Peer reviewed version

Link to published version (if available):
10.1093/biolinnean/blaa105

Link to publication record in Explore Bristol Research
PDF-document

University of Bristol - Explore Bristol Research
General rights

1   **Sexual dimorphism in non-avian dinosaurs and other extinct taxa: the importance of**

2   **effect size statistics in paleontology**

3   *Evan T. Saitta, Maximilian T. Stockdale, Nicholas R. Longrich, Vincent Bonhomme, Michael*

4   *J. Benton, Innes C. Cuthill, and Peter J. Makovicky*

5   RH: SEXUAL DIMORPHISM IN DINOSAURS

6

7   *Evan T. Saitta. Integrative Research Center, Field Museum of Natural History, Chicago, IL,*

8   *60605, USA. E-mail: evansaitta@gmail.com. Maximilian T. Stockdale. Michael J. Benton.*

9   *School of Earth Sciences, University of Bristol, Bristol, BS8 1RJ U.K. E-mail:*

10   *max.stockdale@bristol.ac.uk. Mike.Benton@bristol.ac.uk. Vincent Bonhomme. E-mail:*

11   *vincent.bonhomme@umontpellier.fr. Nicholas R. Longrich. Department of Biology and*

12   *Biochemistry and Milner Centre for Evolution, University of Bath, Bath, BA2 7AY U.K. E-*

13   *mail: n.r.longrich@bath.ac.uk. Innes C. Cuthill. School of Biological Sciences, University of*

14   *Bristol, Bristol, BS8 1TH U.K. E-mail: I.Cuthill@bristol.ac.uk. Peter J. Makovicky.*

15   *Department of Earth and Environmental Sciences, University of Minnesota, Minneapolis, MN,*

16   *55455, USA. E-mail: pmakovic@umn.edu.*

17   *Abstract*.— Despite reports of sexual dimorphism in extinct taxa, such claims in non-avian

18   dinosaurs have been underrepresented recently (~the last decade) and often criticized. Since

19   dimorphism is widespread in sexually reproducing organisms today, underrepresentation might

20   suggest either methodological shortcomings or that this diverse group exhibited highly unusual

21   reproductive biology. Univariate significance testing, especially for bimodality, is ineffective

22   and prone to false negatives. Species recognition and mutual sexual selection hypotheses,

23   therefore, may not be required to explain supposed absence of sexual dimorphism across the

24   grade, likely a type II error. Instead, multiple lines of evidence support sexual selection and

25   variation of structures consistent with secondary sexual characteristics, strongly suggesting

26   sexual dimorphism in non-avian dinosaurs. We propose a framework for studying sexual

27   dimorphism in fossils, focusing on likely secondarily sexual traits and testing against all

28   alternate hypotheses for variation in them using multiple lines of evidence. We use effect size

29   statistics appropriate for low sample sizes, rather than significance testing, to analyze potential

30   divergence of growth curves in traits and constrain estimates for dimorphism magnitude. In

31   many cases, estimates of sexual variation can be reasonably accurate, and further developments

32   in methods to improve sex assignments and account for intrasexual variation (e.g., mixture

33   modelling) will improve accuracy. It is better to compare estimates for the magnitude of and

34   support for dimorphism between datasets than to dichotomously reject or fail to reject

35   monomorphism in a single species, enabling the study of sexual selection across phylogenies

36   and time. We defend our approach with simulated and empirical data, including dinosaur data,

37   showing that even simple approaches can yield fairly accurate estimates of sexual variation in

38   many cases, allowing for comparison of species with high and low support for sexual variation.

39

40   Keywords: dinosaurs; effect size statistics; fossils; sexual dimorphism; sexual selection

Introduction: Nuances of Sexual Selection Pose a Challenge to Paleontologists

When Charles Darwin introduced the concept of sexual selection as a variant of natural selection (Darwin 1888), 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 behaviors 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).

65   In many species, both intrasexual competition and intersexual mate choice can occur

66   and to varying degrees (Hunt et al. 2009), which can make studying sexual selection in the

67   fossil record all the more challenging. For example, male satin bowerbirds (*Ptilonorhynchus*

68   *violaceus*) compete with each other by stealing feathers from rivals' bowers that they use to

69   display to females, and the females then subsequently selectively choose males with which to

70   mate based on the quality of their bowers (Borgia & Gore 1986). Hidden mate preferences can

71   even exist in species whose mating systems do not in practice allow for that preference to be

72   expressed through mate choice; for example, female mosquitofish (*Gambusia holbrooki*), a

73   species in which males 'scramble' for mates without female choice, show mate preferences for

74   males with exaggerated traits under experimental conditions (Gould et al. 1999). Therefore, for

75   any given species, it is imperative to consider the selective pressure within and between each

76   sex, rather than simplifying descriptions of mating systems to entirely male-male competition

77   or female-mate choice (Clutton-Brock 2009). It is now understood that sexual selection is not

78   only an important evolutionary driver, but that its effect is widespread among organisms and

79   can be complex and nuanced.

80   One of the most frequent manifestations of sexual selection is sexual dimorphism.

81   Sexually dimorphic organisms exhibit differences between the sexes in the distributions of

82   certain anatomical, physiological, or behavioral traits (Lande 1980). Dimorphic traits

83   influenced by sexual selection can include some of the most elaborate products of evolution:

84   highly complex behaviors such as bird songs (Catchpole 1987), colors such as 'super black'

85   light-absorbing feathers in some birds of paradise (Paradisaeidae) (McCoy et al. 2018), and

86   exaggerated anatomical structures such as the tusks of elephants (Elephantidae) (Chelliah &

87   Sukumar 2013) or enlarged mandibles of stag beetles (*Cyclommatus metallifer*) (Goyens et al.

88   2015). One of the most common forms of sexual dimorphism is sexual size dimorphism, in

89   which one sex grows to a larger size than the other on average.

90   Sexual dimorphism need not always be expressed as the presence versus absence of a

91   particular characteristic, such as externally protruding tusks of male narwhals (*Monodon*

92   *monoceros*), which are normally absent on females (Gerson & Hickie 1985). Instead, it can

93   often be a difference in degree where one sex is underdeveloped in the trait, such as canine

94   length in gorillas (*Gorilla*) (Schwartz & Dean 2001). The magnitude of dimorphism can vary

95   greatly between different species (i.e., the effect size between male and female distributions,

96   most typically quantified as differences in measures of centrality between the distributions).

97   For example, primates show interspecific variation in the magnitude of sexual dimorphism with

98   respect to body mass and canine tooth length (Clutton-Brock et al. 1977; Harvey et al. 1978).

99   Statistically, a truly monomorphic species (i.e., difference between the male and female

100   distributions is precisely zero) is not expected in finite populations of empirical data

101   (Nakagawa & Cuthill 2007), irrespective of the strength of sexual selection acting on that

102   population. For example, if the measured heights of a group of two people are 1.8 m and 1.6 m

103   and the heights of a second group of two people are 1.8 m and 1.6 m, we would calculate the

104   difference in their average height to be exactly zero. However, few would predict that this

105   difference would remain precisely zero if this empirical data were recorded to the nearest nm.

106   Therefore, it is important to remember that terms like dimorphic and monomorphic are often

107   used subjectively to indicate whether a species shows relatively high or low sexual variation,

108   respectively. It is better to think simply in terms of the magnitude of sexual variation, without

109   forcing species into binary descriptive categories of monomorphic or dimorphic (and this

110   informs our use of effect size statistics below).

111   Thanks to the potential conspicuousness, complexity, and variability of many sexual

112   dimorphisms, as well as the possibility of testing functional hypotheses and their relation to

113   underlying selective pressures in extant organisms, sexual dimorphisms are a major topic of

114   research and an important quantifiable proxy for sexual selection (Fairbairn et al. 2007). Sexual

115　variation can also appear in traits capable of fossilizing in some environments, including certain

116　soft tissues (Parry et al. 2018) (e.g., skeletal or pigmented anatomy).

117　　　　However, sexual selection can act in ways that do not always produce visible variation

118　between the sexes, such as sperm competition (Parker 1970; Birkhead & Møller 1998).

119　Furthermore, factors can work to counter sexual selection and reduce dimorphism, such as

120　female bovids sporting horns when under predation pressure or intrasexual competition over

121　resources (Packer 1983; Caro et al. 2003; Robinson & Kruuk 2007), predation risk countering

122　sexual selection in the coloration of male poeciliid fish (Endler 1984), or male lions (*Panthera*

123　*leo*) reducing their mane thickness in warmer climates (West & Packer 2002). A good example

124　of confounding between sexual functions and secondary functions/biological tradeoffs of

125　sexually selected traits is the dichromatism of the polygynandrous eclectus parrot (*Eclectus*

126　*roratus*); male coloration is a tradeoff between conspicuous sexual display and camouflage

127　during foraging, while female coloration is driven by competition for nest hollows, without

128　any opposing need for camouflage (Heinsohn et al. 2005). Finally, some sexual dimorphisms

129　are difficult to study because of human limitations. For example, some birds once considered

130　monomorphic in color based on trichromatic human vision are actually dimorphic when studied

131　with spectroscopic techniques that reveal ultraviolet color variation, which is detectable by

132　tetrachromatic avian vision (Burkhardt 1989; Hunt et al. 1998; Santos et al. 2006).

133　　　　Here, we 1) highlight the underreporting of sexual dimorphism in non-avian dinosaurs

134　compared to other extinct taxa over the last decade, which reflects the current debate as to

135　whether this group exhibited unusual social/sexual biology or if methodological shortcomings

136　are at play. 2) We then discuss arguments against sexual selection or dimorphism in non-avian

137　dinosaurs, showing that hypotheses explaining a supposed lack of dimorphism throughout the

138　grade can be flawed based on current evolutionary theory. 3) Furthermore, these explanatory

139　hypotheses are likely unnecessary, as we show in our summary of the abundant evidence for

140 sexual selection and probable sexual variation in non-avian dinosaurs. These early sections

141 provide evidence from evolutionary/game theory, modern dimorphisms/extant phylogenetic

142 brackets, and fossils to show that some degree of sexual variation in anisogamic populations,

143 including non-avian dinosaurs, is the expectation, not the exception. In this context, the most

144 appropriate methodologies and statistical approaches can be selected. 4) We then show that

145 significance testing methods used to argue that dimorphism in non-avian dinosaurs lacks

146 evidence are highly prone to type II error. 5) Finally, we present our framework to study sexual

147 dimorphism in extinct taxa that utilizes effect size statistics and controls for alternate

148 hypotheses for observed variation.

149

150 **Non-Avian Dinosaurs: Unique Biology or Methodological Shortcomings?**

151 Detecting sexual dimorphism in the fossil record is complicated by difficulty in distinguishing

152 sexual variation from ontogenetic variation, interspecific variation, and relatively continuous

153 intra-population variation or polymorphisms unrelated to sex (Brusatte 2012). Furthermore,

154 certain characteristics can also show intra-individual variation, such as contour versus flight

155 feathers (Lucas & Stettenheim 1972) or anterior versus posterior osteoderms (Gilmore 1914;

156 Carpenter 1998). Therefore, when fossil specimens are incompletely preserved, within-body

157 variation might be confused for sexual variation. When studying fossils, taphonomic effects

158 must also be considered, such as plastic deformation or partial preservation as a result of

159 scavenging, transport, decay, diagenesis, weathering, or erosion (Parry et al. 2018). Some

160 characteristics that are often sexually selected or dimorphic have limited or no fossilization

161 potential, such as various soft tissues or mating behaviors. Sample sizes of many fossil species

162 are often small as well as geographically and stratigraphically dispersed. For most fossil

163 specimens, it is impossible to assign a sex with certainty, except in exceptional cases such as

164 specimens with eggs (Sato et al. 2005) or embryos (Caldwell & Lee 2001) preserved *in situ* in

165 the body cavity, claspers in chondrichthyans (Maisey 2009), bacula (Abella et al. 2013), or

166 reproductive medullary bone (Lee & Werning 2008). Without destructive sampling for bone

167 histology, it can sometimes be challenging to infer developmental maturity among specimens

168 exhibiting potentially dimorphic traits. Studying sexual selection in fossils is further limited

169 because behavioral observations generally cannot be made, except for minor inferences from

170 trace fossils or pathologies, for example, and behavioral experimentation is entirely precluded

171 (Hone & Faulkes 2014). Furthermore, it is difficult to hypothesize the function of a candidate

172 secondary sexual characteristic or the behavior of an extinct species without close extant

173 relatives or without obvious modern analogs based on ecology, overall body plan, or similar

174 anatomical traits. Many fossils have unusual structures not quite like those of any extant

175 species, such as stegosaur plates.

176   Despite these challenges, there have been many proposed sexual dimorphisms in extinct

177 species, along with discussions of sexual selection in extinct organisms more generally (Knell

178 et al. 2013a). Some of these extinct species are recent with comparable extant relatives and

179 analogs, or are known from many fossil specimens (e.g., invertebrates). Sexual dimorphism in

180 fossil ostracods has even been used to test hypotheses regarding the relation between sexual

181 selection and extinction risk (Martins et al. 2018). However, some examples are relatively

182 ancient, such as ammonoids (Neige et al. 1997), and unique, such as trilobites (Cederström et

183 al. 2011). Vertebrate examples include fossil hominids (Reno et al. 2003) and other primates

184 (Krishtalka et al. 1990), proboscidians (Smith & Fisher 2011), perissodactyls (Gingerich 1981),

185 artiodactyls (Sánchez et al. 2010), pinnipeds (Cullen et al. 2014), felids (Meachen-Samuels &

186 Binder 2010), dicynodonts (Sullivan et al. 2003), pterosaurs (Wang et al. 2014), birds

187 (Chinsamy *et al.* 2013), phytosaurs (Zeigler et al. 2002), basal archosauromorphs (Sengupta et

188 al. 2017), ichthyosauriforms (Motani et al. 2018), pachypleurosaurs (Cheng et al. 2009), and

189 chondrichthyans (Lund 1982), among others. The evolution and function of certain putative

190   secondary sexual characteristics in extinct taxa, such as the antlers of male 'Irish elk'

191   (*Megaloceros giganteus*), have historically been heavily discussed (Gould 1974; Kitchener

192   1987; Lemaître et al. 2014). Reversed dimorphism in moa (*Dinornis*), by definition an extinct

193   dinosaur, is thought to have been so extreme that the sexes were previously considered to be

194   different species (Bunce et al. 2003). Many published claims have not been challenged in the

195   literature or broader media, presumably due to ubiquity of sexual variation in vertebrates.

196   Noticeable exceptions, however, are non-avian dinosaurs, for which claims of sexual

197   dimorphism have recently been highly debated and criticized (Padian & Horner 2011; Hone et

198   al. 2012; Hone & Mallon 2017; Mallon 2017). Hone et al. (2020) state, "To date, no dinosaur

199   has been determined to exhibit sexual dimorphism under rigorous analysis" (p. 13). An

200   examination of over a decade of recent abstracts from the annual meeting of the Society of

201   Vertebrate Paleontology reveals that abstracts proposing or concluding sexual variation in non-

202   avian dinosaurs are highly underrepresented compared with those for all other fossil taxa, in

203   relation to the prevalence of abstracts on non-avian dinosaurs at the meeting (χ^2 goodness of

204   fit test on summed counts over an 11-year period of non-avian dinosaur dimorphism abstracts

205   compared with dimorphism abstracts of all other taxa using the online tool from vassarstats.net:

206   degrees of freedom = 1; expected count = 34, observed count = 15, unadjusted χ^2 = 13.11, *p*-

207   value = 0.0004 for potential dimorphism; expected count = 24, observed count = 2, unadjusted

208   χ^2 = 24.96, *p*-value = <0.0001 for concluded dimorphism) (Table 1). Does this underreporting

209   reflect a highly unusual social/sexual system in non-avian dinosaurs or differences in how

210   dinosaur researchers interpret fossil data compared to other paleontologists? Is the debate

211   around non-avian dinosaur dimorphism a case of biology, or does it stem from methodological

212   shortcomings and/or preconceived notions about a lack of dimorphism?

213

214   **Flawed Alternatives to Sexual Selection and Dimorphism in Non-Avian Dinosaurs**

215   Claims of sexual dimorphism in non-avian dinosaurs have varied in sample size, methodology,

216   and whether or not they were approached quantitatively, with some studies criticized for using

217   relatively little data or analysis (Chapman 1997; Mallon 2017). Published reports include

218   proposals of dimorphism in *Tyrannosaurus rex* (Larson 1994, 2008), *Coelophysis bauri*

219   (Rinehart et al. 2009), *Coelophysis* (= *Syntarsus*) *rhodesiensis* (Raath 1990), *Kentrosaurus*

220   *aethiopicus* (Barden & Maidment 2011), *Plateosaurus* (Weishampel & Chapman 1990),

221   *Stegoceras validum* (Chapman et al. 1981), *Protoceratops andrewsi* (Dodson 1976),

222   *Allosaurus fragilis* (Smith 1998), *Citipati osmolskae* (notably with a sample size of only two

223   [Persons et al. 2015]), and *Hesperosaurus* (= *Stegosaurus*) *mjosi* (Saitta 2015), among others

224   (see table 1 of Mallon 2017 for more examples). Beyond morphological dimorphisms,

225   behavioral dimorphisms that might indirectly result from sexual selection in troodontids and

226   oviraptorids have also been hypothesized in the form of unequal parental care, specifically

227   paternal care (Varricchio et al. 2008; although see Birchard et al. 2013 for a counter). Recently,

228   claims of sexual dimorphism in non-avian dinosaurs, or at least demonstrable evidence for it,

229   have been rejected by some (e.g., Mallon 2017). The postulated absence of sexual dimorphism

230   (or absence of evidence for it) in non-avian dinosaurs has been explained in various ways: an

231   artifact resulting from limited sample size, taphonomic information loss, methodological

232   shortcomings, or, especially regarding 'exaggerated'/'bizarre' traits, as at least partly a result

233   of one of two other biological phenomena termed the species recognition hypothesis and the

234   mutual sexual selection hypothesis.

235

236   Species Recognition Hypothesis

237   Signals can function to discriminate con- from heterospecific individuals in both sexual and

238   non-sexual contexts, such as flocking/shoaling/herding to reduce predation risk (Krause &

239   Ruxton 2002). In the latter context, benefits are likely similar for both sexes, so no dimorphism

240 evolves. Although mixed-species groups certainly form, similarity of morphology and

241 behavior may favor preferential association with one's own species, there being advantages

242 when it comes to group cohesion and escape from predators that seek to separate a prey

243 individual from a group (Croft et al. 2009). A signal evolved to facilitate same-species

244 aggregation could reasonably be described as a trait for species recognition. However, there is

245 no selection driving exaggeration of the trait beyond the minimum for successful detection, so

246 the expectation is that such traits would be relatively low-cost 'road signs' rather than the costly

247 'advertisements' produced through sexual selection (Maynard Smith & Harper 2003; Knell &

248 Sampson 2011).

249   The other context in which species recognition is invoked is in mate choice. Padian &

250 Horner (2011, 2013, 2014) contrast this with sexual selection and have been criticized for doing

251 so (Knell & Sampson 2011; Hone & Naish 2013; Knell et al. 2013b; Knapp et al. 2018). The

252 two concepts are not readily separated (Paterson 1980, 1985; Ryan & Rand 1993). When

253 considering pre-zygotic reproductive barriers, mating with the wrong species is simply an

254 extreme form of sub-optimal mate choice and, because the marginal cost of sperm production

255 is usually lower than that of eggs, selection for mating with the right species will often be

256 higher on females than males. Thus, species recognition for mate choice predicts sexual

257 dimorphism, or lack thereof, in a similar fashion to sexual selection. It is also possible that

258 signals used for species/mate recognition might be exaggerated in order to increase an

259 individual's appeal to the opposite sex (e.g., 'supernormal stimuli' or 'sensory exploitation'

260 hypotheses [Tinbergen 1948; Ryan & Keddy-Hector 1992]). For species recognition as related

261 to mate choice to be separable from sexual selection, the fitness of consequences of mating

262 with different individuals of one's own species would have to be identical. Given the

263 abundance of examples of discriminating mate choice in extant birds, this scenario is unlikely

264 in non-avian dinosaurs (Hone & Naish 2013).

265   One reason why Padian & Horner (2011, 2013, 2014) feel that species recognition for

266   mate choice is readily separable from sexual selection is that they adopt a non-standard

267   definition of what constitutes a sexually selected character. For them, only discrete anatomical

268   traits, such as horns, that are present in one sex and not the other count as sexually selected.

269   This is inconsistent with current sexual selection theory, and indeed Darwin's own writings

270   (Knell & Sampson 2011; Hone & Naish 2013; Knell et al. 2013b; Mendelson & Shaw 2013;

271   Borkovic & Russell 2014). Padian & Horner (2014) claim, "The term "sexual selection" should

272   only be used when one sex uses a feature not present in the other sex to attract mates or repel

273   rivals for mates" (p. 709). If the identification of a sexually selected character requires an

274   extreme dimorphism in the form of binary presence versus absence, then numerous cases of

275   sexual dimorphism seen in modern organisms, expressed as differences in degree, would be

276   rejected. For example, one of the most commonly studied sexually selected traits is body size,

277   which is continuously variable and thus would be excluded under this unreasonably stringent

278   paradigm. Their argument for non-sexual species recognition, therefore, depends on negative

279   evidence, namely that there be no clear examples of presence versus absence sexual

280   dimorphisms in non-avian dinosaurs.

281   There are plenty of examples of species recognition in the context of mate choice that

282   do not require the sorts of exaggerated anatomical structures that are the focus of Padian and

283   Horner's (2011, 2013, 2014) hypothesis. Detailed observations of breeding pedigrees show

284   that some organisms have little difficulty in identifying conspecifics or members of the same

285   newly speciating hybrid lineage and that sufficient character displacement and reproductive

286   isolation can occur rapidly, such as in Galapagos finches (*Geospiza*) which could be argued to

287   lack many of the sorts of conspicuous morphological structures of other species (but do have

288   songs acquired through imprinting) (Grant & Grant 2008, 2010; Lamichhaney et al. 2018).

289   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 signaling. 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/behavioral 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 behavioral (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 versus 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.

314   We do not imply here that various degrees of mutual mate choice or intrasexual

315   competition cannot occur in both sexes of a species. Instead, we discuss a scenario in which

316   mutual sexual selection minimizes sexual variation in a particular trait such that the species

317   might appear to be monomorphic – for example, going beyond a case where both sexes possess

318   an ornament, but where it is expressed to a similar extent in both sexes. Furthermore, this is

319   not to say that previous authors (e.g., Hone et al. 2012) were attempting to propose mutual

320   sexual selection as a 'blanket hypothesis' to be applied to all non-avian dinosaurs, to the

321   exclusion of any dimorphism. However, in order to justify our statistical approach below, we

322   are required to show why mutual sexual selection resulting in minimal to no dimorphism could

323   not be proposed as an alternative to traditional patterns of sexual variation widely across non-

324   avian dinosaurs, as was attempted with the species recognition hypothesis.

325   While the mutual sexual selection hypothesis could explain a supposed lack of sexual

326   dimorphism in non-avian dinosaurs, assuming a lack of dimorphism throughout the grade may

327   be flawed. Abandoning this assumption would then make the mutual sexual selection

328   hypothesis, at least in many cases, unnecessary to invoke. Furthermore, the hypothesis assumes

329   that the effect of any intrasexual competition is nullified by subsequent mate choice processes

330   and that the combined influence of intra- and intersexual selection is equal between the two

331   sexes – a big assumption to apply across all non-avian dinosaurs. A monomorphic equilibrium

332   produced through this sort of mutual sexual selection is likely a rare social/sexual system in

333   modern species (thoroughly demonstrated in a few species [e.g., Jones & Hunter 1993;

334   Kraaijeveld et al. 2004; Nolan et al. 2010]), further compounded by the fact that social/sexual

335   systems can evolve rapidly (Liker et al. 2013). For example, although Kraaijeveld et al. (2007)

336   summarize experiments on 14 bird species with these sorts of mutual ornaments that tested if

337   the ornaments are involved in mate choice (see table 3 therein), these examples may be far

338   exceeded by the number of bird species that lack this precise type of mutual sexual system,

339   given a modern bird diversity of 11,000–18,000 species (Barrowclough et al. 2016). Indeed,

340   some birds are model organisms for sexual dimorphism (e.g., birds of paradise or peafowl,

341   grouse, and pheasants [Phasianidae]) as well as reversed dimorphism (e.g., jacanas [Jacanidae]

342   [Emlen & Wrege 2004]), and it is common knowledge among birders that many species have

343   dimorphic plumage (Sibley 2014). Although this topic is understudied (Lihoreau et al. 2008),

344   the prevalence of such well-balanced mutual sexual selection is unknown, with no indication

345   that it might be as or more common than 'classic' sexual systems (Figure 1). Given the

346   plausible relative scarcity of well-balanced mutual sexual selection among living species, it is

347   unlikely that a highly diverse and disparate grade of animals with at least ~165 million years

348   of evolutionary history and global biogeographic range showed stasis in a social or sexual

349   system, considering the varied and frequent forms of anatomical dimorphism in their living

350   descendants.

351   Another challenge for this hypothesis is that it is strongly helped by *in vivo*

352   experiments/observations of behavior and mate preferences between the two sexes, which is

353   not possible for extinct species. Such experiments can rule out alternative functions for female

354   ornaments/armaments. Selection pressures for female ornaments/armaments other than

355   females competing for mates or mate choice by the males can therefore be tested (vice versa

356   under sex-role reversal). Alternative functions of these female structures could include defense

357   against predators or competition with other females for resources other than mates (Stankowich

358   & Caro 2009; Hone et al. 2012; Tobias et al. 2012). Additionally, genetic correlation

359   (Kraaijeveld et al. 2007) can result in females possessing alleles for a trait that is only actively

360   selected for in males: females produce female offspring with the males whose trait they prefer.

361   For example, females that prefer larger males as mates might produce large daughters when

362   they mate with those males, even if males show no preference for larger females as mates. If

363   expressing those traits is costly to females, expression can be sex-limited (Rice & Chippindale

364   2001; Parker 2006); but if the costs are low, a trait that is only adaptive for males can be

365   expressed in females.

366

367   **Evidence for Sexual Selection and Sexual Variation in Non-Avian Dinosaurs**

368   Under a game theoretic framework, unequal initial investment into reproduction between the

369   sexes is expected to yield different optimal strategies. Therefore, one of the underlying

370   principles of sexual selection theory is that anisogamy tends to result in behavioral,

371   physiological, and anatomical sexual dimorphism (Schärer et al. 2012). Furthermore,

372   statistically, any finite population of males and females is expected to show some non-zero

373   difference between the distributions of the sexes (Nakagawa & Cuthill 2007), regardless of

374   whether sexual selection is acting or not.

375   Non-avian dinosaurs should have exhibited sexual variation according to extant

376   phylogenetic bracketing. Both birds and crocodilians are anisogamic and exhibit various types

377   of sexual dimorphism, including in body size and behaviors (Owens & Hartley 1998; Platt et

378   al. 2009). Even when only examining dimorphism in body masses of extant birds (Dunning

379   2007), without taking into account the prevalent dimorphism in integumentary structures,

380   coloration, and behavior, disruptive selection against monomorphic body mass at the

381   macroevolutionary scale is evident (Figure 1). The $\log_{10}$-transformed distribution of

382   male:female mass is bimodal with peaks corresponding to dimorphism magnitudes

383   approximately +/−10% off from parity between the sexes (i.e., a peak at males ~90% the size

384   of females and a far larger peak at males ~110% the size of females). This serves as a reminder

385   that many sexual dimorphisms are of subtle magnitude. Sexual selection affects rates and

386   directions of phenotypic evolution at the macroevolutionary scale in some birds (Cooney et al.

387   2019). Many crocodilians are known to exhibit size dimorphism, with males growing faster

388   than females to attain larger adult sizes (Wilkinson & Rhodes 1997; Hone & Mallon 2017), in

389　addition to behavioral dimorphism, such as male 'water dance' displays in American alligators

390　(*Alligator mississippiensis*) (Vliet 1989; Moriarty & Holt 2011).

391　　　As an aside, while a non-trivial fraction of bird body mass dimorphism conforms to a

392　reversed pattern with larger females (negative values on the histogram), most birds follow the

393　widespread trend among amniotes toward larger males (Figure 1). Further, a strong positive

394　effect of sexual selection on male, but not female, rates of interspecific divergence of plumage

395　color has been demonstrated in Tyrannida (Cooney et al. 2019). These points weaken the

396　hypothesis that larger, more robust specimens of certain non-avian theropods represent females

397　(Carpenter 1990; Raath 1990; Larson 1994; Chapman 1997). Those hypotheses were based on

398　patterns of body size dimorphism in extant birds of prey. Falconiformes were hypothesized to

399　show reversed size dimorphism due to biomechanical compensation for increases in wing

400　loading during gestation in a clade whose hunting is dependent on flight performance (Wheeler

401　& Greenwood 1983). Such an analogy is likely inappropriate for flightless theropods like

402　*Coelophysis* or *Tyrannosaurus*.

403　　　Non-avian dinosaurs were diverse and morphologically disparate (Barrett et al. 2009;

404　Brusatte et al. 2012) through their global geographic distribution and duration of at least ~165

405　million years. Many lineages of non-avian dinosaurs possessed elaborate or exaggerated

406　structures (e.g., horns, frills in ceratopsians, plates, spikes, spur-like claws in *Iguanodon*,

407　elongated tusk-like teeth in heterodontosaurs, clubs, casques, cranial domes, feathers, bristles,

408　keratinous epidermal spines, or sail-like hyper-elongated neural spines forming a dorsal crest).

409　These resemble modern ornaments or armaments known to be, at least partly, under sexual

410　selection or to exhibit sexual variation (Molnar 1977), such as horns (Bro-Jørgensen 2007),

411　casques (Karsten et al. 2009; Taylor et al. 2017), cranial domes (Wilson 2001), feathers/feather

412　coloration (Møller & Höglund 1991), bristles (Scott & Payne 1934), spurs (Møller 1992), tusks

413　(Cabrera & Stankowich 2018), keratinous epidermal spines (Ord & Stuart-Fox 2005), or sail-

414 like crests (Taylor et al. 2017). Many non-avian dinosaurs show high levels of intraspecific

415 variation of these structures, and some structures developed under positive allometric growth

416 with delayed onset (Hayashi et al. 2009; Hone et al. 2016a), which can be consistent with

417 sexually selection (Bonduriansky 2007). Extreme dimorphisms in non-avian dinosaurs,

418 whereby a trait is present in one sex and absent in the other, might be at risk of being interpreted

419 as entirely ontogenetic or interspecific, as well as obscured by incomplete fossil records. It is

420 also worth noting that even if a structure primarily functions in sexual display/combat and

421 whose evolution is primarily driven by sexual selection, it can still have secondary

422 functions/effects.

423   Although some structures, such as ceratopsian frill epoccipitals, have been suggested

424 to show fluctuating asymmetry (Longrich 2010; Longrich et al. 2010) previously claimed to

425 function as honest signaling of mate quality (Møller & Höglund 1991; Møller 1992; Grammer

426 & Thornhill 1994; Ditchkoff et al. 2001), the connection between fluctuating asymmetry and

427 sexual selection has been doubted due to difficulty in replicating results (Balmford et al. 1993).

428

429 Likely Armaments

430 Beyond the commonly sexually variable trait of body size, there is good evidence (Farke 2014)

431 that non-avian dinosaurs had structures morphologically analogous to armaments of modern

432 animals (i.e., they might have used the structures for combat), with biomechanical analyses

433 suggesting the ability to use the structures as weapons (i.e., they could have used the structures

434 for combat), and with pathologies consistent with intraspecific combat (i.e., they likely did use

435 the structures for combat).

436   Ceratopsian horns developed late in ontogeny and have been hypothesized to be

437 involved in mate competition (Sampson et al. 1997). *Triceratops* and *Centrosaurus* show

438 variation in their horns that mirrors sexual variation in many modern bovids (Poissant et al.

439   2008). In these bovids, males have large, highly curved horns with wide bases whose tips point

440   back towards the skull to allow for non-lethal sparring/head-butting and withstanding

441   associated forces. Female bovids have smaller, thinner horns whose tips point away from the

442   skull to allow for stabbing defense against predators (Packer 1983; Caro et al. 2003). Similarly,

443   while smaller and juvenile *Triceratops* and *Centrosaurus* specimens have thin horns that tend

444   to point upward, larger specimens tend to have large, broad, downward-curving horns (Horner

445   & Goodwin 2006; Frederickson & Tumarkin-Deratzian 2014) that might represent an

446   anteriorly-curving analog to the posteriorly-curving pattern in many bovids, despite the fact

447   that researchers have attributed this horn variation entirely to ontogeny (or taxonomy) and not

448   sex. Since these bony horn cores were likely covered and further extended by a keratin sheath

449   (which can sometimes preserve as calcium phosphate [Brown et al. 2017; Saitta et al. 2018;

450   Saitta & Vinther 2019]), the potential for morphological variation *in vivo* is even greater than

451   the observed skeletal variation. Morphologically complex sheathing can be present on simple

452   horn corns of modern species, like pronghorn antelope (*Antilocapra americana*) whose single

453   tipped horn core supports a two-pronged keratin sheath (Davis et al. 2011). Most extant species

454   use their horns for intraspecific combat, even Jackson's chameleons (*Trioceros jacksonii*),

455   which can exhibit dramatic sexual dimorphism in some subspecies (Waring 1997). Even

456   unusual horn morphologies are used in sparring, such as the spiral horns of greater kudus

457   (*Tragelaphus strepsiceros*) that are at risk of becoming locked together during fights between

458   males (Owen-Smith 1993), the highly inwardly-curved horns of cape buffalo (*Syncerus caffer*)

459   (Turner et al. 2005), or the giant horns of ankole-watusi cattle (*Bos taurus*) (Huber et al. 2008).

460   Pathologies on *Triceratops* skulls are consistent with intraspecific sparring based on the

461   geometry of where horn tips would contact a rival while sparring (Farke 2004; Farke et al.

462   2009). In a study of macroevolutionary and biogeographic trends in ceratopsians, Knapp et al.

463   (2018) could not explain their exaggerated structures simply by non-sexual species recognition.

464   Note that, as Knapp et al. (2018) are careful to qualify, treating all exaggerated structures (e.g.,

465   ceratopsian horns) as ornaments can be problematic given the importance of distinguishing

466   armaments from ornaments in sexual selection (McCullough et al. 2016).

467   Pachycephalosaur cranial domes are morphologically and histologically similar to

468   extant species with domes or thickened skull roofs used for head butting, such as duiker

469   (Cephalophinae), bighorn sheep (*Ovis canadensis*), cape buffalo (*Syncerus caffer*), or musk

470   oxen (*Ovibos moschatus*), and the alignment between their skull and vertebral column are also

471   similar (Galton 1970; Sues 1978; Alexander 1989; Giffin 1989; Carpenter 1997; Snively &

472   Theodor 2011). They show intraspecific variation between flattened and domed crania, which

473   has a strong ontogenetic signal (Schott et al. 2011) but could also be influenced by sex –

474   possibly an example of extreme presence versus absence dimorphism (Galton 1971).

475   Biomechanical analyses of pachycephalosaur domes using finite element analysis and simple

476   physical calculations show that they were capable of withstanding stress and strain from head-

477   butting, especially when a plausible amount of keratin sheathing around the dome is included

478   (Alexander 2006; Snively & Cox 2008; Snively & Theodor 2011). Finally, pathologies on their

479   cranial domes are consistent with injuries sustained from head-butting, as seen in extant head-

480   butting species (Peterson & Vittore 2012; Peterson et al. 2013).

481   Although ankylosaur tail clubs lack obvious modern analogs, some extant species use

482   tail whipping in intraspecific combat, such as the lizard *Agama agama* (Schall et al. 1989;

483   Arbour & Zanno 2018). Ankylosaur tail clubs have been shown through finite element analysis

484   to be capable of withstanding the stress and strain from use as an armament (Arbour & Snively

485   2009). Although limited, some possible pathologies in anterior caudal vertebrae and tail clubs

486   have been noted (Arbour & Currie 2011). Stegosaur tail spikes have been suggested to function

487   in defense against predators based on pathologies in *Allosaurus* bones (Carpenter et al. 2005)

488   and in *Stegosaurus* tail spikes themselves (McWhinney et al. 2001). Like some ankylosaur

489 specimens, anterior caudal vertebrae pathologies have been noted on several North American

490 stegosaur specimens, including the wide-plated hypothesized male of *Hesperosaurus mjosi*

491 (Saitta 2014). Whether stegosaur tail spikes would have been excluded from intraspecific

492 combat due to excessive damage and potential lethality (i.e., the 'total war' avoidance

493 hypothesis [Maynard Smith & Price 1973]) remains to be determined.

494

495 Likely Ornaments

496 Although ornamental function in extinct species is difficult to study given the lack of

497 behavioral data, many non-avian dinosaur structures are consistent with display or inconsistent

498 with mechanical usage as an armament (Hone et al. 2012). Hadrosaur casques house elaborate

499 nasal passages that could have produced low-frequency sounds (Weishampel 1981) and

500 exteriorly (although not in internal structure) resemble casques of modern animals that can

501 exhibit sexual variation or be under sexual selection, such as hornbills (Bucerotidae) (Gamble

502 2007), chameleons (Chamaeleonidae) (Karsten et al. 2009), or casque-headed lizards

503 (Corytophanidae) (Taylor et al. 2017). A soft tissue caruncle, similar to the snoods, wattles,

504 and combs of some modern birds, has been described in *Edmontosaurus* based on preserved

505 skin impressions (Bell et al. 2014), which could represent a sexual ornament, assuming that

506 this is not a taphonomic artefact from decay.

507   Sail-like, hyper-elongated, vertebral neural spines consistent with sexual ornaments

508 (Isles 2009) in spinosaurs (e.g., *Spinosaurus*), sauropods (e.g., *Amargasaurus*), ornithopods

509 (e.g., *Ouranosaurus*), carcharodontosaurs (e.g., *Concavenator* or *Acrocanthosaurus*)*,* and

510 ceratopsians (e.g., *Leptoceratops* or *Koreaceratops*) resemble the sail-like crests and elongated

511 neural spines of various lizards. These include casque-headed lizards like the plumed

512 basilisk (*Basiliscus plumifrons*) (Taylor et al. 2017), sailfin lizards like the Philippine sailfin

513 lizard (*Hydrosaurus pustulatus*) (Ord & Stuart-Fox 2005), or chameleons 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. 2007) 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

2005).

  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). 'Paleo-color' reconstructions of some non-avian

dinosaurs reveal iridescent feather colors consistent with social/sexual signaling (e.g.,

Microraptor), including a 'rainbow' iridescence in *Caihong* analogous to hummingbirds

(Trochilidae) (Li et al. 2012; Hu et al. 2018). Color and color 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 analogs 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

538 plates and frills share at least some functional similarities (e.g., sexual display) to sail-like

539 dorsal crests or cranial casques, respectively.

540   Trace fossils have been used to suggest lekking display behavior in theropod dinosaurs

541 based on footprints and scratch marks (Lockley et al. 2016), although scratch marks can be

542 produced through non-lekking or non-sexual behaviors (e.g., antagonistic displays, territorial

543 marking, or searching for resources).

544

545   **Previous Methods of Investigation into Sexual Dimorphism in Fossils**

546 Methods for detecting sexual dimorphism in fossil taxa vary according to the evidence they

547 invoke, the alternate hypotheses they test, their commitment to quantitative data, and the

548 statistical methods used. In the extreme, some authors (Padian & Horner 2011, 2013, 2014)

549 require binary differences (i.e., presence/absence) to accept dimorphism in fossils. Such a

550 qualitative approach would preclude studying dimorphism in traits such as body size, which

551 may be the most common type of dimorphism among animals. Here, we focus on two recent

552 statistical investigations into sexual dimorphism. Both methods represent univariate

553 significance tests that can be ineffective for detecting a signal of sexual dimorphism and suffer

554 from low statistical power (i.e., a tendency to fail to detect an effect when present) – a serious

555 concern when studying datasets with low sample sizes, as is the case with most vertebrate

556 fossils. These approaches appear to be inconsistent with the American Statistical Association's

557 recent statement about overreliance on p-values that says, "Scientific conclusions and business

558 or policy decisions should not be based only on whether a p-value passes a specific threshold"

559 (Wasserstein & Lazar 2016, p.131).

560

561 The Bimodality Method

562 Hartigans' dip test for unimodality has been used to test for sexual dimorphism in non-avian

563 dinosaurs and to conclude that there is no evidence for it (Mallon 2017). Mallon (2017) writes

564 that "no evidence for sexual dimorphism was found in any of the examined taxa" (p. 495),

565 although is careful to state, "This is not to say that dinosaurs were not sexually dimorphic

566 (phylogenetic inference suggests they may well have been), only that the available evidence

567 precludes its detection" (p. 495). While Mallon (2017) only examined non-avian dinosaur

568 datasets, it is unlikely that other fossil groups would have passed these tests either, because a

569 key point regarding significance testing versus effect size statistics remains. There is a subtle,

570 but important, difference between a claim of 'no dimorphism' in non-avian dinosaurs and a

571 claim of 'no evidence for dimorphism'. It is certainly true, and sometimes acknowledged, that

572 failure to achieve a certain p-value is not evidence for the absence of an effect, but even to say

573 'no evidence' is potentially wrong. This is because such a statement can be the product of

574 excellent data that estimates an effect to be near zero, with tight confidence intervals around

575 that point; or it can be the result of poor data and broad confidence intervals. For the distinction

576 to be clear, effect sizes should be estimated and uncertainty quantified. This is even more

577 sensible when context is included. There are many lines of evidence from the theoretical to the

578 empirical (both modern and fossil) for sexual variation in non-avian dinosaurs, regardless of

579 the weight given to different arguments by different researchers. Our methodologies,

580 particularly our statistical analyses, should reflect this evidence, which does not happen when

581 a null hypothesis is set to monomorphism by convention.

582 Beyond the tempting conclusion of no evidence for dimorphism in non-avian dinosaurs,

583 the bimodality method suffers from further issues. Mallon (2017) reports a method whereby

584 the data are first tested for normality with Shapiro-Wilk and Anderson-Darling tests, followed

585 by Hartigans' dip test for unimodality. The method cannot accommodate ontogenetic effects

586 unless juveniles or sufficiently young individuals are excluded from the dataset (as Mallon

587   [2017] did in some analyses). Especially when lacking histological evidence of growth rate or

588   other indicators of sexual maturity (e.g., medullary bone or *in situ* fetuses/eggs), excluding

589   smaller individuals risks excluding the smaller sex of a potentially dimorphic species and also

590   reduces sample size.

591   A key problem of the Mallon (2017) approach is that sexually variable traits do not

592   always exhibit great enough effect size to produce a bimodal distribution in a single variable,

593   irrespective of sample size. In other words, the magnitude of the dimorphism (i.e., the

594   difference in measures of centrality between male and female distributions) relative to the

595   intrasexual variation (i.e., the spread of the male and female distributions) might not be great

596   enough to produce a negative curvature in the center of the combined male and female

597   distribution. This can be true even for the theoretical population (i.e., a hypothetically infinite

598   sample size), let alone in a random sampling of that population. This statistical problem has

599   been known for some time, as stated by Schilling et al. (2002, p. 233), "a mixture of equally

600   weighted normal distributions with common standard deviation σ is bimodal if and only if the

601   difference between the means of the distributions is greater than 2σ". Schilling et al. (2002)

602   demonstrated this using human height as an example of an accepted sexual dimorphism whose

603   effect size is too small to produce a bimodal distribution, even in theoretical population

604   distributions (i.e., if one were able to measure an infinite number of men and women).

605   As acknowledged by Mallon (2017), the significance testing used (Hartigans' dip test)

606   is highly susceptible to false negatives, or type II errors, and therefore low statistical power.

607   Dip tests are thus an ineffective method for identifying sexual dimorphism in an extinct species.

608   Just how large might the magnitude of dimorphism have to be in order to pass the Hartigans'

609   dip test? Inspired by the illustrative example of Schilling et al. (2002), we have statistically

610   modeled the heights of adult men and women and subjected them to Hartigans' dip test (Figure

611   2), since human height is a commonly accepted and familiar example of sexual dimorphism.

612   We randomly generated data for women using a normal distribution whose height was fixed at

613   an average (μ) of 162 cm with a standard deviation (σ) of 7 cm, a reasonable approximation of

614   observed distributions of women's heights (Schilling et al. 2002). We then randomly generated

615   data for men's height using a normal distribution whose average (μ) was allowed to vary, while

616   keeping the standard deviation (σ) also fixed at 7 cm, in order to study the influence of effect

617   size without confounding the impact of varying standard deviations. After randomly generating

618   male and female data, the data were combined into a single distribution, analogous to not

619   independently knowing the sex of any individual. We also allowed for the total sample sizes to

620   vary from 20 to 20,000 in seven steps (under a log-scale), while keeping the ratio of

621   men:women at 1:1 in each sample to avoid the impact of unequal sex ratios in the samples. At

622   each combination of male average height and total sample size, we generated 1,000 datasets,

623   performed a Hartigan's dip test (10,000 Monte Carlo replicates) on each dataset, and then took

624   the average p-value ($\alpha = 0.05$) produced over those 1,000 iterations. In order to produce a

625   significant average p-value in support of a bimodal distribution at sample sizes consistent with

626   those typical of fossil vertebrates (e.g., 20 and 62 on our log-scale), the average height of men

627   has to be ~192–200 cm – an extreme magnitude of dimorphism relative to natural human

628   populations. Given a more realistic average male height of 176 cm (Schilling et al. 2002), no

629   sample size would yield a significant average p-value in support of bimodality. This is because

630   the underlying theoretical population distribution at this magnitude of dimorphism is in fact

631   unimodal, and so this result is not simply a matter of insufficient sample size. As described by

632   Schilling et al. (2002), this is where the equally weighted (i.e., 1:1 men:women) normal

633   distributions with a common standard deviation ($\sigma = 7$ cm) have a difference in means (176 cm

634   – 162 cm = 14 cm) that is not greater than 2σ (2 x 7 cm = 14 cm). Therefore, any observed p-

635   value under 0.05 at this particular condition would actually represent a true negative, despite

636 the fact that these theoretical parameters are a fairly realistic approximation of naturally

637 occurring sexual dimorphism in the heights of men and women in many human populations.

638   A further problem for the use of bimodality tests to examine sexual dimorphism is that

639 changing the proportion of males to females in the sample or the ratio of standard deviations

640 between the male and female distributions can require even greater differences between the

641 averages of the male and female distributions in order to yield a bimodal theoretical population

642 of combined males and females (Figure 3) (Schilling et al. 2002).

643   Mallon (2017) himself demonstrated that the Hartigans' dip test could not always detect

644 genuine signals of sexual dimorphism when two datasets of extant sexually dimorphic species

645 (alligator and the white-browed coucal bird [*Centropus superciliosus*]) failed to yield a

646 significant result. Rather than concluding that the statistical test was ineffective, Mallon (2017)

647 instead concluded that without independent knowledge of the sex of the individuals,

648 dimorphism might not be detectable. Such a conclusion could effectively prevent any

649 significant research into sexual selection in the fossil record of many taxa, given that it is

650 commonly not possible to identify the sex of a fossil.

651   In addition to Hartigans' dip testing, Mallon (2017) used mixture modeling to fit normal

652 distributions to fossil datasets, and the optimal number of normal distributions was selected

653 using the Akaike information criterion (AIC), with two distributions interpreted as evidence

654 for sexual dimorphism. Although mixture modelling can be extremely useful for a large and

655 representative sample (especially when used alongside other approaches), as with any type of

656 statistical modelling, there are some considerations. Mixture modeling identifies the best ways

657 to explain the observed data by combining multiple distributions, without fully taking

658 uncertainty into account. In other words, deviations in a random sample from the true,

659 theoretical population distribution as a result of noise or small sample sizes are modeled as if

660 they describe the shape of the population distribution, when they are instead 'fitting the noise'.

661   Small sample sizes and preservation biases in fossil data can therefore lead to mixture models

662   that overfit the data. When the data are univariate, overfitting can also be an issue when

663   juvenile/young specimens are included. Finally, although no *p*-values are involved, there is still

664   the risk of dichotomously concluding 'no evidence for dimorphism' based on this particular

665   use of AIC scores. However, in the future, mixture modeling may become an important tool to

666   be incorporated into our framework below with respect to sex assignment and judging

667   intrasexual variation.

668

669   *t*-tests

670   Hone and Mallon (2017) simulated male and female data using parameters derived from

671   empirical growth curve data on alligators and rheas (*Rhea americana*). Their alligator data was

672   based on a von Bertalanffy growth equation, while rhea data was based on a Gompertz growth

673   equation. Their study then randomly sampled from the generated data and used *t*-tests to

674   determine whether the sampled male and female data showed statistically significant

675   differences in mean body size. The rhea data tended to give statistically significant results more

676   often than the alligator data, interpreted as a consequence of protracted growth in alligator.

677   Since many non-avian dinosaurs may have more protracted growth patterns than rheas, they

678   suggested that this could explain the difficulty in detecting dimorphism in non-avian dinosaurs.

679   By using significance testing, Hone and Mallon (2017) assume a statistical framework

680   in which monomorphism is the null hypothesis under the supposed absence of evidence for

681   dimorphism in non-avian dinosaurs; in other words, a binary approach. Furthermore,

682   independent knowledge of the sex of each individual is not possible for most fossil samples,

683   although one could attempt to guess the sex of each datapoint (see discussion of sex assignment

684   below). The test is also univariate, despite the fact that the data being analyzed represent

685   bivariate growth curves, with one variable being the sex-related trait of body size and the other

686 being age (e.g., Wilkinson & Rhodes 1997; Navarro et al. 2005). The fullest signal of

687 dimorphism is the divergence of two growth curves, one for each sex.

688   The results of Hone & Mallon (2017) do not appear to be solely sensitive to the

689 prominence of dimorphism. Despite the fact that the rhea data outperform the alligator data in

690 the *t*-tests, the alligator data showed more prominent dimorphism with a large magnitude of

691 growth curve divergence (i.e., effect size) alongside low intrasexual variation. The alligator

692 data showed clear separation of the largest male and female points, unlike the rhea data. The

693 ease with which sampled males and females pass a *t*-test depends not only on the magnitude of

694 the dimorphism (which was not controlled for in the two-species comparison), but also on a

695 combination of growth rate and life span, as noted by Hone & Mallon (2017). Faster growth

696 rate means that adult size is more quickly attained, but it is the growth rate in relation to life

697 span that dictates the proportion of adults in the sample and distinguishes the rhea dataset from

698 the alligator dataset. By increasing the proportion of the population at adult size, the proportion

699 of the population in which dimorphism is maximally expressed is also increased, meaning that

700 a random sample is more likely to contain a relatively stronger signal of dimorphism

701 independent of the magnitude of dimorphism (Hone & Mallon 2017). Because both species

702 show greater ontogenetic variation than sexual variation, excluding juveniles is key to passing

703 the *t*-test. This is consistent with the lower *p*-values derived from their alligator simulations

704 with a size bias against smaller individuals and higher *p*-values derived from their alligator

705 simulations with altered population structure (i.e., greater proportion of juveniles). As

706 mentioned by Hone & Mallon (2017), the vertebrate fossil record is often not biased in favor

707 of overrepresentation of small juveniles because of taphonomic and collecting factors (Brown

708 et al. 2013), as well as population structures hypothesized for some non-avian dinosaurs

709 (Erickson et al. 2006, 2009; Woodward et al. 2015; although see Varricchio [2011] on

710 aggregations of juveniles in some bonebeds).

711

712 **A Framework for Investigating Sexual Dimorphism in Extinct Taxa**

713 We propose a different method to investigate sexual variation in the fossil record. Rather than

714 performing univariate significance tests on a single dataset to either reject or fail to reject sexual

715 monomorphism in that species (i.e., dichotomous hypothesis testing), this approach focuses on

716 effect size statistics (Nakagawa & Cuthill 2007; Amrhein et al. 2017; Amrhein & Greenland

717 2018; Halsey 2019; Holland 2019), combined with supporting contextual evidence. Our aim is

718 to test alternate hypotheses against the observed variation, as well as estimate its magnitude

719 and constrain uncertainty of that estimate in a given species/dataset compared to other

720 species/datasets. Our framework first involves collecting appropriate data, followed by the

721 consideration of several alternate hypotheses, and finally inferring biological implications of

722 the results, similar to the approach employed by Saitta (2015) on stegosaur fossils. We attempt

723 to provide a fairly exhaustive list of alternate hypotheses, along with lines of evidence that can

724 be used to systematically abandon them one by one. However, some alternate hypotheses are

725 case specific (e.g., intra-individual variation of repeated structures).

726 We introduce our quantitative effect size approach with respect to the first alternate

727 hypothesis of non-sexual variation/polymorphism in a population. Recently, over 800

728 researchers from statisticians, medical scientists, biologists, to psychologists signed a letter

729 calling for an end to overreliance on statistical significance (Amrhein et al. 2019). In 2019, the

730 73rd volume (supplement 1, issue 1) of the academic journal *The American Statistician*

731 published a series of articles echoing this sentiment. Issues with significance testing extend

732 beyond those situations with small sample size and type II error described above. In fact, very

733 large sample sizes can lead to very small *p*-values, even if the effect size is small. This low *p*-

734 value can be falsely interpreted as indicative of great biological importance, even though the

735 observed biological effect is small. We think that the points we raise here are more consistent

736 with this statistical paradigm shift than those from much of the previous work on this topic.

737 We abide by the suggestion that "you can enhance the information provided by frequentist

738 statistics with a focus on effect sizes and a quantified confidence that those effect sizes are

739 accurate" (Halsey 2019, p.1).

740

741 Data collection

742 As with any statistical analysis and scientific study, maximizing the sample size is a priority.

743 While sample sizes of many fossil groups remain small after one or two centuries of scientific

744 collecting, these will undoubtedly increase over time, and science is an intergenerational

745 pursuit. Other than zero or one, from which an estimate of sexual variation cannot be based on

746 data, there is no fixed lower limit on sample size. Sample sizes of two clearly result in

747 extremely high uncertainty, as in Persons et al. (2015).

748   In our approach, we use quantitative bivariate data to produce growth curves of body

749 size or other traits that are commonly secondary sexual characteristics (e.g., putative

750 ornaments/armaments or coloration) (Figure 4). The use of growth trajectories to study sexual

751 variation is precedented (Evans et al. 2018). It is best to avoid traits that are suspected to show

752 negligible sexual variation or be under weak sexual selection (e.g., stegosaur femoral head

753 shape [Barden and Maidment 2015]), unless in comparison to another trait suspected to be

754 sexually variable.

755   Not only is this putative dimorphic trait quantified (e.g., direct measurements, principal

756 component scores, outlines [Bonhomme et al. 2014; Saitta 2014]), it is also collected alongside

757 some measure of/proxy for age (Hone et al. 2016b). For example, this could be histological

758 lines of arrested growth or body size, if body size is not the trait being examined for potential

759 dimorphism. Note that body size is not always a good proxy for age (Hone et al. 2016b) and

760 that the potentially confounding effect of dimorphism in both body size and the other selected

761 trait should be considered. In that case, it might be possible for both sexes to show similar

762 growth curves (i.e., patterns of growth) for a trait when plotted against body size, but the curve

763 for the smaller-trait-bearing sex might terminate at a smaller body size – meaning that

764 individuals of the smaller sex are hidden among younger individuals of the larger sex (Hone &

765 Mallon 2017). If this confounding is suspected, then alternative proxies for age or more precise

766 age estimates (e.g., using lines of arrested growth) can be carried out rather than using body

767 size as an age proxy. Some measures of age might be clade specific. For example, the ages of

768 fossil elephant specimens are often assessed based on tooth eruption and wear (Maschenko

769 2002; Lister 2009).

770 The advantage of regression analysis on bivariate growth curves over univariate

771 analysis on a trait of interest is that the addition of immature individuals in a univariate analysis

772 can lead to greater overlap of male and female distributions, making dimorphism harder to

773 detect (e.g., if females resemble immature males). In contrast, bivariate regression analysis can

774 not only cope with juvenile data points, but juvenile points make curve fitting computationally

775 easier.

776 Researchers should then address alternate hypotheses for the anatomical variation

777 within a sample. If appreciable evidence against all alternate hypotheses can be obtained, then

778 sexual dimorphism is well supported.

779

780 Alternate Hypothesis 1: Controlling for Non-Sexual Variation and Polymorphisms in

781 Populations

782 Other than sexual dimorphism, non-sexual variation/polymorphisms within a population might

783 explain observed variation. Continuing from the efforts to collect data on traits likely to be

784 sexually variable, to rule out this alternate hypothesis, evidence should show that the trait of

785 interest diverges in its growth curve. If, for example, clear patterns of multiple split/separate

786 growth curves suggest three or more morphs, this would indicate that the variation cannot be

787 explained simply by sexual dimorphism. A non-sexual intra-population

788 variation/polymorphism hypothesis can be abandoned under evidence for divergence of a

789 plausibly sexual trait into two growth curves, ideally without overlap of confidence intervals

790 and thereby good estimated support for sex-specific growth models.

791   Controlling for geographic and stratigraphic distribution of the samples can help to rule

792 out geographic or temporal variation of a single species (e.g., biogeographic subspecies or

793 morphological gradients). Datasets from a single locality and horizon are generally better than

794 datasets that span large stratigraphic intervals and geographic areas, as the former are more

795 likely to represent a single population. This approach is comparable to that used to rule out

796 more extreme interspecific variation below.

797   To illustrate the utility of our framework, we first simulate alligator and rhea growth

798 curves under a variety of sample sizes (10–250) and effect sizes (Tables 2–3) over the same

799 age ranges as in Hone and Mallon (2017); multiple runs of this simulation and the code are

800 available as supplemental files. Effect sizes (i.e., difference in L parameters between the sexes)

801 range from true monomorphism (i.e., 0) to double those observed in empirical data for these

802 species (Wilkinson & Rhodes 1997; Navarro et al. 2005). For now, we keep the function

803 (according to size) for population standard deviations of the residuals along the growth curves

804 equal between the sexes (except Figures 5, 12). Note that this still generates greater intrasexual

805 variation in the larger sex. In order to introduce the dynamics at play, our initial simulations

806 keep equal proportions of males and females (except in the appendix and Figure 12). Unlike

807 empirical datasets, simulated data allows for thorough examination of a methodology because

808 the effect size, sex ratio, intrasexual variation, and sample size can all be controlled or

809 experimentally altered as needed. In simulation, the true population-level theoretical values of

810 the parameters are known since these are used to generate the data. This allows for the effect

811   size estimated from a sample using a particular method to be compared to the true population

812   effect size, not simply the observed effect size of a given empirical sample for which the sex

813   of each individual is known.

814   *Identify a Signal Consistent with Growth Curve Divergence.*— A preliminary

815   examination of the data can help determine if further analysis is likely to give a strong signal

816   of dimorphism. A conspicuous signal consistent with divergence of male and female growth

817   curves should show separation or spreading of the data at older ages or larger sizes. Separation

818   might be prominent enough to be noted visually (Saitta 2015). Even if there is no conspicuous

819   separation with a gap, an increase in the spread of the data in older/larger individuals can be

820   consistent with sexual dimorphism. However, the spread of the data could increase along the

821   growth curve even in a truly monomorphic case (i.e., effect size of zero) due to variable growth

822   trajectories of individuals regardless of sex or due to any preferential collection of larger

823   specimens such that juveniles are underrepresented in the sample. Putative sexual traits that are

824   constant or decrease in spread along a growth trajectory, or that are impoverished in

825   adults/large specimens, might be viewed with skepticism, as they likely provide little evidence

826   for sexual variation.

827   Even when both alligator and rhea simulated data are dimorphic (i.e., different growth

828   equations are used to generate males and females), sample sizes are large, and all parameters

829   (including effect size and sex-specific population standard deviations of the residuals) are set

830   to their naturalistic, empirically derived values (Figure 5), only alligator shows clear separation

831   (i.e., large effect size with small intrasexual variation). Rhea only shows an increase in spread

832   of the data without separation. To more easily judge spread, a regression can be fitted to the

833   data as a whole and the residuals can be examined via a residual plot. When a single regression

834   is fitted, naturalistic simulation of alligator data has residuals that not only show an increase in

835   spread, but also clear separation along positive and negative values (indicating even sex ratios

836 and lack of extreme outliers). Naturalistically simulated rhea simply shows a sharp increase in

837 the spread of the residuals under the first ~10% of the life history as evidence of sex-specific

838 growth curve divergence.

839   The type of regression fit to the data (e.g., von Bertalanffy, Gompertz, logistic,

840 logarithmic, exponential, linear, etc.) will vary depending on the type of growth, as dictated by

841 the variables measured and taxon studied, and can be inferred based on R^2 or AIC values as

842 well as known growth patterns in related species. Conspicuousness of data separation or

843 spreading can decrease with smaller sample sizes.

844   *Assign Sex.*— The next step involves assigning the sex of each individual, which can

845 be accomplished in a variety of ways. Here, we focus on the simplest method as an illustrative

846 example (Figure 6), which involves fitting a single best-fit growth curve to the total dataset and

847 assigning sex to specimens based on whether their residuals relative to that curve are positive

848 or negative. One weakness of this method is that it assumes continuous sex-specific growth

849 throughout life, when secondary sexual characteristics might have delayed developmental

850 onset at or near sexual maturity (e.g., facial hair in men coinciding with puberty [Lee 1980]).

851 Additionally, when effect size is small and sexes heavily overlap, there will be bias towards

852 overestimation of effect size (see below). However, an advantage is that fitting such a curve to

853 datasets can be computationally easy.

854   Despite criticism of this sort of division about the center as arbitrary (Mallon 2017), the

855 accuracy (i.e., percentage of true members of a sex that are correctly assigned to that sex) of

856 even this simplistic method in our simulations can be >80% with moderate/naturalistic

857 magnitudes of dimorphism (relative to empirical alligator and rhea data), so long as males and

858 females are similarly represented in the sample. This is because we are not assuming any sex

859 roles (e.g., that males are larger), but rather assigning the points to either a larger or smaller

860 sex, regardless of whether sex-role reversals are present. Avoiding this assumption can prevent

861   unsupported interpretations of mating systems when independent evidence of which sex is

862   larger is absent. Accuracy increases more rapidly with increasing true effect size for the

863   alligator simulations than the rhea simulations due to the smaller intrasexual variation relative

864   to the magnitude of naturalistic dimorphism in alligator growth. This method is expected to be

865   better than random at accurately assigning sex under many circumstances (i.e., if true effect

866   size > 0 and both sexes are equally represented, then sex assignment accuracy > 50% on

867   average) (Figures 7–8). In this specific case (i.e., sex ratio parity), Mallon's (2017) assertion

868   that the technique is arbitrary only holds true on average in the extremely unlikely circumstance

869   that the magnitude of sexual variation is precisely zero. In all other circumstances in which

870   there is any non-zero effect size, the technique will on average (under non-skewed sex ratios)

871   have some degree of useful discriminatory ability in assigning sex. Truly monomorphic

872   datasets (i.e., effect size = 0) with equal proportions of males and females still appear to yield

873   sex assignments with accuracies that center around 50%. Smaller sample sizes lead to greater

874   variability in accuracy between datasets at a given effect size (i.e., spread of accuracy values

875   increases with decreasing sample size) (Figures 7–8).

876   The accuracy of this simplistic approach for sex assignment is consistent with previous

877   studies showing that division about the mean is one of the most robust approaches for

878   estimating dimorphism when accounting for possible variation in sample size, intrasexual

879   variation, and even sex ratio of the sample (Plavcan 1994; Rehg & Leigh 1999). Even simplistic

880   approaches should be satisfactory under many combinations of effect size and sample size, so

881   long as the sample is not highly skewed in sex ratio (Allentoft et al. 2010). Heterogametic sex

882   determination, as in birds, is expected to result in primary sex ratio parity (i.e., at hatching) –

883   relevant to our framework given that we suggest examining a wide ontogenetic range.

884   Assuming no subsequent sex bias in mortality, roughly equal male and female representation

885   would be expected in a random sampling at large sample size. If the probability of an individual

886   fossil datapoint being either male or female is analogous to independent coin flips, then the

887   true relative frequency of each sex (prior to any attempt to assign sex, regardless of method)

888   will approach 50% as more samples are added.

889   Skewed sex ratios could result from biological (Laver et al. 2012) or taphonomic factors

890   (e.g., sex-specific behavior/ecology leading to different burial environments and preservation

891   potential) as well as random sampling at small sample size. Although there is considerable

892   interspecific variation, adult sex ratios can, for example, be somewhat male-biased in birds

893   (Donald 2007; Liker et al. 2014) with the main driver away from a 1:1 primary sex ratio being

894   juvenile mortality (Szekely et al. 2014; Eberhart-Phillips et al. 2018). If a species of interest is

895   suspected to have skewed sex ratios, one possibility would be to introduce a correction factor

896   in the sex assignment step based on observed values in extant members of the clade to see if

897   any trends in estimated sexual variation across species persist once this correction is applied.

898   Some groups, such as crocodilians, show environmental sex determination driven by

899   incubation temperature (Paukstis & Janzen 1990). Theoretical work has suggested that

900   environmental fluctuations from year to year can cancel out fluctuations in sex ratios in long-

901   lived species with environmental sex determination (Bull & Bulmer 1989), and so most fossil

902   datasets of such groups might provide a time-averaged sex ratio closer to parity. A taphonomic

903   example where the sex ratio would differ massively from equality, even at large sample sizes,

904   is a fossilized death assemblage of a single-sex group (Hone & Mallon 2017) due to single-sex

905   socialization or temperature-dependent sex determination. In single-sex-dominant assemblages

906   (Chen et al. 2019), a tradeoff would occur between greater sample size and skewed sex ratio.

907   A dataset that, for example, relies solely on a single bonebed might be better able to control

908   against geographic/stratigraphic confounding, but be at risk of skewed sex ratios. In a single-

909   sex-dominant assemblage, one might expect any variation of traits under sexual selection to be

910 smaller relative to a sample with both sexes represented, and therefore the calculated effect
911 size might be underestimated with respect to measures of uncertainty.

912   In addition to greater overlap of males and females (i.e., greater intrasexual variation
913 relative to effect size), more heavily skewed sex ratios can decrease sex assignment accuracy
914 in this illustrative approach. Fitting a regression to the data may estimate a sex ratio at parity
915 (excepting the impact of outliers/influential observations, which can heavily affect the
916 regression, or datasets with an odd sample size). Skewed sex ratio in a sample can allow sex
917 assignment accuracies for a single sex to fall below 50% on average (appendix). Therefore,
918 future work should compare the accuracy of different sex assignment methods while accouting
919 for growth, perhaps incorporating principal component analysis, cluster analysis, k-means
920 clustering, or Gaussian mixture modeling. In particular, mixture modelling might promise to
921 help reduce bias when the sexes heavily overlap – possibly by 1) removing juvenile points
922 before applying a univariate 2-component Gaussian mixture model such that the distance an
923 adult lies from the single best-fit curve dictates the probability that it is assigned to a particular
924 sex (especially in the asymptotic region of Gompertz curves) or 2) through a more sophisticated
925 mixture modelling that combines regression analysis. Furthermore, when paleontological
926 indicators of sex are present in some individuals (e.g., *in situ* embryos or bacula), then care can
927 be taken not to assign these datapoints to different sexes, thereby increasing sex assignment
928 accuracy. For example, if the sex of individuals with variable skeletal completeness is
929 estimated using the residual-based approach described here, and 90% of the specimens with
930 known bacula are assigned to the larger sex, then the other 10% of specimens with known
931 bacula can be reassigned to the larger sex to improve sex assignment accuracy.

932   Sex assignment goes to the heart of a major misconception about studying sexual
933 selection in the fossil record, which provides the undertone for the statement that "in the
934 absence of a priori knowledge of sex, the ability to detect dimorphism in a fossil sample is

935   likely only in cases of strongly expressed dimorphism (i.e., well-separated peaks on a

936   histogram or discrete character states); weakly expressed dimorphism, where the sexes overlap

937   considerably in morphospace, may be impossible to discriminate in the fossil record" (Mallon

938   2017, p. 502). While it is true that large effect sizes are often easier to detect than smaller effect

939   sizes, abandoning dichotomous detection versus non-detection can allow comparison between

940   large effect sizes with low uncertainty and small effect sizes with high uncertainty in a

941   quantitative continuum. In many cases, it is not necessary to know the sex of each specimen

942   from independent evidence to produce an estimate of effect size that can be compared to other

943   datasets.

944   Similarly, a simple example shows why acceptance or rejection of sexual dimorphism

945   is not dependent upon any single individual's sexually variable trait providing an unambiguous

946   indicator of its sex. If one is told that a person is 178 cm tall, this is insufficient to identify their

947   sex, despite the fact that sexual height dimorphism exists in humans. The ability or inability to

948   assign sex to an individual based solely on a sexually variable trait is an indication of whether

949   or not sexual dimorphism is of an extreme effect size, not whether or not dimorphism is present.

950   *Estimate and Constrain Magnitude of Dimorphism.*— After assigning sex for the data

951   points, a growth curve should be fitted to each sex, allowing for a specific estimate of

952   dimorphism magnitude, together with confidence and prediction intervals. Accepting

953   uncertainty of an estimate is a key tenet of responsible statistical analysis (Wasserstein et al.

954   2019). Confidence intervals attempt to constrain the range in which the theoretical population

955   parameters are expected to lie at a given probability. In this case, 95% confidence intervals

956   constrain our estimate of the fitted sex-specific curves. Specifically, "the 95 % refers only to

957   how often 95 % confidence intervals computed from very many studies would contain the true

958   size *if all the assumptions used to compute the intervals were correct*" (Greenland et al. 2016,

959   p. 343, emphasis theirs). Therefore, the accuracy of our estimate and success at constraining

960   that estimate improves as sex assignment accuracy increases and assumptions are more closely

961   met. Prediction intervals attempt to constrain the range in which a datapoint would be expected

962   to lie at a given probability, accounting for the spread of the data. In this case, 95% prediction

963   intervals attempt to predict the range in which the addition of an individual male or female to

964   the dataset might lie (again dependent upon prior assumptions being met, namely sex

965   assignment accuracy). For our purposes, prediction intervals can provide insight into the

966   estimated degree of overlap between males and females, making them relevant to judging

967   support for dimorphism. Note that statisticians call for careful and nuanced interpretations of

968   such intervals in a similar manner to the caution advised for interpreting *p*-values (Amrhein et

969   al. 2019).

970   We define effect size here as the difference between the male and female *L* parameters

971   of the growth formulae. The accuracy of our estimated sex-specific regressions to the true

972   growth curves (i.e., the accuracy of the estimate for dimorphism magnitude) is expected to be

973   biased in the case of a truly monomorphic dataset because the illustrative method of sex

974   assignment will predict some minor amount of dimorphism. However, even with relatively

975   small true effect sizes, the fitted curves can closely match the true growth curves (Figure 9).

976   Above a certain true effect size used to generate the data, the ratio of true effect size to

977   estimated effect size centers around a value of one. Furthermore, the naturalistic effect size

978   observed in alligators and rheas (i.e., true effect size used in simulation equals the empirical

979   effect size observed in nature) yield model accuracies that approach or are centered near one,

980   with alligator performing better. Smaller sample sizes can lead to greater variability in this

981   measure of model accuracy. As effect size increases, rhea simulations lag behind alligator

982   simulations in their improved model fitting accuracy due to higher intrasexual variation relative

983   to effect size in rhea.

984 If one suspects that sexual traits showing high amounts of intrasexual variability and

985 greater overlap of males and females might lead to overestimates of sexual variation, then the

986 trait suspected to be sexually selected can be compared to a trait in the same sample/species

987 that is less likely to be sexually selected (as in O'Brien et al. 2018). The difference in estimated

988 sexual variation between the putatively sexual and non-sexual trait can then be compared to

989 differences observed in other taxa/datasets. For example, an estimate of sexual variation in

990 horn length relative to an estimate of sexual variation in vertebral or tooth shape could be

991 compared across a horned clade. Again, mixture modelling might have promise in estimating

992 intrasexual variation at a given ontogenetic stage (e.g., adults), since mixture models attempt

993 to estimate standard deviations of the underlying male and female distributions.

994 Here, we examine the separation of the upper bound of the smaller sex from the lower

995 bound of the larger sex and the spread between the lower bound of the smaller sex and the

996 upper bound of the larger sex at the oldest individuals of each assigned sex (Figures 10–11).

997 Note that estimation of the oldest/largest age/size each sex can attain based on the oldest/largest

998 individual of each assigned sex in a given dataset can come with high uncertainty. One may

999 instead wish to measure interval spread and separation at a given age/size class. Our goal here

1000 is simply to be illustrative and internally consistent, but we could have alternatively measured

1001 the intervals at the single oldest age attained by both assigned sexes to better represent this

1002 uncertainty. The measure we use here might instead be more appropriate if using body size as

1003 a proxy for age, since the two sexes might not attain the same maximum body size.

1004 Interval spread and separation increase with increasing true effect size. Confidence and

1005 prediction intervals widen with decreasing sample size, so interval spread increases and

1006 interval separation decreases. Interval separation is important in judging uncertainty in the

1007 estimate for growth curve divergence. Confidence interval separation is more often expected

1008 to be positive when the effect size is truly zero (i.e., monomorphic) than prediction interval

1009   separation because prediction intervals are wider than confidence intervals for a given

1010   regression. Non-zero true effect sizes might still fail to result in interval separation. Rhea

1011   simulations show greater uncertainty in growth curve divergence than alligator simulations due

1012   to relatively higher intrasexual variation with respect to to effect size.

1013   The rhea simulations show many extreme outliers in interval spread/separation due to

1014   difficulty in fitting Gompertz curves to the data. These outliers are datasets in which, after sex

1015   assignment, one or both assigned sexes lack juveniles that allow for easier curve fitting. Even

1016   with large sample sizes , if complex growth models are to be applied to the data, it is important

1017   to include specimens over a range of ages/age proxies. Given the lack of juveniles and extreme

1018   interval spread and separation values (sometimes differing by orders of magnitude), these

1019   outlier datasets are easy to identify and ultimately could be dropped from any comparative

1020   analysis between datasets.

1021   It might be tempting to use interval separation versus overlap as a form of significance

1022   testing. However, we urge researches to avoid the potential pitfall of dichotomous acceptance

1023   versus rejection of the presence of an effect (i.e., black-or-white thinking) (Nakagawa & Cuthill

1024   2007; Amrhein et al. 2017; Amrhein & Greenland 2018; Halsey 2019; Holland 2019). One

1025   simply attempts to estimate sexual variation and constrain the uncertainty in that estimate

1026   without overinterpretation, rather than categorizing a species as either dimorphic or

1027   monomorphic or stating that evidence is either present or absent. Despite the bias towards

1028   overestimation when true effect size is low relative to intrasexual variation, our framework

1029   does not assume dimorphism is necessarily present or strong when intervals overlap. When

1030   constraining estimates with indicators of uncertainty, such as confidence intervals, it is

1031   important to consider the following:

1032   "Accept uncertainty and embrace variation in effects: we can learn much (indeed, more)

1033   about the world by forsaking the false promise of certainty offered by dichotomous

1034   declarations of truth or falsity—binary statements about there being "an effect" or "no

1035   effect"—based on some p-value or other statistical threshold being attained" (McShane

1036   et al. in Wasserstein et al. 2019, p. 15).

1037   When comparing estimated magnitudes of sexual variation between datasets/taxa, we

1038   recommend standardizing estimates of effect size and interval spread/separation (e.g., as a

1039   percentage of the maximum size predicted by a regression fit to the smaller sex). Standardized

1040   estimates of effect size and their confidence/prediction intervals provide a way to compare

1041   different datasets/species in a meta-analytic or phylogenetic framework. This approach takes

1042   into account differences in sample sizes and can be used to look for heterogeneity (e.g., do

1043   some populations differ in dimorphism magnitude as a function of ecology).

1044   *Phylogenetic simulation with more realistic data variability.*— The above simulations

1045   (and the significance testing simulations below) held sex ratio and the function (according to

1046   size) for standard deviation of the residuals constant and at parity between the sexes in order

1047   to show the dynamic between true effect size, sample size, and our estimates. What happens if

1048   parameters/conditions are allowed to vary randomly such that they more closely represent

1049   empirical fossil or modern data? Inspired by natural alligator data, we generated a hypothetical

1050   crocodilian clade containing a grade of 100 'monomorphic' taxa (i.e., low sexual variation)

1051   that evolve a clade of 50 'dimorphic' taxa (i.e., a shift to high sexual variation). Each taxon

1052   was allowed to vary independently in true effect size (i.e., proportionally-applied changes to

1053   male growth parameters, as in the other simulations), sample size, true sex ratio, male standard

1054   deviation of the residuals function, and female standard deviation of the residuals function.

1055   Each taxon was then treated as if it were fossil, whereby the sex of each individual was

1056   unknown and assigned using the simple residual-based method. Effect size was estimated, and

1057   interval spread/separation were calculated at the largest individual of each sex.

1058   The results show that evolutionary shifts in sexual variation of sufficient magnitude can

1059   be detected in a phylogeny (Figure 12A). Intervals are more likely to overlap when true effect

1060   size is low. Although bias at low true effect size is present (to a similar extent to the above

1061   simulations with stricter conditions), in general, estimated effect size tracks true effect size

1062   (Figure 12B). Altering factors such as sex ratio and intrasexual variation contributes noise to

1063   the data away from perfect estimation, but the correlation between true and estimated effect

1064   size is still largely present on average.

1065   *Discussion of Our Quantitative Approach.*— Hone and Mallon (2017) showed that

1066   when similar simulations are done using the empirically derived parameters of naturally

1067   occurring alligator and rhea populations, *t*-tests are not always effective at detecting

1068   dimorphism, even under the unlikely assumption that the sex of all individuals is known. When

1069   the sex of simulated alligator and rhea individuals are assigned using the residual-based

1070   approach (while holding sex ratio and intrasexual variation at parity, as in our earlier

1071   simulations), *p*-values derived from Hartigans' dip tests and *t*-tests can increase with

1072   decreasing sample size and are likely to be non-significant ($\alpha = 0.05$) (Figures 13–14).

1073   Significant *p*-values often require large sample sizes and, more importantly, very large effect

1074   sizes. Rhea simulations are more likely to result in lower *p*-values than alligator simulations

1075   (as in Hone & Mallon 2017), due to the fact that most of the simulated datapoints are fully-

1076   grown adults at maximal dimorphism. Elevated intrasexual variation and skewed sex ratios are

1077   not expected to improve significance testing results when sex is not independently known

1078   (appendix).

1079   Instead, we emphasize effect size statistics rather than significance testing (Nakagawa

1080   & Cuthill 2007; Amrhein et al. 2017; Amrhein & Greenland 2018; Halsey 2019; Holland

1081   2019). While univariate significance tests suffer from low statistical power and can readily give

1082   false negatives (Mallon 2017; Hone & Mallon 2017), our framework is more useful when

1083   dealing with small sample sizes, since reductions in sample size lead to reductions in statistical

1084   power. When sample size decreases, variability in the accuracy of sex assignment and

1085   variability in the accuracy of estimated effect size will increase across repeated samplings,

1086   while confidence intervals of the estimated effect size increase. The intervals (i.e., uncertainty

1087   in the effect size estimate) are also dependent upon the spread of the data (i.e., intrasexual

1088   variation), with larger spreads leading to larger intervals. The use of bivariate data to estimate

1089   sex-specific growth curves means that the inclusion of juveniles does not detract from our

1090   ability to detect dimorphism. Instead, juveniles enhance detection by making regression

1091   analysis computationally easier.

1092   Our statistical method is often expected to reliably quantify sexual variation and, when

1093   data are collected on traits likely to be sexually selected and compared to traits in the same

1094   sample/species unlikely to be sexual, help rule out non-sexual variation/polymorphisms in

1095   populations. Given the prevalence of dimorphism among living animals, paleontologists

1096   should identify species with large estimates for sexual variation (i.e., effect sizes) and high

1097   estimated support for dimorphic growth models compared to other phylogenetically proximate

1098   species, rather than using significance testing on individual datasets to dichotomously reject or

1099   fail to reject monomorphism. Given the evidence for sexual selection in non-avian dinosaurs

1100   and their extant phylogenetic bracket, the question is not whether a given species was

1101   dimorphic, but rather which species show the strongest evidence of dimorphism relative to

1102   others, especially within a clade. Examining sexual selection across a phylogeny (and if fossil

1103   data is used to calibrate node ages, across time) using a measure of sexual dimorphism is

1104   precedented (e.g., plumage dichromatism as a proxy for sexual selection in extant birds

1105   [Cooney et al. 2019]). Estimating effect size allows for phylogenetic comparative methods to

1106   examine evolutionary rates and patterns, as well as ancestral state reconstruction, with respect

1107   to sexual selection so long as evolutionary trends are strong enough to overcome

1108   overestimation bias from small effect sizes.

1109   Some clarifications should be made about our method's inclination (i.e., bias), as

1110   currently presented, to overestimate effect size in datasets with minimal sexual variation and

1111   large overlap between the sexes.

1112   1) Finite sample size statistics, evolutionary/game theory, and fossil evidence of

1113   variation in structures consistent with sexual variation in modern taxa suggests that sexual

1114   variation in anisogamic species is the norm rather than the exception. Since minimally sexually

1115   variable species (at least with respect to certain traits) are not necessarily predicted to be

1116   common, one must approach statistical analyses accordingly:

1117   "Thoughtful research looks ahead to prospective outcomes in the context of theory and

1118   previous research. Researchers would do well to ask, *What do we already know, and*

1119   *how certain are we in what we know?* And building on that and on the field's

1120   theory, *what magnitudes of differences, odds ratios, or other effect sizes are practically*

1121   *important?"* (Wasserstein et al. 2019, p. 4, emphasis theirs).

1122   2) Species with minimal sexual variation are those more likely to have overlapping

1123   confidence intervals. Therefore, some uncertainty is accounted for, according to the degree that

1124   assumptions are met based on sex assignment accuracy.

1125   3) We encourage comparisons of effect size and uncertainty between datasets/taxa,

1126   especially within a phylogenic context. When differences in effect size are high between

1127   datasets/taxa and evolutionary trends are strong, meaningful differences/trends can be detected,

1128   even if species with minimal sexual variation are overestimated.

1129   A dataset with effect size = 0 will result in overestimation under this approach to sex

1130   assignment with the size of the bias dependent on the intrasexual variation. As true effect size

1131   increases, bias is reduced; as overlap between the largest adults of each sex is reduced,

1132   estimated effect size will begin to track true effect size closely (supplemental videos). These

1133   later datasets can be distinguished from those with low effect size affected by bias. Unlike

1134   univariate significance tests that struggle with juvenile data and small effect sizes, our

1135   framework's performance improves as the largest adults become more distinguishable between

1136   sexes. Comparing estimates between datasets is crucial, especially within a clade that might

1137   show similar intrasexual variation across species. If traits with high estimated effect size are

1138   suspected to be overestimated, one can, in addition to examining the calculated uncertainty,

1139   compare plausibly sexual versus non-sexual traits to develop a baseline level of non-sexual

1140   variation in species. Mixture modelling while accounting for growth curves also has promise

1141   in estimating intrasexual variation.

1142   4) Overestimation bias in these instances results from inaccurate sex assignment. We

1143   presented the simplest method of sex assignment as an illustrative example. Future work can

1144   improve statistical sex assignment methods (e.g., incorporating mixture modelling). When

1145   fossil evidence allows for confident sex assignment, all data points with that diagnosable

1146   feature can be assigned to one sex, increasing sex assignment accuracy.

1147   When a distribution with a mean of zero is examined as a magnitude, the mean of

1148   absolute values will be upwardly biased (Hansen 2016; Morrissey 2016). This bias in statistics

1149   of magnitude may not be fully relevant for sexual variation between two distributions, but

1150   researchers have nevertheless proposed corrective methods (Morrissey 2016). We do see this

1151   bias with sex role reversal (i.e., negative effect sizes are overestimated as positive), since our

1152   method simply attempts to distinguish two sexes, not males versus females, unless

1153   unambiguous markers of sex are sufficiently present in the sample.

1154

1155   Alternate Hypothesis 2: Controlling for Ontogenetic Variation

1156   Observed variation can result, at least partly, from differences between juveniles and adults,

1157   with the most extreme examples being species with larval stages and metamorphosis. In a

1158   reanalysis of the morphometric data used by Dodson (1976) to propose sexual dimorphism in

1159   *Protoceratops*, Chapman et al. (2008) found the greatest differences in both principal

1160   coordinate and principal component morphospace are between juveniles and large, adult

1161   specimens. Purported dimorphism among the adult individuals is of much smaller magnitude

1162   and altogether dependent on how missing measurements are treated in the analysis (Chapman

1163   et al. 2008). Even if most variation present in a species is ontogenetic, sexual variation can still

1164   co-occur and even be extreme relative to sexual variation of other species. Complex organisms

1165   would be expected to show greater ontogenetic variation than sexual variation, especially with

1166   respect to body size variation in multicellular animals. Even accepting dimorphism between

1167   adults, juveniles can still form a separate data cluster in a morphospace, creating an impression

1168   that there are three morphs. Therefore, before a recovered signal can be confidently attributed

1169   to sexual selection, potential ontogenetic causes need to be accounted for.

1170   By focusing the quantitative aspects of our framework on detecting divergence between

1171   growth curves, the alternate hypothesis that the observed variation is primarily ontogenetic is

1172   already largely accounted for. Because growth curves for fossils are often generated from

1173   growth markers in histological cross sections of bones, these provide a means for establishing

1174   corresponding degrees of maturity between putative sexual morphs. There may be considerable

1175   individual variation of growth curves in a population (Sander & Klein 2005; Woodward et al.

1176   2015); such variation among growth curves of a single population is taken into account under

1177   the intrasexual variation modeled here (i.e., population standard deviations of the residuals).

1178   However, further observations can be made to more thoroughly test an ontogenetic

1179   alternate hypothesis. Some evidence, such as external fundamental systems (EFS) in bone

1180   marking the cessation of somatic growth, can be useful to demonstrate that a smaller or less

1181   developed morph is not growing into the other morph, especially when size is used as a proxy

1182   for age in the quantitative analysis (Saitta 2015).

1183   Other skeletal indicators that may be useful in establishing age or maturity (Hone et al.

1184   2016b) include surface bone texture (Brown et al. 2009), markers of reproductive maturity like

1185   the presence of medullary bone (Lee & Werning 2008), tooth eruption and wear in mammals

1186   (Maschenko 2002; Lister 2009), or fusion of bone sutures (Brochu 1996), although none of

1187   these indicators is fully reliable across a wide range of taxa (Tumarkin-Deretzian et al. 2006;

1188   Cerda et al. 2014, Irmis 2007). Therefore, careful consideration of which ontogenetic indicators

1189   are appropriate depends on the clade of interest.

1190   Some methods to detect sexual maturity (or sex) may not work with fossils; for

1191   example, immunochemical techniques that report binding to specific endogenous, intact

1192   molecular organic markers of reproductive medullary bone (Schweitzer et al. 2016) may be

1193   false positives in that they appear to bind to organic material unlikely to survive fossilization

1194   (Saitta et al. 2018; Saitta & Vinther 2019). Other methods of ontogenetic analysis, such as the

1195   parsimony-based, size-independent method of ontogenetic sequence analysis (OSA) that

1196   accounts for developmental sequence polymorphism, have also been proposed (Colbert 1999;

1197   Colbert & Rowe 2008; Griffin & Nesbitt 2016), but should be further examined for

1198   compatibility with our framework. For quantitative aspects of our approach, continuous

1199   age/age proxy data work best for regression analysis. Discrete age proxies are still useful in

1200   identifying specimens of each putative morph with ceased growth, providing further evidence

1201   against the alternate hypothesis of ontogenetic variation.

1202

1203   Alternate Hypothesis 3: Controlling for Interspecific Variation

1204   Paleontologists mostly use diagnostic morphological features to work out taxonomic

1205   relationships and delineate species, but traits that can greatly affect an organism's ecology

1206   should be considered regardless of whether or not they have been used historically as

1207   taxonomic indicators. To rule out interspecific variation, putative sexual morphs must coexist

1208   temporally and geographically without evidence for niche partitioning in other parts of their

1209   anatomy, indicating that they are not two species under the influence of the competitive

1210   exclusion principle (Hardin 1960). An example is a multi-individual bonebed in which

1211   dimorphism is exhibited in the putative secondary sexual characteristic or body size, but not

1212   readily apparent in other traits associated with diet or locomotion (Saitta 2015), further urging

1213   comparison of estimated sexual variation between plausibly sexual and non-sexual traits in a

1214   population. A corollary is that potential secondary sexual characteristics, especially those with

1215   high variability, should be carefully considered before their use as diagnostic taxonomic

1216   characters. Extreme body size variation might arise from interspecific niche partitioning (e.g.,

1217   as has traditionally been quantified as 'Hutchinson's rule/ratios', but challenged as artefactual

1218   [Eadie et al. 1987]).

1219   Chronological constraint is important given the morphological changes that can occur

1220   in closely related specimens through a stratigraphic section. Stratigraphic discrepancies

1221   between supposed dimorphs have been used to reject hypothesized sexual dimorphism in

1222   *Lambeosaurus* (Evans & Reisz 2007), *Chasmosaurus* (Mallon & Holmes 2006), and

1223   *Centrosaurus*/*Styracosaurus* (Ryan et al. 2007). Note that it is difficult to ascribe

1224   morphological changes through time to anagenesis when changing biogeographic ranges and

1225   migrations into and out of depositional basins cannot be ruled out, (e.g., the suggestion that

1226   short-nasal-horned *Triceratops horridus* of the lower Hell Creek Formation evolved into long-

1227   nasal-horned *Triceratops prorsus* of the upper Hell Creek Formation [Scannella et al. 2014]).

1228   Given the form-function link in biology, sexual dimorphism can be linked to sex

1229   differences in behavior or ecology arising from dimorphic morphology or size (Pearson et al.

1230   2002; Radford & Du Plessis 2003), meaning that evidence for shared niches among the sexes

1231   might vary between taxa. Researchers should still strive to show that putative dimorphs with

1232   similar taxonomic and ecological diagnostic features co-existed in time and space.

1233

1234   Alternate Hypothesis 4: Controlling for Intra-Individual or Intra-Structural Variation

1235   Certain structures that occur multiple times in a single individual can exhibit extensive intra-

1236   individual variation (e.g., osteoderms [Saitta 2015] or feathers [Lucas & Stettenheim 1972]).

1237   Structures can show multiple occurrences within an individual sequentially over time as well

1238   as in multiple positions along the body, such as molting feathers or annually shed antlers. Such

1239   variation can be controlled for in quantitative analysis by directly comparing equivalent,

1240   topologically homologous structures (e.g., only measuring the main shoulder spike of nodosaur

1241   specimens) or by comparing whole-individual morphometric profiles of repeating structures

1242   (e.g., measuring spine length along the lateral-most row of spines over the entire body length

1243   of nodosaur specimens). In incomplete/disarticulated fossils, it might be required to

1244   demonstrate that intra-individual variation in the structures is limited (i.e., lesser than inter-

1245   individual variation) or an appropriate approximation for ontogenetic variation (e.g., smaller

1246   and larger structures within an individual vary in a manner that mirrors variation between the

1247   structures on smaller/younger and larger/older individuals).

1248   Although disorders of sexual developmental can result in extreme intra-individual

1249   variation (e.g., bilateral gynandromorphism [Lillie 1931]), the occurrence of intersex

1250   individuals can be very rare in species selected for discrete male and female individuals

1251   (Stockholm et al. 2006; Bojesen et al. 2003). Unless dealing with a clade that exhibits high

1252   prevalence of simultaneous or sequential (i.e., sex change) hermaphroditism (Warner 1975;

1253   Heath 1977), intra-individual sex-related variation due to developmental disorders or selective

1254   pressures for hermaphroditism can likely be assumed to be rare in the fossil record. Fossil

1255   specimens suspected to show dramatic developmental disorders (e.g., Buffetaut et al. 2006) or

1256   unexpected instances of hermaphroditism should be tested for authenticity (Rowe *et al.* 2001)

1257   before inclusion in a dataset.

1258   Complex structures might show intra-structural variation that should be accounted for

1259   (e.g., barb morphology along the rachis in a feather) by comparing equivalent/homologous

1260   portions of the structure (e.g., the very apical end of a feather), equivalent metrics between

1261   different specimens (e.g., a measure of maximum fractal branching order of the feather), or

1262   whole-structure morphometric profiles (e.g., a profile of barb angle from base to apex along

1263   the entire rachis of a feather).

1264

1265   Alternate Hypothesis 5: Controlling for Pathological Variation

1266   Pathologies can add variation to a fossil dataset, but they can usually be identified by atypical

1267   frequencies (i.e., unlike ~1:1 male:female ratio in large, random samples) and distinctive

1268   diagnostic anatomical features (Moodie 1918, 1923; Tanke & Rothschild 2002). Although

1269   some secondary sexual characteristics can show asymmetry, one should consider whether

1270   extreme asymmetry might indicate pathology. Pathological specimens can be retro-deformed

1271   or dropped from a dataset.

1272

1273   Alternate Hypothesis 6: Controlling for Taphonomic Variation

1274   Other sources of variation might be taphonomic (e.g., burial, transport, scavenging, decay,

1275   weathering, or erosion). Weathering of organically preserved soft tissues (Vinther 2015) or

1276   plastic deformation of bones (Boyd & Motani 2008), for example, could introduce noise to a

1277   dataset. Hedrick and Dodson (2013) suggested that taphonomy alone might explain differences

1278   between purported species or morphotypes of *Psittacosaurus lujiatunesis*. Taphonomic

1279   consideration might include examining fossils for deformation, incompleteness, or breakage.

1280   Scanning electron microscopy on carbonaceous compression fossils can identify mouldic

1281 melanosome impressions in sediment within areas where dark organic stains of melanin were

1282 lost through oxidative weathering (Vinther 2015). One can reduce taphonomic noise in

1283 quantitative analyses by eliminating highly altered specimens from the dataset, rerunning

1284 analyses with only the best-preserved specimens, or attempting to retro-deform or reconstruct

1285 specimens (Boyd & Motani 2008; Saitta 2015).

1286

1287 Biological Implications

1288 Once alternate hypotheses for observed variation are weakened, then the type and magnitude

1289 of estimated sexual variation can be used to draw biological inferences. The degree of

1290 dimorphism might indicate the extent to which males and females exhibited similar behaviors

1291 (e.g., mating behavior) and, in extreme cases, ecologies. Dimorphism magnitude might provide

1292 insight into the social/sexual system (e.g., polygamy versus monogamy, social group size, or

1293 sexual segregation). Monogamous species often show less body size dimorphism than

1294 polygynous species (Lack 1968; Gautier-Hion 1975; Clutton-Brock et al. 1977; Clutton-Brock

1295 & Harvey 1978; Alexander et al. 1979; Shine 1979; Clutton-Brock 1985; Heske & Ostfeld

1296 1990). Bird species with polygynous or lek mating systems have greater dimorphism in body

1297 mass, plumage, wing length, and tail length compared to monogamous species (Dunn et al.

1298 2001). Mating system inferences from estimated magnitudes of sexual dimorphism have been

1299 made using fossils – with *Canis dirus* hypothesized as monogamous, similar to most extant

1300 canids, and *Smilodon fatalis* hypothesized as more monogamous compared to modern lions

1301 (Van Valkenburgh & Sacco 2002). Sexual segregation is seen in dimorphic ruminants (Bowyer

1302 2004). Increased group size in bovids is correlated with increased body size dimorphism and

1303 male horn length (Bro-Jørgensen 2007). In cervids, male antler size, polygyny, and breeding

1304 group size are correlated (Clutton-Brock et al. 1980). Fossil evidence might reveal whether a

1305   species was sex-role reversed (e.g., unequivocal markers of female sex, such as *in situ* eggs, in

1306   individuals with more exaggerated secondary sexual characteristics).

1307   Simplistic inference into the relative importance of intrasexual competition versus

1308   intersexual mate choice can be hypothesized depending on whether the dimorphic structure is

1309   consistent with an armament or ornament, respectively (Sullivan et al. 2003; Saitta 2015). For

1310   example, some stegosaurs and ceratopsians possess structures that consistent with ornaments

1311   (e.g., plates and frills) alongside structures possibly consistent with armaments (e.g., spikes

1312   and horns). Could this indicate the presence of both intrasexual combat to compete for mates

1313   in addition to, or followed by, intersexual display and mate choice (Molnar 1977; Farke et al.

1314   2009; Krauss et al. 2010)?

1315   Dimorphism magnitude and the nature of the anatomical structure might help identify

1316   primary versus secondary functions of traits, such as females using traits sexually selected for

1317   in males as predator deterrents (e.g., possibly stegosaur osteoderms [Saitta 2015] or ceratopsian

1318   horns). Even when the estimated magnitude of and support for dimorphism are great,

1319   multiple/secondary functions for structures should be considered (Farke 2014), particularly

1320   when the sex with the less exaggerated trait still possesses the trait (e.g., uniquely among

1321   cervids, female reindeer [*Rangifer tarandus*] possess antlers, presumably for scraping away

1322   snow during winter feeding).

1323   The fossil record can illuminate sexual selection's effect on macroevolution, and certain

1324   evolutionary trends might in turn provide evidence that a trait is sexually selected and likely

1325   dimorphic. Although there is mixed evidence that sexual selection can increase speciation rates,

1326   for example through the formation of reproductive barriers (Panhuis et al. 2001; Kraaijeveld et

1327   al. 2011), our effect size approach may allow fossils to supplement extant data. When

1328   reconstructing evolutionary histories of sexual dimorphism, habitat, and social behavior

1329 (Pérez-Barbería et al. 2002), fossil data could supplement phylogenetic and ecological analyses

1330 (Pringle 2020).

1331

1332 **Example: Empirical Non-avian Dinosaur Datasets**

1333 Figure 15 shows the results of our divergence analysis on data from *Maiasaura*

1334 *peeblesorum* (Woodward et al. 2015), *Psittacosaurus lujiatunensis* (Erickson et al. 2015), and

1335 *Tyrannosaurus rex* (Horner & Padian 2004; Erickson et al. 2004; Lee & Werning 2008). One

1336 should not categorically conclude that any one of these taxa does or does not exhibit sexual

1337 dimorphism. Instead, it is better to conclude simply that given the current data and assumptions

1338 (e.g., logistic growth, independence of datapoints, and reliable estimation/assignment of body

1339 mass, age, and sex), *Maiasaura* exhibits the largest standardized estimate of sexual variation

1340 (i.e., percent change in asymptotic size from smaller to larger assigned sex) and the lowest

1341 uncertainty in that estimate (i.e., tightest confidence intervals), a pattern otherwise unapparent

1342 in univariate histograms. *Maiasaura* data also derive from a single bonebed (unlike the other

1343 two datasets), providing tight stratigraphic and geographic control, and perform similarly well

1344 under Gompertz regression (appendix). Furthermore, both assigned sexes contain at least one

1345 individual with an EFS, indicating that these individuals have ceased growth. *Psittacosaurus*

1346 data suffer from a lack of older individuals, while *Tyrannosaurus* data suffer from small sample

1347 size. Therefore, *Maiasaura* currently shows better evidence for sexual variation than both

1348 *Psittacosaurus* and *Tyrannosaurus*.

1349

1350 **Conclusions**

1351 Sexual variation is prevalent in animals generally and archosaurs specifically and is expected

1352 in extinct archosaurs, such as non-avian dinosaurs. Hypotheses to explain a supposedly unique

1353 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 centers 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 analyze growth curve divergence of plausible 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 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 paleontologists, 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 approaches parity as sample size increases. Future work should improve

statistical methods of sex assignment and paleontological 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 versus 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

1403   dataset and onset of putative sexual maturity/pubescence might be identified using multiple

1404   regression.

1405   We recommend attempting to constrain sexual dimorphism magnitude using

1406   confidence interval spread/separation. Rather than binary rejection or non-rejection of sexual

1407   monomorphism on a species-by-species basis in a manner prone to false negatives, effect size

1408   statistics enables investigation of sexual selection across phylogenies and geologic time.

1409   The ideas that non-avian dinosaurs do not show evidence of sexual dimorphism or that

1410   independent knowledge of the sex of all specimens might be required to investigate the problem

1411   are needlessly stringent in many cases, and non-significant p-values cannot be interpreted as

1412   'no evidence' for an effect. Returning to a previous quote, Mallon (2017) stated that "…in the

1413   absence of a priori knowledge of sex, the ability to detect dimorphism in a fossil sample is

1414   likely only in cases of strongly expressed dimorphism (i.e., well-separated peaks on a

1415   histogram or discrete character states)…" (p. 502). Our simulations show that, in many cases,

1416   reasonably accurate estimates of sexual variation can be obtained even if datasets do not show

1417   bimodality and the sex of all individuals is unknown. We take an opposing viewpoint of Mallon

1418   (2017) when he states, "Such a conservative approach [the bimodality method] is subject to

1419   type II error, but this is preferable to attributing sexual dimorphism to every perceived instance

1420   of intraspecific variation in the fossil record, particularly if there is no compelling reason to

1421   think it should exist in the first place" (p. 501). Rather, sexual selection is too influential a

1422   process to be disregarded (as a result of type II error) as one of the major sources of variation

1423   in paleobiology.

1424

1442

Literature Cited

1444   Abella, J., Valenciano, A., Pérez-Ramos, A., Montoya, P., and Morales, J., 2013. On the socio-

1445   sexual behaviour of the extinct ursid *Indarctos arctoides*: an approach based on its

1446   baculum size and morphology. PLoS One *8*(9):e73711.

1447   Alexander, R. D., Hoogland, J. L., Howard, R. D., Noonan, M., and Sherman, P. W. 1979.

1448   Sexual dimorphisms, and breeding systems in pinnipeds, ungulates, primates and

1449   humans, in: *Evolutionary Biology and Human Social Behavior: An Anthropological*

1450   *Perspective* (N. A. Chagnon and W. Irons, eds.), pp. 402–604, Duxbury Press, North

1451   Scituate, Massachusetts.

1452   Alexander, R.M., 1989. Dynamics of dinosaurs and other extinct giants. Columbia University

1453   Press.

1454   Alexander, R.M., 2006. Dinosaur biomechanics. Proceedings of the Royal Society B:

1455   Biological Sciences 273:1849-1855.

1456   Allentoft, M.E., Bunce, M., Scofield, R.P., Hale, M.L., and Holdaway, R.N., 2010. Highly

1457   skewed sex ratios and biased fossil deposition of moa: ancient DNA provides new

1458   insight on New Zealand's extinct megafauna. Quaternary Science Reviews 29:753-762.

1459   Amrhein, V. and Greenland, S., 2018. Remove, rather than redefine, statistical

1460   significance. Nature Human Behaviour 2:4.

1461   Amrhein, V., Korner-Nievergelt, F., and Roth, T., 2017. The earth is flat (p> 0.05): significance

1462   thresholds and the crisis of unreplicable research. PeerJ 5:e3544.

1463   Amrhein, V., Greenland, S., and McShane, B., 2019. Scientists rise up against statistical

1464   significance. Nature 567:305-307.

1465   Andersson, M.B., 1994. Sexual selection. Princeton University Press.

1466   Andersson, M. and Simmons, L.W., 2006. Sexual selection and mate choice. Trends in

1467   Ecology & Evolution 21:296-302.

1468   Arbour, V.M. and Currie, P.J., 2011. Tail and pelvis pathologies of ankylosaurian

1469   dinosaurs. Historical Biology 23:375-390.

1470   Arbour, V.M. and Snively, E., 2009. Finite element analyses of ankylosaurid dinosaur tail club

1471   impacts. The Anatomical Record: Advances in Integrative Anatomy and Evolutionary

1472   Biology: Advances in Integrative Anatomy and Evolutionary Biology 292:1412-1426.

1473   Arbour, V.M. and Zanno, L.E., 2018. The evolution of tail weaponization in

1474   amniotes. Proceedings of the Royal Society B: Biological Sciences 285:20172299.

1475   Autier-Dérian, D., Deputte, B.L., Chalvet-Monfray, K., Coulon, M., and Mounier, L., 2013.

1476   Visual discrimination of species in dogs (*Canis familiaris*). Animal Cognition 16:637-

1477   651.

1478   Balmford, A., Jones, I.L., and Thomas, A.L., 1993. On avian asymmetry: evidence of natural

1479   selection for symmetrical tails and wings in birds. Proceedings of the Royal Society of

1480   London. Series B: Biological Sciences 252:245-251.

1481   Barden, H.E. and Maidment, S.C., 2011. Evidence for sexual dimorphism in the stegosaurian

1482   dinosaur Kentrosaurus aethiopicus from the Upper Jurassic of Tanzania. Journal of

1483   Vertebrate Paleontology 31:641-651.

1484   Barlow, G.W., 2005. How do we decide that a species is sex-role reversed? The Quarterly

1485   review of biology 80:28-35.

1486   Barrett, P.M., McGowan, A.J., and Page, V., 2009. Dinosaur diversity and the rock

1487   record. Proceedings of the Royal Society of London B: Biological Sciences 276:2667-

1488   2674.

1489   Barrowclough, G.F., Cracraft, J., Klicka, J., and Zink, R.M., 2016. How many kinds of birds

1490   are there and why does it matter? PLoS One 11:e0166307.

1491   Barsbold, R., Osmólska, H., Watabe, M., Currie, P.J., and Tsogtbaatar, K., 2000. A new

1492   oviraptorosaur [Dinosauria, Theropoda] from Mongolia: the first dinosaur with a

1493   pygostyle. Acta Palaeontologica Polonica 45:97-106.

1494   Bell, P.R., Fanti, F., Currie, P.J. and Arbour, V.M., 2014. A mummified duck-billed dinosaur

1495   with a soft-tissue cock's comb. Current Biology 24:70-75.

1496   Birchard, G.F., Ruta, M., and Deeming, D.C., 2013. Evolution of parental incubation behaviour

1497   in dinosaurs cannot be inferred from clutch mass in birds. Biology Letters 9:20130036.

1498   Birkhead, T.R. and Møller, A.P. eds., 1998. *Sperm competition and sexual selection*. Elsevier.

1499 Blackburn, G., 2004. Ornaments and overflight behaviour of tufted puffins (*Fratercula*
1500 *cirrhata*) breeding on Triangle Island, British Columbia (Doctoral dissertation,
1501 Science: Biological Sciences Department).

1502 Bojesen, A., Juul, S. and Gravholt, C.H., 2003. Prenatal and postnatal prevalence of Klinefelter
1503 syndrome: a national registry study. The Journal of Clinical Endocrinology &
1504 Metabolism 88:622-626.

1505 Bonduriansky, R., 2007. Sexual selection and allometry: a critical reappraisal of the evidence
1506 and ideas. *Evolution:* International Journal of Organic Evolution 61:838-849.

1507 Bonhomme, V., Picq, S., Gaucherel, C., and J. Claude. 2014. Momocs: Outline analysis using
1508 R. Journal of Statistical Software 56:1–24.

1509 Borgia, G. and Gore, M.A., 1986. Feather stealing in the satin bowerbird (*Ptilonorhynchus*
1510 *violaceus*): male competition and the quality of display. Animal Behaviour 34:727-738.

1511 Borkovic, B. and Russell, A., 2014. Sexual selection according to Darwin: A response to
1512 Padian and Horner's interpretation. Comptes Rendus Palevol 13:701-707.

1513 Bowyer, R.T., 2004. Sexual segregation in ruminants: definitions, hypotheses, and implications
1514 for conservation and management. Journal of Mammalogy 85:1039-1052.

1515 Boyd, A.A. and Motani, R., 2008. Three-dimensional re-evaluation of the deformation removal
1516 technique based on ''jigsaw puzzling''. Palaeontologia Electronica 11:7A.

1517 Brochu, C.A., 1996. Closure of neurocentral sutures during crocodilian ontogeny: implications
1518 for maturity assessment in fossil archosaurs. Journal of Vertebrate Paleontology 16:49-
1519 62.

1520 Bro-Jørgensen J., 2007. The intensity of sexual selection predicts weapon size in male bovids.
1521 Evolution 61:1316-26.

1522   Brown, C.M., Russell, A.P., and Ryan, M.J., 2009. Pattern and transition of surficial bone
1523   texture of the centrosaurine frill and their ontogenetic and taxonomic
1524   implications. Journal of Vertebrate Paleontology 29:132-141.

1525   Brown, C.M., Evans, D.C., Campione, N.E., O'Brien, L.J., and Eberth, D.A., 2013. Evidence
1526   for taphonomic size bias in the Dinosaur Park Formation (Campanian, Alberta), a
1527   model Mesozoic terrestrial alluvial-paralic system. Palaeogeography,
1528   Palaeoclimatology, Palaeoecology 372:108-122.

1529   Brown, C.M., Henderson, D.M., Vinther, J., Fletcher, I., Sistiaga, A., Herrera, J. and Summons,
1530   R.E., 2017. An exceptionally preserved three-dimensional armored dinosaur reveals
1531   insights into coloration and Cretaceous predator-prey dynamics. Current
1532   Biology 27:2514-2521.

1533   Brusatte, S.L., 2012. *Dinosaur paleobiology*. John Wiley & Sons.

1534   Brusatte, S.L., Butler, R.J., Prieto-Márquez, A., and Norell, M.A., 2012. Dinosaur
1535   morphological diversity and the end-Cretaceous extinction. Nature
1536   Communications 3:804.

1537   Buffetaut, E., Li, J., Tong, H. and Zhang, H., 2006. A two-headed reptile from the Cretaceous
1538   of China. Biology Letters 3:81-82.

1539   Bull, J.J. and Bulmer, M.G., 1989. Longevity enhances selection of environmental sex
1540   determination. Heredity 63:315-320.

1541   Bunce, M., Worthy, T.H., Ford, T., Hoppitt, W., Willerslev, E., Drummond, A., and Cooper,
1542   A., 2003. Extreme reversed sexual size dimorphism in the extinct New Zealand moa
1543   *Dinornis*. Nature 425:172.

1544   Burkhardt, D., 1989. UV vision: a bird's eye view of feathers. Journal of Comparative
1545   Physiology A 164:787-796.

1546   Burnham, K.P. and Anderson, D.R., 1998. Model Selection and Multimodel Inference.

1547   Springer, Berlin.

1548   Cabrera, D. and Stankowich, T., 2018. Stabbing slinkers: tusk evolution among

1549   Artiodactyls. Journal of Mammalian Evolution:1-8.

1550   Caldwell, M.W. and Lee, M.S., 2001. Live birth in Cretaceous marine lizards

1551   (mosasauroids). Proceedings of the Royal Society of London B: Biological

1552   Sciences 268:2397-2401.

1553   Caro, T.M., Graham, C.M., Stoner, C.J., and Flores, M.M., 2003. Correlates of horn and antler

1554   shape in bovids and cervids. Behavioral Ecology and Sociobiology 55:32-41.

1555   Carpenter, K., 1990. Variation in *Tyrannosaurus rex*. Dinosaur Systematics: Perspectives and

1556   Approaches:141-145.

1557   Carpenter, K., 1997. Agonistic behavior in pachycephalosaurs (Ornithischia, Dinosauria); a

1558   new look at head-butting behavior. Rocky Mountain Geology 32:19-25.

1559   Carpenter, K., 1998. Armor of *Stegosaurus stenops*, and the taphonomic history of a new

1560   specimen from Garden Park, Colorado. Modern Geology 23:127-144.

1561   Carpenter, K., Sanders, F., McWhinney, L., and Wood, L. 2005. Evidence for predator–prey

1562   relationships: example for *Allosaurus* and *Stegosaurus*. In Carpenter K. (ed.). The

1563   Carnivorous Dinosaurs, 325–350. Indiana University Press, Bloomington.

1564   Catchpole, C.K., 1987. Bird song, sexual selection and female choice. Trends in Ecology &

1565   Evolution 2:94-97.

1566   Cederström, P., Ahlberg, P., Nilsson, C.H., Ahlgren, J., and Eriksson, M.E., 2011. Moulting,

1567   ontogeny and sexual dimorphism in the Cambrian ptychopariid trilobite *Strenuaeva*

1568   *inflata* from the northern Swedish Caledonides. Palaeontology 54:685-703.

1569   Cerda, I.A., Chinsamy, A., and Pol, D., 2014. Unusual endosteally formed bone tissue in a

1570   Patagonian basal sauropodomorph dinosaur. The Anatomical Record 297:1385-1391.

1571   Chapman, R.E., 1997. Sexual dimorphism in dinosaurs. In Dinofest international:83-93.

1572   Chapman, R.E., Galton, P.M., Sepkoski Jr, J.J., and Wall, W.P., 1981. A morphometric study

1573   of the cranium of the pachycephalosaurid dinosaur *Stegoceras*. Journal of Paleontology

1574   55:608-618.

1575   Chapman, R.E., Sadleir, A., Dodson, P., and Makovicky, P., 2008. Handling missing data in

1576   paleontological matrices: Approaches for exploratory multivariate analyses in

1577   morphometrics. Journal of Vertebrate Paleontology 28:62A.

1578   Chelliah, K. and Sukumar, R., 2013. The role of tusks, musth and body size in male–male

1579   competition among Asian elephants, *Elephas maximus*. Animal Behaviour 86:1207-

1580   1214.

1581   Chen, J., Wang, B., Zheng, Y., Jarzembowski, E., Jiang, T., Wang, X., Zheng, X., and Zhang,

1582   H., 2019. Female-biased froghoppers (Hemiptera, Cercopoidea) from the Mesozoic of

1583   China and phylogenetic reconstruction of early Cercopoidea. Journal of Systematic

1584   Palaeontology:1-13.

1585   Cheng, Y.N., Holmes, R., Wu, X.C., and Alfonso, N., 2009. Sexual dimorphism and life history

1586   of *Keichousaurus hui* (Reptilia: Sauropterygia). Journal of Vertebrate

1587   Paleontology 29:401-408.

1588   Chinsamy, A., Chiappe, L.M., Marugán-Lobón, J., Chunling, G., and Fengjiao, Z., 2013.

1589   Gender identification of the Mesozoic bird *Confuciusornis sanctus*. Nature

1590   Communications 4:1381.

1591   Clutton-Brock, T.H., 1985. Size, sexual dimorphism, and polygyny in primates. In *Size and*

1592   *scaling in primate biology* (pp. 51-60). Springer, Boston, MA.

1593   Clutton-Brock, T., 2007. Sexual selection in males and females. Science 318:1882-1885.

1594   Clutton-Brock, T.H. and Harvey, P.H., 1978. Mammals, resources and reproductive

1595   strategies. Nature 273:191-195.

1596   Clutton-Brock, T.H., Harvey, P.H., and Rudder, B., 1977. Sexual dimorphism, socionomic sex

1597   ratio and body weight in primates. Nature 269:797.

1598   Clutton-Brock, T.H., Albon, S.D. and Harvey, P.H., 1980. Antlers, body size and breeding

1599   group size in the Cervidae. Nature 285:565-567.

1600   Colbert, M.W., 1999. Patterns of evolution and variation in the Tapiroidea (Mammalia:

1601   Perissodactyla). Doctoral dissertation.

1602   Colbert, M.W. and Rowe, T., 2008. Ontogenetic sequence analysis: using parsimony to

1603   characterize developmental sequences and sequence polymorphism. Journal of

1604   Experimental Zoology Part B: Molecular and Developmental Evolution 310:398-416.

1605   Cooney, C.R., Varley, Z.K., Nouri, L.O., Moody, C.J.A., Jardine, M.D., and Thomas, G.H.,

1606   2019. Sexual selection predicts the rate and direction of colour divergence in a large

1607   avian radiation. Nature Communications 10:1773.

1608   Croft, D.P., Darden, S.K., and Ruxton, G.D., 2009. Predation risk as a driving force for

1609   phenotypic assortment: a cross-population comparison. Proceedings of the Royal

1610   Society of London B 276:1899-1904.

1611   Cullen, T.M., Fraser, D., Rybczynski, N., and Schröder-Adams, C., 2014. Early evolution of

1612   sexual dimorphism and polygyny in Pinnipedia. Evolution 68:1469-1484.

1613   Czerkas, S.A., 1992. Discovery of dermal spines reveals a new look for sauropod

1614   dinosaurs. Geology 20:1068-1070.

1615   Darwin, C., 1888. The descent of man and selection in relation to sex. Vol. 1. Murray.

1616   Davis, E.B., Brakora, K.A., and Lee, A.H., 2011. Evolution of ruminant headgear: a

1617   review. Proceedings of the Royal Society B: Biological Sciences 278:2857-2865.

1618   Ditchkoff, S.S., Lochmiller, R.L., Masters, R.E., Starry, W.R., and Leslie Jr, D.M., 2001. Does

1619   fluctuating asymmetry of antlers in white–tailed deer (*Odocoileus virginianus*) follow

1620   patterns predicted for sexually selected traits? Proceedings of the Royal Society of

1621   London. Series B: Biological Sciences 268:891-898.

1622   Dodson, P., 1976. Quantitative aspects of relative growth and sexual dimorphism in

1623   *Protoceratops*. Journal of Paleontology 50:929-940.

1624   Donald, P.F., 2007. Adult sex ratios in wild bird populations. Ibis 149:671-692.

1625   Dunn, P.O., Whittingham, L.A. and Pitcher, T.E., 2001. Mating systems, sperm competition,

1626   and the evolution of sexual dimorphism in birds. Evolution 55:161-175.

1627   Dunning, J.B., 2007. CRC handbook of avian body masses. 2nd edition. CRC press.

1628   Eadie, J.M., Broekhoven, L., and Colgan, P., 1987. Size ratios and artifacts: Hutchinson's rule

1629   revisited. The American Naturalist 129:1-17.

1630   Eberhart-Phillips, L.J., Küpper, C., Carmona-Isunza, M.C., Vincze, O., Zefania, S., Cruz-

1631   López, M., Kosztolányi, A., Miller, T.E., Barta, Z., Cuthill, I.C., and Burke, T., 2018.

1632   Demographic causes of adult sex ratio variation and their consequences for parental

1633   cooperation. Nature Communications 9:1651.

1634   Emlen, S.T., 1972. An experimental analysis of the parameters of bird song eliciting species

1635   recognition. Behaviour 41:130-171.

1636   Emlen, S.T. and Wrege, P.H., 2004. Size dimorphism, intrasexual competition, and sexual

1637   selection in Wattled Jacana (*Jacana jacana*), a sex-role-reversed shorebird in

1638   Panama. The Auk 121:391-403.

1639   Endler, J.A., 1984. Natural and sexual selection on color patterns in poeciliid fishes.

1640   In Evolutionary ecology of neotropical freshwater fishes (pp. 95-111). Springer,

1641   Dordrecht.

1642   Endler, J.A., 1988. Sexual selection and predation risk in guppies. Nature 332:593-594.

1643   Erickson, G.M., Currie, P.J., Inouye, B.D., and Winn, A.A., 2006. Tyrannosaur life tables: an

1644   example of nonavian dinosaur population biology. Science 313:213-217.

1645   Erickson, G.M., Makovicky, P.J., Inouye, B.D., Zhou, C.F., and Gao, K.Q., 2009. A life table

1646   for *Psittacosaurus lujiatunensis*: initial insights into ornithischian dinosaur population

1647   biology. The Anatomical Record: Advances in Integrative Anatomy and Evolutionary

1648   Biology: Advances in Integrative Anatomy and Evolutionary Biology 292:1514-1521.

1649   Evans, D.C. and Reisz, R.R., 2007. Anatomy and relationships of *Lambeosaurus*

1650   *magnicristatus*, a crested hadrosaurid dinosaur (Ornithischia) from the Dinosaur Park

1651   Formation, Alberta. Journal of Vertebrate Paleontology 27:373-393.

1652   Evans, K.M., Bernt, M.J., Kolmann, M.A., Ford, K.L., and Albert, J.S., 2018. Why the long

1653   face? Static allometry in the sexually dimorphic phenotypes of Neotropical electric

1654   fishes. Zoological Journal of the Linnean Society 186:633-649.

1655   Fairbairn, D.J., Blanckenhorn, W.U., and Székely, T. eds., 2007. Sex, size and gender roles:

1656   evolutionary studies of sexual size dimorphism. Oxford University Press.

1657   Farke, A.A., 2004. Horn use in *Triceratops* (Dinosauria: Ceratopsidae): testing behavioral

1658   hypotheses using scale models. Palaeontologia Electronica 7:1-10.

1659   Farke, A.A., 2014. Evaluating combat in ornithischian dinosaurs. Journal of Zoology 292:242-

1660   249.

1661   Farke, A.A., Wolff, E.D., and Tanke, D.H., 2009. Evidence of combat in *Triceratops*. PLoS

1662   One 4:e4252.

1663   Frederickson, J.A. and Tumarkin-Deratzian, A.R., 2014. Craniofacial ontogeny in

1664   *Centrosaurus apertus*. PeerJ 2:e252.

1665   Galton, P.M., 1970. Pachycephalosaurids: dinosaurian battering rams. Discovery 6:23-32.

1666   Galton, P.M., 1971. A primitive dome-headed dinosaur (Ornithischia: Pachycephalosauridae)

1667   from the Lower Cretaceous of England and the function of the dome of

1668   pachycephalosaurids. Journal of Paleontology 45:40-47.

1669   Gamble, K.C., 2007. Internal anatomy of the hornbill casque described by radiography,

1670   contrast radiography, and computed tomography. Journal of avian medicine and

1671   surgery 21:38-49.

1672   Gautier-Hion, A. 1975. Dimorphisme sexuel et organisation sociale chez les cercopithécinés

1673   africains. Mammalia 39:365–374.

1674   Gerson, H.B. and Hickie, J.P., 1985. Head scarring on male narwhals (*Monodon monoceros*):

1675   evidence for aggressive tusk use. Canadian Journal of Zoology 63:2083-2087.

1676   Giffin, E.B., 1989. Pachycephalosaur paleoneurology (Archosauria: Ornithischia). Journal of

1677   Vertebrate Paleontology 9:67-77.

1678   Gilmore, C.W., 1914. Osteology of the armored Dinosauria in the United States National

1679   museum: with special reference to the genus *Stegosaurus* (No. 89). Govt. print. off.

1680   Gingerich, P.D., 1981. Variation, sexual dimorphism, and social structure in the early Eocene

1681   horse *Hyracotherium* (Mammalia, Perissodactyla). Paleobiology 7:443-455.

1682   Godfrey, L.R., Lyon, S.K., and Sutherland, M.R., 1993. Sexual dimorphism in large-bodied

1683   primates: The case of the subfossil lemurs. American Journal of Physical

1684   Anthropology 90:315-334.

1685   Gould, J.L. and Gould, C.G., 1989. Sexual selection (No. 591.56 G6).

1686   Gould, J.L., Elliott, S.L., Masters, C.M. and Mukerji, J., 1999. Female preferences in a fish

1687   genus without female mate choice. Current Biology 9:497-500.

1688   Gould, S.J., 1974. The origin and function of 'bizarre' structures: antler size and skull size in

1689   the 'Irish Elk,' *Megaloceros giganteus*. Evolution 28:191-220.

1690   Goyens, J., Dirckx, J., and Aerts, P., 2015. Costly sexual dimorphism in *Cyclommatus*

1691   *metallifer* stag beetles. Functional Ecology 29:35-43.

1692   Grammer, K. and Thornhill, R., 1994. Human (*Homo sapiens*) facial attractiveness and sexual

1693   selection: the role of symmetry and averageness. Journal of Comparative

1694   Psychology 108:233.

1695   Grant, P.R. and Grant, B.R., 2008. Pedigrees, assortative mating and speciation in Darwin's

1696   finches. Proceedings of the Royal Society B: Biological Sciences 275:661-668.

1697   Grant, B.R. and Grant, P.R., 2010. Songs of Darwin's finches diverge when a new species

1698   enters the community. Proceedings of the National Academy of Sciences 107:20156-

1699   20163.

1700   Greenland, S., Senn, S.J., Rothman, K.J., Carlin, J.B., Poole, C., Goodman, S.N. and Altman,

1701   D.G., 2016. Statistical tests, P values, confidence intervals, and power: a guide to

1702   misinterpretations. European Journal of Epidemiology 31:337-350.

1703   Griffin, C.T. and Nesbitt, S.J., 2016. The femoral ontogeny and long bone histology of the

1704   Middle Triassic (? late Anisian) dinosauriform *Asilisaurus kongwe* and implications for

1705   the growth of early dinosaurs. Journal of Vertebrate Paleontology 36:e1111224.

1706   Halsey, L.G., 2019. The reign of the p-value is over: what alternative analyses could we employ

1707   to fill the power vacuum? Biology Letters 15:20190174.

1708   Hansen, T.F., 2016. On bias and precision in meta-analysis: the error in the error. Journal of

1709   Evolutionary Biology 29:1919-1921.

1710   Hardin, G., 1960. The competitive exclusion principle. *Science* 131:1292-1297.

1711   Harvey, P.H., Kavanagh, M., and Clutton-Brock, T.H., 1978. Sexual dimorphism in primate

1712   teeth. Journal of Zoology 186:475-485.

1713   Hayashi, S., Carpenter, K., Suzuki, D., 2009. Different growth patterns between the skeleton

1714   and osteoderms of *Stegosaurus* (Ornithischia: Thyreophora). Journal of Vertebrate

1715   Paleontology 29:123-31.

1716   Heath, D.J., 1977. Simultaneous hermaphroditism; cost and benefit. Journal of Theoretical

1717   Biology 64:363-373.

1718   Heinsohn, R., Legge, S., and Endler, J.A., 2005. Extreme reversed sexual dichromatism in a

1719   bird without sex role reversal. Science 309:617-619.

1720   Heske, E.J. and Ostfeld, R.S., 1990. Sexual dimorphism in size, relative size of testes, and

1721   mating systems in North American voles. Journal of Mammalogy 71:510-519.

1722   Holland, S.M., 2019. Estimation, not significance. Paleobiology 45:1-6.

1723   Hone, D.W.E. and Faulkes, C.G., 2014. A proposed framework for establishing and evaluating

1724   hypotheses about the behaviour of extinct organisms. Journal of Zoology 292:260-267.

1725   Hone, D.W. and Mallon, J.C., 2017. Protracted growth impedes the detection of sexual

1726   dimorphism in non-avian dinosaurs. Palaeontology 60:535-545.

1727   Hone, D.W.E. and Naish, D., 2013. The 'species recognition hypothesis' does not explain the

1728   presence and evolution of exaggerated structures in non-avialan dinosaurs. Journal of

1729   Zoology 290:172-180.

1730   Hone, D.W., Naish, D., and Cuthill, I.C., 2012. Does mutual sexual selection explain the

1731   evolution of head crests in pterosaurs and dinosaurs? Lethaia 45:139-156.

1732   Hone, D.W., Farke, A.A., and Wedel, M.J., 2016b. Ontogeny and the fossil record: what, if

1733   anything, is an adult dinosaur? Biology letters 12:20150947.

1734   Hone, D.W., Wood, D., and Knell, R.J., 2016a. Positive allometry for exaggerated structures

1735   in the ceratopsian dinosaur *Protoceratops andrewsi* supports socio-sexual

1736   signaling. Palaeontologia Electronica 19:1-13.

1737   Hone, D., Mallon, J.C., Hennessey, P. and Witmer, L.M., 2020. Ontogeny of a sexually

1738   selected structure in an extant archosaur *Gavialis gangeticus* (Pseudosuchia:

1739   Crocodylia) with implications for sexual dimorphism in dinosaurs. PeerJ 8:e9134.

1740   Hu, D., Clarke, J.A., Eliason, C.M., Qiu, R., Li, Q., Shawkey, M.D., Zhao, C., D'Alba, L.,

1741   Jiang, J., and Xu, X., 2018. A bony-crested Jurassic dinosaur with evidence of

1742   iridescent plumage highlights complexity in early paravian evolution. Nature

1743   Communications 9:217.

1744   Huber, R., Baumung, R., Wurzinger, M., Semambo, D., Mwai, O., and Winckler, C., 2008.

1745   Grazing, social and comfort behaviour of Ankole and crossbred (Ankole× Holstein)

1746   heifers on pasture in south western Uganda. Applied Animal Behaviour

1747   Science 112:223-234.

1748   Hunt, J., Breuker, C.J., Sadowski, J.A., and Moore, A.J., 2009. Male–male competition, female

1749   mate choice and their interaction: determining total sexual selection. Journal of

1750   evolutionary biology 22:13-26.

1751   Hunt, S., Bennett, A.T.D., Cuthill, I.C., and Griffiths, R., 1998. Blue tits are ultraviolet

1752   tits. Proceedings of the Royal Society of London. Series B: Biological

1753   Sciences 265:451-455.

1754   Hurvich, C.M. and Tsai, C.L., 1989. Regression and time series model selection in small

1755   samples. Biometrika 76:297-307.

1756   Hedrick, B.P. and Dodson, P., 2013. Lujiatun psittacosaurids: understanding individual and

1757   taphonomic variation using 3D geometric morphometrics. PLoS One 8:e69265.

1758   Irmis, R.B., 2007. Axial skeleton ontogeny in the Parasuchia (Archosauria: Pseudosuchia) and

1759   its implications for ontogenetic determination in archosaurs. Journal of Vertebrate

1760   Paleontology 27:350-361.

1761   Isles, T.E., 2009. The socio-sexual behaviour of extant archosaurs: implications for

1762   understanding dinosaur behaviour. Historical Biology 21:139-214.

1763   Jones, I.L. and Hunter, F.M., 1993. Mutual sexual selection in a monogamous

1764   seabird. Nature 362:238.

Karsten, K.B., Andriamandimbiarisoa, L.N., Fox, S.F., and Raxworthy, C.J., 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. and 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, R.J., Farke, A.A., Loewen, M.A., and Hone, D.W., 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, R.J. and 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, R.J., Naish, D., Tomkins, J.L., and Hone, D.W., 2013b. Is sexual selection defined by dimorphism alone? A reply to Padian and Horner. Trends in Ecology & Evolution 28:250-251.

Knell, R.J., Naish, D., Tomkins, J.L., and Hone, D.W., 2013a. Sexual selection in prehistoric animals: detection and implications. Trends in Ecology & Evolution 28:38-47.

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

Kraaijeveld, K., Kraaijeveld-Smit, F.J., and Maan, M.E., 2011. Sexual selection and speciation: the comparative evidence revisited. Biological Reviews 86:367-377.

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

Krause, J. and Ruxton, G.D., 2002. Living in Groups. Oxford: Oxford University Press.

1790 Krauss, D., Pezon, A.N., Nguyen, P.E., Salame, I.S. and Rywkin, S., 2010. Evolutionary

1791 interactions between horn and frill morphology in chasmosaurine ceratopsians. In *New*

1792 *Perspectives on Horned Dinosaurs: The Royal Tyrrell Museum Ceratopsian*

1793 *Symposium* (pp. 282-292). Indiana University Press.

1794 Krishtalka, L., Stucky, R.K. and Beard, K.C., 1990. The earliest fossil evidence for sexual

1795 dimorphism in primates. Proceedings of the National Academy of Sciences 87:5223-

1796 5226.

1797 Lack D. 1968 Ecological adaptations for breeding in birds. London, UK: Methuen & Co.

1798 Laver, R.J., Purwandana, D., Ariefiandy, A., Imansyah, J., Forsyth, D., Ciofi, C. and Jessop,

1799 T.S., 2012. Life-history and spatial determinants of somatic growth dynamics in

1800 Komodo dragon populations. PLoS One 7:e45398.

1801 Lemaître, J.F., Vanpé, C., Plard, F. and Gaillard, J.M., 2014. The allometry between secondary

1802 sexual traits and body size is nonlinear among cervids. Biology Letters 10:20130869.

1803 Lamichhaney, S., Han, F., Webster, M.T., Andersson, L., Grant, B.R., and Grant, P.R., 2018.

1804 Rapid hybrid speciation in Darwin's finches. Science 359:224-228.

1805 Lande, R., 1980. Sexual dimorphism, sexual selection, and adaptation in polygenic

1806 characters. Evolution 34:292-305.

1807 Larson, P.L., 1994. *Tyrannosaurus* sex. The Paleontological Society Special

1808 Publications 7:139-156.

1809 Larson, P.L., 2008. Variation and sexual dimorphism in *Tyrannosaurus rex*. *Tyrannosaurus*

1810 *rex*:103-130.

1811 Lee, A.H. and Werning, S., 2008. Sexual maturity in growing dinosaurs does not fit reptilian

1812 growth models. Proceedings of the National Academy of Sciences 105:582-587.

1813 Lee, P.A., 1980. Normal ages of pubertal events among American males and females. Journal

1814 of Adolescent Health Care 1:26-29.

1815 Lewis, S.M., Cratsley, C.K., and Demary, K., 2004, January. Mate recognition and choice in

1816   Photinus fireflies. In Annales Zoologici Fennici (pp. 809-821). Finnish Zoological and

1817   Botanical Publishing Board.

1818 Li, Q., Gao, K.Q., Meng, Q., Clarke, J.A., Shawkey, M.D., D'Alba, L., Pei, R., Ellison, M.,

1819   Norell, M.A., and Vinther, J., 2012. Reconstruction of *Microraptor* and the evolution

1820   of iridescent plumage. Science 335:1215-1219.

1821 Lihoreau, M., Zimmer, C. and Rivault, C., 2008. Mutual mate choice: when it pays both sexes

1822   to avoid inbreeding. PLoS One 3:e3365.

1823 Liker, A., Freckleton, R.P., and Székely, T., 2013. The evolution of sex roles in birds is related

1824   to adult sex ratio. Nature Communications 4:1587.

1825 Liker, A., Freckleton, R.P., and Székely, T. 2014. Divorce and infidelity are associated with

1826   skewed adult sex ratios in birds. Current Biology 24:880-884.

1827 Lillie, F.R., 1931. Bilateral gynandromorphism and lateral hemihypertrophy in

1828   birds. Science 74:387-390.

1829 Lister, A.M., 2009. Late-glacial mammoth skeletons (*Mammuthus primigenius*) from

1830   Condover (Shropshire, UK): anatomy, pathology, taphonomy and chronological

1831   significance. Geological Journal 44:447-479.

1832 Lockley, M.G., McCrea, R.T., Buckley, L.G., Lim, J.D., Matthews, N.A., Breithaupt, B.H.,

1833   Houck, K.J., Gierliński, G.D., Surmik, D., Kim, K.S., and Xing, L., 2016. Theropod

1834   courtship: large scale physical evidence of display arenas and avian-like scrape

1835   ceremony behaviour by Cretaceous dinosaurs. Scientific Reports 6:18952.

1836 Löfstedt, C., 1993. Moth pheromone genetics and evolution. Philosophical Transactions of the

1837   Royal Society of London. Series B: Biological Sciences 340:167-177.

1838 Longrich, N.R., 2010. *Mojoceratops perifania*, a new chasmosaurine ceratopsid from the late

1839   Campanian of western Canada. Journal of Paleontology 84:681-694.

1840   Longrich, N.R., Sankey, J., and Tanke, D., 2010. *Texacephale langstoni*, a new genus of

1841   pachycephalosaurid (Dinosauria: Ornithischia) from the upper Campanian Aguja

1842   Formation, southern Texas, USA. Cretaceous Research 31:274-284.

1843   Lucas, A.M. and Stettenheim, P.R., 1972. Avian Anatomy Integument (No. QL697 L8).

1844   Lund, R., 1982. *Harpagofututor volsellorhinus* new genus and species (Chondrichthyes,

1845   Chondrenchelyiformes) from the Namurian Bear Gulch Limestone, Chondrenchelys

1846   problematica Traquair (Visean), and their sexual dimorphism. Journal of Paleontology

1847   56:938-958.

1848   Maisey, J.G., 2009. The spine-brush complex in symmoriiform sharks (Chondrichthyes;

1849   Symmoriiformes), with comments on dorsal fin modularity. Journal of Vertebrate

1850   Paleontology 29:14-24.

1851   Mallon, J.C., 2017. Recognizing sexual dimorphism in the fossil record: lessons from nonavian

1852   dinosaurs. Paleobiology 43:495-507.

1853   Mallon, J.C. and Holmes, R.B., 2006. A reevaluation of sexual dimorphism in the postcranium

1854   of the chasmosaurine ceratopsid *Chasmosaurus belli* (Dinosauria: Ornithischia). The

1855   Canadian Field-Naturalist 120:403-412.

1856   Martins, M.J.F., Puckett, T.M., Lockwood, R., Swaddle, J.P. and Hunt, G., 2018. High male

1857   sexual investment as a driver of extinction in fossil ostracods. Nature 556:366-369.

1858   Maschenko, E.N., 2002. Individual development, biology and evolution of the woolly

1859   mammoth. Cranium 19:4-120.

1860   Maynard Smith, J. and Harper, D., 2003. Animal Signals. Oxford: Oxford University Press.

1861   Maynard Smith, J. and Price, G.R., 1973. The logic of animal conflict. Nature 246:15.

1862   Mayr, G., Pittman, M., Saitta, E., Kaye, T.G., and Vinther, J., 2016. Structure and homology

1863   of *Psittacosaurus* tail bristles. Palaeontology 59:793-802.

1864   McCullough, E.L., Miller, C.W., and Emlen, D.J., 2016. Why sexually selected weapons are

1865   not ornaments. Trends in Ecology & Evolution 31:742-751.

1866   McCoy, D.E., Feo, T., Harvey, T.A., and Prum, R.O., 2018. Structural absorption by barbule

1867   microstructures of super black bird of paradise feathers. Nature Communications 9:1.

1868   McWhinney, L. A., B. M. Rothschild, and K. Carpenter., 2001. Posttraumatic chronic

1869   osteomyelitis in *Stegosaurus* dermal spikes; pp. 141-156 in K. Carpenter (ed.), The

1870   Armored Dinosaurs. Indiana University Press, Bloomington.

1871   Meachen-Samuels, J.A. and Binder, W.J., 2010. Sexual dimorphism and ontogenetic growth

1872   in the American lion and sabertoothed cat from Rancho La Brea. Journal of

1873   Zoology 280:271-279.

1874   Mendelson, T.C. and Shaw, K.L., 2013. Further misconceptions about species recognition: a

1875   reply to Padian and Horner. Trends in Ecology & Evolution 28:252-253.

1876   Møller, A.P., 1992. Patterns of fluctuating asymmetry in weapons: evidence for reliable

1877   signalling of quality in beetle horns and bird spurs. Proceedings of the Royal Society

1878   of London B 248:199-206.

1879   Møller, A.P. and Höglund, J., 1991. Patterns of fluctuating asymmetry in avian feather

1880   ornaments: implications for models of sexual selection. Proceedings of the Royal

1881   Society of London B 245:1-5.

1882   Molnar, R.E., 1977. Analogies in the evolution of combat and display structures in ornithopods

1883   and ungulates. Evolutionary Theory 3:165-190.

1884   Moodie, R.L., 1918. Studies in Paleopathology: General Consideration of the Evidences of

1885   Pathological Conditions Found Among Fossil Animals. I. Paul B. Hoeber.

1886   Moodie, R.L., 1923. Paleopathology. An Introduction to the Study of Ancient Evidences of

1887   Disease. Urbana, Illinois: University of Illinois Press.

1888   Moriarty, P. and Holt, R.G., 2011. Faraday waves produced by periodic substrates: Mimicking
1889   the alligator water dance. The Journal of the Acoustical Society of America 129:2411-
1890   2411.

1891   Morrissey, M.B., 2016. Meta-analysis of magnitudes, differences and variation in evolutionary
1892   parameters. Journal of Evolutionary Biology 29:1882-1904.

1893   Motani, R., Huang, J., Jiang, D.Y., Tintori, A., Rieppel, O., You, H., Hu, Y.C., and Zhang, R.,
1894   2018. Separating sexual dimorphism from other morphological variation in a specimen
1895   complex of fossil marine reptiles (Reptilia, Ichthyosauriformes,
1896   Chaohusaurus). Scientific Reports 8:14978.

1897   Murphy, N.L., Trexler, D. and Thompson, M., 2007. "Leonardo," a mummified
1898   *Brachylophosaurus* from the Judith River Formation. Horns and beaks: Ceratopsian
1899   and Ornithopod Dinosaurs:117-33.

1900   Nakagawa, S. and Cuthill, I.C., 2007. Effect size, confidence interval and statistical
1901   significance: a practical guide for biologists. Biological Reviews 82:591-605.

1902   Navarro, J.L., Vignolo, P.E., Demaria, M.R., Maceira, N.O., and Martella, M.B., 2005. Growth
1903   curves of farmed Greater Rheas (*Rhea americana albescens*) from central
1904   Argentina. Archiv fur Geflugelkunde 69:90-93.

1905   Neige, P., Marchand, D. and Laurin, B., 1997. Heterochronic differentiation of sexual
1906   dimorphs among Jurassic ammonite species. Lethaia 30:145-155.

1907   Nelson, D.A., 1989. The importance of invariant and distinctive features in species recognition
1908   of bird song. The Condor 91:120-130.

1909   Nolan, P.M., Stephen Dobson, F., Nicolaus, M., Karels, T.J., McGraw, K.J., and Jouventin, P.,
1910   2010. Mutual mate choice for colorful traits in king penguins. Ethology 116:635-644.

1911   Nottebohm, F. and Arnold, A.P., 1976. Sexual dimorphism in vocal control areas of the
1912   songbird brain. Science 194:211-213.

1913   Nowicki, S., Peters, S., and Podos, J., 1998. Song learning, early nutrition and sexual selection

1914   in songbirds. American Zoologist 38:179-190.

1915   O'Brien, D.M., Allen, C.E., Van Kleeck, M.J., Hone, D., Knell, R., Knapp, A., Christiansen,

1916   S., and Emlen, D.J., 2018. On the evolution of extreme structures: static scaling and the

1917   function of sexually selected signals. Animal Behaviour 144:95-108.

1918   Ord, T.J. and Stuart-Fox, D., 2006. Ornament evolution in dragon lizards: multiple gains and

1919   widespread losses reveal a complex history of evolutionary change. Journal of

1920   Evolutionary Biology 19:797-808.

1921   Owens, I.P. and Hartley, I.R., 1998. Sexual dimorphism in birds: why are there so many

1922   different forms of dimorphism? Proceedings of the Royal Society of London B:

1923   Biological Sciences 265:397-407.

1924   Owen-Smith, N., 1993. Comparative mortality rates of male and female kudus: the costs of

1925   sexual size dimorphism. Journal of Animal Ecology 62:428-440.

1926   Packer, C., 1983. Sexual dimorphism: the horns of African antelopes. Science 221:1191-1193.

1927   Padian, K. and Horner, J.R., 2011. The evolution of 'bizarre structures' in dinosaurs:

1928   biomechanics, sexual selection, social selection or species recognition? Journal of

1929   Zoology 283:3-17.

1930   Padian, K. and Horner, J.R., 2013. Misconceptions of sexual selection and species recognition:

1931   a response to Knell et al. and to Mendelson and Shaw. Trends in Ecology &

1932   Evolution 28:249-250.

1933   Padian, K. and Horner, J.R., 2014. The species recognition hypothesis explains exaggerated

1934   structures in non-avialan dinosaurs better than sexual selection does. Comptes Rendus

1935   Palevol 13:97-107.

1936   Panhuis, T.M., Butlin, R., Zuk, M., and Tregenza, T., 2001. Sexual selection and

1937   speciation. Trends in Ecology & Evolution 16:364-371.

1938   Parker, G.A., 1970. Sperm competition and its evolutionary consequences in the

1939   insects. Biological Reviews 45:525-567.

1940   Parker, G.A., 2006. Sexual conflict over mating and fertilization: an overview. Philosophical

1941   Transactions of the Royal Society of London B Biological Sciences 361:235-259

1942   Parry, L.A., Smithwick, F., Nordén, K.K., Saitta, E.T., Lozano-Fernandez, J., Tanner, A.R.,

1943   Caron, J.B., Edgecombe, G.D., Briggs, D.E., and Vinther, J., 2018. Soft-Bodied Fossils

1944   Are Not Simply Rotten Carcasses–Toward a Holistic Understanding of Exceptional

1945   Fossil Preservation: Exceptional Fossil Preservation Is Complex and Involves the

1946   Interplay of Numerous Biological and Geological Processes. Bioessays 40:1700167.

1947   Paterson, H.E., 1980. A comment on "mate recognition systems." Evolution 34:330–331.

1948   Paterson H.E., 1985. The recognition concept of species. In *Species and Speciation* (ed. E.

1949   Vrba), pp. 21–29. Transvaal Museum Monograph. Pretoria, South Africa: Transvaal

1950   museum.

1951   Paukstis, G.L. and Janzen, F.J., 1990. Sex determination in reptiles: summary of effects of

1952   constant temperatures of incubation on sex ratios of offspring. Smithsonian

1953   Herpetological Information Service 83:1-28.

1954   Pearson, D., Shine, R., and How, R., 2002. Sex-specific niche partitioning and sexual size

1955   dimorphism in Australian pythons (Morelia spilota imbricata). Biological Journal of

1956   the Linnean Society 77:113-125.

1957   Pérez-Barbería, F.J., Gordon, I.J. and Pagel, M., 2002. The origins of sexual dimorphism in

1958   body size in ungulates. Evolution 56:1276-1285.

1959   Persons IV, W.S., Funston, G.F., Currie, P.J., and Norell, M.A., 2015. A possible instance of

1960   sexual dimorphism in the tails of two oviraptorosaur dinosaurs. Scientific

1961   Reports 5:9472.

1962 Peterson, J.E. and Vittore, C.P., 2012. Cranial pathologies in a specimen of
1963   Pachycephalosaurus. PloS one 7:e36227.

1964 Peterson, J.E., Dischler, C., and Longrich, N.R., 2013. Distributions of cranial pathologies
1965   provide evidence for head-butting in dome-headed dinosaurs
1966   (Pachycephalosauridae). PloS one 8:e68620.

1967 Petrie, M., Tim, H., and Carolyn, S., 1991. Peahens prefer peacocks with elaborate
1968   trains. Animal Behaviour 41:323-331.

1969 Platt, S.G., Rainwater, T.R., Thorbjarnarson, J.B., Finger, A.G., Anderson, T.A., and
1970   McMurry, S.T., 2009. Size estimation, morphometrics, sex ratio, sexual size
1971   dimorphism, and biomass of Morelet's crocodile in northern Belize. Caribbean Journal
1972   of Science 45:80-93.

1973 Plavcan, J.M., 1994. Comparison of four simple methods for estimating sexual dimorphism in
1974   fossils. American Journal of Physical Anthropology 94:465-476.

1975 Poissant, J., Wilson, A.J., Festa-Bianchet, M., Hogg, J.T., and Coltman, D.W., 2008.
1976   Quantitative genetics and sex-specific selection on sexually dimorphic traits in bighorn
1977   sheep. Proceedings of the Royal Society of London B: Biological Sciences 275:623-
1978   628.

1979 Pringle, R.M., 2020. Paleoecology: the functional uniqueness of ancient megafauna. Current
1980   Biology 30:R32-R35.

1981 Raath, M.A., 1990. Morphological variation in small theropods and its meaning in systematics:
1982   evidence from *Syntarsus rhodesiensis*, in Carpenter, K. & Currie, P.J. (Eds.), Dinosaur
1983   Systematics: Approaches and Perspectives, Cambridge Univ. Press, Cambridge, pp. 91-
1984   105.

1985 Rehg, J.A. and Leigh, S.R., 1999. Estimating sexual dimorphism and size differences in the
1986   fossil record: a test of methods. American Journal of Physical Anthropology 110:95-
1987   104.

1988   Reno, P.L., Meindl, R.S., McCollum, M.A., and Lovejoy, C.O., 2003. Sexual dimorphism in
1989   Australopithecus afarensis was similar to that of modern humans. Proceedings of the
1990   National Academy of Sciences 100:9404-9409.

1991   Rice, W.R. and Chippindale, A.K., 2001. Intersexual ontogenetic conflict. Journal of
1992   Evolutionary Biology 14:685-693.

1993   Rico-Guevara, A. and Hurme, K.J., 2019. Intrasexually selected weapons. Biological
1994   Reviews 94:60-101.

1995   Rinehart, L.F., Lucas, S.G., Heckert, A.B., Spielmann, J.A., and Celeskey, M.D., 2009. The
1996   Paleobiology of *Coelophysis bauri* (Cope) from the Upper Triassic (Apachean)
1997   Whitaker quarry, New Mexico, with detailed analysis of a single quarry block: Bulletin
1998   45. New Mexico Museum of Natural History and Science.

1999   Robinson, M.R. and Kruuk, L.E., 2007. Function of weaponry in females: the use of horns in
2000   intrasexual competition for resources in female Soay sheep. Biology Letters 3:651-654.

2001   Rowe, T., Ketcham, R.A., Denison, C., Colbert, M., Xu, X. and Currie, P.J., 2001. Forensic
2002   palaeontology: the *Archaeoraptor* forgery. Nature 410:539.

2003   Roy, A., Pittman, M., Saitta, E.T., Kaye, T.G. and Xu, X., 2019. Recent advances in amniote
2004   palaeocolour reconstruction and a framework for future research. Biological Reviews
2005   doi:10.1111/brv.12552.

2006   Ryan, M.J. and Keddy-Hector, A., 1992. Directional patterns of female mate choice and the
2007   role of sensory biases. The American Naturalist 139:S4-S35.

2008   Ryan, M.J. and Rand, A.S., 1993. Species recognition and sexual selection as a unitary problem
2009   in animal communication. Evolution 47:647-657.

2010   Ryan, M.J., Holmes, R. and Russell, A.P., 2007. A revision of the late Campanian
2011   centrosaurine ceratopsid genus *Styracosaurus* from the Western Interior of North
2012   America. Journal of Vertebrate Paleontology 27:944-962.

2013 Saitta, E. T., 2014. Paleobiology of North American stegosaurs: Evidence for sexual
2014   dimorphism. Princeton University Senior Thesis.

2015 Saitta, E. T. 2015. Evidence for Sexual Dimorphism in the Plated Dinosaur *Stegosaurus mjosi*
2016   (Ornithischia, Stegosauria) from the Morrison Formation (Upper Jurassic) of Western
2017   USA. PLoS ONE 10:e0123503.

2018 Saitta, E.T. and Vinther, J., 2019. A perspective on the evidence for keratin protein preservation
2019   in fossils: An issue of replication versus validation. Palaeontologia Electronica 22.3.2E.

2020 Saitta, E.T., Fletcher, I., Martin, P., Pittman, M., Kaye, T.G., True, L.D., Norell, M.A., Abbott,
2021   G.D., Summons, R.E., Penkman, K., and Vinther, J., 2018. Preservation of feather
2022   fibers from the Late Cretaceous dinosaur *Shuvuuia deserti* raises concern about
2023   immunohistochemical analyses on fossils. Organic Geochemistry 125:142-151.

2024 Sampson, S.D., Ryan, M.J. and Tanke, D.H., 1997. Craniofacial ontogeny in centrosaurine
2025   dinosaurs (Ornithischia: Ceratopsidae): taxonomic and behavioral
2026   implications. Zoological Journal of the Linnean Society 121:293-337.

2027 Sánchez, I.M., Quiralte, V., Morales, J., Azanza, B., and Pickford, M., 2010. Sexual
2028   dimorphism of the frontal appendages of the early Miocene African pecoran
2029   *Prolibytherium* Arambourg, 1961 (Mammalia, Ruminantia). Journal of Vertebrate
2030   Paleontology 30:1306-1310.

2031 Sander, P.M. and Klein, N., 2005. Developmental plasticity in the life history of a prosauropod
2032   dinosaur. Science 310:1800-1802.

2033 Santos, S.I.O., Elward, B., and Lumeij, J.T., 2006. Sexual dichromatism in the Blue-fronted
2034   Amazon Parrot (*Amazona aestiva*) revealed by multiple-angle spectrometry. Journal of
2035   Avian Medicine and Surgery 20:8-15.

2036 Sato, T., Cheng, Y.N., Wu, X.C., Zelenitsky, D.K., and Hsiao, Y.F., 2005. A pair of shelled
2037   eggs inside a female dinosaur. Science 308:375-375.

2038   Scannella, J.B., Fowler, D.W., Goodwin, M.B., and Horner, J.R., 2014. Evolutionary trends in *Triceratops* from the Hell Creek Formation, Montana. Proceedings of the National Academy of Sciences 111:10245-10250.

2041   Schall, J.J., Bromwich, C.R., Werner, Y.L., and Midlege, J., 1989. Clubbed regenerated tails in *Agama agama* and their possible use in social interactions. Journal of Herpetology 23:303-305.

2044   Schärer, L., Rowe, L., and Arnqvist, G., 2012. Anisogamy, chance and the evolution of sex roles. Trends in Ecology & Evolution 27:260-264.

2046   Schilling, M.F., Watkins, A.E., and Watkins, W., 2002. Is human height bimodal? The American Statistician 56:223-229.

2048   Schott, R.K., Evans, D.C., Goodwin, M.B., Horner, J.R., Brown, C.M., and Longrich, N.R., 2011. Cranial ontogeny in *Stegoceras validum* (Dinosauria: Pachycephalosauria): a quantitative model of pachycephalosaur dome growth and variation. PLoS One 6:e21092.

2052   Schwartz, G.T. and Dean, C., 2001. Ontogeny of canine dimorphism in extant hominoids. American Journal of Physical Anthropology 115:269-283.

2054   Schweitzer, M.H., Zheng, W., Zanno, L., Werning, S., and Sugiyama, T., 2016. Chemistry supports the identification of gender-specific reproductive tissue in *Tyrannosaurus rex*. Scientific Reports 6:23099.

2057   Scott, H.M. and Payne, L.F., 1934. The effect of gonadectomy on the secondary sexual characters of the bronze turkey (*M. gallopavo*). Journal of Experimental Zoology 69:123-136.

2060   Seddon, N., 2005. Ecological adaptation and species recognition drives vocal evolution in neotropical suboscine birds. Evolution 59:200-215.

2062 Sengupta, S., Ezcurra, M.D., and Bandyopadhyay, S., 2017. A new horned and long-necked

2063   herbivorous stem-archosaur from the Middle Triassic of India. Scientific

2064   Reports 7:8366.

2065 Shine, R. 1979. Sexual selection and sexual dimorphism in the Amphibia. Copeia 2:297–306.

2066 Sibley, D.A., 2014. *The Sibley Guide to Birds*. Second edition. Knopf. 624 p.

2067 Skeate, S.T., 1984. Courtship and reproductive behaviour of captive white-fronted amazon

2068   parrots *Amazona albifrons.* Bird Behavior 5:103-109.

2069 Smith, D.K., 1998. A morphometric analysis of *Allosaurus*. Journal of Vertebrate

2070   Paleontology 18:126-142.

2071 Smith, J.N., Goldizen, A.W., Dunlop, R.A., and Noad, M.J., 2008. Songs of male humpback

2072   whales, *Megaptera novaeangliae*, are involved in intersexual interactions. Animal

2073   Behaviour 76:467-477.

2074 Smith, K.M. and Fisher, D.C., 2011. Sexual dimorphism of structures showing indeterminate

2075   growth: tusks of American mastodons (*Mammut americanum*). Paleobiology 37:175-

2076   194.

2077 Snively, E. and Cox, A., 2008. Structural mechanics of pachycephalosaur crania permitted

2078   head-butting behavior. Palaeontologia Electronica 11:3A.

2079 Snively, E. and Theodor, J.M., 2011. Common functional correlates of head-strike behavior in

2080   the pachycephalosaur *Stegoceras validum* (Ornithischia, Dinosauria) and combative

2081   artiodactyls. PLoS One 6:e21422.

2082 Stankowich, T. and Caro, T., 2009. Evolution of weaponry in female bovids. Proceedings of

2083   the Royal Society of London B: Biological Sciences 276:20091256.

2084 Stochholm, K., Juul, S., Juel, K., Naeraa, R.W. and Højbjerg Gravholt, C., 2006. Prevalence,

2085   incidence, diagnostic delay, and mortality in Turner syndrome. The Journal of Clinical

2086   Endocrinology & Metabolism 91:3897-3902.

2087 Sues, H.D., 1978. Functional morphology of the dome in pachycephalosaurid dinosaurs. Neues
2088   Jahrbuch für Geologie und Paläontologie Monatshefte 8:459-472.

2089 Sugiura, N., 1978. Further analysts of the data by akaike's information criterion and the finite
2090   corrections: Further analysts of the data by akaike's. Communications in Statistics-
2091   Theory and Methods 7:13-26.

2092 Sullivan, C., Reisz, R.R., and Smith, R.M., 2003. The Permian mammal-like herbivore
2093   *Diictodon*, the oldest known example of sexually dimorphic armament. Proceedings of
2094   the Royal Society of London. Series B: Biological Sciences 270:173-178.

2095 Székely, T., Webb, J.N., and Cuthill, I.C., 2000. Mating patterns, sexual selection and parental
2096   care: an integrative approach. Pp. 159-185 in M. Apollonio, M. Festa-Bianchet & D.
2097   Mainardi (eds) Vertebrate Mating Systems. World Scientific Publishing Inc., London.

2098 Székely, T., Liker, A., Freckleton, R.P., Fichtel, C., and Kappeler, P.M., 2014. Sex-biased
2099   survival predicts adult sex ratio variation in wild birds. Proceedings of the Royal
2100   Society B: Biological Sciences 281:20140342.

2101 Tanke, D.H. and Rothschild, B.M., 2002. DINOSORES: An Annotated Bibliography of
2102   Dinosaur Paleopathology and Related Topics—1838-2001: Bulletin 20. New Mexico
2103   Museum of Natural History and Science.

2104 Taylor, G.W., Santos, J.C., Perrault, B.J., Morando, M., Vásquez Almazán, C.R. and Sites Jr,
2105   J.W., 2017. Sexual dimorphism, phenotypic integration, and the evolution of head
2106   structure in casque-headed lizards. Ecology and Evolution 7:8989-8998.

2107 Tinbergen, N., 1948. Social releasers and the experimental method required for their study. The
2108   Wilson Bulletin 60:6-51.

2109 Tobias, J.A., Montgomerie, R., and Lyon, B.E., 2012. The evolution of female ornaments and
2110   weaponry: social selection, sexual selection and ecological competition. Philosophical
2111   Transactions of the Royal Society B: Biological Sciences 367:2274-2293.

2112 Tumarkin-Deratzian, A.R., Vann, D.R., and Dodson, P., 2006. Bone surface texture as an

2113   ontogenetic indicator in long bones of the Canada goose *Branta canadensis*

2114   (Anseriformes: Anatidae). Zoological Journal of the Linnean Society 148:133-168.

2115 Turner, A.H., Makovicky, P.J., and Norell, M.A., 2007. Feather quill knobs in the dinosaur

2116   Velociraptor. Science 317:1721-1721.

2117 Turner, W.C., Jolles, A.E., and Owen-Smith, N., 2005. Alternating sexual segregation during

2118   the mating season by male African buffalo (*Syncerus caffer*). Journal of

2119   Zoology 267:291-299.

2120 Vanpé, C., Gaillard, J.M., Kjellander, P., Mysterud, A., Magnien, P., Delorme, D., Laere, G.V.,

2121   Klein, F., Liberg, O., and Mark Hewison, A.J., 2007. Antler size provides an honest

2122   signal of male phenotypic quality in roe deer. The American Naturalist 169:481-493.

2123 Van Valkenburgh, B. and Sacco, T., 2002. Sexual dimorphism, social behavior, and intrasexual

2124   competition in large Pleistocene carnivorans. Journal of Vertebrate

2125   Paleontology 22:164-169.

2126 Varricchio, D.J., 2011. A distinct dinosaur life history? Historical Biology 23:91-107.

2127 Varricchio, D.J., Moore, J.R., Erickson, G.M., Norell, M.A., Jackson, F.D., and Borkowski,

2128   J.J., 2008. Avian paternal care had dinosaur origin. Science 322:1826-1828.

2129 Vinther, J., 2015. A guide to the field of palaeo colour. BioEssays 37:643-656.

2130 Vliet, K.A., 1989. Social displays of the American alligator (*Alligator*

2131   *mississippiensis*). American Zoologist 29:1019-1031.

2132 Waas, J.R., 1988. Acoustic displays facilitate courtship in little blue penguins, *Eudyptula*

2133   *minor*. Animal Behaviour 36:366-371.

2134 Wang, X., Kellner, A.W., Jiang, S., Wang, Q., Ma, Y., Paidoula, Y., Cheng, X., Rodrigues, T.,

2135   Meng, X., Weishampel, D.B., and Chapman, R.E., 1990. Morphometric study of

2136 Plateosaurus from Trossingen (Baden-Württemberg, Federal Republic of

2137 Germany). Dinosaur Systematics: Approaches and Perspectives:43-51.

2138 Waring, G.H., 1997. Preliminary study of the behavior and ecology of Jackson's chameleons

2139 of Maui, Hawaii. Report for USGS/BRD/PIERC Haleakala Field Station presented by

2140 Hawaii Ecosystems at Risk (HEAR) project. Located at: www. hear. org.

2141 Warner, R.R., 1975. The adaptive significance of sequential hermaphroditism in animals. The

2142 American Naturalist 109:61-82.

2143 Wasserstein, R.L. and Lazar, N.A., 2016. The ASA's statement on p-values: context, process,

2144 and purpose. The American Statistician 70:129-133.

2145 Wasserstein, R.L., Schirm, A.L., and Lazar, N.A., 2019. Moving to a world beyond "p < 0.05".

2146 American Statistician 73(Supplement 1):1-19.

2147 Weishampel, D.B., 1981. Acoustic analyses of potential vocalization in lambeosaurine

2148 dinosaurs (Reptilia: Ornithischia). Paleobiology 7:252-261.

2149 Weishampel, D.B. and Chapman, R.E., 1990. Morphometric study of Plateosaurus from

2150 Trossingen (Baden-Württemberg, Federal Republic of Germany). *Dinosaur*

2151 *Systematics: Approaches and Perspectives*:43-51.

2152 West, P.M. and Packer, C., 2002. Sexual selection, temperature, and the lion's

2153 mane. Science 297:1339-1343.

2154 Wheeler, P. and Greenwood, P.J., 1983. The evolution of reversed sexual dimorphism in birds

2155 of prey. Oikos 40:145-149.

2156 Wilkinson, P.M. and Rhodes, W.E., 1997. Growth rates of American alligators in coastal South

2157 Carolina. The Journal of Wildlife Management 61:397-402.

2158 Wilson V.J., 2001. Duikers of Africa: masters of the African forest floor: a study of Duikers-

2159 people-hunting and bushmeat. Zimbi books.Woodward, H.N., Freedman Fowler, E.A.,

2160 Farlow, J.O., and Horner, J.R., 2015. *Maiasaura*, a model organism for extinct

2161 vertebrate population biology: a large sample statistical assessment of growth dynamics

2162 and survivorship. Paleobiology 41:503-527.

2163 Zeigler, K.E., Lucas, S.G., and Heckert, A.B., 2002. The Late Triassic Canjilon quarry (Upper

2164 Chinle Group, New Mexico) phytosaur skulls: evidence of sexual dimorphism in

2165 phytosaurs. New Mexico Museum of Natural History and Science Bulletin 21:179-188.

2166 Zelenitsky, D.K., Therrien, F., Erickson, G.M., DeBuhr, C.L., Kobayashi, Y., Eberth, D.A.,

2167 and Hadfield, F., 2012. Feathered non-avian dinosaurs from North America provide

2168 insight into wing origins. Science 338:510-514.

2169 Zhang, J. and Li, N., 2014. Sexually dimorphic tridimensionally preserved pterosaurs and their

2170 eggs from China. Current Biology 24:1323-1330.

2171

2172 **Figure and Table Captions**

2173
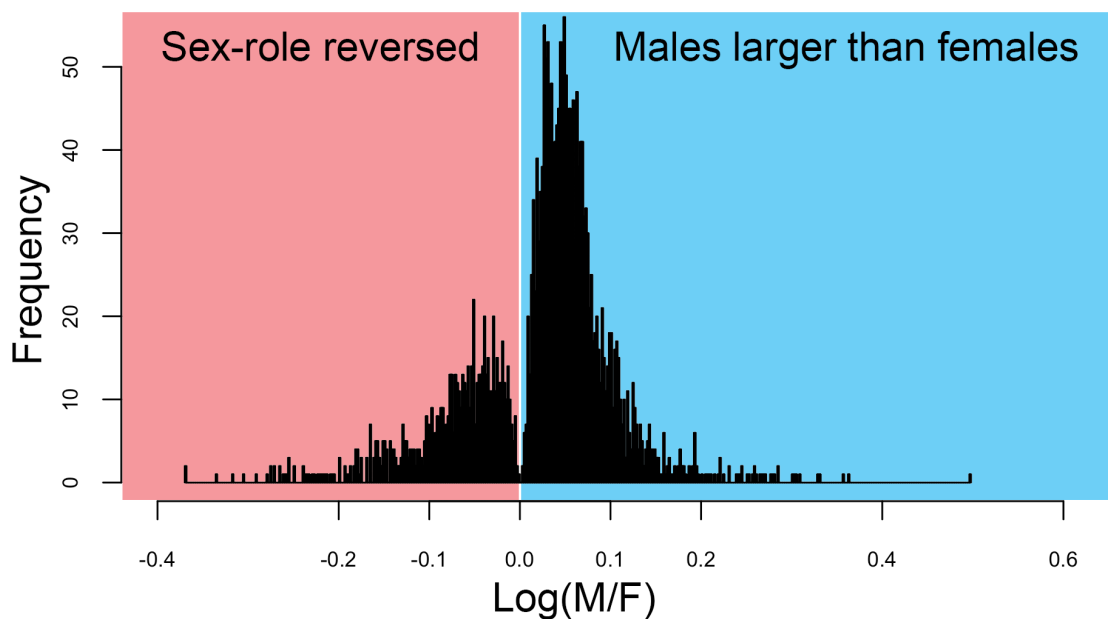
2174 Figure 1. Extant bird body mass sexual dimorphism. Data from Dunning (2007) and presented

2175 as average male body mass divided by average female body mass (M/F). Sample size = 2,576

2176 taxa. Sample sizes and data quality vary widely between taxa. All entries with both male and

2177 female values in Dunning (2007) were included.

2178
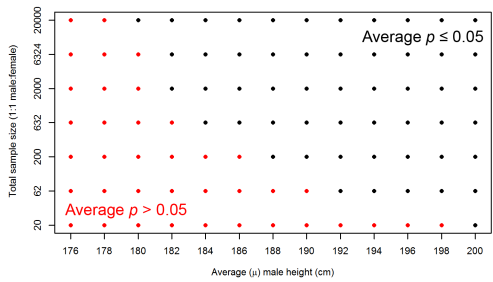
2179   Figure 2. Hartigans' dip testing in simulated human

2180   height dimorphism. Each point at a given magnitude of dimorphism and sample size represents

2181   the average *p*-value of 1,000 iterations of data simulation. Red indicates average *p*-values >

2182   0.05. Black indicates average p-values ≤ 0.05.

2183
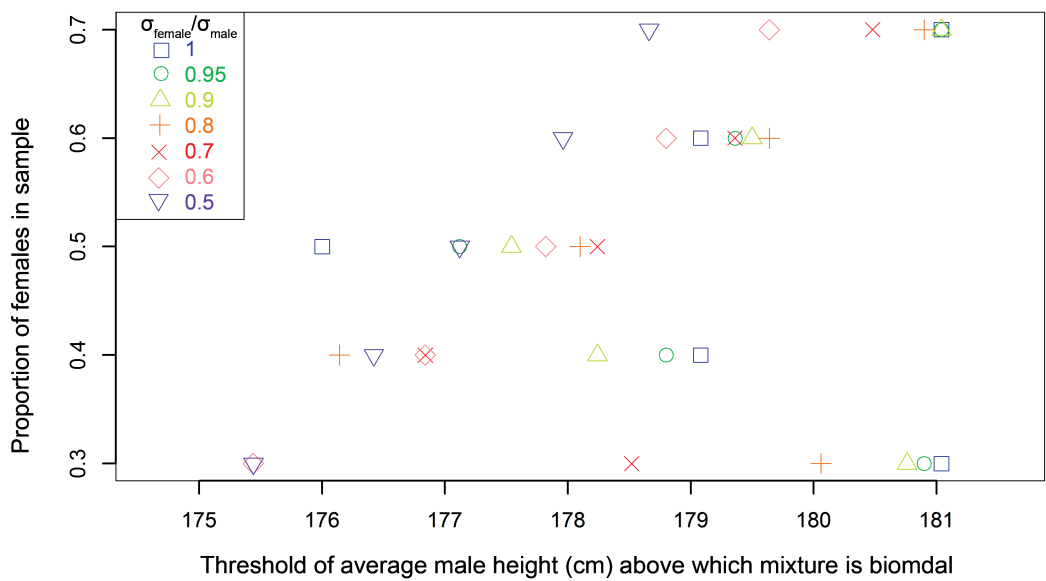
2184

2185   Figure 3. Average male height (μ) above which the theoretical combined (male and female)

2186   population distribution modelled in Figure 2 is bimodal, given certain proportions of females

2187   in the sample and ratios of female to male standard deviations ($\sigma_{female}/\sigma_{male}$). Data is converted

2188   from the values in table 2 of Schilling et al. (2001). Different symbols directly correlate with

2189   the color key and are used simply to show overlapping points.

2190
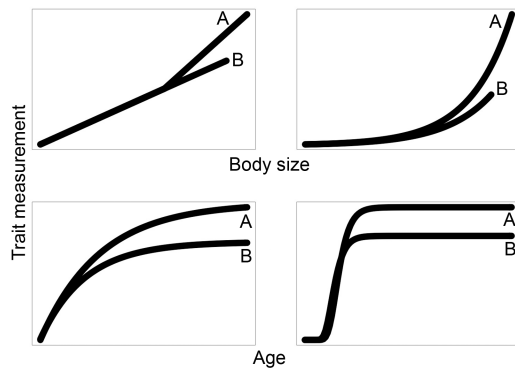
Figure 4. Hypothetical examples of growth curve divergence. Sex A and B are the sexes with larger and smaller trait measurements, respectively. Sex A is not necessarily male, but in many cases, this might be the expectation. Trait development can be examined according to age or a proxy for age (e.g., 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).
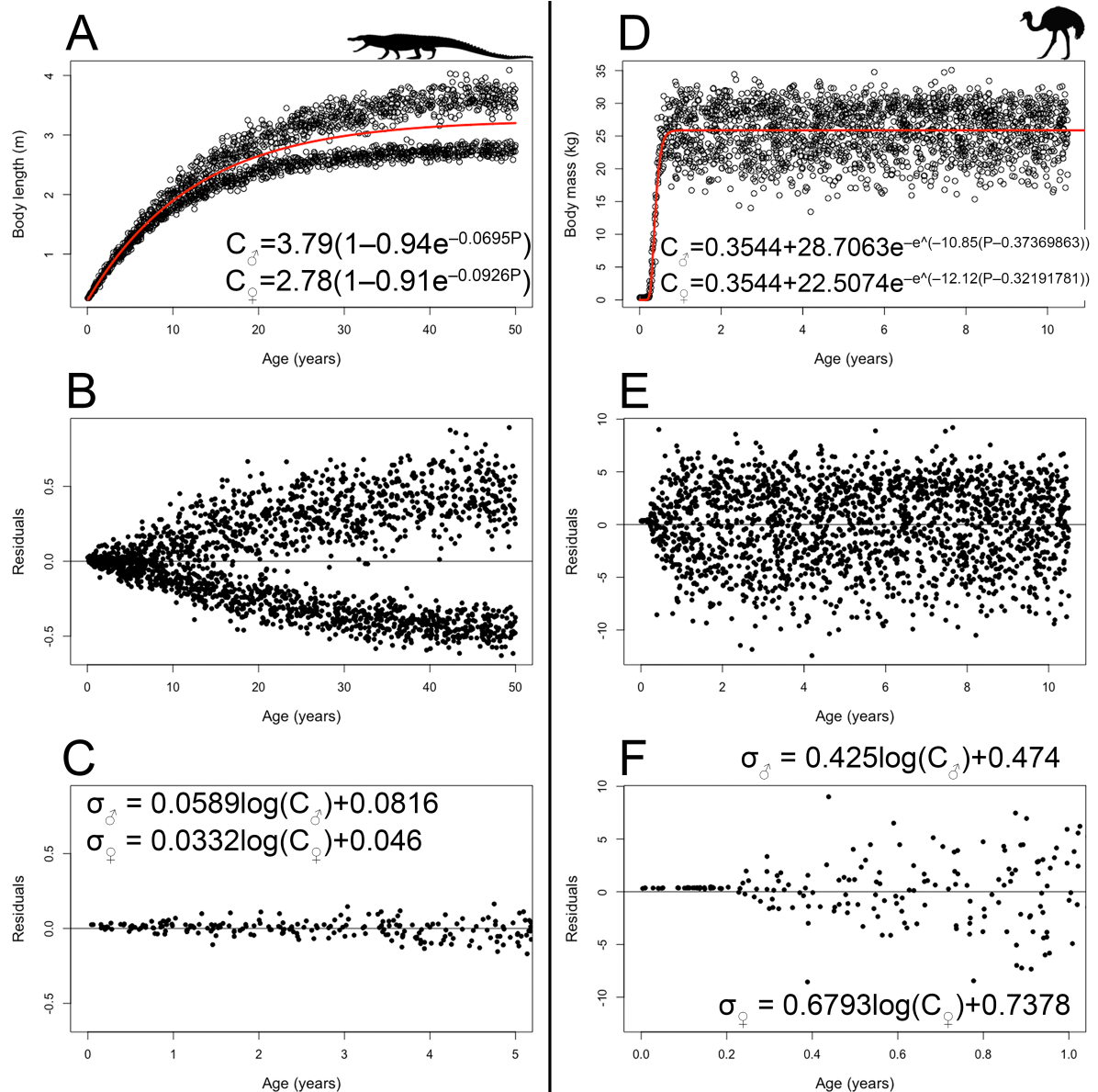
Figure 5. Preliminary visual examinations of potential growth curve divergence in simulated dimorphic male and female data (see code of Hone & Mallon 2017) based on empirical observations from Wilkinson and Rhodes (1997) for alligator (A–C) and Navarro et al. (2005) for rhea (D–F). Sex of individuals is unknown, as in fossil samples. 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.

2209
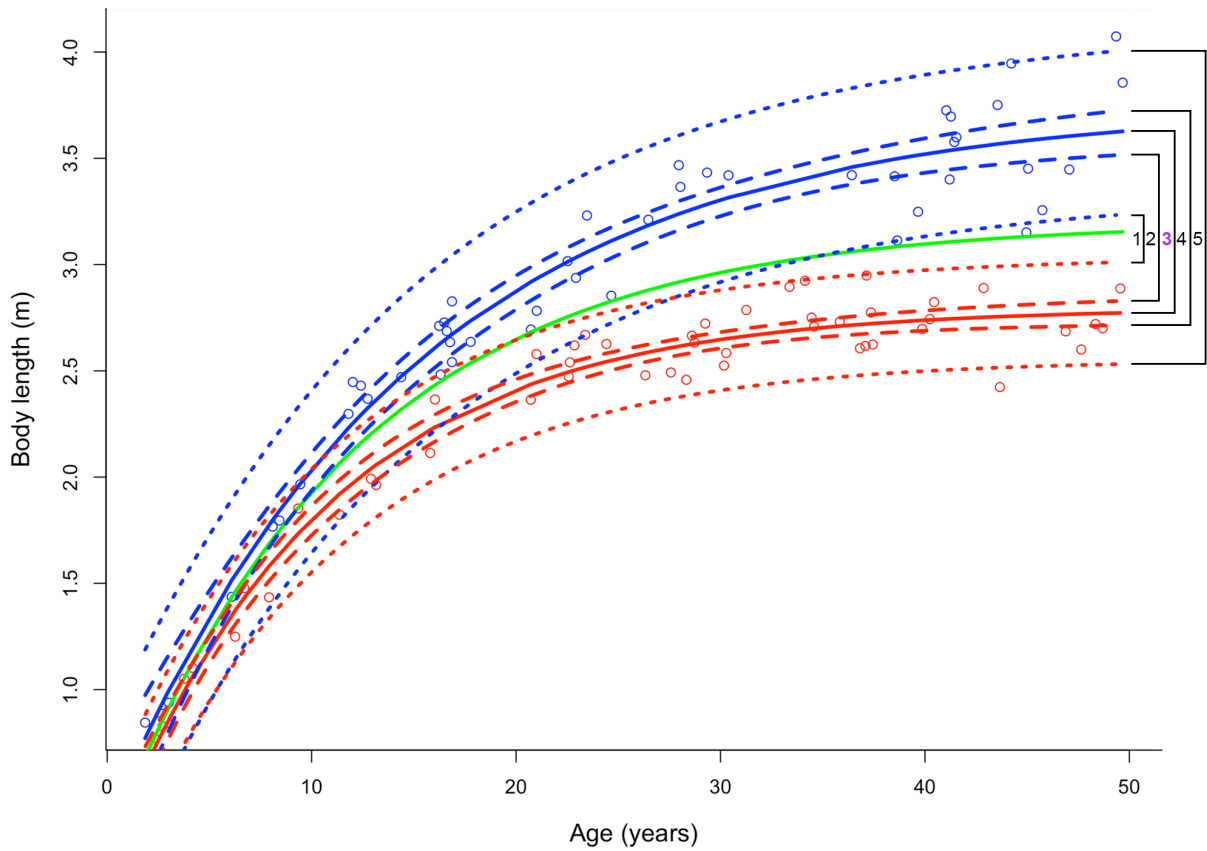
2210

2211   Figure 6. Summary of our statistical approach. Alligator data simulated with functions for male

2212   and female population standard deviations of the residuals kept equal to each other and effect

2213   size equal to that of natural populations (Wilkinson & Rhodes 1997). A curve (green) is fit to

2214   the whole dataset. Points with positive (blue) and negative (red) residuals are assigned to

2215   opposite sexes. Sex specific curves are fit (3) to estimate effect size, E_E, measured as $L_{\infty M}$-$L_{\infty F}$

2216   for von Bertalanffy curves and L_M-L_F for Gompertz curves. 95% confidence (thick dashes) and

2217   prediction (thin dashes) intervals are calculated for each sex. Separation of prediction

2218   (1)/confidence (2) intervals and spread of confidence (4)/prediction (5) intervals can be

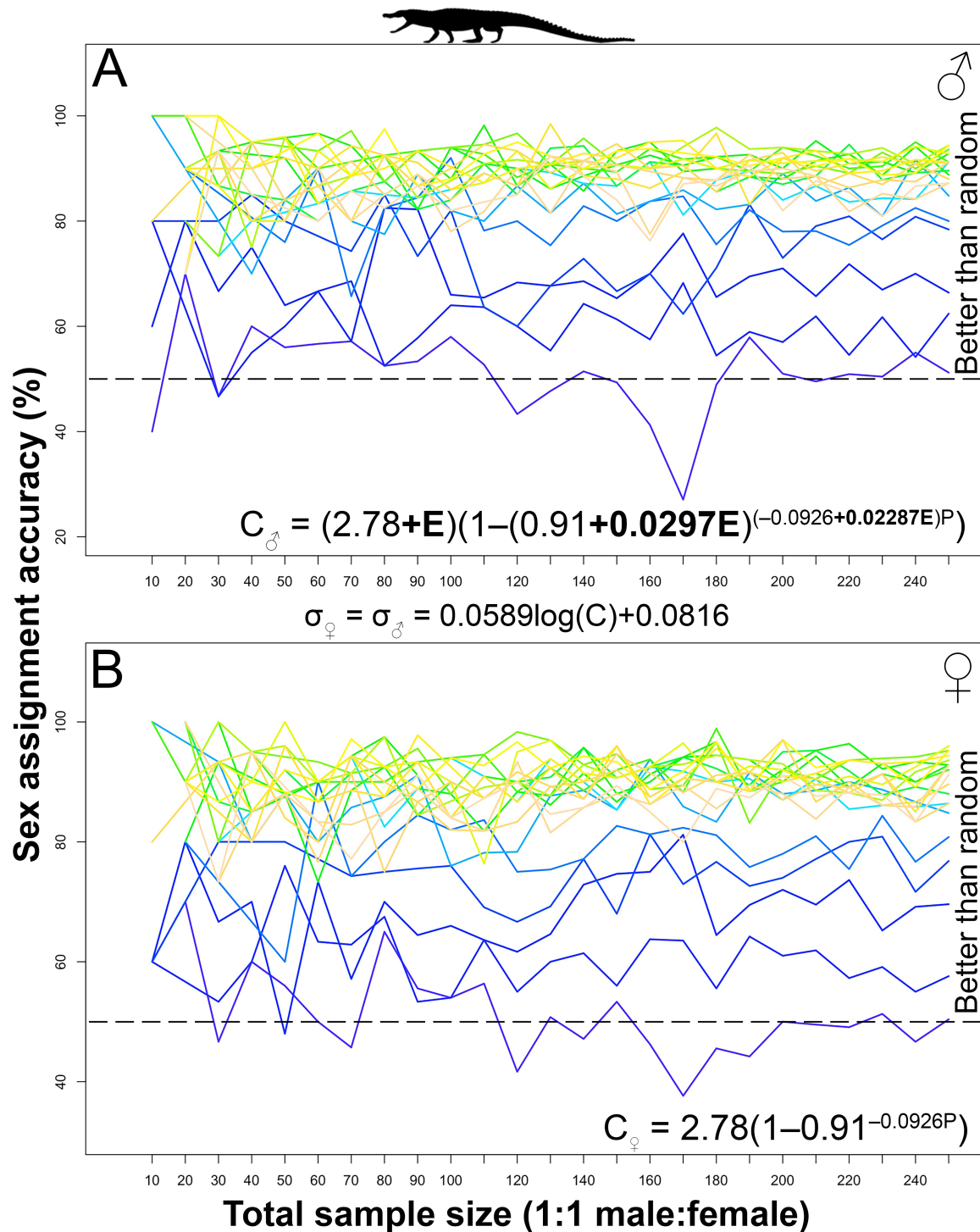2219   calculated at the largest individuals of each assigned sex.

2220
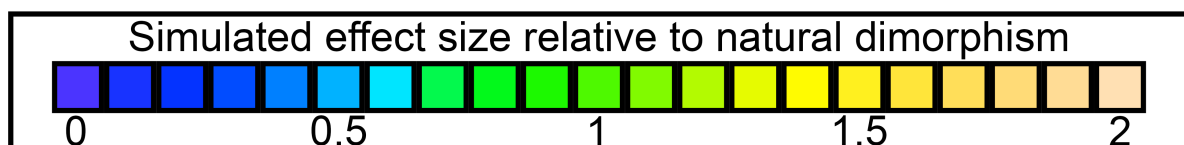
Figure 7. Sex assignment accuracy for simulated male (A) and female (B) alligator data. Dashed line indicates an accuracy of 50%. Color coding represents the true effect size used to generate the data divided by observed empirical dimorphism (Wilkinson & Rhodes 1997): zero

2225   is monomorphism (true effect size, $E = 0$), one is dimorphism matching natural dimorphism (E

2226 $= 1.01$), and two is dimorphism twice as great as natural dimorphism ($E = 2.02$). E is applied

2227   proportionally to each growth curve parameter based on natural dimorphism. Function for

2228   population standard deviation of the residuals is kept equal between males and females, based

2229   on the natural values of the sex with greater intrasexual variation. These results are from a

2230   single run of our code. Each time the code is run, precise values will vary while overall trends
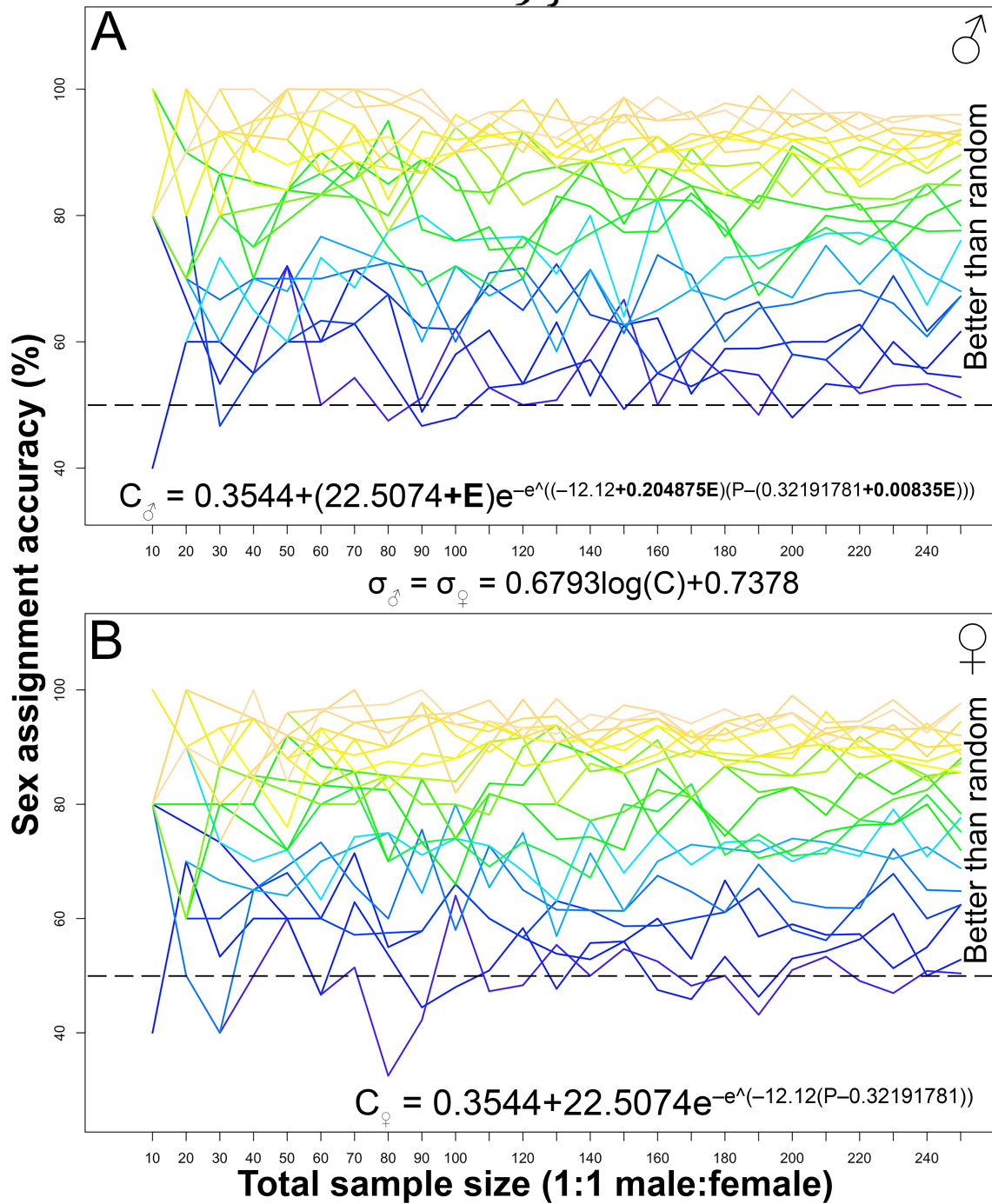
2231   persist.

2232

Figure 8. Sex assignment accuracy for simulated male (A) and female (B) rhea data. See Figure 7 legend for further details. Observed empirical data from Navarro et al. (2005). For effect size ratio of one, $E = 6.199$. For effect size ratio of two, $E = 12.398$.
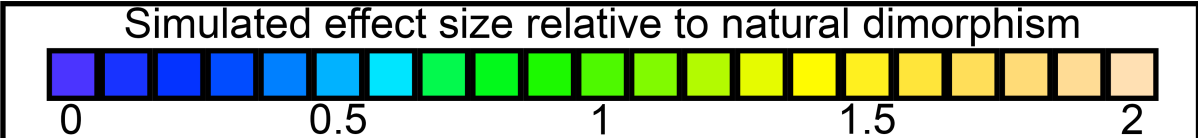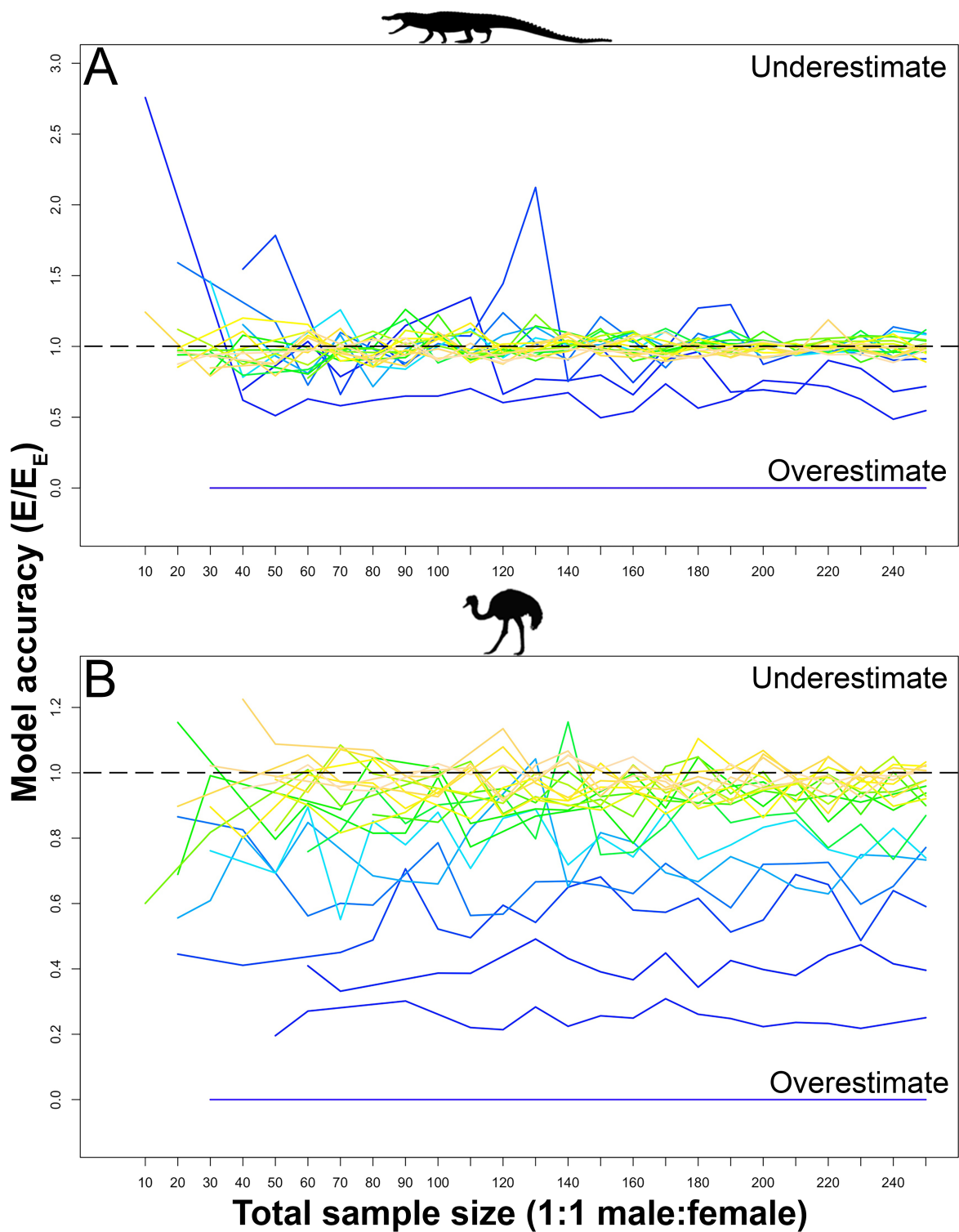
Figure 9. Fitting dimorphic models to simulated alligator (A) and rhea (B) data. Accuracy

measured as true simulated effect size, *E*, divided by estimated effect size, E_E. Dashed line

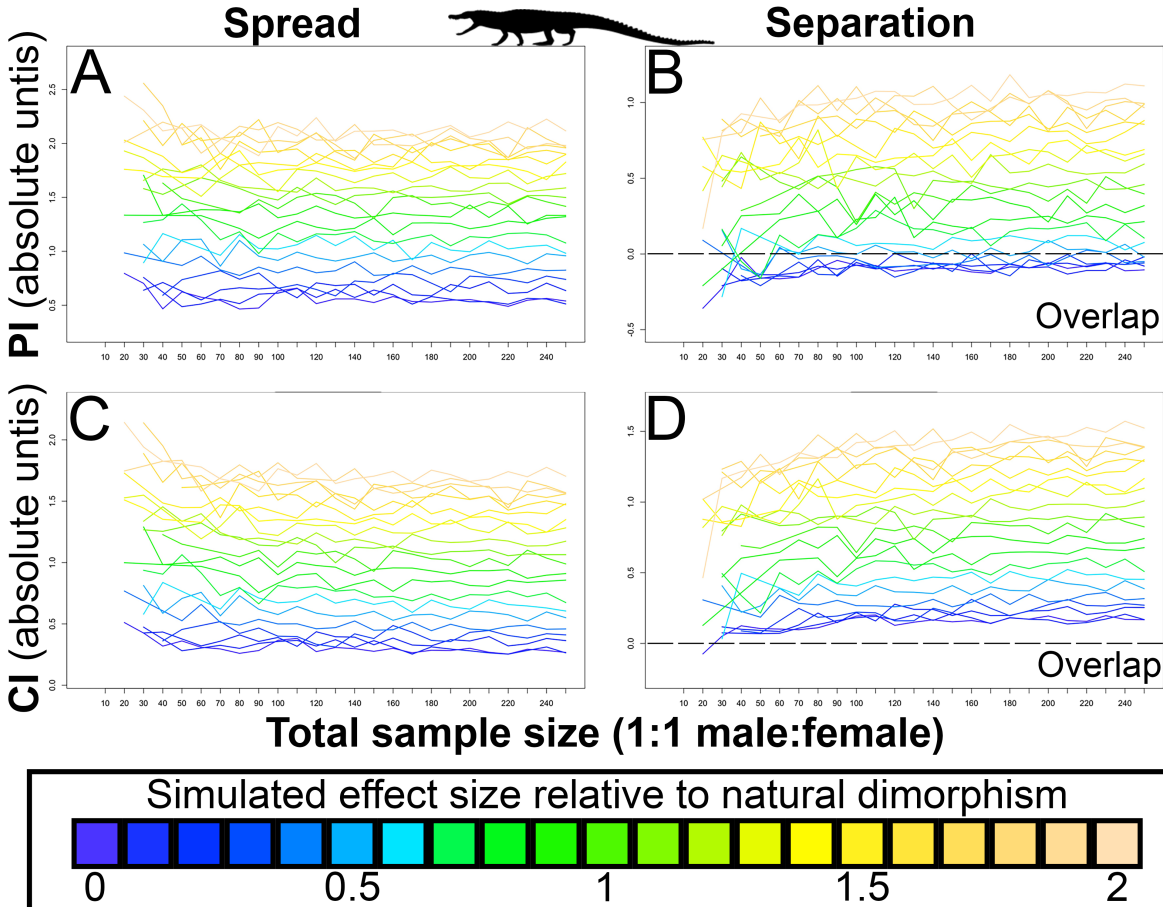2241   indicates a ratio of one ($E = E_E$). See Figure 7 legend for further details.

2242

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.
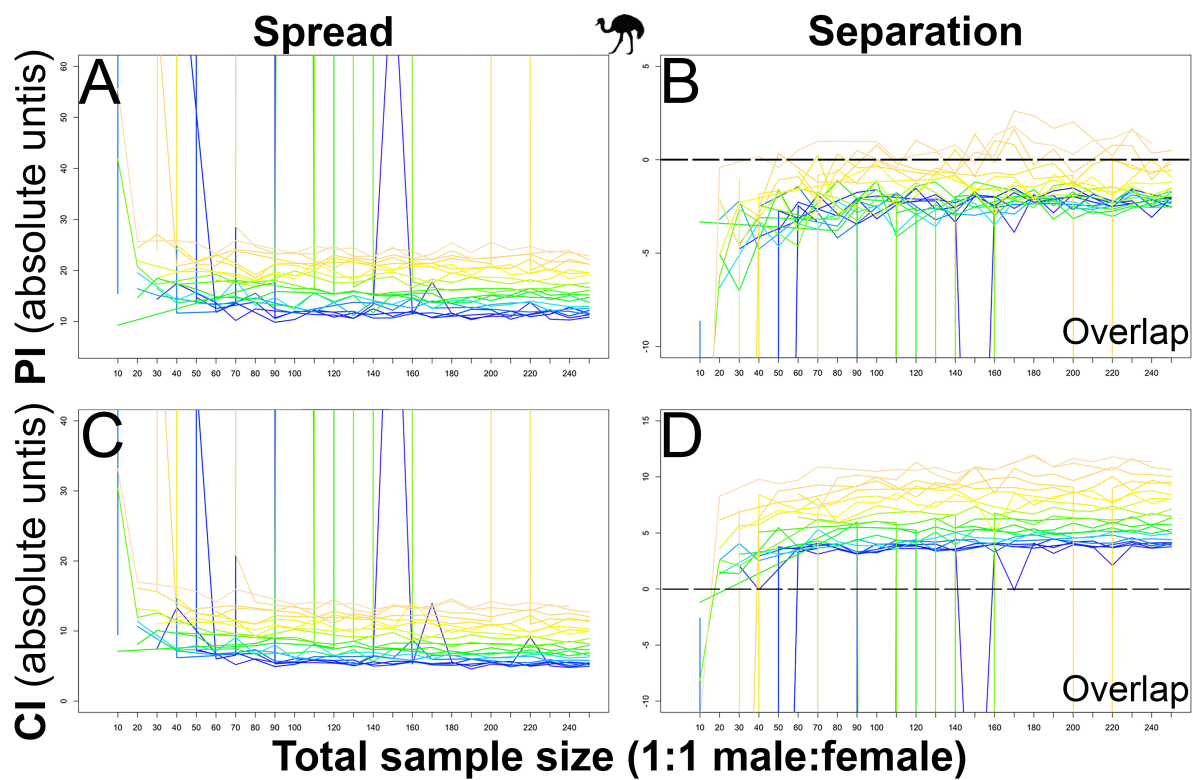
2247

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: m) for interval separation. See Figure 7 legend for further details.
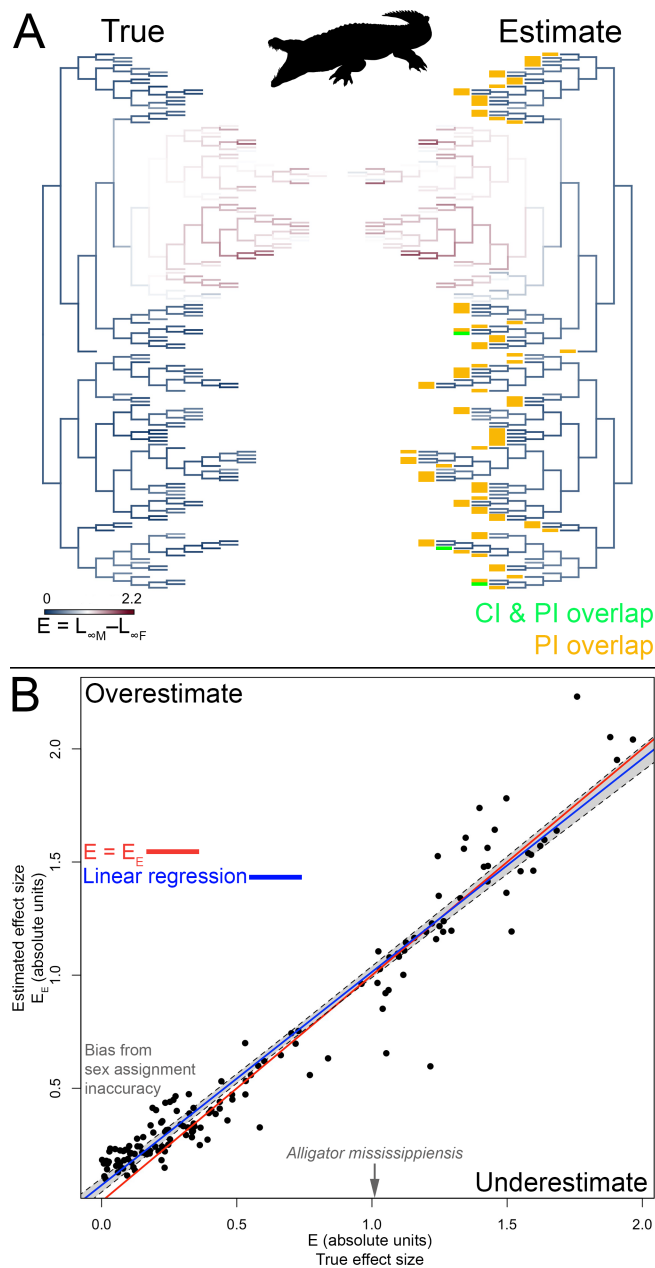
Figure 12. Phylogenetic simulation of a hypothetical crocodilian clade. Each taxon randomly and independently varies in true effect size ($\sim$0.00 $\leq E \leq \sim$1.97), sample size (16–99), true sex ratio ($\sim$25–74% males), male standard deviation of the residuals ($\sim$0.63 $\leq d \leq \sim$1.51), and female standard deviation of the residuals ($\sim$0.55 $\leq d \leq \sim$1.51), where d is multiplied to the average of natural male and female alligator functions for standard deviation of residuals (0.04605log(C)+0.0638) such that both simulated sexes cover the natural intrasexual variation in alligator and also extend beyond it. A, the phylogeny with color-coded true (left) and estimated (right), E_E, effect size (from dark blue at low effect size, to white, to dark red at high effect size) and interval overlap (confidence and

2262   prediction in green, prediction only in orange). Sex assigned by the residual method to yield

2263 E_E. Interval spread/separation measured at the largest individual of each assigned sex. B, plot

2264   of E versus E_E of each taxon. Perfect estimation in red. Linear regression in blue with 95%

2265   confidence intervals (black dashes). Natural alligator $E = 1.01$ (Wilkinson & Rhodes 1997).
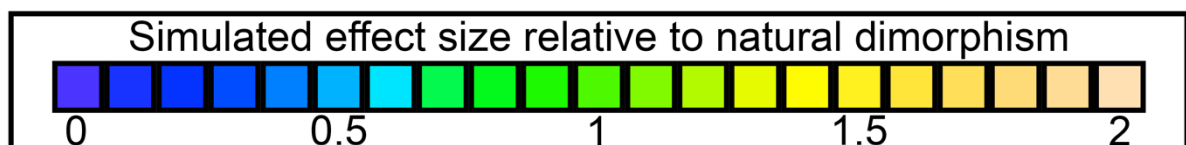
2266

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, 9–10.
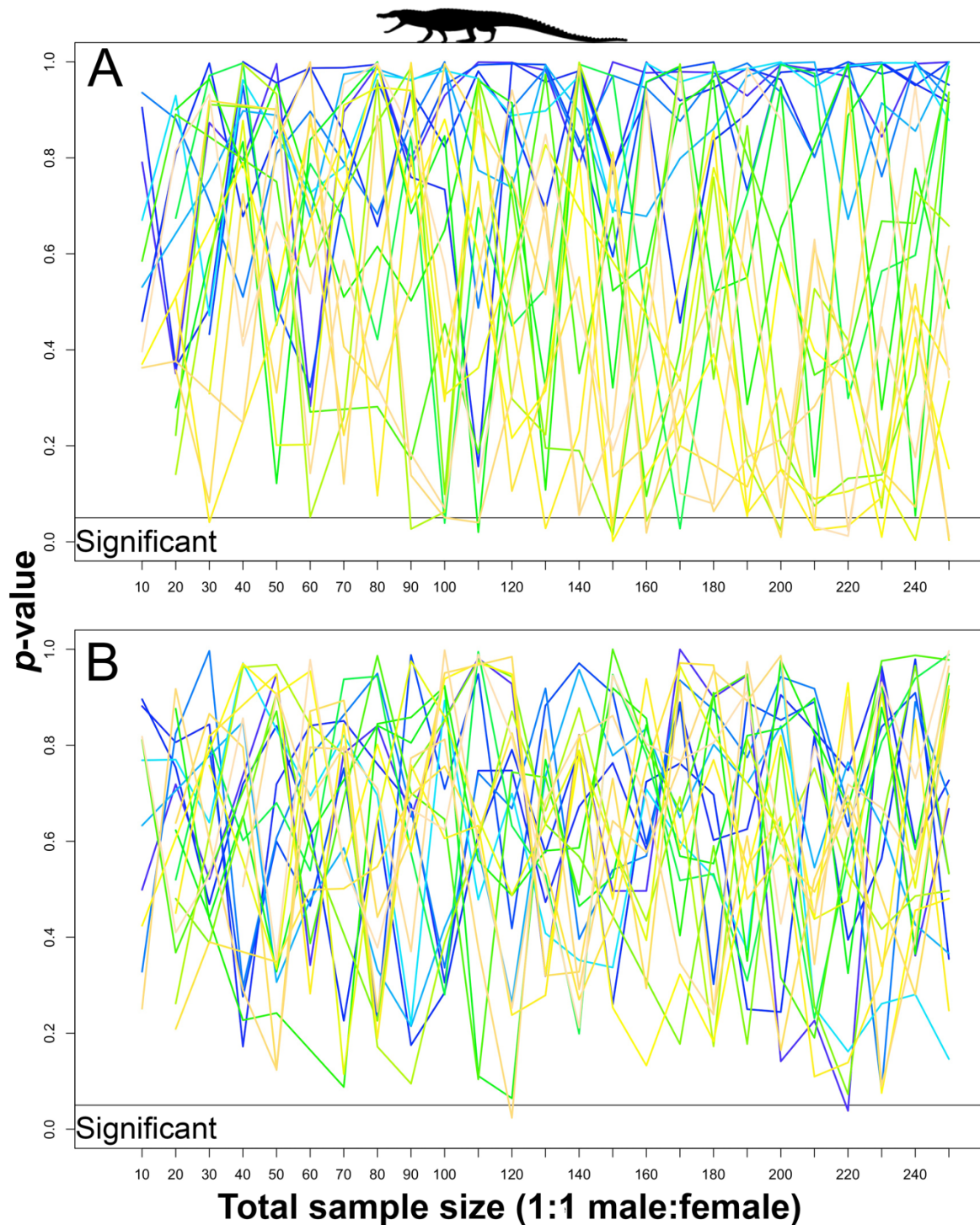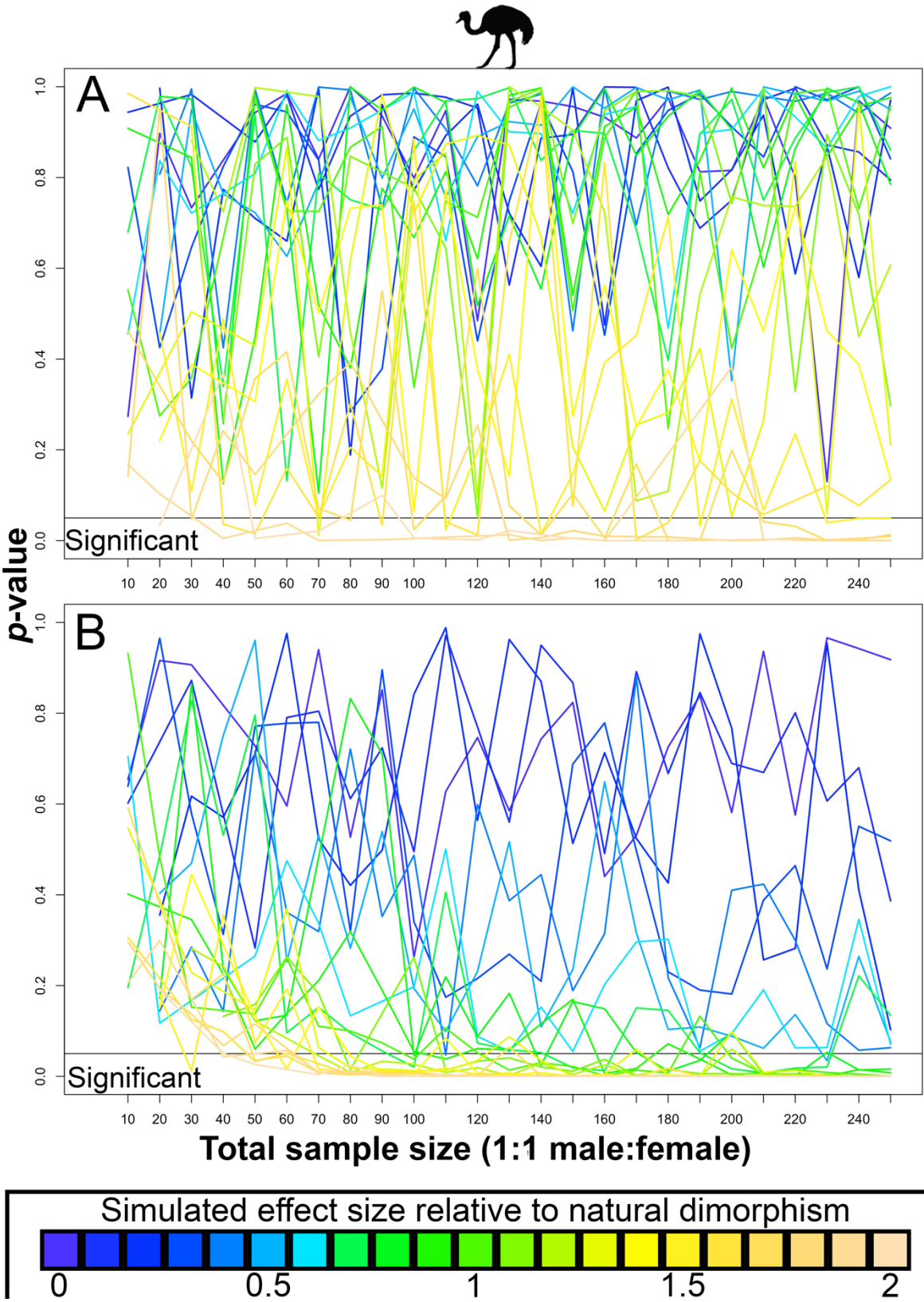
Figure 14. Hartigans' dip tests for unimodality (A) and *t*-tests (B) using the residual sex

2274   assignment method on simulated rhea data. Black line indicates $\alpha = 0.05$. See Figure 7 legend

2275   for further details. This run of the code is different from the run displayed in Figures 8–9, 11.

2276

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

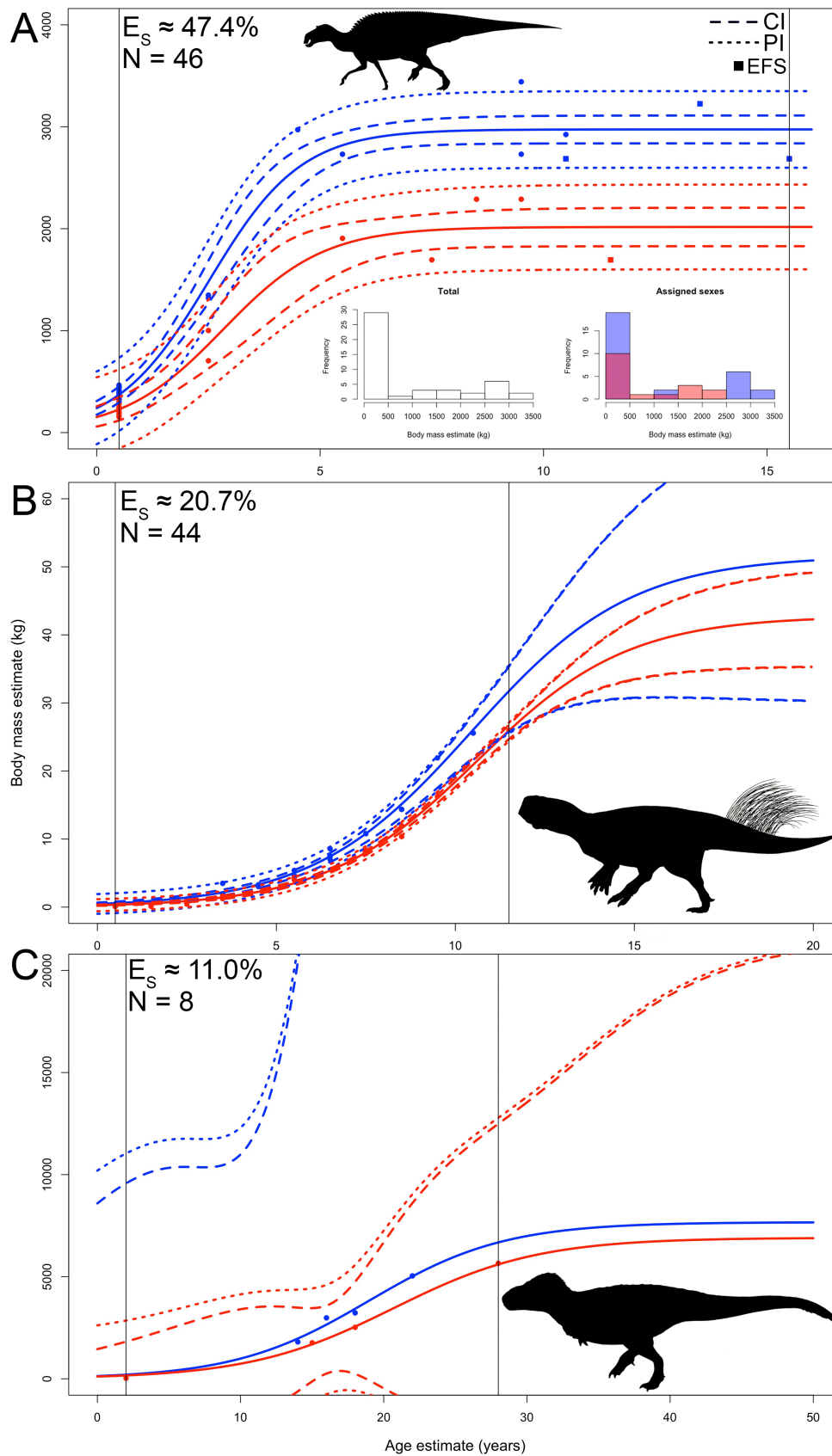2280 et al. 2015), and (C) *Tyrannosaurus rex* (Horner & Padian 2004; Erickson et al. 2004; Lee &

2281 Werning 2008). Sex assigned by residual method. 95% confidence (CI) and prediction (PI)

2282 intervals shown. N = sample size. E_S = standardized effect size (i.e., percent change in

2283 asymptote L parameter from smaller to larger assigned sex). Vertical lines delineate youngest

2284 and oldest specimens.

2285

	2008	2009	2010	2011	2012	2013	2014
% conclusive abstracts — Non-avian dinosaurs	100	NA	33	0	0	0	NA
% conclusive abstracts — Other taxa	60	64	69	74	73	94	65
Expected count other (anatomical)	6	7	8	12	6	13	11
Expected count non- (anatomical)	1	2	2	2	2	3	2
Observed other taxa (Abstracts more conclusively implying anatomical)	6	9	9	14	8	16	13
Observed non-avian (Abstracts more conclusively implying anatomical)	1	0	1	0	0	0	0
Observed other taxa (Abstracts implying anatomical)	10	14	13	19	11	17	20
Observed non-avian (Abstracts implying anatomical)	1	0	3	3	1	2	0
Expected count other (anatomical)	6	7	8	12	6	13	11
Expected count non- (secondary anatomical)	2	2	3	4	2	4	4
Expected count other (secondary sexual)	9	12	13	18	10	15	16
Total abstracts	601	750	695	774	628	777	889
Abstracts focusing on non-avian dinosaurs	113	133	126	136	127	151	161
Year	2008	2009	2010	2011	2012	2013	2014

	2015	2016	2017	2018	Total
NA	NA	0	0	0	13
	50	82	85	91	75
	0	0	0	0	2
	4	23	11	10	123
	1	5	2	2	24
	3	18	9	8	101
	0	2	1	2	15
	8	28	13	11	164
	1	6	3	3	34
	7	24	11	10	145
	774	779	649	744	8060
	145	156	136	161	1545

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 χ^2 goodness of fit tests. Each abstract considered an independent observation regardless of repeated authorship.

Growth model
Logistic
Gompertz
von Bertalanffy

Table 2 (rotated):

Taxon	Trait	Parameters	Monomorphic model	Dimorphic model	Effect size
Maiasaura, *Psittacosaurus*, *Tyrannosaurus*	Body mass	L, q, k	$C = L/(1+e^{q+kP})$	$C = SL_M/(1+e^{(q_M+k_MP)}) + (1-S)L_F/(1+e^{(q_F+k_FP)})$	L_M-L_F
Rhea	Body mass	L, k, i	$C = Le^{-e^{(-k(P-i))}}$	$C = SL_Me^{(-e^{(-k_M(P-i_M))})} + (1-S)L_Fe^{(-e^{(-k_F(P-\dots))})}$	L_M-L_F
Alligator	Body length	L_∞, K, t	$C = L_\infty(1-e^{-K(P-t)})$	$C = SL_{\infty M}(1-e^{(-K_M(P-t_M))}) + (1-S)L_{\infty F}(1-e^{(-\dots)})$	$L_{\infty M}-L_{\infty F}$

2296 Table 2. General description of growth models. *C* = trait measured as potentially sexually

2297 variable. *P* = age or proxy for age (e.g., size). M subscript indicates sex with more trait

2298 development; F subscript indicates sex with less trait development. *S* = sex (e.g., either 1 [for

2299 M] or 0 [for F]).

2300

Second table (rotated):

Taxon	Data simulation				Curve for mean	residual-based sex assignment	Sex
	Female	Male	Rhea	Stat			
Rhea					$C = Le^{-e^{(-k(P-i))}}$	$C = Le^{-e^{(-k(P-i))}}$	L
Alligator					$C = L_\infty(1-e^{-K(P-t)})$	$C = L_\infty(1-e^{-K(P-t)})$	L

2301   Table 3. Workflow of data simulation and analysis in Figures 7–11, 13–14. *E* = effect size. *E*

2302   applied proportionally to each parameter based on natural dimorphism for that species to avoid

2303   unrealistically shaped curves. See Table 2 legend for further details. *Function for the standard

2304   deviation of the residuals of each sex independently varied between each hypothetical taxa in

2305   the phylogenetic simulation (Figure 12), set to sex-specific natural values in Figure 5, and

2306   doubled for each sex a simulation in the appendix.