Hartigans' dip test could not always detect genuine signals of sexual dimorphism when two datasets of extant sexually dimorphic species [alligator and the white-browed coucal bird (*Centropus superciliosus*)] failed to yield a significant result. Rather than concluding that the statistical test was ineffective, Mallon (2017) instead concluded that without independent knowledge of the sex of the individuals, dimorphism might not be detectable. Such a conclusion could effectively prevent any significant research into sexual selection in the fossil record of many taxa, given that it is commonly not possible to identify the sex of a fossil.

In addition to Hartigans' dip testing, Mallon (2017) used mixture modelling to fit normal distributions to fossil datasets, and the optimal number of normal distributions was selected using the Akaike information criterion (AIC), with two distributions interpreted as evidence for sexual dimorphism. Although mixture modelling can be extremely useful for a large and representative sample (especially when used alongside other approaches), as with any type of statistical modelling, there are some considerations. Mixture modelling identifies the best ways to explain the observed data by combining multiple distributions, without fully taking uncertainty into account. In other words, deviations in a random sample from the true, theoretical population distribution as a result of noise or small sample sizes are modelled as if they describe the shape of the population distribution, when they are instead 'fitting the noise'. Small sample sizes and preservation biases in fossil data can therefore

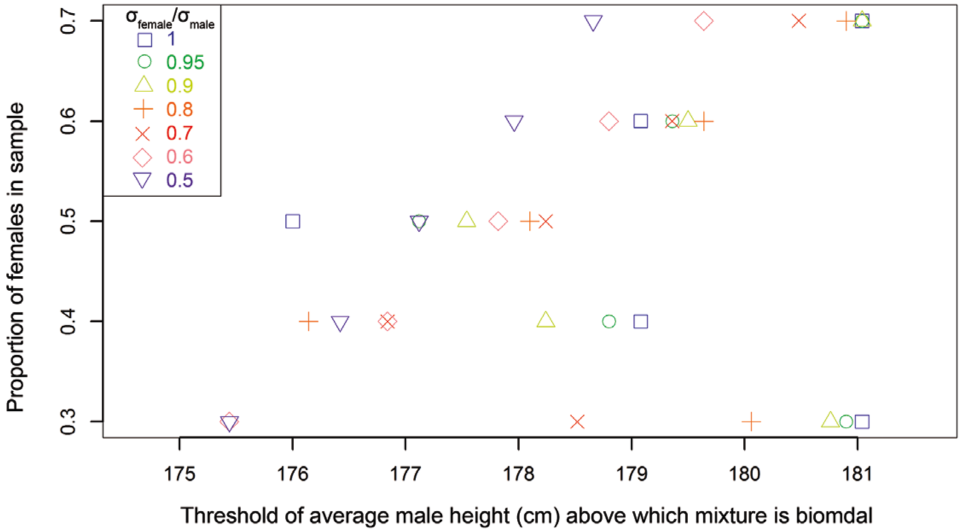

**Figure 3.** Average male height ($\mu$) above which the theoretical combined (male and female) population distribution modelled in Fig. 2 is bimodal, given certain proportions of females in the sample and ratios of female to male standard deviations ($\sigma_{female}/\sigma_{male}$). Data is converted from the values in Table 2 of Schilling *et al*. (2002). Different symbols directly correlate with the colour key and are used simply to show overlapping points.

lead to mixture models that overfit the data. When the data are univariate, overfitting can also be an issue when juvenile/young specimens are included. Finally, although no *P*-values are involved, there is still the risk of dichotomously concluding 'no evidence for dimorphism' based on this particular use of AIC scores. However, in the future, mixture modelling may become an important tool to be incorporated into our framework below with respect to sex assignment and judging intrasexual variation.

### *t-tests*

Hone & Mallon (2017) simulated male and female data using parameters derived from empirical growth curve data on alligators and rheas (*Rhea americana*). Their alligator data was based on a von Bertalanffy growth equation, whereas rhea data was based on a Gompertz growth equation. Their study then randomly sampled from the generated data and used *t*-tests to determine whether the sampled male and female data showed statistically significant differences in mean body size. The rhea data tended to give statistically significant results more often than the alligator data, interpreted as a consequence of protracted growth in alligator. Since many non-avian dinosaurs may have more protracted growth patterns than rheas, they suggested that this could explain the difficulty in detecting dimorphism in non-avian dinosaurs.

By using significance testing, Hone & Mallon (2017) assume a statistical framework in which monomorphism is the null hypothesis under the supposed absence of evidence for dimorphism in non-avian dinosaurs; in other words, a binary approach. Furthermore, independent knowledge of the sex of each individual is not possible for most fossil samples, although one could attempt to guess the sex of each data point (see discussion of sex assignment below). The test is also univariate, despite the fact that the data being analysed represent bivariate growth curves, with one variable being the sex-related trait of body size and the other being age (e.g. Wilkinson & Rhodes, 1997; Navarro *et al.*, 2005). The fullest signal of dimorphism is the divergence of two growth curves, one for each sex.

The results of Hone & Mallon (2017) do not appear to be solely sensitive to the prominence of dimorphism. Despite the fact that the rhea data outperform the alligator data in the *t*-tests, the alligator data showed more prominent dimorphism with a large magnitude of growth curve divergence (i.e. effect size) alongside low intrasexual variation. The alligator data showed clear separation of the largest male and female points, unlike the rhea data. The ease with which sampled males and females pass a *t*-test depends not only on the magnitude of the dimorphism (which was not controlled for in the two-species comparison), but also on a combination of growth rate and life span, as noted by Hone & Mallon (2017). Faster growth rate means that adult size is more quickly attained, but it is the growth rate in relation to life span that dictates the proportion of adults in the sample and distinguishes the rhea dataset from the alligator dataset. By increasing the proportion of the population at adult size, the proportion of the population in which dimorphism is maximally expressed is also increased, meaning that a random sample is more likely to contain a relatively stronger signal of dimorphism independent of the magnitude of dimorphism (Hone & Mallon, 2017). Because both species show greater ontogenetic variation than sexual variation, excluding juveniles is key to passing the *t*-test. This is consistent with the lower *P*-values derived from their alligator simulations with a size bias against smaller individuals and higher *P*-values derived from their alligator simulations with altered population structure (i.e. greater proportion of juveniles). As mentioned by Hone & Mallon (2017), the vertebrate fossil record is often not biased in favour of overrepresentation of small juveniles because of taphonomic and collecting factors (Brown *et al.*, 2013), as well as population structures hypothesized for some non-avian dinosaurs [Erickson *et al.*, 2006, 2009; Woodward *et al.*, 2015; although see Varricchio (2011) on aggregations of juveniles in some bonebeds].

## A FRAMEWORK FOR INVESTIGATING SEXUAL DIMORPHISM IN EXTINCT TAXA

We propose a different method to investigate sexual variation in the fossil record. Rather than performing univariate significance tests on a single dataset to either reject or fail to reject sexual monomorphism in that species (i.e. dichotomous hypothesis testing), our approach focuses on effect size statistics (Nakagawa & Cuthill, 2007; Amrhein *et al.*, 2017; Amrhein & Greenland, 2018; Halsey, 2019; Holland, 2019), combined with supporting contextual evidence. Our aim is to test alternate hypotheses against the observed variation, as well as estimate its magnitude and constrain uncertainty of that estimate in a given species/dataset compared to other species/datasets. Our framework first involves collecting appropriate data, followed by the consideration of several alternate hypotheses, and finally inferring biological implications of the results, similar to the approach employed by Saitta (2015) on stegosaur fossils. We attempt to provide a fairly exhaustive list of alternate hypotheses, along with lines of evidence that can be used to systematically abandon them one by one. However, some alternate hypotheses are case specific (e.g. intra-individual variation of repeated structures).

We introduce our quantitative effect size approach with respect to the first alternate hypothesis of non-sexual variation/polymorphism in a population. Recently, over 800 researchers including statisticians, medical scientists, biologists and psychologists signed a letter calling for an end to over-reliance on statistical significance (Amrhein *et al*., 2019). In 2019, the 73rd volume (supplement 1, issue 1) of the academic journal *The American Statistician* published a series of articles echoing this sentiment. Issues with significance testing extend beyond those situations with small sample size and type II error described above. In fact, very large sample sizes can lead to very small *P*-values, even if the effect size is small. This low *P*-value can be falsely interpreted as indicative of great biological importance, even though the observed biological effect is small. We think that the points we raise here are more consistent with this statistical paradigm shift than those from much of the previous work on this topic. We abide by the suggestion that "you can enhance the information provided by frequentist statistics with a focus on effect sizes and a quantified confidence that those effect sizes are accurate" (Halsey, 2019: p. 1).

### Data collection

As with any statistical analysis and scientific study, maximizing the sample size is a priority. Although sample sizes of many fossil groups remain small after one or two centuries of scientific collecting, these will undoubtedly increase over time, and science is an intergenerational pursuit. Other than zero or one, from which an estimate of sexual variation cannot be based on data, there is no fixed lower limit on sample size. Sample sizes of two clearly result in extremely high uncertainty, as in Persons *et al*. (2015).

In our approach, we use quantitative bivariate data to produce growth curves of body size or other traits that are commonly secondary sexual characteristics (e.g. putative ornaments/armaments or coloration) (Fig. 4). The use of growth trajectories to study sexual variation is precedented (Evans *et al*., 2018). It is best to avoid traits that are suspected to show negligible sexual variation or be under weak sexual selection [e.g. stegosaur femoral head shape (Barden and Maidment, 2015)], unless in comparison to another trait suspected to be sexually variable.

Not only is this putative dimorphic trait quantified [e.g. direct measurements, principal component scores, outlines (Bonhomme *et al*., 2014; Saitta, 2014)], it is also collected alongside some measure of/proxy for age (Hone *et al*., 2016b). For example, this could be histological lines of arrested growth or body size, if body size is not the trait being examined for potential dimorphism. Note that body size is not always a good proxy for age (Hone *et al*., 2016b) and that the potentially confounding effect of dimorphism in both body size and the other selected trait should be considered. In that case, it might be possible for both sexes to show similar growth curves (i.e. patterns of growth) for a trait when plotted against body size, but the curve for the smaller-trait-bearing sex might terminate at a smaller body size—meaning that individuals of the smaller sex are hidden among younger individuals of the larger sex (Hone & Mallon, 2017). If this confounding is suspected, then alternative proxies for age or more precise age estimates (e.g. using lines of arrested growth) can be carried out rather than using body size as an age proxy. Some measures of age might be clade specific. For example, the ages of fossil elephant specimens are often assessed based on tooth eruption and wear (Maschenko, 2002; Lister, 2009).

The advantage of regression analysis on bivariate growth curves over univariate analysis on a trait of interest is that the addition of immature individuals in a univariate analysis can lead to greater overlap of male and female distributions, making dimorphism harder to detect (e.g. if females resemble immature males). In contrast, bivariate regression analysis can not only cope with juvenile data points, but juvenile points make curve fitting computationally easier.

Researchers should then address alternate hypotheses for the anatomical variation within a sample. If appreciable evidence against all alternate

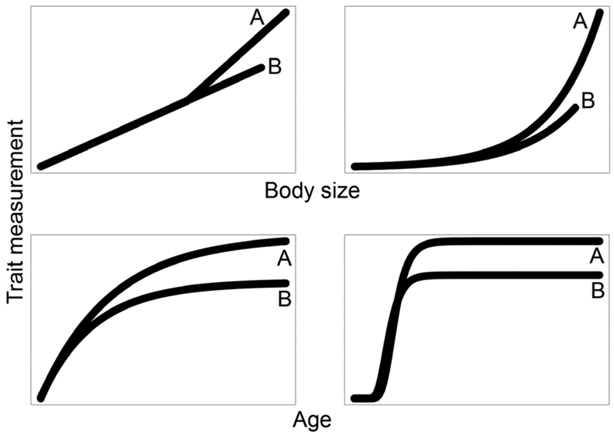

**Figure 4.** Hypothetical examples of growth curve divergence. Sex A and B are the sexes with larger and smaller trait measurements, respectively. Sex A is not necessarily male, but in many cases, this might be the expectation. Trait development can be examined according to age or a proxy for age (e.g. body size). Depending on trait growth pattern, different regression models (i.e. different formulae) may be appropriate. The growth curve of one sex might terminate earlier than the other, particularly when age proxies like body size are used (e.g. dimorphism is expressed in both the trait and body size).

hypotheses can be obtained, then sexual dimorphism is well supported.

### ALTERNATE HYPOTHESIS 1: CONTROLLING FOR NON-SEXUAL VARIATION AND POLYMORPHISMS IN POPULATIONS

Other than sexual dimorphism, non-sexual variation/polymorphisms within a population might explain observed variation. Continuing from the efforts to collect data on traits likely to be sexually variable, to rule out this alternate hypothesis, evidence should show that the trait of interest diverges in its growth curve. If, for example, clear patterns of multiple split/separate growth curves suggest three or more morphs, this would indicate that the variation cannot be explained simply by sexual dimorphism. A non-sexual intra-population variation/polymorphism hypothesis can be abandoned under evidence for divergence of a plausibly sexual trait into two growth curves, ideally without overlap of confidence intervals and thereby good estimated support for sex-specific growth models.

Controlling for geographic and stratigraphic distribution of the samples can help to rule out geographic or temporal variation of a single species (e.g. biogeographic subspecies or morphological gradients). Datasets from a single locality and horizon are generally better than datasets that span large stratigraphic intervals and geographic areas, as the former are more likely to represent a single population. This approach is comparable to that used to rule out more extreme interspecific variation below.

To illustrate the utility of our framework, we first simulate alligator and rhea growth curves under a variety of sample sizes (10–250) and effect sizes (Tables 2–3) over comparable age ranges as in Hone & Mallon (2017); multiple runs of this simulation and the code are available as Supporting Information (Appendix S1). Effect sizes (i.e. difference in $L$ parameters between the sexes) range from true monomorphism (i.e. 0) to double those observed in empirical data for these species (Wilkinson & Rhodes, 1997; Navarro *et al.*, 2005). For now, we keep the function (according to size) for population standard deviations of the residuals along the growth curves equal between the sexes (except Figs 5, 12). Note that this still generates greater intrasexual variation in the larger sex. In order to introduce the dynamics at play, our initial simulations keep equal proportions of males and females [except in Supporting Information (Appendix S1) and Fig. 12]. Unlike empirical datasets, simulated data allows for thorough examination of a methodology because the effect size, sex ratio, intrasexual variation and sample size can all be controlled or experimentally altered as needed. In simulation, the true population-level theoretical values of the parameters are known since these are used to generate the data. This allows for the effect size estimated from a sample using a particular method to be compared to the true population effect size, not simply to the observed effect size of a given empirical sample for which the sex of each individual is known.

### *Identify a signal consistent with growth curve divergence*

A preliminary examination of the data can help determine if further analysis is likely to give a strong signal of dimorphism. A conspicuous signal consistent with divergence of male and female growth curves should show separation or spreading of the data at older ages or larger sizes. Separation might be prominent enough to be noted visually (Saitta, 2015). Even if there is no conspicuous separation with a gap, an increase in the spread of the data in older/larger individuals can be consistent with sexual dimorphism. However, the spread of the data could increase along the growth curve even in a truly monomorphic case (i.e. effect size of zero) due to variable growth trajectories of individuals regardless of sex or due to any preferential collection of larger specimens such that juveniles are under-represented in the sample. Putative sexual traits that are constant or decrease in spread along a growth trajectory, or that are impoverished in adults/large specimens, might be viewed with scepticism, as they likely provide little evidence for sexual variation.

Even when both alligator and rhea simulated data are dimorphic (i.e. different growth equations are used to generate males and females), sample sizes are large, sex ratio is at parity, and all parameters (including effect size and sex-specific population standard deviations of the residuals) are set to their naturalistic, empirically derived values (Fig. 5), only alligator shows clear separation (i.e. large effect size with small intrasexual variation). Rhea only shows an increase in spread of the data without separation. To more easily judge spread, a regression can be fitted to the data as a whole and the residuals can be examined via a residual plot. When a single regression is fitted, naturalistic simulation of alligator data has residuals that not only show an increase in spread, but also clear separation along positive and negative values (indicating even sex ratios and lack of extreme outliers). Naturalistically simulated rhea simply shows a sharp increase in the spread of the residuals over the first ~10% of the life history.

The type of regression fit to the data (e.g. von Bertalanffy, Gompertz, logistic, logarithmic, exponential, linear, etc.) will vary depending on

**Table 2.** General description of growth models.

| | Growth model | | |
|---|---|---|---|
| | von Bertalanffy | Gompertz | Logistic |
| Taxon | *Alligator* | *Rhea* | *Maiasaura*<br>*Psittacosaurus*<br>*Tyrannosaurus* |
| Trait | Body length | Body mass | Body mass |
| Parameters | $L_\infty$ | L | L |
| | K | k | q |
| | t | i | k |
| Monomorphic model | $C^* = L_\infty(1\text{-}e^{-K(P-t)})$ | $C = Le^{-e\text{\textasciicircum}(-k(P-i))}$ | $C = L/(1+e^{q+kP})$ |
| Dimorphic model | $C = SL_{\infty M}(1\text{-}e\text{\textasciicircum}(-K_M(P\text{-}t_M)))$<br>$+ (1\text{-}S)L_{\infty F}(1\text{-}e\text{\textasciicircum}(-K_F(P\text{-}t_F)))$ | $C = SL_Me\text{\textasciicircum}(-e\text{\textasciicircum}(-k_M(P\text{-}i_M)))$<br>$+ (1\text{-}S)L_Fe\text{\textasciicircum}(-e\text{\textasciicircum}(-k_F(P\text{-}i_F)))$ | $C = SL_M/(1+e\text{\textasciicircum}(q_M+k_MP))$<br>$+ (1\text{-}S)L_F/(1+e\text{\textasciicircum}(q_F+k_FP))$ |
| Effect size | $L_{\infty M}\text{-}L_{\infty F}$ | $L_M\text{-}L_F$ | $L_M\text{-}L_F$ |

**C* = trait measured as potentially sexually variable. *P* = age or proxy for age. M subscript indicates sex with more trait development; F subscript indicates sex with less trait development. *S* = sex (e.g. either 1 [for M] or 0 [for F])

the type of growth, as dictated by the variables measured and taxon studied, and can be inferred based on $R^2$ or AIC values as well as known growth patterns in related species. Conspicuousness of data separation or spreading can decrease with smaller sample sizes.

### *Assign sex*

The next step involves assigning the sex of each individual, which can be accomplished in a variety of ways. Here, we focus on the simplest method as an illustrative example (Fig. 6), which involves fitting a single best-fit growth curve to the total dataset and assigning sex to specimens based on whether their residuals relative to that curve are positive or negative. One weakness of this method is that it assumes continuous sex-specific growth throughout life, when secondary sexual characteristics might have delayed developmental onset at or near sexual maturity [e.g. facial hair in men coinciding with puberty (Lee, 1980)]. Additionally, when effect size is small and sexes heavily overlap, there will be bias towards overestimation of effect size (see below). However, an advantage is that fitting such a curve to datasets can be computationally easy.

Despite criticism of this sort of division about the centre as arbitrary (Mallon, 2017), the accuracy (i.e. percentage of true members of a sex that are correctly assigned to that sex) of even this simplistic method in our simulations can be > 80% with moderate/naturalistic magnitudes of dimorphism (relative to empirical alligator and rhea data), so long as males and females are similarly represented in the sample. This is because we are not assuming any sex roles (e.g. that males are larger), but rather assigning the points to either a larger or smaller sex, regardless of whether sex-role reversals are present. Avoiding this assumption can prevent unsupported interpretations of mating systems when independent evidence of which sex is larger is absent. Accuracy increases more rapidly with increasing true effect size for the alligator simulations than the rhea simulations due to the smaller intrasexual variation relative to the magnitude of naturalistic dimorphism in alligator growth. This method is expected to be better than random at accurately assigning sex under many circumstances (i.e. if true effect size > 0 and both sexes are equally represented, then sex assignment accuracy > 50% on average) (Figs 7, 8). In this specific case (i.e. sex ratio parity), Mallon's (2017) assertion that the technique is arbitrary only holds true on average in the extremely unlikely circumstance that the magnitude of sexual variation is precisely zero. In all other circumstances in which there is any non-zero effect size, the technique will on average (under non-skewed sex ratios) have some degree of useful discriminatory ability in assigning sex. Truly monomorphic datasets (i.e. effect size = 0) with equal proportions of males and females still appear to yield sex assignments with accuracies that centre around 50%. Smaller sample sizes lead to greater variability in accuracy between datasets at a given effect size (i.e. spread of accuracy values increases with decreasing sample size) (Figs 7, 8).

The accuracy of this simplistic approach for sex assignment is consistent with previous studies showing that division about the mean is one of the most robust approaches for estimating dimorphism when accounting for possible variation in sample size,

**Table 3.** Workflow of data simulation and analysis in Figures 7–11 and 13–14.

| | | *Alligator* | *Rhea* |
|---|---|---|---|
| Data simulation | Female | $C = 2.78(1-0.91^{-0.0926P})$ | $C = 0.3544+22.5074e^{-e^\wedge(-12.12(P-0.32191781))}$ |
| | Male | $C = (2.78\mathbf{+E})(1-(0.91\mathbf{+0.02970297E})^{(-0.0926\mathbf{+0.022871287E})P})$ | $C = 0.3544+(22.5074\mathbf{+E})e^{-e^\wedge((-12.12\mathbf{+0.204875058E})(P-(0.32191781\mathbf{+0.008353227E})))}$ |
| | Range | $1 \leq P \leq 50$ | $0 \leq P \leq 10.5$ |
| | Standard deviation of the residuals for each sex | $0.0589\log(C)+0.0816$ | $0.6793\log(C)+0.7378$ |
| | Curve for residual-based sex assignment | $C = L_\infty(1-e^{-K(P-t)})$ | $C = Le^{-e^\wedge(-k(P-i))}$ |
| Sex-specific curve fitting | Female | $C = L_{\infty F}(1-e^\wedge(-K_F(P-t_F)))$ | $C = L_F e^\wedge(-e^\wedge(-k_F(P-i_F)))$ |
| | Male | $C = L_{\infty M}(1-e^\wedge(-K_M(P-t_M)))$ | $C = L_M e^\wedge(-e^\wedge(-k_M(P-i_M)))$ |

Bold font indicates the proportionally applied effect size incorporated into the male growth curve.
*$E$ = effect size. $E$ applied proportionally to each parameter based on natural dimorphism for that species to avoid unrealistically shaped curves. See Table 2 legend for further details. Function for the standard deviation of the residuals of each sex independently varied between each hypothetical taxon in the phylogenetic simulation (Fig. 12), set to sex-specific natural values in Fig. 5, and doubled for both sexes in a simulation in the Supporting Information (Appendix S1)

intrasexual variation and even sex ratio of the sample (Plavcan, 1994; Rehg & Leigh, 1999). Even simplistic approaches should be satisfactory under many combinations of effect size and sample size, so long as the sample is not highly skewed in sex ratio (Allentoft *et al.*, 2010). Heterogametic sex determination, as in birds, is expected to result in primary sex ratio parity (i.e. at hatching)—relevant to our framework given that we suggest examining a wide ontogenetic range. Assuming no subsequent sex bias in mortality, roughly equal male and female representation would be expected in a random sampling at large sample size. If the probability of an individual fossil data point being either male or female is analogous to independent coin flips, then the true relative frequency of each sex (prior to any attempt to assign sex, regardless of method) will approach 50% as more samples are added.

Skewed sex ratios could result from biological (Laver *et al.*, 2012) or taphonomic factors (e.g. sex-specific behaviour/ecology leading to different burial environments and preservation potential) as well as random sampling at small sample size. Although there is considerable interspecific variation, adult sex ratios can, for example, be somewhat male-biased in birds (Donald, 2007; Liker *et al.*, 2014) with the main driver away from a 1:1 primary sex ratio being juvenile mortality (Szekely *et al.*, 2014; Eberhart-Phillips *et al.*, 2018). If a species of interest is suspected to have skewed sex ratios, one possibility would be to introduce a correction factor in the sex assignment step based on observed values in extant members of the clade to see if any trends in estimated sexual variation across species persist once this correction is applied. Some groups, such as crocodilians, show environmental sex determination driven by incubation temperature (Paukstis & Janzen, 1990). Theoretical work has suggested that environmental fluctuations from year to year can cancel out fluctuations in sex ratios in long-lived species with environmental sex determination (Bull & Bulmer, 1989), and so most fossil datasets of such groups might provide a time-averaged sex ratio closer to parity. A taphonomic example where the sex ratio would differ massively from equality, even at large sample sizes, is a fossilized death assemblage of a single-sex group (Hone & Mallon, 2017) due to single-sex socialization or temperature-dependent sex determination. In single-sex-dominant assemblages (Chen *et al.*, 2019), a trade-off would occur between greater sample size and skewed sex ratio. A dataset that, for example, relies solely on a single bonebed might be better able to control against geographic/stratigraphic confounding, but be at risk of skewed sex ratios. In a single-sex-dominant assemblage, one might expect any variation of traits under sexual selection to be smaller relative to a sample with both sexes represented, and therefore the calculated effect size might be underestimated with respect to measures of uncertainty.

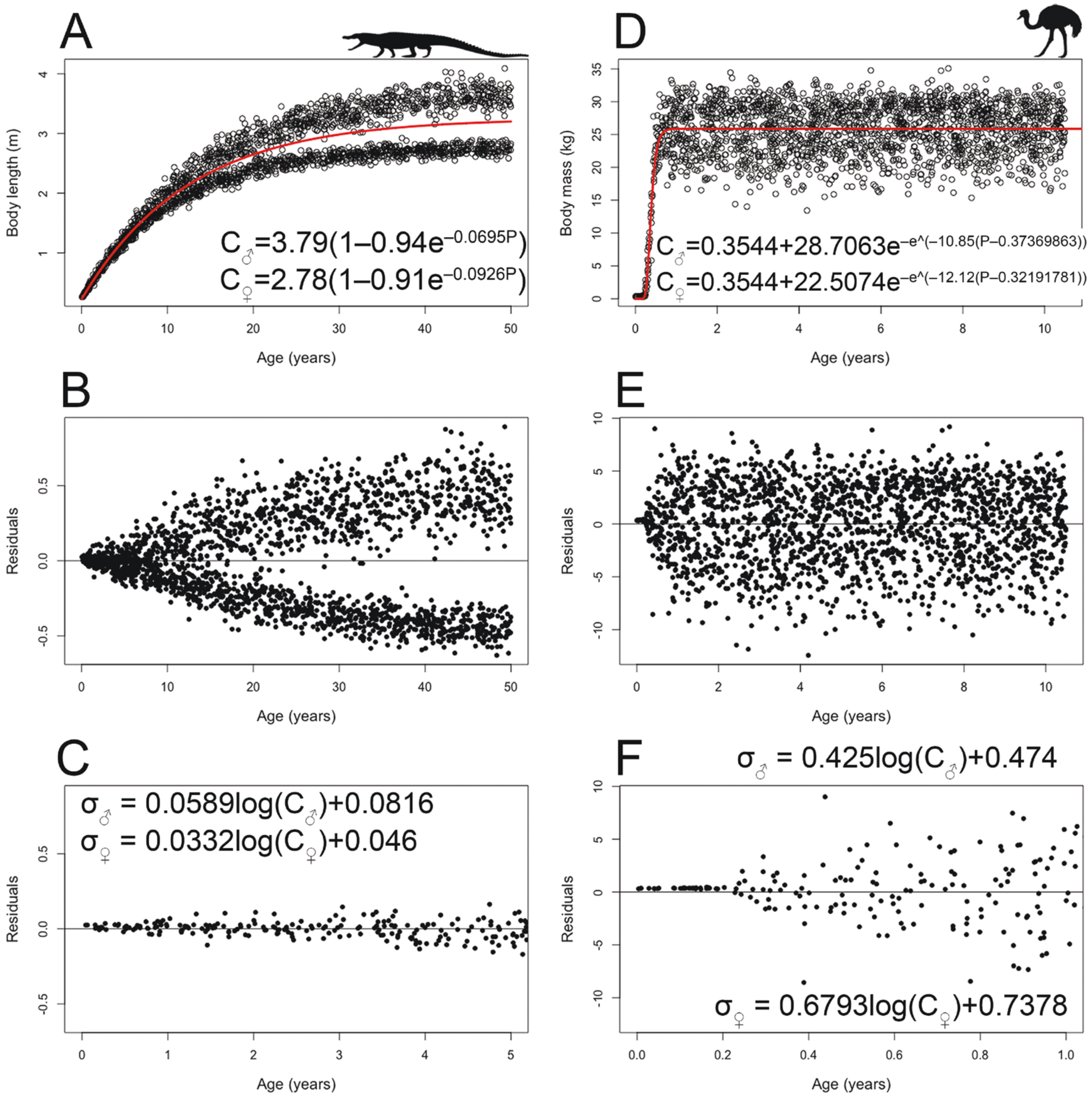

**Figure 5.** Preliminary visual examinations of potential growth curve divergence in simulated dimorphic male and female data (see code of Hone & Mallon, 2017) based on empirical observations from Wilkinson & Rhodes (1997) for alligator (A–C) and Navarro *et al*. (2005) for rhea (D–F). Results of data simulation (A, D). Sex of individuals is unknown, as in fossil samples, and sex ratio is at parity. This simulation is fully empirically derived (i.e. function for population standard deviation of the residuals (σ) along growth curves is unequal between the sexes), unlike our other simulations. Single regressions (von Bertalanffy for alligator, Gompertz for rhea) are fit to the combined male and female data. Residual plots (B, E) are examined for evidence of dimorphic growth. Residuals also examined over the first 10% of lifespan (C, F).

In addition to greater overlap of males and females (i.e. greater intrasexual variation relative to effect size), more heavily skewed sex ratios can also decrease sex assignment accuracy in this illustrative approach. Fitting a regression to the data may estimate a sex ratio at parity (excepting the impact of outliers/influential observations, which can heavily affect the regression, or datasets with an odd sample size).

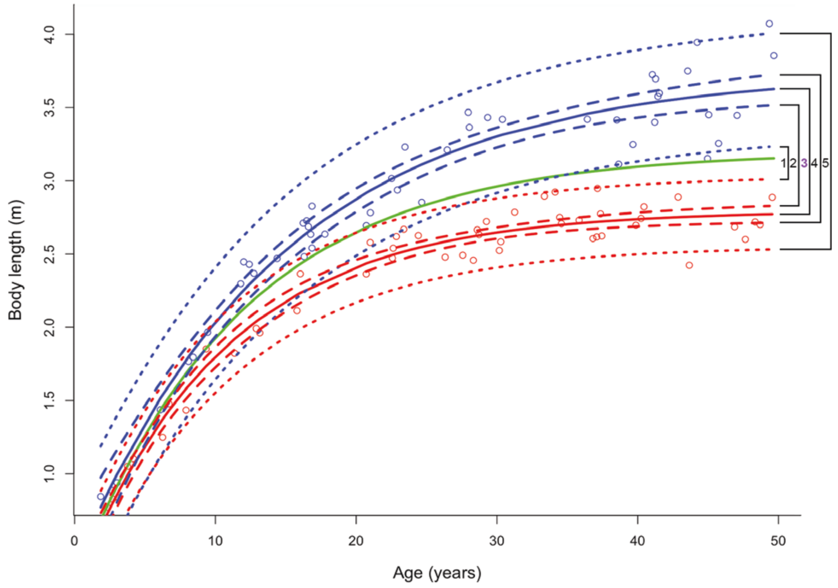

**Figure 6.** Summary of our statistical approach. Alligator data simulated here with functions for male and female population standard deviations of the residuals kept equal to each other and effect size equal to that of natural populations (Wilkinson & Rhodes, 1997). A curve (green) is fit to the whole dataset. Points with positive (blue) and negative (red) residuals are assigned to opposite sexes. Sex specific curves are fit (3) to estimate effect size, $E_E$, measured as $L_{\infty M}$-$L_{\infty F}$ for von Bertalanffy curves and $L_M$-$L_F$ for Gompertz curves. Ninety-five percent confidence (thick dashes) and prediction (thin dashes) intervals are calculated for each sex. Separation of prediction (1)/confidence (2) intervals and spread of confidence (4)/prediction (5) intervals can be calculated at the largest individuals of each assigned sex.

Skewed sex ratio in a sample can allow sex assignment accuracies for a single sex to fall below 50% on average (Supporting Information, Appendix S1). Therefore, future work should compare the accuracy of different sex assignment methods while accounting for growth, perhaps incorporating principal component analysis, cluster analysis, k-means clustering, or Gaussian mixture modelling. In particular, mixture modelling might promise to help reduce bias when the sexes heavily overlap—possibly by (1) removing juvenile points before applying a univariate 2-component Gaussian mixture model such that the distance an adult lies from the single best-fit curve dictates the probability that it is assigned to a particular sex (especially in the asymptotic region of Gompertz curves) or (2) through a more sophisticated mixture modelling that combines regression analysis. Furthermore, when palaeontological indicators of sex are present in some individuals (e.g. *in situ* embryos or bacula), then care can be taken not to assign these data points to different sexes, thereby increasing sex assignment accuracy. For example, if the sex of individuals with variable skeletal completeness is estimated using the residual-based approach described here, and 90% of the specimens with known bacula are assigned to the larger sex, then the other 10% of specimens with known bacula can be reassigned to the larger sex to improve sex assignment accuracy.

Sex assignment goes to the heart of a major misconception about studying sexual selection in the fossil record, which provides the undertone for the statement that "in the absence of a priori knowledge of sex, the ability to detect dimorphism in a fossil sample is likely only in cases of strongly expressed dimorphism (i.e. well-separated peaks on a histogram or discrete character states); weakly expressed dimorphism, where the sexes overlap considerably in morphospace, may be impossible to discriminate in the fossil record" (Mallon, 2017: p. 502). Although it is true that large effect sizes are often easier to detect than smaller effect sizes, abandoning dichotomous detection vs. non-detection can allow comparison between large effect sizes with low uncertainty and small effect sizes with high uncertainty in a quantitative continuum. In many cases, it is not necessary to know the sex of each specimen from independent evidence to produce an estimate of effect size that can be compared to other datasets.

Similarly, a simple example shows why acceptance or rejection of sexual dimorphism is not dependent upon any single individual's sexually variable trait providing an unambiguous indicator of its sex. If one is told that a person is 178 cm tall, this is insufficient to identify their sex, despite the fact that sexual height dimorphism exists in humans. The ability or inability to assign sex to an individual based solely on a sexually

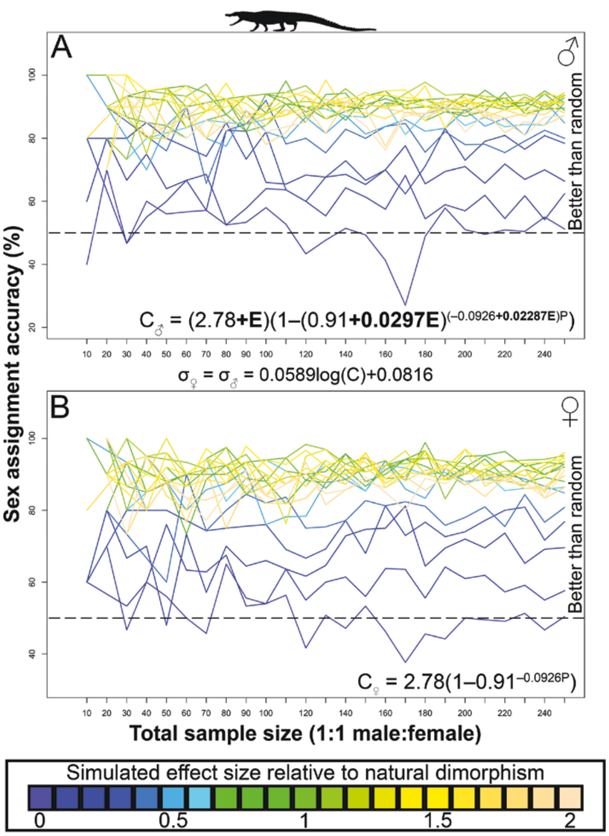

**Figure 7.** Sex assignment accuracy for simulated male (A) and female (B) alligator data. Dashed line indicates an accuracy of 50%. Colour coding represents the true effect size used to generate the data divided by observed empirical dimorphism (Wilkinson & Rhodes, 1997): zero is monomorphism (true effect size, $E$ = 0), one is dimorphism matching natural dimorphism ($E$ = 1.01), and two is dimorphism twice as great as natural dimorphism ($E$ = 2.02). $E$ is applied proportionally to each growth curve parameter based on natural dimorphism. Function for population standard deviation of the residuals is kept equal between males and females, based on the natural values of the sex with greater intrasexual variation. Sex ratio at parity. These results are from a single run of our code. Each time the code is run, precise values will vary while overall trends persist.

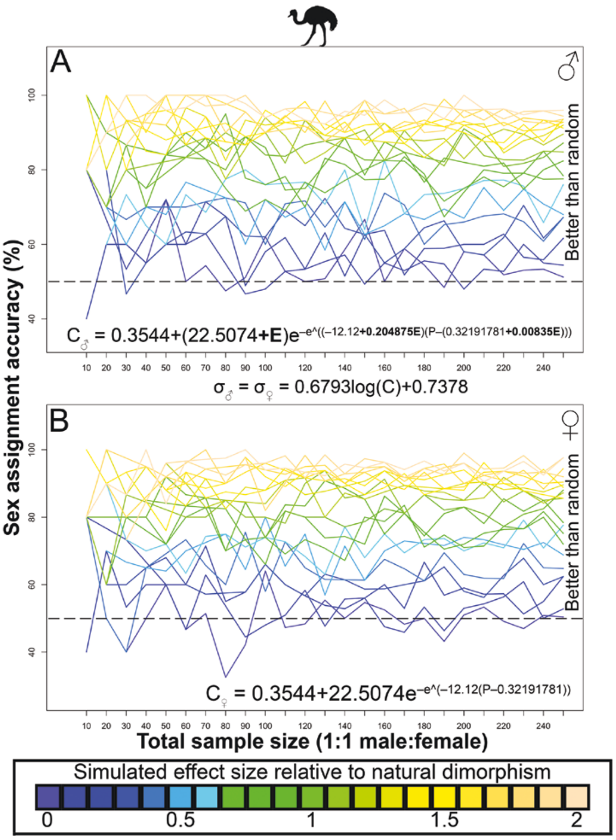

**Figure 8.** Sex assignment accuracy for simulated male (A) and female (B) rhea data. See Figure 7 legend for further details. Observed empirical dimorphism from Navarro *et al.* (2005). For effect size ratio of one, $E$ = 6.199. For effect size ratio of two, $E$ = 12.398.

variable trait is an indication of whether or not sexual dimorphism is of an extreme effect size, not whether or not dimorphism is present.

### *Estimate and constrain magnitude of dimorphism*

After assigning sex for the data points, a growth curve should be fitted to each sex, allowing for a specific estimate of dimorphism magnitude, together with confidence and prediction intervals. Accepting uncertainty of an estimate is a key tenet of responsible statistical analysis (Wasserstein *et al.*, 2019). Confidence intervals attempt to constrain the range in which the theoretical population parameters are expected to lie at a given probability. In this case, 95% confidence intervals constrain our estimates of the fitted sex-specific curve. Specifically, "the 95% refers only to how often 95% confidence intervals computed from very many studies would contain the true size *if all the assumptions used to compute the intervals were correct*" (Greenland *et al.*, 2016: p. 343, emphasis theirs). Therefore, the accuracy of our estimate and success at constraining that estimate improves as sex assignment accuracy increases and assumptions are more closely met. Prediction intervals attempt to constrain the range in which a data point would be expected to lie at a given probability, accounting for the spread of the data. In this case, 95% prediction intervals attempt to predict the range in which the addition of an individual male or female to the dataset might lie (again dependent upon prior assumptions being met, namely sex assignment accuracy). For our purposes, prediction intervals can provide insight into the estimated degree of overlap between males and females, making them relevant to judging support for dimorphism. Note that statisticians call for careful and nuanced interpretations of such intervals in a similar manner to the caution advised for interpreting *P*-values (Amrhein *et al*., 2019).

We define effect size here as the difference between the male and female $L$ parameters of the growth formulae. The accuracy of our estimated sex-specific regressions to the true growth curves (i.e. the accuracy of the estimate for dimorphism magnitude) is expected to be biased in the case of a truly monomorphic dataset because the illustrative method of sex assignment will predict some minor amount of dimorphism. However, even with relatively small true effect sizes, the fitted curves can closely match the true growth curves (Fig. 9). Above a certain true effect size used to generate the data, the ratio of true effect size to estimated effect size centres around a value of one. Furthermore, the naturalistic effect size observed in alligators and rheas (i.e. true effect size used in simulation equals the empirical effect size observed in nature) yields model accuracies that approach or are centred near one, with alligator performing better. Smaller sample sizes can lead to greater variability in this measure of model accuracy. As effect size increases, rhea simulations lag behind alligator simulations in their improved model fitting accuracy due to higher intrasexual variation relative to effect size in rhea.

If one suspects that sexual traits showing high amounts of intrasexual variability and greater overlap of males and females might lead to overestimates of sexual variation, then the trait suspected to be sexually selected can be compared to a trait in the same sample/species that is less likely to be sexually selected (as in O'Brien *et al*., 2018). The difference in estimated sexual variation between the putatively sexual and non-sexual trait can then be compared to differences observed in other taxa/datasets. For example, an estimate of sexual variation in horn length relative to an estimate of sexual variation in vertebral or tooth shape could be compared across a horned clade. Again, mixture modelling might have promise in estimating intrasexual variation at a given ontogenetic stage (e.g. adults), since mixture models attempt to estimate standard deviations of the underlying male and female distributions.

Here, we examine the separation of the upper bound of the smaller sex from the lower bound of the larger sex and the spread between the lower bound of the smaller sex and the upper bound of the larger sex at the oldest individuals of each assigned sex (Figs 10, 11). Note that estimation of the oldest/largest age/size each sex can attain based on the oldest/largest individual of each assigned sex in a given dataset can come with high uncertainty. One may instead wish to measure interval spread and separation at a given age/size class. Our goal here is simply to be illustrative and internally consistent, but we could have alternatively measured the intervals at the single oldest age attained by both assigned sexes to better represent this uncertainty. The measure we use here might instead be more appropriate if using body size as a proxy for age, since the two sexes might not attain the same maximum body size.

Interval spread and separation increase with increasing true effect size. Confidence and prediction intervals widen with decreasing sample size, so interval spread increases and interval separation decreases. Interval separation is important in judging uncertainty in the estimate for growth curve divergence. Confidence interval separation is more often expected to be positive when the effect size is truly zero (i.e. monomorphic) than prediction interval separation because prediction intervals are wider than confidence intervals for a given regression. Non-zero true effect sizes might still fail to result in interval separation. Rhea simulations show greater uncertainty in growth curve divergence than alligator simulations due to relatively higher intrasexual variation with respect to effect size.

The rhea simulations show many extreme outliers in interval spread/separation due to difficulty in fitting Gompertz curves to the data. These outliers are datasets in which, after sex assignment, one or both assigned sexes lack juveniles that allow for easier curve fitting. Even with large sample sizes, if complex growth models are to be applied to the data, it is important to include specimens over a range of ages/age proxies. Given the lack of juveniles and extreme interval spread and separation values (sometimes differing by orders of magnitude), these outlier datasets are easy to identify and ultimately could be dropped from any comparative analysis between datasets.

It might be tempting to use interval separation vs. overlap as a form of significance testing. However, we urge researches to avoid the potential pitfall of dichotomous acceptance vs. rejection of the presence of an effect (i.e. black-or-white thinking) (Nakagawa & Cuthill, 2007; Amrhein *et al*., 2017; Amrhein & Greenland, 2018; Halsey, 2019; Holland, 2019). One simply attempts to estimate sexual variation and constrain the uncertainty in that estimate without overinterpretation, rather than categorizing a species as either dimorphic or monomorphic or stating that evidence is either present or absent. Despite the bias towards overestimation when true effect size is low relative to intrasexual variation, our framework does not assume dimorphism is necessarily present or strong when intervals overlap. When constraining estimates with indicators of uncertainty, such as confidence intervals, it is important to consider the following:

> “Accept uncertainty and embrace variation in effects: we can learn much (indeed, more) about the world by forsaking the false promise of certainty offered by dichotomous declarations of truth or falsity—binary statements about there being “an effect” or “no effect”—based on some *P*-value or other statistical threshold being attained” (McShane *et al.* in Wasserstein *et al.*, 2019: p. 15).

When comparing estimated magnitudes of sexual variation between datasets/taxa, we recommend standardizing estimates of effect size and interval

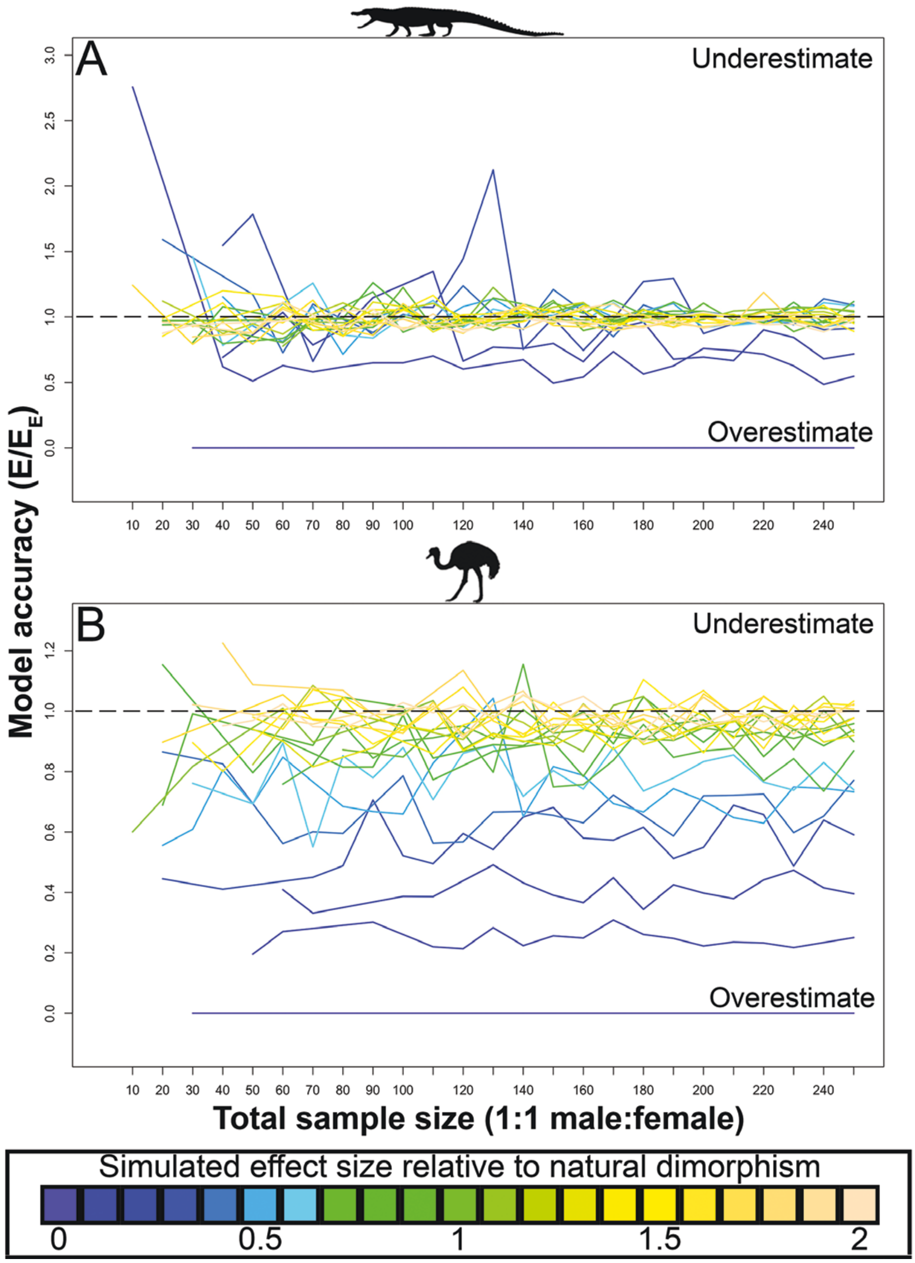

**Figure 9.** Fitting dimorphic models to simulated alligator (A) and rhea (B) data. Accuracy measured as true simulated effect size, $E$, divided by estimated effect size, $E_E$. Dashed line indicates a ratio of one ($E = E_E$). See Figure 7 legend for further details.

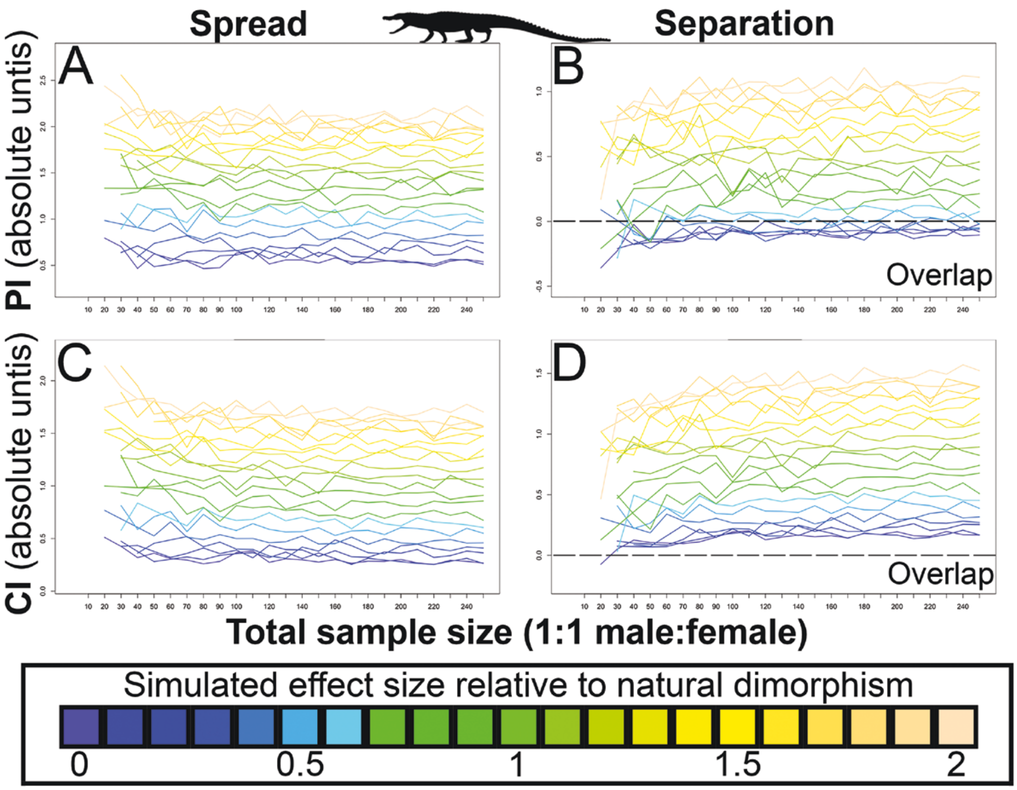

**Figure 10.** Prediction (PI) (A, B) and confidence (CI) (C, D) interval spread (A, C) and separation (B, D) for simulated alligator data. Dashed line indicates a value of zero (in absolute units: m) for interval separation. See Figure 7 legend for further details.

spread/separation (e.g. as a percentage of the maximum size predicted by a regression fit to the smaller sex). Standardized estimates of effect size and their confidence/prediction intervals provide a way to compare different datasets/species in a meta-analytic or phylogenetic framework. This approach takes into account differences in sample sizes and can be used to look for heterogeneity (e.g. if some populations differ in dimorphism magnitude as a function of ecology).

### *Phylogenetic simulation with more realistic data variability*

The above simulations (and the significance testing simulations below) held sex ratio and the function (according to size) for standard deviation of the residuals constant and at parity between the sexes in order to show the dynamic between true effect size, sample size, and our estimates. What happens if parameters/conditions are allowed to vary randomly such that they more closely represent empirical fossil or modern data? Inspired by natural alligator data, we generated a hypothetical crocodilian clade containing a grade of 100 'monomorphic' taxa (i.e. low sexual variation) that evolve a clade of 50 'dimorphic' taxa (i.e. a shift to high sexual variation). Each taxon was allowed to vary randomly and independently in true effect size (i.e. proportionally applied changes to male growth parameters, as in the other simulations), sample size, true sex ratio, male standard deviation of the residuals function, and female standard deviation of the residuals function. Each taxon was then treated as if it were fossil, whereby the sex of each individual was unknown and assigned using the simple residual-based method. Effect size was estimated, and interval spread/separation were calculated at the largest individual of each sex.

The results show that evolutionary shifts in sexual variation of sufficient magnitude can be detected in a phylogeny (Fig. 12A). Intervals are more likely to overlap when true effect size is low. Although bias at low true effect size is present (to a similar extent to the above simulations with stricter conditions), in general, estimated effect size tracks true effect size (Fig. 12B). Altering factors such as sex ratio and intrasexual variation contributes noise to the data away from perfect estimation, but the correlation between true and estimated effect size is still largely present on average.

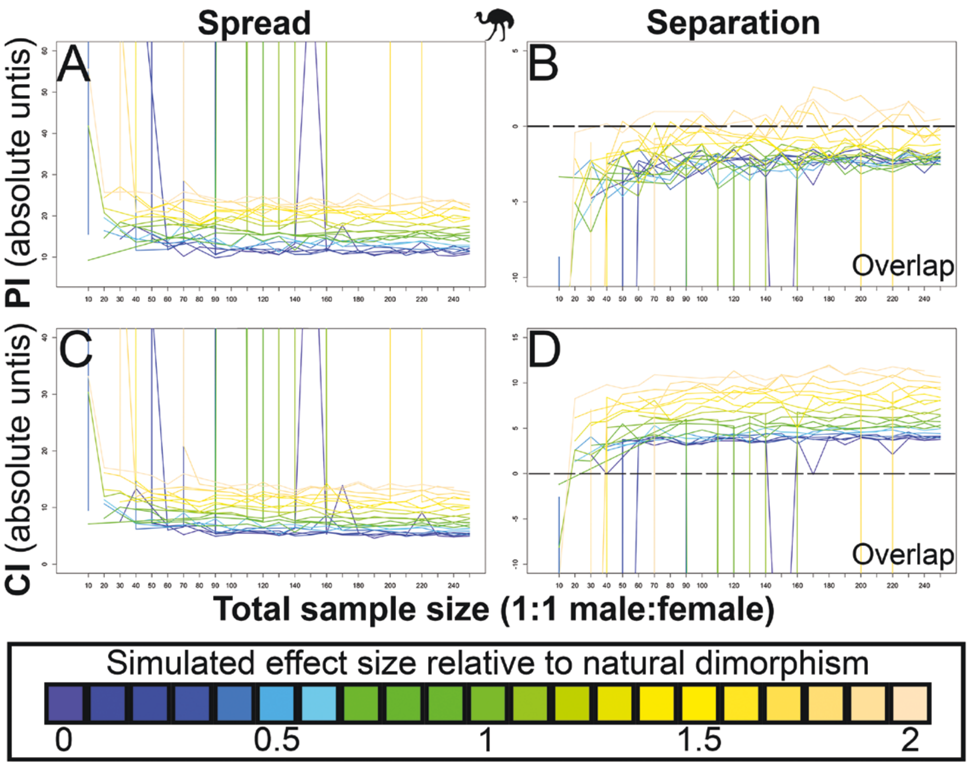

**Figure 11.** Prediction (PI) (A, B) and confidence (CI) (C, D) interval spread (A, C) and separation (B, D) for simulated rhea data. Dashed line indicates a value of zero (in absolute units: kg) for interval separation. See Figure 7 legend for further details.

### *Discussion of our quantitative approach*

Hone & Mallon (2017) showed that when similar simulations are done using the empirically derived parameters of naturally occurring alligator and rhea populations, *t*-tests are not always effective at detecting dimorphism, even under the unlikely assumption that the sex of all individuals is known. When the sex of simulated alligator and rhea individuals are assigned using the residual-based approach (while holding sex ratio and intrasexual variation at parity, as in our earlier simulations), *P*-values derived from Hartigans' dip tests and *t*-tests can increase with decreasing sample size and are likely to be non-significant ($\alpha = 0.05$) (Figs 13, 14). Significant *P*-values often require large sample sizes and, more importantly, very large effect sizes. Rhea simulations are more likely to result in lower *P*-values than alligator simulations (as in Hone & Mallon, 2017), due to the fact that most of the simulated data points are fully-grown adults at maximal dimorphism. Elevated intrasexual variation and skewed sex ratios are not expected to improve significance testing results when sex is not independently known (Supporting Information, Appendix S1).

Instead, we emphasize effect size statistics rather than significance testing (Nakagawa & Cuthill, 2007; Amrhein *et al*., 2017; Amrhein & Greenland, 2018; Halsey, 2019; Holland, 2019). While univariate significance tests suffer from low statistical power and can readily give false negatives (Hone & Mallon, 2017; Mallon, 2017), our framework is more useful when dealing with small sample sizes, since reductions in sample size lead to reductions in statistical power. When sample size decreases, variability in the accuracy of sex assignment and variability in the accuracy of estimated effect size will increase across repeated samplings, while confidence intervals of the estimated effect size increase. The intervals (i.e. uncertainty in the effect size estimate) are also dependent upon the spread of the data (i.e. intrasexual variation), with larger spreads leading to larger intervals. The use of bivariate data to estimate sex-specific growth curves means that the inclusion of juveniles does not detract from our ability to detect dimorphism. Instead, juveniles enhance detection by making regression analysis computationally easier.

Our statistical method is often expected to reliably quantify sexual variation and, when data are collected

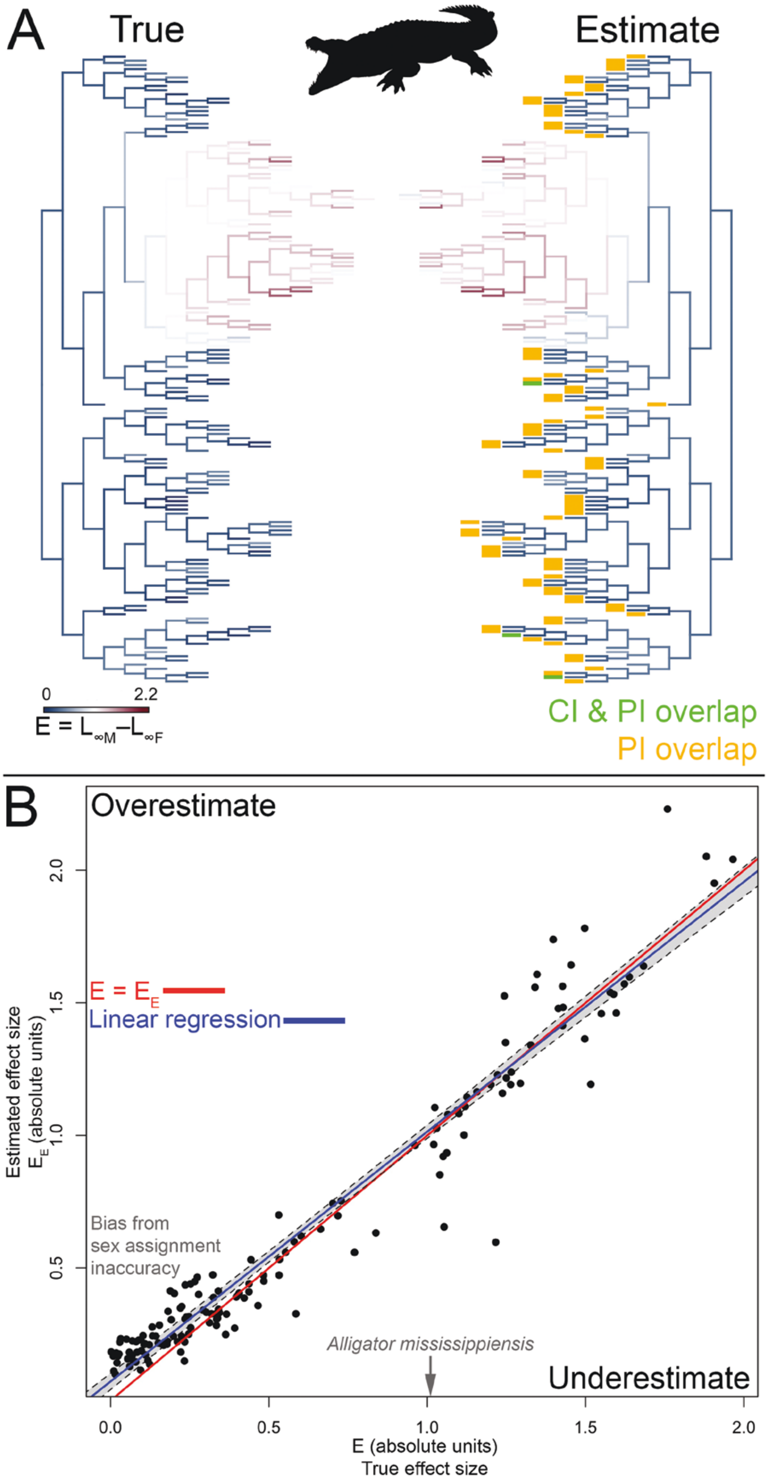

**Figure 12.** Phylogenetic simulation of a hypothetical crocodilian clade. Each taxon randomly and independently varies in true effect size ($\sim 0.00 \le E \le \sim 1.97$), sample size (16–99), true sex ratio (~25–74% males), male standard deviation of the residuals ($\sim 0.63 \le d \le \sim 1.51$), and female standard deviation of the residuals ($\sim 0.55 \le d \le \sim 1.51$), where $d$ is multiplied to the average of natural male and female alligator functions for standard deviation of residuals ($0.04605\log(C) + 0.0638$) such that both simulated sexes cover the natural intrasexual variation in alligator and also extend beyond it. A, the phylogeny with colour-coded true (left) and estimated (right, $E_E$) effect size (from dark blue at low effect size, to white, to dark red at

on traits likely to be sexually selected and compared to traits in the same sample/species unlikely to be sexual, help rule out non-sexual variation/polymorphisms in populations. Given the prevalence of dimorphism among living animals, palaeontologists should identify species with large estimates for sexual variation (i.e. effect sizes) and high estimated support for dimorphic growth models compared to other phylogenetically proximate species, rather than using significance testing on individual datasets to dichotomously reject or fail to reject monomorphism. Given the evidence for sexual selection in non-avian dinosaurs and their extant phylogenetic bracket, the question is not whether a given species was dimorphic, but rather which species show the strongest evidence of dimorphism relative to others, especially within a clade. Examining sexual selection across a phylogeny (and if fossil data is used to calibrate node ages, across time) using a measure of sexual dimorphism is precedented [e.g. plumage dichromatism as a proxy for sexual selection in extant birds (Cooney *et al*., 2019)]. Estimating effect size allows for phylogenetic comparative methods to examine evolutionary rates and patterns, as well as ancestral state reconstruction, with respect to sexual selection so long as evolutionary trends are strong enough to overcome overestimation bias from small effect sizes.

Some clarifications should be made about our method's inclination (i.e. bias), as currently presented, to overestimate effect size in datasets with minimal sexual variation and large overlap between the sexes:

(1) Finite sample size statistics, evolutionary/game theory, and fossil evidence of variation in structures consistent with sexual variation in modern taxa suggests that sexual variation in anisogamic species is the norm rather than the exception. Since minimally sexually variable species (at least with respect to certain traits) are not necessarily predicted to be common, one must approach statistical analyses accordingly:

> "Thoughtful research looks ahead to prospective outcomes in the context of theory and previous research. Researchers would do well to ask, *What do we already know, and how certain are we in what we know?* And building on that and on the field's theory, *what magnitudes of differences, odds ratios, or other effect sizes are practically important?"* (Wasserstein *et al*., 2019: p. 4, emphasis theirs).

(2) Species with minimal sexual variation are those more likely to have overlapping confidence intervals. Therefore, some uncertainty is accounted for, according to the degree that assumptions are met based on sex assignment accuracy.

(3) We encourage comparisons of effect size and uncertainty between datasets/taxa, especially within a phylogenic context. When differences in effect size are high between datasets/taxa and evolutionary trends are strong, meaningful differences/trends can be detected, even if species with minimal sexual variation are overestimated.

A dataset with effect size = 0 will result in overestimation under this approach to sex assignment with the size of the bias dependent on the intrasexual variation. As true effect size increases, bias is reduced; as overlap between the largest/oldest adults of each sex is reduced, estimated effect size will begin to track true effect size closely (Supporting Information, Appendix S1). These later datasets can be distinguished from those with low effect size affected by bias. Unlike univariate significance tests that struggle with juvenile data and small effect sizes, our framework's performance improves as the largest/oldest adults become more distinguishable between sexes. Comparing estimates between datasets is crucial, especially within a clade that might show similar intrasexual variation across species. If traits with high estimated effect size are suspected to be overestimated, one can, in addition to examining the calculated uncertainty, compare plausibly sexual vs. non-sexual traits to develop a baseline level of non-sexual variation in species. Mixture modelling while accounting for growth curves also has promise in estimating intrasexual variation.

(4) Overestimation bias in these instances results from inaccurate sex assignment. We presented the simplest method of sex assignment as an illustrative example. Future work can improve statistical sex assignment methods (e.g. incorporating mixture modelling). When fossil evidence allows for confident sex assignment, all data points with that diagnosable feature can be assigned to one sex, increasing sex assignment accuracy.

high effect size) and interval overlap (confidence and prediction in green, prediction only in orange). Sex assigned by the residual method to yield $E_E$. Interval spread/separation measured at the largest individual of each assigned sex. B, plot of $E$ vs. $E_E$ of each taxon. Perfect estimation in red. Linear regression in blue with 95% confidence intervals (black dashes). Natural alligator $E = 1.01$ (Wilkinson & Rhodes, 1997).

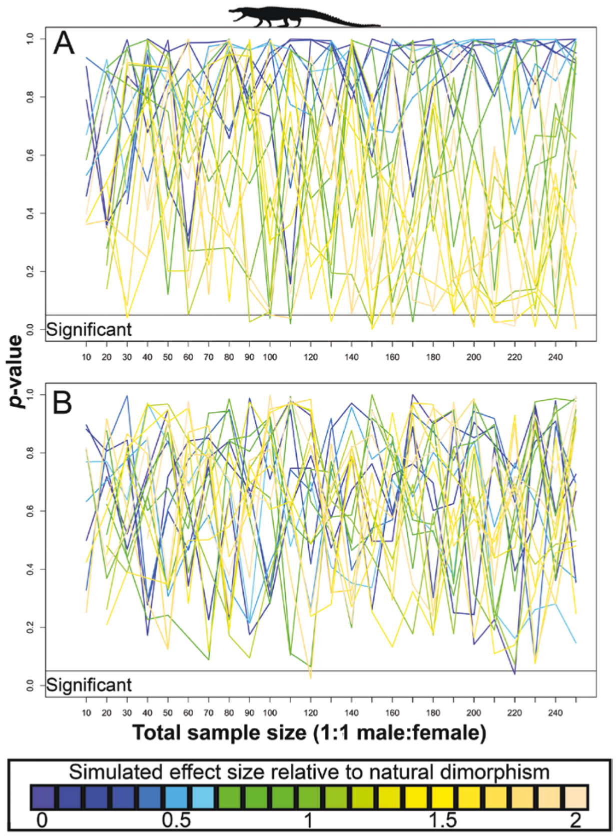

**Figure 13.** Hartigans' dip tests for unimodality (A) and *t*-tests (B) using residual sex assignment method on simulated alligator data. Black line indicates $\alpha$ = 0.05. See Figure 7 legend for further details. This run of the code is different from the run displayed in Figures 7, 9 and 10.

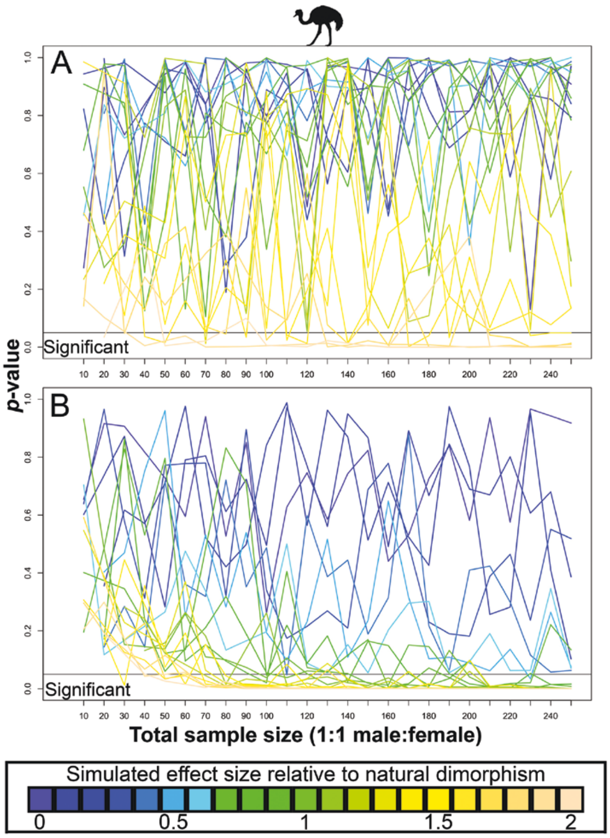

**Figure 14.** Hartigans' dip tests for unimodality (A) and *t*-tests (B) using the residual sex assignment method on simulated rhea data. Black line indicates $\alpha$ = 0.05. See Figure 7 legend for further details. This run of the code is different from the run displayed in Figures 8, 9 and 11.

When a distribution with a mean of zero is examined as a magnitude, the mean of absolute values will be upwardly biased (Hansen, 2016; Morrissey, 2016). This bias in statistics of magnitude may not be fully relevant for sexual variation between two distributions, but researchers have nevertheless proposed corrective methods (Morrissey, 2016). We do see this bias with sex role reversal (i.e. negative effect sizes are overestimated as positive), since our method simply attempts to distinguish two sexes, not males vs. females, unless unambiguous markers of sex are sufficiently present in the sample.

### ALTERNATE HYPOTHESIS 2: CONTROLLING FOR ONTOGENETIC VARIATION

Observed variation can result, at least partly, from differences between juveniles and adults, with the most extreme examples being species with larval stages and metamorphosis. In a reanalysis of the morphometric data used by Dodson (1976) to propose sexual dimorphism in *Protoceratops*, Chapman *et al*. (2008) found the greatest differences in both principal coordinate and principal component morphospace are between juveniles and large, adult specimens. Purported dimorphism among the adult individuals is of much smaller magnitude and altogether dependent on how missing measurements are treated in the analysis (Chapman *et al*., 2008). Even if most variation present in a species is ontogenetic, sexual variation can still co-occur and even be extreme relative to sexual variation of other species. Complex organisms would be expected to show greater ontogenetic variation than sexual variation, especially with respect to body size variation in multicellular animals. Even accepting dimorphism between adults, juveniles can still form a separate data cluster in a morphospace, creating an impression that there are three morphs. Therefore, before a recovered signal can be confidently attributed to sexual selection, potential ontogenetic causes need to be accounted for.

By focusing the quantitative aspects of our framework on detecting divergence between growth curves, the alternate hypothesis that the observed variation is primarily ontogenetic is already largely accounted for. Because growth curves for fossils are often generated from growth markers in histological cross-sections of bones, these provide a means for establishing corresponding degrees of maturity between putative sexual morphs. There may be considerable individual variation of growth curves in a population (Sander & Klein, 2005; Woodward *et al*., 2015); such variation among growth curves of a single population is taken into account under the intrasexual variation modeled here (i.e. population standard deviations of the residuals).

However, further observations can be made to more thoroughly test an ontogenetic alternate hypothesis. Some evidence, such as external fundamental systems (EFS) in bone marking the cessation of somatic growth, can be useful to demonstrate that a smaller or less developed morph is not growing into the other morph, especially when size is used as a proxy for age in the quantitative analysis (Saitta, 2015).

Other skeletal indicators that may be useful in establishing age or maturity (Hone *et al*., 2016b) include surface bone texture (Brown *et al*., 2009), markers of reproductive maturity like the presence of a medullary bone (Lee & Werning, 2008), tooth eruption and wear in mammals (Maschenko, 2002; Lister, 2009), or fusion of bone sutures (Brochu, 1996), although none of these indicators are fully reliable across a wide range of taxa (Tumarkin-Deretzian *et al*., 2006; Irmis, 2007; Cerda *et al*., 2014). Therefore, careful consideration of which ontogenetic indicators are appropriate depends on the clade of interest.

Some methods to detect sexual maturity (or sex) may not work with fossils; for example, immunochemical techniques that report binding to specific endogenous, intact molecular organic markers of reproductive medullary bone (Schweitzer *et al*., 2016) may be false positives in that they appear to bind to organic material unlikely to survive fossilization (Saitta *et al*., 2018; Saitta & Vinther, 2019). Other methods of ontogenetic analysis, such as the parsimony-based, size-independent method of ontogenetic sequence analysis (OSA) that accounts for developmental sequence polymorphism, have also been proposed (Colbert, 1999; Colbert & Rowe, 2008; Griffin & Nesbitt, 2016), but should be further examined for compatibility with our framework. For quantitative aspects of our approach, continuous age/age proxy data work best for regression analysis. Discrete age proxies are still useful in identifying specimens of each putative morph with ceased growth, providing further evidence against the alternate hypothesis of ontogenetic variation.

### ALTERNATE HYPOTHESIS 3: CONTROLLING FOR INTERSPECIFIC VARIATION

Palaeontologists mostly use diagnostic morphological features to work out taxonomic relationships and delineate species; however, traits that can greatly affect an organism's ecology should be considered regardless of whether or not they have been used historically as taxonomic indicators. To rule out interspecific variation, putative sexual morphs must coexist temporally and geographically without evidence for niche partitioning in other parts of their anatomy, indicating that they are not two species under the influence of the competitive exclusion principle (Hardin, 1960). An example is a multi-individual bonebed in which dimorphism is exhibited in the putative secondary sexual characteristic or body size, but not readily apparent in other traits associated with diet or locomotion (Saitta, 2015), further urging comparison of estimated sexual variation between plausibly sexual and non-sexual traits in a population. A corollary is that potential secondary sexual characteristics, especially those with high variability, should be carefully considered before their use as diagnostic taxonomic characters. Extreme body size variation might arise from interspecific niche partitioning [e.g. as has traditionally been quantified as 'Hutchinson's rule/ratios', but challenged as artefactual (Eadie *et al*., 1987)].

Chronological constraint is important given the morphological changes that can occur in closely related specimens through a stratigraphic section. Stratigraphic discrepancies between supposed dimorphs have been used to reject hypothesized sexual dimorphism in *Lambeosaurus* (Evans & Reisz, 2007), *Chasmosaurus* (Mallon & Holmes, 2006) and *Centrosaurus*/*Styracosaurus* (Ryan *et al*., 2007). Note that it is difficult to ascribe morphological changes through time to anagenesis when changing biogeographic ranges and migrations into and out of depositional basins cannot be ruled out [e.g. the suggestion that short-nasal-horned *Triceratops horridus* of the lower Hell Creek Formation evolved into long-nasal-horned *Triceratops prorsus* of the upper Hell Creek Formation (Scannella *et al*., 2014)].

Given the form-function link in biology, sexual dimorphism can be linked to sex differences in behaviour or ecology arising from dimorphic morphology or size (Pearson *et al*., 2002; Radford & Du Plessis, 2003), meaning that evidence for shared niches among the sexes might vary between taxa. Researchers should still strive to show that putative dimorphs with similar taxonomic and ecological diagnostic features co-existed in time and space.

### ALTERNATE HYPOTHESIS 4: CONTROLLING FOR INTRA-INDIVIDUAL OR INTRA-STRUCTURAL VARIATION

Certain structures that occur multiple times in a single individual can exhibit extensive intra-individual variation [e.g. osteoderms (Saitta, 2015) or feathers (Lucas & Stettenheim, 1972)]. Structures can show multiple occurrences within an individual sequentially over time as well as in multiple positions along the body, such as moulting feathers or annually shed antlers. Such variation can be controlled for in quantitative analysis by directly comparing equivalent, topologically homologous structures (e.g. only measuring the main shoulder spike of nodosaur specimens) or by comparing whole-individual morphometric profiles of repeating structures (e.g. measuring spine length along the lateral-most row of spines over the entire body length of nodosaur specimens). In incomplete/disarticulated fossils, it might be required to demonstrate that intra-individual variation in the structures is limited (i.e. lesser than inter-individual variation) or an appropriate approximation for ontogenetic variation (e.g. smaller and larger structures within an individual vary in a manner that mirrors variation between the structures on smaller/younger and larger/older individuals).

Although disorders of sexual developmental can result in extreme intra-individual variation [e.g. bilateral gynandromorphism (Lillie, 1931)], the occurrence of intersex individuals can be very rare in species selected for discrete male and female individuals (Bojesen *et al.*, 2003; Stockholm *et al.*, 2006). Unless dealing with a clade that exhibits high prevalence of simultaneous or sequential (i.e. sex change) hermaphroditism (Warner, 1975; Heath, 1977), intra-individual sex-related variation due to developmental disorders or selective pressures for hermaphroditism can likely be assumed to be rare in the fossil record. Fossil specimens suspected to show dramatic developmental disorders (e.g. Buffetaut *et al.*, 2006) or unexpected instances of hermaphroditism should be tested for authenticity (Rowe *et al.*, 2001) before inclusion in a dataset.

Complex structures might show intra-structural variation that should be accounted for (e.g. barb morphology along the rachis in a feather) by comparing equivalent/homologous portions of the structure (e.g. the very apical end of a feather), equivalent metrics between different specimens (e.g. a measure of maximum fractal branching order of the feather), or whole-structure morphometric profiles (e.g. a profile of barb angle from base to apex along the entire rachis of a feather).

### Alternate hypothesis 5: controlling for pathological variation

Pathologies can add variation to a fossil dataset, but they can usually be identified by atypical frequencies (i.e. unlike ~1:1 male:female ratio in some large, random samples) and distinctive diagnostic anatomical features (Moodie, 1918, 1923; Tanke & Rothschild, 2002). Although some secondary sexual characteristics can show asymmetry, one should consider whether extreme asymmetry might indicate pathology. Pathological specimens can be retro-deformed or dropped from a dataset.

### Alternate hypothesis 6: controlling for taphonomic variation

Other sources of variation might be taphonomic (e.g. burial, transport, scavenging, decay, weathering, or erosion). Weathering of organically preserved soft tissues (Vinther, 2015) or plastic deformation of bones (Boyd & Motani, 2008), for example, could introduce noise to a dataset. Hedrick & Dodson (2013) suggested that taphonomy alone might explain differences between purported species or morphotypes of *Psittacosaurus lujiatunesis*. Taphonomic consideration might include examining fossils for deformation, incompleteness or breakage. Scanning electron microscopy on carbonaceous compression fossils can identify mouldic melanosome impressions in sediment within areas where dark organic stains of melanin were lost through oxidative weathering (Vinther, 2015). One can reduce taphonomic noise in quantitative analyses by eliminating highly altered specimens from the dataset, re-running analyses with only the best-preserved specimens, or attempting to retro-deform or reconstruct specimens (Boyd & Motani, 2008; Saitta, 2015).

## Biological implications

Once alternate hypotheses for observed variation are weakened, then the type and magnitude of estimated sexual variation can be used to draw biological inferences. The degree of dimorphism might indicate the extent to which males and females exhibited similar behaviours (e.g. mating behaviour) and, in extreme cases, ecologies. Dimorphism magnitude might provide insight into the social/sexual system (e.g. polygamy vs. monogamy, social group size or sexual segregation). Monogamous species often show less body size dimorphism than polygynous species (Lack, 1968; Gautier-Hion, 1975; Clutton-Brock *et al.*, 1977; Clutton-Brock & Harvey, 1978; Alexander *et al.*, 1979; Shine, 1979; Clutton-Brock, 1985; Heske & Ostfeld, 1990). Bird species with polygynous or lek mating systems have greater dimorphism in body mass, plumage, wing length and tail length compared to monogamous species (Dunn *et al.*, 2001). Mating system inferences from estimated magnitudes of sexual dimorphism have been made using fossils—with *Canis dirus* hypothesized as monogamous, similar to most extant canids, and *Smilodon fatalis* hypothesized as more monogamous compared to modern lions (Van Valkenburgh & Sacco, 2002). Sexual segregation is seen in dimorphic ruminants (Bowyer, 2004). Increased group size in bovids is correlated with increased body size dimorphism and male horn length (Bro-Jørgensen, 2007). In cervids, male antler size, polygyny and breeding group size are correlated (Clutton-Brock *et al.*, 1980). Fossil evidence might reveal whether a species was sex-role reversed (e.g. unequivocal markers of female sex, such as *in situ* eggs, in individuals with more exaggerated secondary sexual characteristics).

Simplistic inference into the relative importance of intrasexual competition vs. intersexual mate choice can be hypothesized depending on whether the dimorphic structure is consistent with an armament or ornament, respectively (Sullivan *et al.*, 2003; Saitta, 2015). For example, some stegosaurs and ceratopsians possess structures that are consistent with ornaments (e.g. plates and frills) alongside structures possibly consistent with armaments (e.g. spikes and horns). Could this indicate the presence of both intrasexual combat to compete for mates in addition to, or followed by, intersexual display and mate choice (Molnar, 1977; Farke *et al.*, 2009; Krauss *et al.*, 2010)?

Dimorphism magnitude and the nature of the anatomical structure might help identify primary vs. secondary functions of traits, such as females using traits sexually selected for in males as predator deterrents [e.g. possibly stegosaur osteoderms (Saitta, 2015) or ceratopsian horns]. Even when the estimated magnitude of and support for dimorphism are great, multiple/secondary functions for structures should be considered (Farke, 2014), particularly when the sex with the less exaggerated trait still possesses the trait [e.g. uniquely among cervids, female reindeer (*Rangifer tarandus*) possess antlers, presumably for scraping away snow during winter feeding].

The fossil record can illuminate sexual selection's effect on macroevolution, and certain evolutionary trends might in turn provide evidence that a trait is sexually selected and likely dimorphic. Although there is mixed evidence that sexual selection can increase speciation rates, for example through the formation of reproductive barriers (Panhuis *et al.*, 2001; Kraaijeveld *et al.*, 2011), our effect size approach may allow fossils to supplement extant data. When reconstructing evolutionary histories of sexual dimorphism, habitat and social behaviour (Pérez-Barbería *et al.*, 2002), fossil data could supplement phylogenetic and ecological analyses (Pringle, 2020).

### EXAMPLE: EMPIRICAL NON-AVIAN DINOSAUR DATASETS

Figure 15 shows the results of our divergence analysis on data from *Maiasaura peeblesorum* (Woodward *et al.*, 2015), *Psittacosaurus lujiatunensis* (Erickson *et al.*, 2015) and *Tyrannosaurus rex* (Erickson *et al.*, 2004; Horner & Padian, 2004; Lee & Werning, 2008). One should not categorically conclude that any one of these taxa does or does not exhibit sexual dimorphism. Instead, it is better to conclude simply that given the current data and assumptions (e.g. logistic growth, independence of data points, and reliable estimation/assignment of body mass, age and sex), *Maiasaura* exhibits the largest standardized estimate of sexual variation (i.e. percent change in asymptotic size from smaller to larger assigned sex) and the lowest uncertainty in that estimate (i.e. tightest confidence intervals), a pattern otherwise unapparent in univariate histograms. *Maiasaura* data also derive from a single bonebed (unlike the other two datasets), providing tight stratigraphic and geographic control, and perform similarly well under Gompertz regression (Supporting Information, Appendix S1). Furthermore, both assigned sexes contain at least one individual with an EFS, indicating that these individuals have ceased growth. *Psittacosaurus* data suffer from a lack of older individuals, while *Tyrannosaurus* data suffer from small sample size. Therefore, *Maiasaura* currently shows better evidence for sexual variation than both *Psittacosaurus* and *Tyrannosaurus*.

## CONCLUSION

Sexual variation is prevalent in animals generally and archosaurs specifically and is expected in extinct archosaurs, such as non-avian dinosaurs. Hypotheses to explain a supposedly unique absence of sexual dimorphism in non-avian dinosaurs among other fossil groups (either for certain traits/species or more broadly across the grade) are either inconsistent with sexual selection theory or invoke rare social/sexual equilibria unlikely to be applicable across a long-lived, diverse grade. Non-avian dinosaurs show many examples, along multiple lines of evidence, of sexual selection, structures consistent with secondary sexual characteristics, and variation in those structures.

Much discussion of sexual dimorphism in non-avian dinosaurs centres around unreliable significance testing, most recently tests for bimodality. Univariate significance tests suffer from low power and fail to investigate the truest signal of sexual dimorphism (e.g. they struggle with confounding ontogenetic variation if the data contains juveniles). Here, we analyse growth curve divergence of plausibly sexual traits (which can be compared to plausibly non-sexual traits in the same species). Data simulation demonstrates an appreciable amount of success in our approach. This divergence analysis should be coupled

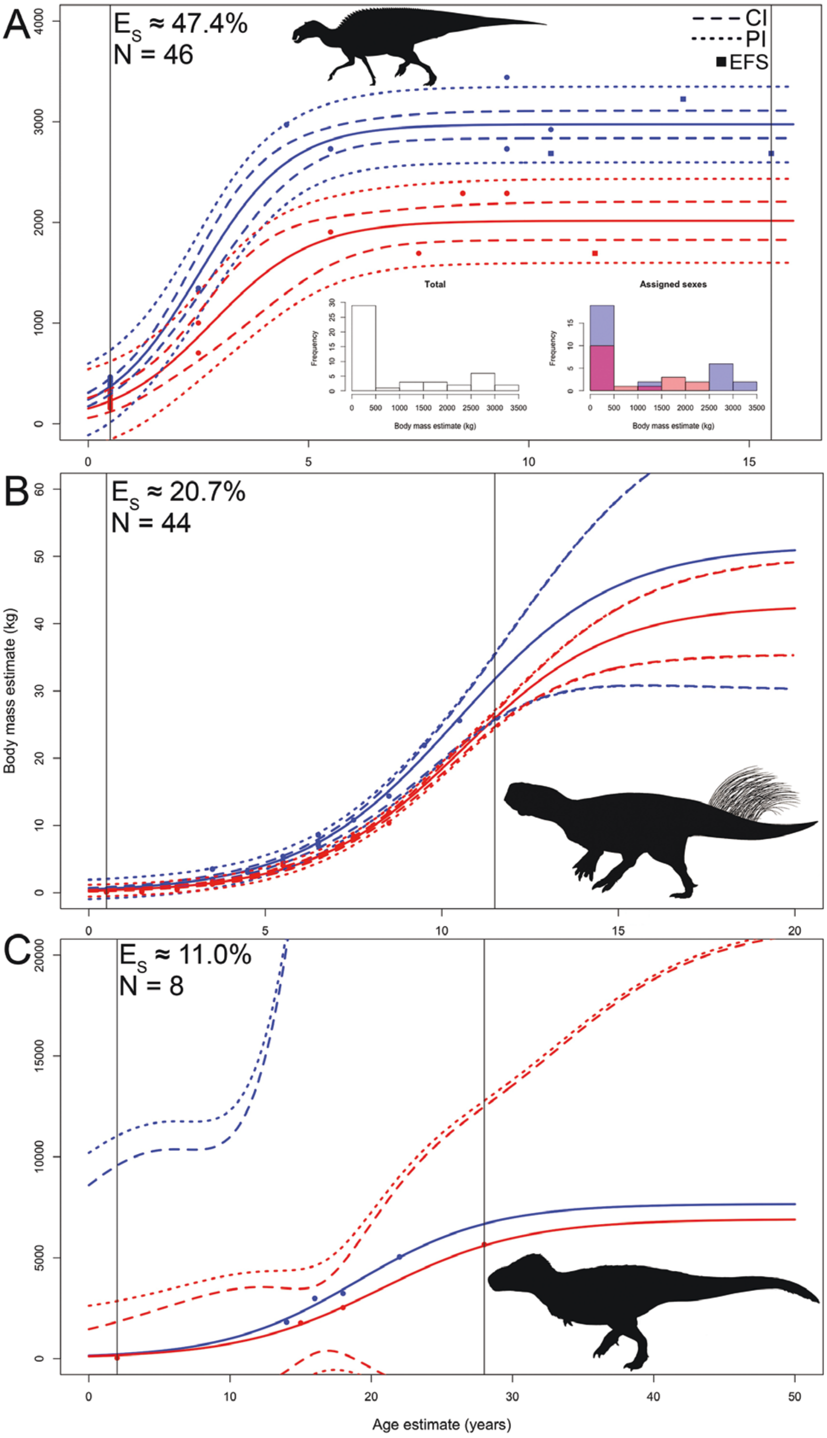

**Figure 15.** Sex-specific logistic regressions of (A) *Maiasaura peeblesorum* (Woodward *et al.*, 2015) with EFS-bearing *Maiasaura* indicated and histograms inset, (B) *Psittacosaurus lujiatunensis* (Erickson *et al.*, 2015), and (C) *Tyrannosaurus rex* (Erickson *et al.*, 2004; Horner & Padian, 2004; Lee & Werning, 2008). Sex assigned by residual method. Ninety-five percent confidence (CI) and prediction (PI) intervals shown. $N$ = sample size. $E_S$ = standardized effect size (i.e. percent change in asymptote L parameter from smaller to larger assigned sex). Vertical lines delineate youngest and oldest specimens.

with the testing of alternate hypotheses for observed variation (e.g. controlling for non-sexual variation/polymorphisms in populations, ontogeny, interspecific variation, pathology, taphonomy, intra-individual or intra-structural variation) and utilizes effect size statistics. By attempting to assign sex of individuals, one can compare the estimated degree of and support for sexual variation in a fossil dataset to that of other datasets. The question is not whether a given species was sexually dimorphic, but which species show better evidence for and larger estimated magnitudes of sexual variation than others. Our approach is appropriate given the immense influence of sexual selection on sexually reproducing species and low sample sizes typical of the fossil record.

Our illustrative sex assignment method is consistent with studies showing that dividing a sample about the mean is the method least sensitive to variation in sample size, intrasexual variation, or sex ratio and is one of the better ways to estimate dimorphism (Plavcan, 1994; Rehg & Leigh, 1999). When dimorphism magnitude is small, one is limited to estimating a maximum amount of possible dimorphism in a univariate unimodal distribution (Godfrey *et al*., 1993). Similarly in our approach, overestimation bias occurs when effect size is small, but which only requires separation of the largest/oldest adults to produce reasonably accurate estimates. These older studies highlight how nearly two decades of research on this topic by vertebrate palaeontologists, particularly dinosaur researchers, have in some cases been influenced by poor interpretations of sexual selection theory or unrealistic statistical expectations.

The crucial bottleneck to our approach is sex assignment accuracy. Even simplistic approaches can be expected to be better than random in many circumstances and no worse than random on average when randomly sampling non-sex-skewed populations, with relative frequency of each sex approaching parity as sample size increases. Future work should improve statistical methods of sex assignment and palaeontological methods of sex identification. It would similarly be good to develop approaches to account for intrasexual variation in order to better choose between monomorphic or dimorphic models, since it is the combination of effect size and intrasexual variation that dictates whether the data clearly diverges along growth curves. Focusing on a single clade suspected to show similar amounts of intrasexual variation between species might be one approach, as well as establishing species-specific baselines of intrasexual variation by comparing plausibly sexual vs. non-sexual traits (O'Brien *et al*., 2018). Mixture modelling while accounting for growth curves might improve sex assignments and estimates of intrasexual variation.

Standardizing effect sizes of dimorphism is crucial when comparing across disparate datasets. Future work could determine whether other measures of effect size (e.g. Cohen's d), model selection (e.g. AIC), or comparisons between correlations of monomorphic and dimorphic models (e.g. $R^2$) are compatible with our framework. The divergence point in a dataset and onset of putative sexual maturity/pubescence might be identified using multiple regression.

We recommend attempting to constrain sexual dimorphism magnitude using confidence interval spread/separation. Rather than binary rejection or non-rejection of sexual monomorphism on a species-by-species basis in a manner prone to false negatives, effect size statistics enables investigation of sexual selection across phylogenies and geologic time.

The ideas that non-avian dinosaurs do not show evidence of sexual dimorphism or that independent knowledge of the sex of all specimens might be required to investigate the problem are needlessly stringent in many cases, and non-significant *P*-values cannot be interpreted as 'no evidence' for an effect. Returning to a previous quote, Mallon (2017) stated that "…in the absence of a priori knowledge of sex, the ability to detect dimorphism in a fossil sample is likely only in cases of strongly expressed dimorphism (i.e. well-separated peaks on a histogram or discrete character states)…" (p. 502). Our simulations show that, in many cases, reasonably accurate estimates of sexual variation can be obtained even if datasets do not show bimodality and the sex of all individuals is unknown. We take an opposing viewpoint of Mallon (2017) when he states, "Such a conservative approach [the bimodality method] is subject to type II error, but this is preferable to attributing sexual dimorphism to every perceived instance of intraspecific variation in the fossil record, particularly if there is no compelling reason to think it should exist in the first place" (p. 501). Rather, sexual selection is too influential a process to be disregarded (as a result of type II error) as one of the major sources of variation in palaeobiology.

## ACKNOWLEDGEMENTS

We thank Steven Holland, Gene Hunt, Jordan Mallon and four anonymous reviewers for their helpful comments on the manuscript, Sam Houston (Stirling University, Scotland) for helpful comments on the R code, and John Allen for his editorial efforts.

Data sources are acknowledged as follows. *Alligator mississippiensis* [Scott Hartman, Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported

(CC BY-NC-SA 3.0), https://creativecommons.org/licenses/by-nc-sa/3.0/legalcode], *Rhea americana* [Darren Naish, vectorized by T. Michael Keesey, Creative Commons Attribution 3.0 Unported (CC BY 3.0), https://creativecommons.org/licenses/by/3.0/legalcode], *Crocodylus porosus* [Steven Traver, CC0 1.0 Unviersal (CC0 1.0) Public Domain Dedication, https://creativecommons.org/publicdomain/zero/1.0/legalcode], *Maiasaura peeblesorum* [Scott Hartman, Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported (CC BY-NC-SA 3.0), https://creativecommons.org/licenses/by-nc-sa/3.0/legalcode], and *Tyrannosaurus rex* [Emily Willoughby, Creative Commons Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0), https://creativecommons.org/licenses/by-sa/3.0/legalcode] silhouettes were obtained through phylopic.org. *Psittacosaurus* silhouette modified from the sculpture by Robert Nicholls. P.J.M. is currently supported by National Science Foundation awards PLR 1341645 and FRES 1925884.

## REFERENCES

**Abella J, Valenciano A, Pérez-Ramos A, Montoya P, Morales J. 2013.** On the socio-sexual behaviour of the extinct ursid *Indarctos arctoides*: an approach based on its baculum size and morphology. *PLoS One* **8:** e73711.

**Alexander RD, Hoogland JL, Howard RD, Noonan M, Sherman PW. 1979.** Sexual dimorphisms, and breeding systems in pinnipeds, ungulates, primates and humans. In: Chagnon NA, Irons W, eds. *Evolutionary biology and human social behavior: an anthropological perspective*. North Scituate: Duxbury Press, 402–604.

**Alexander RM. 1989.** *Dynamics of dinosaurs and other extinct giants*. New York: Columbia University Press.

**Alexander RM. 2006.** Dinosaur biomechanics. *Proceedings of the Royal Society B: Biological Sciences* **273:** 1849–1855.

**Allentoft ME, Bunce M, Scofield RP, Hale ML, Holdaway RN. 2010.** Highly skewed sex ratios and biased fossil deposition of moa: ancient DNA provides new insight on New Zealand's extinct megafauna. *Quaternary Science Reviews* **29:** 753–762.

**Amrhein V, Greenland S. 2018.** Remove, rather than redefine, statistical significance. *Nature Human Behaviour* **2:** 4.

**Amrhein V, Greenland S, McShane B. 2019.** Scientists rise up against statistical significance. *Nature* **567:** 305–307.

**Amrhein V, Korner-Nievergelt F, Roth T. 2017.** The earth is flat (p> 0.05): significance thresholds and the crisis of unreplicable research. *PeerJ* **5:** e3544.

**Andersson MB. 1994.** *Sexual selection*. Princeton: Princeton University Press.

**Andersson M, Simmons LW. 2006.** Sexual selection and mate choice. *Trends in Ecology & Evolution* **21:** 296–302.

**Arbour VM, Currie PJ. 2011.** Tail and pelvis pathologies of ankylosaurian dinosaurs. *Historical Biology* **23:** 375–390.

**Arbour VM, Snively E. 2009.** Finite element analyses of ankylosaurid dinosaur tail club impacts. *The Anatomical Record* **292:** 1412–1426.

**Arbour VM, Zanno LE. 2018.** The evolution of tail weaponization in amniotes. *Proceedings of the Royal Society B: Biological Sciences* **285:** 20172299.

**Autier-Dérian D, Deputte BL, Chalvet-Monfray K, Coulon M, Mounier L. 2013.** Visual discrimination of species in dogs (*Canis familiaris*). *Animal Cognition* **16:** 637–651.

**Balmford A, Jones IL, Thomas AL. 1993.** On avian asymmetry: evidence of natural selection for symmetrical tails and wings in birds. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **252:** 245–251.

**Barden HE, Maidment SC. 2011.** Evidence for sexual dimorphism in the stegosaurian dinosaur *Kentrosaurus aethiopicus* from the Upper Jurassic of Tanzania. *Journal of Vertebrate Paleontology* **31:** 641–651.

**Barlow GW. 2005.** How do we decide that a species is sex-role reversed? *The Quarterly Review of Biology* **80:** 28–35.

**Barrett PM, McGowan AJ, Page V. 2009.** Dinosaur diversity and the rock record. *Proceedings of the Royal Society of London B: Biological Sciences* **276:** 2667–2674.

**Barrowclough GF, Cracraft J, Klicka J, Zink RM. 2016.** How many kinds of birds are there and why does it matter? *PLoS One* **11:** e0166307.

**Barsbold R, Osmólska H, Watabe M, Currie PJ, Tsogtbaatar K. 2000.** A new oviraptorosaur [Dinosauria, Theropoda] from Mongolia: the first dinosaur with a pygostyle. *Acta Palaeontologica Polonica* **45:** 97–106.

**Bell PR, Fanti F, Currie PJ, Arbour VM. 2014.** A mummified duck-billed dinosaur with a soft-tissue cock's comb. *Current Biology* **24:** 70–75.

**Birchard GF, Ruta M, Deeming DC. 2013.** Evolution of parental incubation behaviour in dinosaurs cannot be inferred from clutch mass in birds. *Biology Letters* **9:** 20130036.

**Birkhead TR, Møller AP, eds. 1998.** *Sperm competition and sexual selection*. Elsevier.

**Blackburn G. 2004.** *Ornaments and overflight behaviour of tufted puffins* (Fratercula cirrhata) *breeding on Triangle Island, British Columbia*. Unpublished Doctoral Thesis, Simon Fraser University.

**Bojesen A, Juul S, Gravholt CH. 2003.** Prenatal and postnatal prevalence of Klinefelter syndrome: a national registry study. *The Journal of Clinical Endocrinology & Metabolism* **88:** 622–626.

**Bonduriansky R. 2007.** Sexual selection and allometry: a critical reappraisal of the evidence and ideas. *Evolution* **61:** 838–849.

**Bonhomme V, Picq S, Gaucherel C, Claude J. 2014.** Momocs: outline analysis using R. *Journal of Statistical Software* **56:** 1–24.

**Borgia G, Gore MA. 1986.** Feather stealing in the satin bowerbird (*Ptilonorhynchus violaceus*): male competition and the quality of display. *Animal Behaviour* **34:** 727–738.

**Borkovic B, Russell A. 2014.** Sexual selection according to Darwin: a response to Padian and Horner's interpretation. *Comptes Rendus Palevol* **13:** 701–707.

**Bowyer RT. 2004.** Sexual segregation in ruminants: definitions, hypotheses, and implications for conservation and management. *Journal of Mammalogy* **85:** 1039–1052.

**Boyd AA, Motani R. 2008.** Three-dimensional re-evaluation of the deformation removal technique based on "jigsaw puzzling". *Palaeontologia Electronica* **11:** 7A.

**Brochu CA. 1996.** Closure of neurocentral sutures during crocodilian ontogeny: implications for maturity assessment in fossil archosaurs. *Journal of Vertebrate Paleontology* **16:** 49–62.

**Bro-Jørgensen J. 2007.** The intensity of sexual selection predicts weapon size in male bovids. *Evolution* **61:** 1316–1326.

**Brown CM, Russell AP, Ryan MJ. 2009.** Pattern and transition of surficial bone texture of the centrosaurine frill and their ontogenetic and taxonomic implications. *Journal of Vertebrate Paleontology* **29:** 132–141.

**Brown CM, Evans DC, Campione NE, O'Brien LJ, Eberth DA. 2013.** Evidence for taphonomic size bias in the Dinosaur Park Formation (Campanian, Alberta), a model Mesozoic terrestrial alluvial-paralic system. *Palaeogeography, Palaeoclimatology, Palaeoecology* **372:** 108–122.

**Brown CM, Henderson DM, Vinther J, Fletcher I, Sistiaga A, Herrera J, Summons RE. 2017.** An exceptionally preserved three-dimensional armored dinosaur reveals insights into coloration and Cretaceous predator-prey dynamics. *Current Biology* **27:** 2514–2521.

**Brusatte SL. 2012.** *Dinosaur paleobiology*. Chichester: John Wiley & Sons.

**Brusatte SL, Butler RJ, Prieto-Márquez A, Norell MA. 2012.** Dinosaur morphological diversity and the end-Cretaceous extinction. *Nature Communications* **3:** 804.

**Buffetaut E, Li J, Tong H, Zhang H. 2006.** A two-headed reptile from the Cretaceous of China. *Biology Letters* **3:** 81–82.

**Bull JJ, Bulmer MG. 1989.** Longevity enhances selection of environmental sex determination. *Heredity* **63:** 315–320.

**Bunce M, Worthy TH, Ford T, Hoppitt W, Willerslev E, Drummond A, Cooper A. 2003.** Extreme reversed sexual size dimorphism in the extinct New Zealand moa *Dinornis*. *Nature* **425:** 172.

**Burkhardt D. 1989.** UV vision: a bird's eye view of feathers. *Journal of Comparative Physiology A* **164:** 787–796.

**Cabrera D, Stankowich T. 2018.** Stabbing slinkers: tusk evolution among artiodactyls. *Journal of Mammalian Evolution* **27:** 1–8.

**Caldwell MW, Lee MS. 2001.** Live birth in Cretaceous marine lizards (mosasauroids). *Proceedings of the Royal Society of London B: Biological Sciences* **268:** 2397–2401.

**Caro TM, Graham CM, Stoner CJ, Flores MM. 2003.** Correlates of horn and antler shape in bovids and cervids. *Behavioral Ecology and Sociobiology* **55:** 32–41.

**Carpenter K. 1990.** Variation in *Tyrannosaurus rex*. In: Carpenter K, Currie PJ, eds. *Dinosaur systematics: perspectives and approaches*. Cambridge: Cambridge University Press, 141–145.

**Carpenter K. 1997.** Agonistic behavior in pachycephalosaurs (Ornithischia, Dinosauria); a new look at head-butting behavior. *Rocky Mountain Geology* **32:** 19–25.

**Carpenter K. 1998.** Armor of *Stegosaurus stenops*, and the taphonomic history of a new specimen from Garden Park, Colorado. *Modern Geology* **23:** 127–144.

**Carpenter K, Sanders F, McWhinney L, Wood L. 2005.** Evidence for predator–prey relationships: example for *Allosaurus* and *Stegosaurus*. In: Carpenter K, ed. *The carnivorous dinosaurs*. Bloomington: Indiana University Press, 325–350.

**Catchpole CK. 1987.** Bird song, sexual selection and female choice. *Trends in Ecology & Evolution* **2:** 94–97.

**Cederström P, Ahlberg P, Nilsson CH, Ahlgren J, Eriksson ME. 2011.** Moulting, ontogeny and sexual dimorphism in the Cambrian ptychopariid trilobite *Strenuaeva inflata* from the northern Swedish Caledonides. *Palaeontology* **54:** 685–703.

**Cerda IA, Chinsamy A, Pol D. 2014.** Unusual endosteally formed bone tissue in a Patagonian basal sauropodomorph dinosaur. *The Anatomical Record* **297:** 1385–1391.

**Chapman RE, Weishampel DB, Hunt G, Rasskin-Gutman D. 1997.** Sexual dimorphism in dinosaurs. In: Wolberg DL, Stump E, Rosenberg GD, eds. *Dinofest International: proceedings of a symposium held at Arizona State University, Philadelphia, Academy of Natural Sciences*, 83–93.

**Chapman RE, Galton PM, Sepkoski JJ Jr, Wall WP. 1981.** A morphometric study of the cranium of the pachycephalosaurid dinosaur *Stegoceras*. *Journal of Paleontology* **55:** 608–618.

**Chapman RE, Sadleir A, Dodson P, Makovicky P. 2008.** Handling missing data in paleontological matrices: approaches for exploratory multivariate analyses in morphometrics. *Journal of Vertebrate Paleontology* **28:** 62A.

**Chelliah K, Sukumar R. 2013.** The role of tusks, musth and body size in male–male competition among Asian elephants, *Elephas maximus*. *Animal Behaviour* **86:** 1207–1214.

**Chen J, Wang B, Zheng Y, Jarzembowski E, Jiang T, Wang X, Zheng X, Zhang H. 2019.** Female-biased froghoppers (Hemiptera, Cercopoidea) from the Mesozoic of China and phylogenetic reconstruction of early Cercopoidea. *Journal of Systematic Palaeontology* **17:** 1–13.

**Cheng YN, Holmes R, Wu XC, Alfonso N. 2009.** Sexual dimorphism and life history of *Keichousaurus hui* (Reptilia: Sauropterygia). *Journal of Vertebrate Paleontology* **29:** 401–408.

**Chinsamy A, Chiappe LM, Marugán-Lobón J, Chunling G, Fengjiao Z. 2013.** Gender identification of the Mesozoic bird *Confuciusornis sanctus*. *Nature Communications* **4:** 1381.

**Clutton-Brock TH. 1985.** Size, sexual dimorphism, and polygyny in primates. In: Jungers WL, ed. *Size and scaling in primate biology.* Advances in Primatology. Boston, MA: Springer.

**Clutton-Brock TH. 1985.** Size, sexual dimorphism, and polygyny in primates. In: *Size and scaling in primate biology*. Boston: Springer, 51–60.

**Clutton-Brock T. 2007.** Sexual selection in males and females. *Science* **318:** 1882–1885.

**Clutton-Brock TH, Albon SD, Harvey PH. 1980.** Antlers, body size and breeding group size in the Cervidae. *Nature* **285:** 565–567.

**Clutton-Brock TH, Harvey PH. 1978.** Mammals, resources and reproductive strategies. *Nature* **273:** 191–195.

**Clutton-Brock TH, Harvey PH, Rudder B. 1977.** Sexual dimorphism, socionomic sex ratio and body weight in primates. *Nature* **269:** 797.

**Colbert MW. 1999.** *Patterns of evolution and variation in the Tapiroidea (Mammalia: Perissodactyla)*. Unpublished Doctoral Thesis, The University of Texas at Austin.

**Colbert MW, Rowe T. 2008.** Ontogenetic sequence analysis: using parsimony to characterize developmental sequences and sequence polymorphism. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **310:** 398–416.

**Cooney CR, Varley ZK, Nouri LO, Moody CJA, Jardine MD, Thomas GH. 2019.** Sexual selection predicts the rate and direction of colour divergence in a large avian radiation. *Nature Communications* **10:** 1773.

**Croft DP, Darden SK, Ruxton GD. 2009.** Predation risk as a driving force for phenotypic assortment: a cross-population comparison. *Proceedings of the Royal Society of London B* **276:** 1899–1904.

**Cullen TM, Fraser D, Rybczynski N, Schröder-Adams C. 2014.** Early evolution of sexual dimorphism and polygyny in Pinnipedia. *Evolution* **68:** 1469–1484.

**Czerkas SA. 1992.** Discovery of dermal spines reveals a new look for sauropod dinosaurs. *Geology* **20:** 1068–1070.

**Darwin C. 1871.** *The descent of man and selection in relation to sex*, **Vol. 1**. London: John Murray.

**Davis EB, Brakora KA, Lee AH. 2011.** Evolution of ruminant headgear: a review. *Proceedings of the Royal Society B: Biological Sciences* **278:** 2857–2865.

**Ditchkoff SS, Lochmiller RL, Masters RE, Starry WR, Leslie DM Jr. 2001.** Does fluctuating asymmetry of antlers in white–tailed deer (*Odocoileus virginianus*) follow patterns predicted for sexually selected traits? *Proceedings of the Royal Society of London. Series B: Biological Sciences* **268:** 891–898.

**Dodson P. 1976.** Quantitative aspects of relative growth and sexual dimorphism in *Protoceratops*. *Journal of Paleontology* **50:** 929–940.

**Donald PF. 2007.** Adult sex ratios in wild bird populations. *Ibis* **149:** 671–692.

**Dunn PO, Whittingham LA, Pitcher TE. 2001.** Mating systems, sperm competition, and the evolution of sexual dimorphism in birds. *Evolution* **55:** 161–175.

**Dunning JB. 2007.** *CRC handbook of avian body masses*, *2nd edn*. Boca Raton: CRC Press.

**Eadie JM, Broekhoven L, Colgan P. 1987.** Size ratios and artifacts: Hutchinson's rule revisited. *The American Naturalist* **129:** 1–17.

**Eberhart-Phillips LJ, Küpper C, Carmona-Isunza MC, Vincze O, Zefania S, Cruz-López M, Kosztolányi A, Miller TE, Barta Z, Cuthill IC, Burke T. 2018.** Demographic causes of adult sex ratio variation and their consequences for parental cooperation. *Nature Communications* **9:** 1651.

**Emlen ST. 1972.** An experimental analysis of the parameters of bird song eliciting species recognition. *Behaviour* **41:** 130–171.

**Emlen ST, Wrege PH. 2004.** Size dimorphism, intrasexual competition, and sexual selection in wattled jacana (*Jacana jacana*), a sex-role-reversed shorebird in Panama. *The Auk* **121:** 391–403.

**Endler JA. 1984.** Natural and sexual selection on color patterns in poeciliid fishes. In: Zaret TM, eds. *Evolutionary ecology of neotropical freshwater fishes*. Developments in environmental biology of fishes, Vol. **3**. Dordrecht: Springer, 95–111.

**Endler JA. 1988.** Sexual selection and predation risk in guppies. *Nature* **332:** 593–594.

**Erickson GM, Currie PJ, Inouye BD, Winn AA. 2006.** Tyrannosaur life tables: an example of nonavian dinosaur population biology. *Science* **313:** 213–217.

**Erickson GM, Makovicky PJ, Inouye BD, Zhou CF, Gao KQ. 2009.** A life table for *Psittacosaurus lujiatunensis*: initial insights into ornithischian dinosaur population biology. *The Anatomical Record* **292:** 1514–1521.

**Erickson GM, Makovicky PJ, Inouye BD, Zhou CF, Gao KQ. 2015.** Flawed Analysis? A Response to Myhrvold. *The Anatomical Record* **298:** 1669–1672.

**Evans DC, Reisz RR. 2007.** Anatomy and relationships of *Lambeosaurus magnicristatus*, a crested hadrosaurid dinosaur (Ornithischia) from the Dinosaur Park Formation, Alberta. *Journal of Vertebrate Paleontology* **27:** 373–393.

**Evans KM, Bernt MJ, Kolmann MA, Ford KL, Albert JS. 2018.** Why the long face? Static allometry in the sexually dimorphic phenotypes of Neotropical electric fishes. *Zoological Journal of the Linnean Society* **186:** 633–649.

**Fairbairn DJ, Blanckenhorn WU, Székely T, eds. 2007.** *Sex, size and gender roles: evolutionary studies of sexual size dimorphism*. Oxford: Oxford University Press.

**Farke AA. 2004.** Horn use in *Triceratops* (Dinosauria: Ceratopsidae): testing behavioral hypotheses using scale models. *Palaeontologia Electronica* **7:** 1–10.

**Farke AA. 2014.** Evaluating combat in ornithischian dinosaurs. *Journal of Zoology* **292:** 242–249.

**Farke AA, Wolff ED, Tanke DH. 2009.** Evidence of combat in *Triceratops*. *PLoS One* **4:** e4252.

**Frederickson JA, Tumarkin-Deratzian AR. 2014.** Craniofacial ontogeny in *Centrosaurus apertus*. *PeerJ* **2:** e252.

**Galton PM. 1970.** Pachycephalosaurids: dinosaurian battering rams. *Discovery* **6:** 23–32.

**Galton PM. 1971.** A primitive dome-headed dinosaur (Ornithischia: Pachycephalosauridae) from the Lower Cretaceous of England and the function of the dome of pachycephalosaurids. *Journal of Paleontology* **45:** 40–47.

**Gamble KC. 2007.** Internal anatomy of the hornbill casque described by radiography, contrast radiography, and computed tomography. *Journal of Avian Medicine and Surgery* **21:** 38–49.

**Gautier-Hion A. 1975.** Dimorphisme sexuel et organisation sociale chez les cercopithécinés africains. *Mammalia* **39:** 365–374.

**Gerson HB, Hickie JP. 1985.** Head scarring on male narwhals (*Monodon monoceros*): evidence for aggressive tusk use. *Canadian Journal of Zoology* **63:** 2083–2087.

**Giffin EB. 1989.** Pachycephalosaur paleoneurology (Archosauria: Ornithischia). *Journal of Vertebrate Paleontology* **9:** 67–77.

**Gilmore CW. 1914.** *Osteology of the armored Dinosauria in the United States National Museum, with special reference to the genus Stegosaurus*. Bulletin of the United States National Museum **89:** 1–136.

**Gingerich PD. 1981.** Variation, sexual dimorphism, and social structure in the early Eocene horse *Hyracotherium* (Mammalia, Perissodactyla). *Paleobiology* **7:** 443–455.

**Godfrey LR, Lyon SK, Sutherland MR. 1993.** Sexual dimorphism in large-bodied primates: the case of the subfossil lemurs. *American Journal of Physical Anthropology* **90:** 315–334.

**Gould J, Gould C. 1989.** *Sexual selection*. New York: Scientific American Library.

**Gould JL, Elliott SL, Masters CM, Mukerji J. 1999.** Female preferences in a fish genus without female mate choice. *Current Biology* **9:** 497–500.

**Gould SJ. 1974.** The origin and function of 'bizarre' structures: antler size and skull size in the 'Irish elk,' *Megaloceros giganteus*. *Evolution* **28:** 191–220.

**Goyens J, Dirckx J, Aerts P. 2015.** Costly sexual dimorphism in *Cyclommatus metallifer* stag beetles. *Functional Ecology* **29:** 35–43.

**Grammer K, Thornhill R. 1994.** Human (*Homo sapiens*) facial attractiveness and sexual selection: the role of symmetry and averageness. *Journal of Comparative Psychology* **108:** 233.

**Grant BR, Grant PR. 2010.** Songs of Darwin's finches diverge when a new species enters the community. *Proceedings of the National Academy of Sciences* **107:** 20156–20163.

**Grant PR, Grant BR. 2008.** Pedigrees, assortative mating and speciation in Darwin's finches. *Proceedings of the Royal Society B: Biological Sciences* **275:** 661–668.

**Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, Altman DG. 2016.** Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology* **31:** 337–350.

**Griffin CT, Nesbitt SJ. 2016.** The femoral ontogeny and long bone histology of the Middle Triassic (? late Anisian) dinosauriform *Asilisaurus kongwe* and implications for the growth of early dinosaurs. *Journal of Vertebrate Paleontology* **36:** e1111224.

**Halsey LG. 2019.** The reign of the p-value is over: what alternative analyses could we employ to fill the power vacuum? *Biology Letters* **15:** 20190174.

**Hansen TF. 2016.** On bias and precision in meta-analysis: the error in the error. *Journal of Evolutionary Biology* **29:** 1919–1921.

**Hardin G. 1960.** The competitive exclusion principle. *Science* **131:** 1292–1297.

**Harvey PH, Kavanagh M, Clutton-Brock TH. 1978.** Sexual dimorphism in primate teeth. *Journal of Zoology* **186:** 475–485.

**Hayashi S, Carpenter K, Suzuki D. 2009.** Different growth patterns between the skeleton and osteoderms of *Stegosaurus* (Ornithischia: Thyreophora). *Journal of Vertebrate Paleontology* **29:** 123–31.

**Heath DJ. 1977.** Simultaneous hermaphroditism; cost and benefit. *Journal of Theoretical Biology* **64:** 363–373.

**Heinsohn R, Legge S, Endler JA. 2005.** Extreme reversed sexual dichromatism in a bird without sex role reversal. *Science* **309:** 617–619.

**Heske EJ, Ostfeld RS. 1990.** Sexual dimorphism in size, relative size of testes, and mating systems in North American voles. *Journal of Mammalogy* **71:** 510–519.

**Holland SM. 2019.** Estimation, not significance. *Paleobiology* **45:** 1–6.

**Hone DWE, Faulkes CG. 2014.** A proposed framework for establishing and evaluating hypotheses about the behaviour of extinct organisms. *Journal of Zoology* **292:** 260–267.

**Hone DW, Mallon JC. 2017.** Protracted growth impedes the detection of sexual dimorphism in non-avian dinosaurs. *Palaeontology* **60:** 535–545.

**Hone DWE, Naish D. 2013.** The 'species recognition hypothesis' does not explain the presence and evolution of exaggerated structures in non-avialan dinosaurs. *Journal of Zoology* **290:** 172–180.

**Hone DW, Naish D, Cuthill IC. 2012.** Does mutual sexual selection explain the evolution of head crests in pterosaurs and dinosaurs? *Lethaia* **45:** 139–156.

**Hone DW, Wood D, Knell RJ. 2016a.** Positive allometry for exaggerated structures in the ceratopsian dinosaur *Protoceratops andrewsi* supports socio-sexual signaling. *Palaeontologia Electronica* **19:** 1–13.

**Hone DW, Farke AA, Wedel MJ. 2016b.** Ontogeny and the fossil record: what, if anything, is an adult dinosaur? *Biology Letters* **12:** 20150947.

**Hone D, Mallon JC, Hennessey P, Witmer LM. 2020.** Ontogeny of a sexually selected structure in an extant archosaur *Gavialis gangeticus* (Pseudosuchia: Crocodylia) with implications for sexual dimorphism in dinosaurs. *PeerJ* **8:** e9134.

**Horner JR, Goodwin MB. 2006.** Major cranial changes during Triceratops ontogeny. *Proceedings of the Royal Society B: Biological Sciences* **273:** 2757–2761.

**Horner JR, Padian K. 2004.** Age and growth dynamics of Tyrannosaurus rex. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **271:** 1875–1880.

**Hu D, Clarke JA, Eliason CM, Qiu R, Li Q, Shawkey MD, Zhao C, D'Alba L, Jiang J, Xu X. 2018.** A bony-crested Jurassic dinosaur with evidence of iridescent plumage highlights complexity in early paravian evolution. *Nature Communications* **9:** 217.

**Huber R, Baumung R, Wurzinger M, Semambo D, Mwai O, Winckler C. 2008.** Grazing, social and comfort behaviour of Ankole and crossbred (Ankole× Holstein) heifers on pasture in south western Uganda. *Applied Animal Behaviour Science* **112:** 223–234.

**Hunt J, Breuker CJ, Sadowski JA, Moore AJ. 2009.** Male–male competition, female mate choice and their interaction: determining total sexual selection. *Journal of Evolutionary Biology* **22:** 13–26.

**Hunt S, Bennett ATD, Cuthill IC, Griffiths R. 1998.** Blue tits are ultraviolet tits. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **265:** 451–455.

**Hedrick BP, Dodson P. 2013.** Lujiatun psittacosaurids: understanding individual and taphonomic variation using 3D geometric morphometrics. *PLoS One* **8:** e69265.

**Irmis RB. 2007.** Axial skeleton ontogeny in the Parasuchia (Archosauria: Pseudosuchia) and its implications for ontogenetic determination in archosaurs. *Journal of Vertebrate Paleontology* **27:** 350–361.

**Isles TE. 2009.** The socio-sexual behaviour of extant archosaurs: implications for understanding dinosaur behaviour. *Historical Biology* **21:** 139–214.

**Jones IL, Hunter FM. 1993.** Mutual sexual selection in a monogamous seabird. *Nature* **362:** 238.

**Karsten KB, Andriamandimbiarisoa LN, Fox SF, Raxworthy CJ. 2009.** Sexual selection on body size and secondary sexual characters in 2 closely related, sympatric chameleons in Madagascar. *Behavioral Ecology* **20:** 1079–1088.

**Kitchener A. 1987.** Fighting behaviour of the extinct Irish elk. *Modern Geology* **11:** 1–28.

**Klaver C, Böhme W. 1992.** The species of the *Chamaeleo cristatus* group from Cameroon and adjacent countries, West Africa. *Bonner Zoologische Beiträge* **43:** 433–476.

**Knapp A, Knell RJ, Farke AA, Loewen MA, Hone DW. 2018.** Patterns of divergence in the morphology of ceratopsian dinosaurs: sympatry is not a driver of ornament evolution. *Proceedings of the Royal Society B* **285:** 20180312.

**Knell RJ, Sampson S. 2011.** Bizarre structures in dinosaurs: species recognition or sexual selection? A response to Padian and Horner. *Journal of Zoology* **283:** 18–22.

**Knell RJ, Naish D, Tomkins JL, Hone DW. 2013a.** Sexual selection in prehistoric animals: detection and implications. *Trends in Ecology & Evolution* **28:** 38–47.

**Knell RJ, Naish D, Tomkins JL, Hone DW. 2013b.** Is sexual selection defined by dimorphism alone? A reply to Padian and Horner. *Trends in Ecology & Evolution* **28:** 250–251.

**Kraaijeveld K, Gregurke J, Hall C, Komdeur J, Mulder RA. 2004.** Mutual ornamentation, sexual selection, and social dominance in the black swan. *Behavioral Ecology* **15:** 380–389.

**Kraaijeveld K, Kraaijeveld-Smit FJ, Komdeur J. 2007.** The evolution of mutual ornamentation. *Animal Behaviour* **74:** 657–677.

**Kraaijeveld K, Kraaijeveld-Smit FJ, Maan ME. 2011.** Sexual selection and speciation: the comparative evidence revisited. *Biological Reviews* **86:** 367–377.

**Krause J, Ruxton GD. 2002.** *Living in groups*. Oxford: Oxford University Press.

**Krauss D, Pezon AN, Nguyen PE, Salame IS, Rywkin S. 2010.** Evolutionary interactions between horn and frill morphology in chasmosaurine ceratopsians. In: Ryan MJ, Chinnery-Allgeier BJ, Eberth DA, eds. *New Perspectives on Horned Dinosaurs: The Royal Tyrrell Museum Ceratopsian Symposium*. Bloomington: Indiana University Press, 282–292.

**Krishtalka L, Stucky RK, Beard KC. 1990.** The earliest fossil evidence for sexual dimorphism in primates. *Proceedings of the National Academy of Sciences* **87:** 5223–5226.

**Lack D. 1968.** *Ecological adaptations for breeding in birds*. London: Methuen & Co.

**Laver RJ, Purwandana D, Ariefiandy A, Imansyah J, Forsyth D, Ciofi C, Jessop TS. 2012.** Life-history and spatial determinants of somatic growth dynamics in Komodo dragon populations. *PLoS One* **7:** e45398.

**Lemaître JF, Vanpé, C, Plard F, Gaillard JM. 2014.** The allometry between secondary sexual traits and body size is nonlinear among cervids. *Biology Letters* **10:** 20130869.

**Lamichhaney S, Han F, Webster MT, Andersson L, Grant BR, Grant PR. 2018.** Rapid hybrid speciation in Darwin's finches. *Science* **359:** 224–228.

**Lande R. 1980.** Sexual dimorphism, sexual selection, and adaptation in polygenic characters. *Evolution* **34:** 292–305.

**Larson PL. 1994.** *Tyrannosaurus* sex. *The Paleontological Society Special Publications* **7:** 139–156.

**Larson P. 2008.** Variation and sexual dimorphism in *Tyrannosaurus rex*. In: Larson P, Carpenter K, eds. *Tyrannosaurus rex, The Tyrant King*. Bloomington: Indiana University Press, 102–128.

**Lee AH, Werning S. 2008.** Sexual maturity in growing dinosaurs does not fit reptilian growth models. *Proceedings of the National Academy of Sciences* **105:** 582–587.

**Lee PA. 1980.** Normal ages of pubertal events among American males and females. *Journal of Adolescent Health Care* **1:** 26–29.

**Lewis SM, Cratsley CK, Demary K. 2004.** Mate recognition and choice in *Photinus* fireflies. *Annales Zoologici Fennici*. **41:** 809–821. https://www.jstor.org/stable/23736144?casa.

**Li Q, Gao KQ, Meng Q, Clarke JA, Shawkey MD, D'Alba L, Pei R, Ellison M, Norell MA, Vinther J. 2012.** Reconstruction of *Microraptor* and the evolution of iridescent plumage. *Science* **335:** 1215–1219.

**Lihoreau M, Zimmer C, Rivault C. 2008.** Mutual mate choice: when it pays both sexes to avoid inbreeding. *PLoS One* **3:** e3365.

**Liker A, Freckleton RP, Székely T. 2013.** The evolution of sex roles in birds is related to adult sex ratio. *Nature Communications* **4:** 1587.

**Liker A, Freckleton RP, Székely T. 2014.** Divorce and infidelity are associated with skewed adult sex ratios in birds. *Current Biology* **24:** 880–884.

**Lillie FR. 1931.** Bilateral gynandromorphism and lateral hemihypertrophy in birds. *Science* **74:** 387–390.

**Lister AM. 2009.** Late-glacial mammoth skeletons (*Mammuthus primigenius*) from Condover (Shropshire, UK): anatomy, pathology, taphonomy and chronological significance. *Geological Journal* **44:** 447–479.

**Lockley MG, McCrea RT, Buckley LG, Lim JD, Matthews NA, Breithaupt BH, Houck KJ, Gierliński GD, Surmik D, Kim KS, Xing L. 2016.** Theropod courtship: large scale physical evidence of display arenas and avian-like scrape ceremony behaviour by Cretaceous dinosaurs. *Scientific Reports* **6:** 18952.

**Löfstedt C. 1993.** Moth pheromone genetics and evolution. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **340:** 167–177.

**Longrich NR. 2010.** *Mojoceratops perifania*, a new chasmosaurine ceratopsid from the late Campanian of western Canada. *Journal of Paleontology* **84:** 681–694.

**Longrich NR, Sankey J, Tanke D. 2010.** *Texacephale langstoni*, a new genus of pachycephalosaurid (Dinosauria: Ornithischia) from the upper Campanian Aguja Formation, southern Texas, USA. *Cretaceous Research* **31:** 274–284.

**Lucas AM, Stettenheim PR. 1972.** Avian anatomy--integument, part II. Agriculture handbook 362. Washington DC: US Government Printing Office.

**Lund R. 1982.** *Harpagofututor volsellorhinus* new genus and species (Chondrichthyes, Chondrenchelyiformes) from the Namurian Bear Gulch Limestone, Chondrenchelys problematica Traquair (Visean), and their sexual dimorphism. *Journal of Paleontology* **56:** 938–958.

**Maisey JG. 2009.** The spine-brush complex in symmoriiform sharks (Chondrichthyes; Symmoriiformes), with comments on dorsal fin modularity. *Journal of Vertebrate Paleontology* **29:** 14–24.

**Mallon JC. 2017.** Recognizing sexual dimorphism in the fossil record: lessons from nonavian dinosaurs. *Paleobiology* **43:** 495–507.

**Mallon JC, Holmes RB. 2006.** A reevaluation of sexual dimorphism in the postcranium of the chasmosaurine ceratopsid *Chasmosaurus belli* (Dinosauria: Ornithischia). *The Canadian Field-Naturalist* **120:** 403–412.

**Martins MJF, Puckett TM, Lockwood R, Swaddle JP, Hunt G. 2018.** High male sexual investment as a driver of extinction in fossil ostracods. *Nature* **556:** 366–369.

**Maschenko EN. 2002.** Individual development, biology and evolution of the woolly mammoth. *Cranium* **19:** 4–120.

**Maynard Smith J, Harper D. 2003.** *Animal signals*. Oxford: Oxford University Press.

**Maynard Smith J, Price GR. 1973.** The logic of animal conflict. *Nature* **246:** 15.

**Mayr G, Pittman M, Saitta E, Kaye TG, Vinther J. 2016.** Structure and homology of *Psittacosaurus* tail bristles. *Palaeontology* **59:** 793–802.

**McCullough EL, Miller CW, Emlen DJ. 2016.** Why sexually selected weapons are not ornaments. *Trends in Ecology & Evolution* **31:** 742–751.

**McCoy DE, Feo T, Harvey TA, Prum RO. 2018.** Structural absorption by barbule microstructures of super black bird of paradise feathers. *Nature Communications* **9:** 1.

**McWhinney LA, Rothschild BM, Carpenter K. 2001.** Posttraumatic chronic osteomyelitis in *Stegosaurus* dermal spikes. In: Carpenter K, ed. *The armored dinosaurs*. Bloomington: Indiana University Press, 141–156.

**Meachen-Samuels JA, Binder WJ. 2010.** Sexual dimorphism and ontogenetic growth in the American lion and sabertoothed cat from Rancho La Brea. *Journal of Zoology* **280:** 271–279.

**Mendelson TC, Shaw KL. 2013.** Further misconceptions about species recognition: a reply to Padian and Horner. *Trends in Ecology & Evolution* **28:** 252–253.

**Møller AP. 1992.** Patterns of fluctuating asymmetry in weapons: evidence for reliable signalling of quality in beetle horns and bird spurs. *Proceedings of the Royal Society of London B* **248:** 199–206.

**Møller AP, Höglund J. 1991.** Patterns of fluctuating asymmetry in avian feather ornaments: implications for models of sexual selection. *Proceedings of the Royal Society of London B* **245:** 1–5.

**Molnar RE. 1977.** Analogies in the evolution of combat and display structures in ornithopods and ungulates. *Evolutionary Theory* **3:** 165–190.

**Moodie RL. 1918.** *Studies in paleopathology: general consideration of the evidences of pathological conditions found among fossil animals. I.* Urbana: Paul B. Hoeber.

**Moodie RL. 1923.** *Paleopathology. An introduction to the study of ancient evidences of disease*. Urbana: University of Illinois Press.

**Moriarty P, Holt RG. 2011.** Faraday waves produced by periodic substrates: mimicking the alligator water dance. *The Journal of the Acoustical Society of America* **129:** 2411.

**Morrissey MB. 2016.** Meta-analysis of magnitudes, differences and variation in evolutionary parameters. *Journal of Evolutionary Biology* **29:** 1882–1904.

**Motani R, Huang J, Jiang DY, Tintori A, Rieppel O, You H, Hu YC, Zhang R. 2018.** Separating sexual dimorphism from other morphological variation in a specimen complex of fossil marine reptiles (Reptilia, Ichthyosauriformes, *Chaohusaurus*). *Scientific Reports* **8:** 14978.

**Murphy NL, Trexler D, Thompson M. 2006.** "Leonardo," a mummified Brachylophosaurus from the Judith River Formation. In: Carpenter K, eds. *Horns and beaks: ceratopsian and ornithopod dinosaurs*. Bloomington: Indiana University Press, 117–133.

**Nakagawa S, Cuthill IC. 2007.** Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews* **82:** 591–605.

**Navarro JL, Vignolo PE, Demaria MR, Maceira NO, Martella MB. 2005.** Growth curves of farmed greater rheas (*Rhea americana albescens*) from Central Argentina. *Archiv fur Geflugelkunde* **69:** 90–93.

**Neige P, Marchand D, Laurin B. 1997.** Heterochronic differentiation of sexual dimorphs among Jurassic ammonite species. *Lethaia* **30:** 145–155.

**Nelson DA. 1989.** The importance of invariant and distinctive features in species recognition of bird song. *The Condor* **91:** 120–130.

**Nolan PM, Dobson FS, Nicolaus M, Karels TJ, McGraw KJ, Jouventin P. 2010.** Mutual mate choice for colorful traits in king penguins. *Ethology* **116:** 635–644.

**Nottebohm F, Arnold AP. 1976.** Sexual dimorphism in vocal control areas of the songbird brain. *Science* **194:** 211–213.

**Nowicki S, Peters S, Podos J. 1998.** Song learning, early nutrition and sexual selection in songbirds. *American Zoologist* **38:** 179–190.

**O'Brien DM, Allen CE, Van Kleeck MJ, Hone D, Knell R, Knapp A, Christiansen S, Emlen DJ. 2018.** On the evolution of extreme structures: static scaling and the function of sexually selected signals. *Animal Behaviour* **144:** 95–108.

**Ord TJ, Stuart-Fox D. 2006.** Ornament evolution in dragon lizards: multiple gains and widespread losses reveal a complex history of evolutionary change. *Journal of Evolutionary Biology* **19:** 797–808.

**Owens IP, Hartley IR. 1998.** Sexual dimorphism in birds: why are there so many different forms of dimorphism? *Proceedings of the Royal Society of London B: Biological Sciences* **265:** 397–407.

**Owen-Smith N. 1993.** Comparative mortality rates of male and female kudus: the costs of sexual size dimorphism. *Journal of Animal Ecology* **62:** 428–440.

**Packer C. 1983.** Sexual dimorphism: the horns of African antelopes. *Science* **221:** 1191–1193.

**Padian K, Horner JR. 2011.** The evolution of 'bizarre structures' in dinosaurs: biomechanics, sexual selection, social selection or species recognition? *Journal of Zoology* **283:** 3–17.

**Padian K, Horner JR. 2013.** Misconceptions of sexual selection and species recognition: a response to Knell *et al*. and to Mendelson and Shaw. *Trends in Ecology & Evolution* **28:** 249–250.

**Padian K, Horner JR. 2014.** The species recognition hypothesis explains exaggerated structures in non-avialan dinosaurs better than sexual selection does. *Comptes Rendus Palevol* **13:** 97–107.

**Panhuis TM, Butlin R, Zuk M, Tregenza T. 2001.** Sexual selection and speciation. *Trends in Ecology & Evolution* **16:** 364–371.

**Parker GA. 1970.** Sperm competition and its evolutionary consequences in the insects. *Biological Reviews* **45:** 525–567.

**Parker GA. 2006.** Sexual conflict over mating and fertilization: an overview. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **361:** 235–259

**Parry LA, Smithwick F, Nordén KK, Saitta ET, Lozano-Fernandez J, Tanner AR, Caron JB, Edgecombe GD, Briggs DE, Vinther J. 2018.** Soft-bodied fossils are not simply rotten carcasses–toward a holistic understanding of exceptional fossil preservation: exceptional fossil preservation is complex and involves the interplay of numerous biological and geological processes. *Bioessays* **40:** 1700167.

**Paterson HE. 1980.** A comment on "mate recognition systems." *Evolution* **34:** 330–331.

**Paterson HE. 1985.** The recognition concept of species. In: Vrba E, ed. *Species and speciation. Transvaal Museum Monograph*. Pretoria: Transvaal Museum, 21–29.

**Paukstis GL, Janzen FJ. 1990.** Sex determination in reptiles: summary of effects of constant temperatures of incubation on sex ratios of offspring. *Smithsonian Herpetological Information Service* **83:** 1–28.

**Pearson D, Shine R, How R. 2002.** Sex-specific niche partitioning and sexual size dimorphism in Australian pythons (*Morelia spilota imbricata*). *Biological Journal of the Linnean Society* **77:** 113–125.

**Pérez-Barbería FJ, Gordon IJ, Pagel M. 2002.** The origins of sexual dimorphism in body size in ungulates. *Evolution* **56:** 1276–1285.

**Persons WS IV, Funston GF, Currie PJ, Norell MA. 2015.** A possible instance of sexual dimorphism in the tails of two oviraptorosaur dinosaurs. *Scientific Reports* **5:** 9472.

**Peterson JE, Vittore CP. 2012.** Cranial pathologies in a specimen of *Pachycephalosaurus*. *PLoS One* **7:** e36227.

**Peterson JE, Dischler C, Longrich NR. 2013.** Distributions of cranial pathologies provide evidence for head-butting in dome-headed dinosaurs (Pachycephalosauridae). *PLoS One* **8:** e68620.

**Petrie M, Tim H, Carolyn S. 1991.** Peahens prefer peacocks with elaborate trains. *Animal Behaviour* **41:** 323–331.

**Platt SG, Rainwater TR, Thorbjarnarson JB, Finger AG, Anderson TA, McMurry ST. 2009.** Size estimation, morphometrics, sex ratio, sexual size dimorphism, and biomass of Morelet's crocodile in northern Belize. *Caribbean Journal of Science* **45:** 80–93.

**Plavcan JM. 1994.** Comparison of four simple methods for estimating sexual dimorphism in fossils. *American Journal of Physical Anthropology* **94:** 465–476.

**Poissant J, Wilson AJ, Festa-Bianchet M, Hogg JT, Coltman DW. 2008.** Quantitative genetics and sex-specific selection on sexually dimorphic traits in bighorn sheep. *Proceedings of the Royal Society of London B: Biological Sciences* **275:** 623–628.

**Pringle RM. 2020.** Paleoecology: the functional uniqueness of ancient megafauna. *Current Biology* **30:** R32–R35.

**Raath MA. 1990.** Morphological variation in small theropods and its meaning in systematics: evidence from *Syntarsus rhodesiensis*. In: Carpenter K, Currie PJ, eds. *Dinosaur systematics: approaches and perspectives*. Cambridge: Cambridge University Press, 91–105.

**Radford AN, Du Plessis MA. 2003.** Bill dimorphism and foraging niche partitioning in the green woodhoopoe. *Journal of Animal Ecology* **72:** 258–269.

**Rehg JA, Leigh SR. 1999.** Estimating sexual dimorphism and size differences in the fossil record: a test of methods. *American Journal of Physical Anthropology* **110:** 95–104.

**Reno PL, Meindl RS, McCollum MA, Lovejoy CO. 2003.** Sexual dimorphism in *Australopithecus afarensis* was similar to that of modern humans. *Proceedings of the National Academy of Sciences* **100:** 9404–9409.

**Rice WR, Chippindale AK. 2001.** Intersexual ontogenetic conflict. *Journal of Evolutionary Biology* **14:** 685–693.

**Rico-Guevara A, Hurme KJ. 2019.** Intrasexually selected weapons. *Biological Reviews* **94:** 60–101.

**Rinehart LF, Lucas SG, Heckert AB, Spielmann JA, Celeskey MD. 2009.** *The Paleobiology of Coelophysis bauri (Cope) from the Upper Triassic (Apachean) Whitaker quarry, New Mexico, with detailed analysis of a single quarry block: Bulletin 45*, 1–260. Albuquerque: New Mexico Museum of Natural History and Science.

**Robinson MR, Kruuk LE. 2007.** Function of weaponry in females: the use of horns in intrasexual competition for resources in female Soay sheep. *Biology Letters* **3:** 651–654.

**Rowe T, Ketcham RA, Denison C, Colbert M, Xu X, Currie PJ. 2001.** Forensic palaeontology: the *Archaeoraptor* forgery. *Nature* **410:** 539.

**Roy A, Pittman M, Saitta ET, Kaye TG, Xu X. 2019.** Recent advances in amniote palaeocolour reconstruction and a framework for future research. *Biological Reviews* doi:10.1111/brv.12552.

**Ryan MJ, Holmes R, Russell AP. 2007.** A revision of the late Campanian centrosaurine ceratopsid genus *Styracosaurus* from the western Interior of North America. *Journal of Vertebrate Paleontology* **27:** 944–962.

**Ryan MJ, Keddy-Hector A. 1992.** Directional patterns of female mate choice and the role of sensory biases. *The American Naturalist* **139:** S4–S35.

**Ryan MJ, Rand AS. 1993.** Species recognition and sexual selection as a unitary problem in animal communication. *Evolution* **47:** 647–657.

**Saitta ET. 2014.** *Paleobiology of North American stegosaurs: evidence for sexual dimorphism*. Unpublished Thesis, Princeton University.

**Saitta ET. 2015.** Evidence for sexual dimorphism in the plated dinosaur *Stegosaurus mjosi* (Ornithischia, Stegosauria) from the Morrison Formation (Upper Jurassic) of western USA. *PLoS One* **10:** e0123503.

**Saitta ET, Vinther J. 2019.** A perspective on the evidence for keratin protein preservation in fossils: an issue of replication versus validation. *Palaeontologia Electronica* **22.3.2E**.

**Saitta ET, Fletcher I, Martin P, Pittman M, Kaye TG, True LD, Norell MA, Abbott GD, Summons RE, Penkman K, Vinther J. 2018.** Preservation of feather fibers from the Late Cretaceous dinosaur *Shuvuuia deserti* raises concern about immunohistochemical analyses on fossils. *Organic Geochemistry* **125:** 142–151.

**Saitta ET, Stockdale MT, Longrich NR, Bonhomme V, Benton MJ, Cuthill IC, Makovicky PJ. 2020.** Data from: An effect size statistical framework for investigating sexual dimorphism in non-avian dinosaurs and other extinct taxa. Dryad, Dataset, https://doi.org/10.5061/dryad.kkwh70s2k.

**Sampson SD, Ryan MJ, Tanke DH. 1997.** Craniofacial ontogeny in centrosaurine dinosaurs (Ornithischia: Ceratopsidae): taxonomic and behavioral implications. *Zoological Journal of the Linnean Society* **121:** 293–337.

**Sánchez IM, Quiralte V, Morales J, Azanza B, Pickford M. 2010.** Sexual dimorphism of the frontal appendages of the early Miocene African pecoran *Prolibytherium* Arambourg, 1961 (Mammalia, Ruminantia). *Journal of Vertebrate Paleontology* **30:** 1306–1310.

**Sander PM, Klein N. 2005.** Developmental plasticity in the life history of a prosauropod dinosaur. *Science* **310:** 1800–1802.

**Santos SIO, Elward B, Lumeij JT. 2006.** Sexual dichromatism in the blue-fronted Amazon parrot (*Amazona aestiva*) revealed by multiple-angle spectrometry. *Journal of Avian Medicine and Surgery* **20:** 8–15.

**Sato T, Cheng YN, Wu XC, Zelenitsky DK, Hsiao YF. 2005.** A pair of shelled eggs inside a female dinosaur. *Science* **308:** 375.

**Scannella JB, Fowler DW, Goodwin MB, Horner JR. 2014.** Evolutionary trends in *Triceratops* from the Hell Creek Formation, Montana. *Proceedings of the National Academy of Sciences* **111:** 10245–10250.

**Schall JJ, Bromwich CR, Werner YL, Midlege J. 1989.** Clubbed regenerated tails in *Agama agama* and their possible use in social interactions. *Journal of Herpetology* **23:** 303–305.

**Schärer L, Rowe L, Arnqvist G. 2012.** Anisogamy, chance and the evolution of sex roles. *Trends in Ecology & Evolution* **27:** 260–264.

**Schilling MF, Watkins AE, Watkins W. 2002.** Is human height bimodal? *The American Statistician* **56:** 223–229.

**Schott RK, Evans DC, Goodwin MB, Horner JR, Brown CM, Longrich NR. 2011.** Cranial ontogeny in *Stegoceras validum* (Dinosauria: Pachycephalosauria): a quantitative model of pachycephalosaur dome growth and variation. *PLoS One* **6:** e21092.

**Schwartz GT, Dean C. 2001.** Ontogeny of canine dimorphism in extant hominoids. *American Journal of Physical Anthropology* **115:** 269–283.

**Schweitzer MH, Zheng W, Zanno L, Werning S, Sugiyama T. 2016.** Chemistry supports the identification of gender-specific reproductive tissue in *Tyrannosaurus rex*. *Scientific Reports* **6:** 23099.

**Scott HM, Payne LF. 1934.** The effect of gonadectomy on the secondary sexual characters of the bronze turkey (*M. gallopavo*). *Journal of Experimental Zoology* **69:** 123–136.

**Seddon N. 2005.** Ecological adaptation and species recognition drives vocal evolution in neotropical suboscine birds. *Evolution* **59:** 200–215.

**Sengupta S, Ezcurra MD, Bandyopadhyay S. 2017.** A new horned and long-necked herbivorous stem-archosaur from the Middle Triassic of India. *Scientific Reports* **7:** 8366.

**Shine R. 1979.** Sexual selection and sexual dimorphism in the Amphibia. *Copeia* **2:** 297–306.

**Sibley DA. 2014.** *The Sibley guide to birds*, *2nd edn*. New York: Alfred A. Knopf.

**Skeate ST. 1984.** Courtship and reproductive behaviour of captive white-fronted amazon parrots *Amazona albifrons*. *Bird Behavior* **5:** 103–109.

**Smith DK. 1998.** A morphometric analysis of *Allosaurus*. *Journal of Vertebrate Paleontology* **18:** 126–142.

**Smith JN, Goldizen AW, Dunlop RA, Noad MJ. 2008.** Songs of male humpback whales, *Megaptera novaeangliae*, are involved in intersexual interactions. *Animal Behaviour* **76:** 467–477.

**Smith KM, Fisher DC. 2011.** Sexual dimorphism of structures showing indeterminate growth: tusks of American mastodons (*Mammut americanum*). *Paleobiology* **37:** 175–194.

**Snively E, Cox A. 2008.** Structural mechanics of pachycephalosaur crania permitted head-butting behavior. *Palaeontologia Electronica* **11:** 3A.

**Snively E, Theodor JM. 2011.** Common functional correlates of head-strike behavior in the pachycephalosaur *Stegoceras validum* (Ornithischia, Dinosauria) and combative artiodactyls. *PLoS One* **6:** e21422.

**Stankowich T, Caro T. 2009.** Evolution of weaponry in female bovids. *Proceedings of the Royal Society of London B: Biological Sciences* **276:** 20091256.

**Stochholm K, Juul S, Juel K, Naeraa RW, Højbjerg Gravholt C. 2006.** Prevalence, incidence, diagnostic delay, and mortality in Turner syndrome. *The Journal of Clinical Endocrinology & Metabolism* **91:** 3897–3902.

**Sues HD. 1978.** Functional morphology of the dome in pachycephalosaurid dinosaurs. *Neues Jahrbuch für Geologie und Paläontologie Monatshefte* **8:** 459–472.

**Sullivan C, Reisz RR, Smith RM. 2003.** The Permian mammal-like herbivore *Diictodon*, the oldest known example of sexually dimorphic armament. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270:** 173–178.

**Székely T, Webb JN, Cuthill IC. 2000.** Mating patterns, sexual selection and parental care: an integrative approach. In: Apollonio M Festa-Bianchet M, Mainardi D, eds. *Vertebrate mating systems*. London: World Scientific Publishing Inc., 159–185.

**Székely T, Liker A, Freckleton RP, Fichtel C, Kappeler PM. 2014.** Sex-biased survival predicts adult sex ratio variation in wild birds. *Proceedings of the Royal Society B: Biological Sciences* **281:** 20140342.

**Tanke DH, Rothschild BM. 2002.** *DINOSORES: an annotated bibliography of dinosaur paleopathology and related topics—1838–2001: Bulletin 20*. Albuquerque: New Mexico Museum of Natural History and Science.

**Taylor GW, Santos JC, Perrault BJ, Morando M, Vásquez Almazán CR, Sites JW Jr. 2017.** Sexual dimorphism, phenotypic integration, and the evolution of head structure in casque-headed lizards. *Ecology and Evolution* **7:** 8989–8998.

**Tinbergen N. 1948.** Social releasers and the experimental method required for their study. *The Wilson Bulletin* **60:** 6–51.

**Tobias JA, Montgomerie R, Lyon BE. 2012.** The evolution of female ornaments and weaponry: social selection, sexual selection and ecological competition. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367:** 2274–2293.

**Tumarkin-Deratzian AR, Vann DR, Dodson P. 2006.** Bone surface texture as an ontogenetic indicator in long bones of the Canada goose *Branta canadensis* (Anseriformes: Anatidae). *Zoological Journal of the Linnean Society* **148:** 133–168.

**Turner AH, Makovicky PJ, Norell MA. 2007.** Feather quill knobs in the dinosaur *Velociraptor*. *Science* **317:** 1721.

**Turner WC, Jolles AE, Owen-Smith N. 2005.** Alternating sexual segregation during the mating season by male African buffalo (*Syncerus caffer*). *Journal of Zoology* **267:** 291–299.

**Vanpé C, Gaillard JM, Kjellander P, Mysterud A, Magnien P, Delorme D, Laere GV, Klein F, Liberg O, Hewison AJM. 2007.** Antler size provides an honest signal of male phenotypic quality in roe deer. *The American Naturalist* **169:** 481–493.

**Van Valkenburgh B, Sacco T. 2002.** Sexual dimorphism, social behavior, and intrasexual competition in large Pleistocene carnivorans. *Journal of Vertebrate Paleontology* **22:** 164–169.

**Varricchio DJ. 2011.** A distinct dinosaur life history? *Historical Biology* **23:** 91–107.

**Varricchio DJ, Moore JR, Erickson GM, Norell MA, Jackson FD, Borkowski JJ. 2008.** Avian paternal care had dinosaur origin. *Science* **322:** 1826–1828.

**Vinther J. 2015.** A guide to the field of palaeo colour. *BioEssays* **37:** 643–656.

**Vliet KA. 1989.** Social displays of the American alligator (*Alligator mississippiensis*). *American Zoologist* **29:** 1019–1031.

**Waas JR. 1988.** Acoustic displays facilitate courtship in little blue penguins, *Eudyptula minor*. *Animal Behaviour* **36:** 366–371.

**Wang X, Kellner AWA, Jiang S, Wang Q, Ma Y, Paidoula Y, Cheng X, Rodrigues T, Meng X, Zhang J, Li N, Zhou Z. 2014.** Sexually dimorphic tridimensionally preserved pterosaurs and their eggs from China. *Current Biology*. **24:** 1323–1330.

**Waring GH. 1997.** *Preliminary study of the behavior and ecology of Jackson's chameleons of Maui, Hawaii*. Report for USGS/BRD/PIERC Haleakala Field Station presented by Hawaii Ecosystems at Risk (HEAR) project. Available at: http://www.hear.org

**Warner RR. 1975.** The adaptive significance of sequential hermaphroditism in animals. *The American Naturalist* **109:** 61–82.

**Wasserstein RL, Lazar NA. 2016.** The ASA's statement on *p*-values: context, process, and purpose. *The American Statistician* **70:** 129–133.

**Wasserstein RL, Schirm AL, Lazar NA. 2019.** Moving to a world beyond "p < 0.05". *American Statistician* **73**(Supplement 1): 1–19.

**Weishampel DB. 1981.** Acoustic analyses of potential vocalization in lambeosaurine dinosaurs (Reptilia: Ornithischia). *Paleobiology* **7:** 252–261.

**Weishampel DB, Chapman RE. 1990.** Morphometric study of Plateosaurus from Trossingen (Baden-Wurttemburg, Federal Republic of Germany). In: Carpenter K, Currie PJ, eds. *Dinosaur systematics: approaches and perspectives*. Cambridge: Cambridge University Press, 43–52.

**West PM, Packer C. 2002.** Sexual selection, temperature, and the lion's mane. *Science* **297:** 1339–1343.

**Wheeler P, Greenwood PJ. 1983.** The evolution of reversed sexual dimorphism in birds of prey. *Oikos* **40:** 145–149.

**Wilkinson PM, Rhodes WE. 1997.** Growth rates of American alligators in coastal South Carolina. *The Journal of Wildlife Management* **61:** 397–402.

**Wilson VJ. 2001.** *Duikers of Africa: masters of the African forest floor: a study of Duikers-people-hunting and bushmeat.* Pretoria: Zimbi Books.

**Woodward HN, Freedman Fowler EA, Farlow JO, Horner JR. 2015.** *Maiasaura*, a model organism for extinct vertebrate population biology: a large sample statistical assessment of growth dynamics and survivorship. *Paleobiology* **41:** 503–527.

**Zeigler KE, Lucas SG, Heckert AB. 2002.** The Late Triassic Canjilon quarry (Upper Chinle Group, New Mexico) phytosaur skulls: evidence of sexual dimorphism in phytosaurs. *New Mexico Museum of Natural History and Science Bulletin* **21:** 179–188.

**Zelenitsky DK, Therrien F, Erickson GM, DeBuhr CL, Kobayashi Y, Eberth DA, Hadfield F. 2012.** Feathered non-avian dinosaurs from North America provide insight into wing origins. *Science* **338:** 510–514.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site (see also SHARED DATA):

**Appendix S1.** The appendix contains descriptions of the following:
Annotated SVP abstracts uploaded to Dryad.
Dunning (2007) data uploaded to Dryad.
Human height bimodality simulation.
Alligator and rhea simulations in the main text and files uploaded to Dryad, alongside additional simulation results of alligator under high intrasexual variation or skewed sex ratio.
Non-avian dinosaur analyses and additional results using Gompertz regression on these data.

## SHARED DATA

The full supplementary files of the raw data (referred to in Supporting Information above) are available from the Dryad Digital Repository (Saitta *et al*., 2020).