On the Evolution of Extreme Structures: Static Scaling and the Function of Sexually Selected Signals

Devin M. O'Brien[1]*, Cerisse E. Allen[1], Melissa J. Van Kleeck[2], David Hone[3], Robert Knell[3], Andrew Knapp[3], Stuart Christiansen[4], Douglas J. Emlen[1]

[1] Division of Biological Sciences, University of Montana, Heath Sciences 104, 32 Campus Drive, Missoula, MT, 59812, USA

[2] Department of Biology, University of Hawaiʻi at Mānoa, 2538 McCarthy Mall, Edmondson Hall 216, Honolulu, HI, 96822, USA

[3] School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK

[4] Sentinel High School, 901 South Ave W, Missoula, MT, 59801, USA

*Corresponding author:

Devin M. O'Brien
Heath Sciences 104
32 Campus Drive
Missoula, MT 59812
devin.m.obrien@gmail.com
(518) 366 2831

Running head: Static Scaling of Extreme Structures

Keywords: Sexual selection, Scaling, Animal signals, Fossils

Word count (total): 10,608
Word count (text body): 5,410

Introduction

Understanding how morphology scales with body size is one of the most pervasive topics in organismal biology (Dial, Greene, & Irschick, 2008; Gould, 1966, 1974b, 1974a; J. Huxley, 1932; Schmidt-Nielsen, 1984; Templeton, Greene, & Davis, 2005; Thompson, 1917; Voje, 2016; West & Brown, 2005; West, Brown, & Enquist, 1997). The reason for this is simple - virtually every measurable aspect of an organism scales with body size. Some relationships hold across hundreds of species, spanning multiple orders of magnitude in overall size (e.g., Kleiber's Law (Kleiber, 1932); Rubner's Surface Rule (Rubner, 1883; Von Bertalanffy, 1957); Cope's Rule (Stanley, 1973); Rensch's Rule (Abouheif & Fairbairn, 1997; Wolf U. Blanckenhorn, Meier, & Teder, 2007; Fairbairn, 1997)). Others account for transformations in shape arising during ontogeny (e.g., brain/body weight (Cock, 1966; Gould, 1974a, 1977); Dyar's Law (Dyar, 1890)). Here we focus on 'static' allometry, scaling that occurs among individuals of the same age sampled from within populations (Cheverud, 1982; sensu Cock, 1966; Pélabon et al., 2013).

Perhaps the most striking pattern in the study of static scaling is the observation that many extreme products of sexual selection – ornaments of choice and weapons of intrasexual competition – scale steeply with body size (Bonduriansky & Day, 2003; Eberhard, 1998; Egset et al., 2012; Emlen, 1996; Emlen & Allen, 2003; Fromhage & Kokko, 2014; Gould, 1974b; Hongo, 2007; Kelly, 2005; Kodric-Brown, Sibly, & Brown, 2006; Miller & Emlen, 2010; Painting & Holwell, 2013; Shingleton, Frankino, Flatt, Nijhout, & Emlen, 2007; Shingleton, Mirth, & Bates, 2008; L. W. Simmons & Tomkins, 1996; Stern & Emlen, 1999; Voje, 2016; Wilkinson, 1993). Specifically, when examined on a log scale, the relationship between the size of these structures and body size is greater than one ('positive allometry') (Gould, 1966; J. S. Huxley & Teissier, 1936; Kerkhoff & Enquist, 2009; Shingleton & Frankino, 2013; Voje, 2016). These steep scaling relationships cause ornaments and weapons to attain extraordinary proportions in the largest individuals, inspiring descriptions such as 'extreme', 'exaggerated' (Darwin, 1871) and 'bizarre' (Gould, 1974b) (Fig. 1).

Early studies of static scaling often focused on the products of sexual selection, including cervid antlers (Gould, 1973; J. Huxley, 1932; Thompson, 1917), fiddler crab (*Uca*) chelae (J. Huxley, 1932), and beetle (Scarabaeidae) horns (Bateson & Brindley, 1892; Paulian, 1935). Since then, hundreds of sexually selected structures have been examined, and the overwhelming majority scale steeply with body size (Emlen, 2008; Emlen & Nijhout, 2000; Knell, Naish, Tomkins, & Hone, 2013b; Kodric-Brown et al., 2006; e.g., Otte & Stayman, 1979; Petrie, 1988, 1992; Voje, 2016). In fact, the link between steep scaling and exaggerated ornaments and weapons is so widespread that many consider the steepness of static allometry indicative of the intensity of sexual selection acting on a structure (e.g., stalk-eyed fly (Diopsidae) eyestalks (Baker & Wilkinson, 2001); frog (Anura) forelimbs (Schulte-Hostedde, Kuula, Martin, Schank, & Lesbarrères, 2011); earwig forceps (L. W. Simmons & Tomkins, 1996)), and testing of this 'positive allometry' hypothesis is frequently used to infer a sexual selection function when natural observation is unattainable (e.g., trilobite spines (Knell & Fortey, 2005)).

The positive allometry hypothesis has, however, been met with resistance. Bonduriansky (2007) noted that the near universality of this pattern may be an artefact of the structures researchers elect to study. That is, when studies focus on morphological scaling, scientists seek the extremes, so the literature is biased in favour of steep scaling relationships (Emlen, 2008; Emlen & Nijhout, 2000; Kodric-Brown et al., 2006). Some extreme structures known to function as sexually selected ornaments, such as elaborate plumage in birds, do not scale positively with body size (José Javier Cuervo & Møller, 2001), nor do many genitalic traits, despite the fact that some experience strong selection for increased size (Bertin & Fairbairn, 2007; W. U. Blanckenhorn, Kraushaar, Teuschl, & Reim, 2004; Voje, 2016). Indeed, considering the full range of sexually selected structures, including those that are not extreme in size, reveals that slopes are frequently shallow or negative (Bonduriansky, 2007).

Furthermore, at least a few naturally selected structures, such as long bones in large mammals (Bertram & Biewener, 1990; Christiansen, 1999) and cranial horns in lizards (Bergmann & Berk, 2012), also scale positively with body size (Voje, 2016). Clearly, sexual selection need not lead to the evolution of steep scaling, and other agents of selection, such as locomotion and predator defence, occasionally lead to positive static scaling. Where, then, does this leave the positive allometry hypothesis?

We argue that steep static scaling relationship slopes can be powerful clues to trait function, particularly when combined with other morphological measures of among-individual variation (e.g., trait-specific coefficients of variation; see below). In this context, we suggest much of the controversy and inconsistency in the literature stems from two sources. First, the positive allometry hypothesis has been applied to all sexually selected structures, when, in fact, the logic holds only for a particular subset: sexually selected signal structures where the size of the structure functions as an honest signal of the body size *or* resource holding potential of their bearers. Second, tests of the positive allometry hypothesis often rely on demonstrating a slope significantly greater than one. While rich in historical precedent, this approach fails to incorporate the signalling function of these structures. We propose future studies ask not whether the slope is greater than one, but rather whether the slope is relatively steeper for the focal signal structure than it is for other, more typically proportioned, non-signal related body parts. It is the relative increase in slope that allows these structures to function effectively as signals, and appropriate tests should incorporate this into their methods.

We summarize literature on animal signalling to show why positive allometry is likely when structures evolve as signals of body size, and why these structures are predicted to scale more steeply with body size than other, non-signal structures measured in the same individuals. By the same logic, we explain why other types of extreme structures, such as those used in prey capture or locomotion, should not scale more steeply than other body parts.

We test these predictions by comparing the slopes of a suite of extreme morphological structures (14 signal, 15 non-signal; Table 1) to slopes of more typically proportioned 'reference' structures within the same organism (rather than the traditional comparison to isometry, see below), and show that relatively steep slopes are common for structures that function as sexually selected signals but not for comparably extreme structures that function in other, non-signalling contexts.

Methods

Specimen/structure selection and morphological measures

All species with putatively 'extreme' structures – hereafter referred to as 'focal structures' (see Appendix 1 for our classification of 'extreme') – and adequate sample size (n ≈ 10) were surveyed from the Phillip L. Wright Zoological Museum at the University of Montana (MT, USA), the Museum of Comparative Zoology at Harvard (MA, USA), and the Emlen Lab Entomological Collection (MT, USA). Surveying all species that met these criteria allowed for a relatively unbiased sample of both taxa and structure type. However, since most sexually selected structures in insects are beetle horns (reviewed in Emlen, 2008), the invertebrates surveyed here appear somewhat Coleoptera-biased. Six additional datasets were sourced specifically for this analysis – Jackson's chameleons (*Triceros jacksonii*) for the presence of both an extreme signal (horns) and non-signal (tongue) structure, large bee flies (*Bombylius major*), sabre wasps (*Rhyssa persuasoria*), and peacock moths (*Saturnia pyri*), for the presence of sexually selected non-signal structures, and ceratopsids (*Protoceratops andrewsi*) and pterosaurs (*Rhamphorhynchus muensteri*) to test the described methods on fossil datasets. Finally, it should be noted that while the species/structures surveyed here were unbiased relative to the sampled collections, the collections may have been biased either in taxa or in favour of particularly exaggerated structures. If true, then the results presented here, and their interpretation, may be limited to a particular subset of extreme morphology.

Focal structures of extant species were categorized as a 'sexually selected signals'(i.e., structures used by potential mates *or* competitive rivals as visual signals of the bearer's overall condition/quality (M. B. Andersson, 1994; Bradbury & Vehrencamp, 1998)) or 'non-signal' structures using relevant behavioural studies from the literature (see Table 1). When literature on the focal species was unavailable, studies in closely related species were used to infer trait function. Bill function in the American pelican (*Pelecanus erythrorphychos*) was inferred from its sister species, *P. occidentalis* (Bels et al., 2012; Kennedy, Taylor, Nádvorník, & Spencer, 2013; Orians, 1969; Schreiber, Woolfenden, & Curtsinger, 1975). Lantern function in the Malagasy lantern bug (*Zanna madagascariensis*) was inferred from several other Fulgoridae species with similar head morphology (Hogue, 1984; Urban & Cryan, 2009). Snout function in the elephant shrew (*Elephantulus fuscus*) was inferred from two species of the same genus with similar rostral morphology, *E. brachyrhynchus* and *E. myurus* (Kingdon, 1974; Kratzing & Woodall, 1988). Horn function in dung beetles (*Sulcophanaeus menelas*, *Phanaeus saphirinus*, *Othophagus lanista*) was inferred from both a comprehensive review of horn function in beetles (Eberhard, 1980) and empirical studies of dung beetle mating systems (e.g., Emlen, Marangelo, Ball, & Cunningham, 2005; Moczek & Emlen, 2000). Hindleg function in frog legged beetles (*Sagra buqueti*) was inferred from a closely related species with similar leg morphology and mating behaviour (Katsuki, Yokoi, Funakoshi, & Oota, 2014; O'Brien, Katsuki, & Emlen, 2017). Finally, the function of focal traits in extinct species were inferred from key publications focused on 'bizarre' morphology in the fossil record (Knell & Sampson, 2011; Knell, Naish, Tomkins, & Hone, 2013a; D. W. Hone, Wood, & Knell, 2016; but see Padian & Horner, 2011, 2013, 2014).

Reference structures were then chosen for each species as structures that could be consistently measured across all samples and lacked obvious functional connection with the focal structure. These criteria appear adequate in choosing reference structures. However, the authors recognize the limitation of using a single reference structure and encourage the use of multiple reference structures per organism in future application of the described methods. Doing so will better capture the scaling relationship of 'typical' (i.e., non-signal) traits and help mitigate impact of choosing inappropriate reference structures.

Measures of overall body size were based on established, taxon specific methods for estimating body size. For species where established estimates of body size were unavailable, methods were adopted from closely related taxa. A summary of study species names, sample sizes, relevant morphological information (e.g., focal structure, reference structure, body size measures), and literature used to establish sexually selected signal/naturally selected non-signal function is provided in Table 1.

Dung beetles (*Sulcophanaeus menelas*), earwigs, mantidflies (*Climaciella brunnea*), large bee flies, sabre wasps, and wildebeest (*Connochaetes tourinus*) were measured using photographs (including scale bars) and ImageJ 1.50i software (NIH, USA). *S. menelas*, earwigs, and mantidflies, large bee flies, and sabre wasps were photographed using a 16.2 megapixel Nikon D5100 DSLR camera mounted on a binocular stereo microscope (Leica S6D) set at a fixed distance. Wildebeest were photographed using a 14.2 megapixel Nikon D3100 DSLR camera set at a fixed distance designated to minimize perspective effects (i.e., approximating orthographic projection). All other extant species were measured using digital callipers.

Measures of ceratopsians (*Protoceratops andrewsi*) and pterosaurs (*Rhamphorhynchus muensteri*) were collected directly using digital callipers, from photographs of specimens including scale bars, or from the literature when appropriate, to maximize the number of available specimens (see Appendices 2 and 3).

Statistical analyses

Statistical analyses were performed in R 3.3.2 (R Core Development Team 2016). Measurements were $\log_{10}$ transformed and mean standardized prior to analysis. Ordinary least squares (OLS) regression was used to assess scaling relationship slope (Kilmer & Rodríguez, 2016; Smith, 2009; Warton, Duursma, Falster, & Taskinen, 2012; Warton, Wright, Falster, & Westoby, 2006). For every species, focal structure size and reference structure size were regressed on body size in separate models. Analyses of covariance (ANCOVA) were then used to compare regression slopes of focal structure size on body size (β_{focal}) to regression slopes of reference structure size on body size ($\beta_{reference}$) within the same species (i.e., to determine whether or not there was a significant interaction between body size and trait group (focal/reference) in explaining trait size). (Differences in intercept were not analysed, since all data were mean-standardized prior to analysis.) In addition, slope estimates (β_{focal} and $\beta_{reference}$) were collected from each model and 95% confidence intervals constructed. These 95% confidence intervals were then compared between focal and reference structures within the same species.

Mean β_{focal} was calculated for sexually selected signal structures and compared to mean β_{focal} calculated for non-signal structures using Welch's t test. Mean $\beta_{reference}$ was calculated for species with sexually selected signal structures and compared to mean $\beta_{reference}$ for species with exaggerated non-signal structures using Welch's t-test. 95% confidence intervals were constructed around mean $\beta_{reference}$ for species with sexually selected signal structures and mean $\beta_{reference}$ for species with non-signal structures and compared. The difference between β_{focal} and $\beta_{reference}$ ($\Delta\beta_{focal\text{-}reference}$) was calculated for each species. Mean $\Delta\beta_{focal\text{-}reference}$ for species with sexually selected signal structures was compared to mean $\Delta\beta_{focal\text{-}reference}$ for species with non-signal structures using Welch's t-test. 95% confidence intervals were constructed around mean $\Delta\beta_{focal\text{-}reference}$ for sexually selected signal structures and mean $\Delta\beta_{focal\text{-}reference}$ for non-signal structures and compared.

Coefficients of variation were calculated for every structure. Mean coefficient of variation was calculated across all signal structures and compared to the mean coefficient of variation compared across all non-signal structures using 95% confidence intervals and Welch's t test.

Results

Results of species-level analyses are summarized in Table 1, including slope estimates (β_{focal} and $\beta_{reference}$) and adjusted R^2 values for all models, differences between β_{focal} and $\beta_{reference}$ ($\Delta\beta_{focal\text{-}reference}$), ANCOVA results, 95% confidence intervals surrounding β_{focal}, $\beta_{reference}$, and $\Delta\beta_{focal\text{-}reference}$, and coefficients of variation. For the majority of species with sexually selected signal structures, β_{focal} was significantly greater than $\beta_{reference}$ (Table 1; Appendix 4). For two of these species, whitetail deer and wildebeest, β_{focal} was greater than $\beta_{reference}$, but 95% confidence intervals surrounding these estimates were overlapping and the ANCOVA showed no significant difference between β_{focal} and $\beta_{reference}$. In pronghorn antelope, 95% confidence intervals surrounding β_{focal} and $\beta_{reference}$ were overlapping, but ANCOVA showed a (slightly) significant difference between β_{focal} and $\beta_{reference}$. Earwigs, on the other hand, displayed non-overlapping 95% confidence intervals surrounding β_{focal} and $\beta_{reference}$, but the ANCOVA showed no significant difference between β_{focal} and $\beta_{reference}$. For all species with exaggerated, non-signal structures, β_{focal} and $\beta_{reference}$ were either not significantly different, or $\beta_{reference}$ was significantly higher than β_{focal} (Table 1; Appendix 5). Unlike extreme sexually selected signal structures, extreme non-signal structures appear to scale similarly to reference structures within the same organism. Mean slope (β_{focal}) of all exaggerated sexually selected signal structures was greater than the mean slope (β_{focal}) of all non-signal structures ($t_{13.543} = -3.835$, $p < 0.01$) and 95% confidence intervals were non-overlapping (95% CI mean β_{focal} for sexually selected signal structures [1.709, 4.56]; 95% CI mean β_{focal} for non-signal structures [0.374, 0.783]). Mean $\Delta\beta_{focal\text{-}reference}$ for sexually selected signal structures was greater than mean $\Delta\beta_{focal\text{-}reference}$ for non-signal structures ($t_{14.164} = 4.079$, $p = 0.001$; Appendix 6) and 95% confidence intervals did not overlap (95% CI mean $\Delta\beta_{focal\text{-}reference}$ for sexually selected signal structures [1.072, 3.831]; 95% CI mean $\Delta\beta_{focal\text{-}reference}$ for non-signal structures [-0.501, 0.078]).

Coefficients of variation were significantly higher for extreme, sexually selected signal structures (mean = 15.444, 95% CI [9.325, 21.562]) than for non-signal structures (mean = 5.351, 95% CI [3.263, 7.438]) ($t_{16.043} = 3.37$, $p < 0.01$; Appendix 7).

Discussion

Within species, sexually selected signal structures scaled steeply with body size (Table 1; Appendix 4). In the majority of sexually selected species surveyed here, the scaling relationship of the signal (β_{focal}) was significantly steeper than that of the reference structure ($\beta_{reference}$). Surprisingly, this pattern did not hold for whitetail deer (*Odocoileus virginianus*) or wildebeest. In these species, β_{focal} was greater than $\beta_{reference}$, but there was no significant difference between β_{focal} and $\beta_{reference}$. Similarly, for earwigs, the ANCOVA showed no significant difference between β_{focal} and $\beta_{reference}$, but β_{focal} was greater than $\beta_{reference}$ and 95% confidence intervals surrounding these estimates were non-overlapping (Table 1; Appendix 4). These results may be an artefact of relatively small sample size (e.g., n < 18 for whitetail deer) and/or biased sampling (e.g., hunters favouring largest antlered males in sampled populations), since previous work has shown positive allometry and/or strong selection for these, and similar, weapons (e.g., Kruuk et al., 2002; Melnycky, Weladji, Holand, & Nieminen, 2013; Lundrigan, 1996; L. W. Simmons & Tomkins, 1996). Alternatively, these structures may function strictly as weapons (i.e., tools) of intrasexual competition, not as visual signals of quality. If true, then steep scaling between weapon and body size is not expected (McCullough, Miller, & Emlen, 2016, see below). Overall, our results for sexually selected signal structures are consistent with previous work showing that these types of extreme structures tend to be positively allometric (Bonduriansky & Day, 2003; Emlen, 2008; Green, 1992; Kodric-Brown & Brown, 1984; Kodric-Brown et al., 2006; Petrie, 1988, 1992; L. W. Simmons & Tomkins, 1996; Voje, 2016).

Every exaggerated non-signal structure measured scaled with a slope that was either less than, or not significantly different from, that of the reference structure (Table 1; Appendix 5). In addition, across species, the scaling relationship (β_{focal}) of sexually selected signal structures was significantly steeper than that of non-signal structures ($t_{11.902} = -3.23$, $p < 0.01$). Even within the same organism, non-signal structures scaled at a shallower rate than sexually selected signals. In Jackson's chameleon, for example, where both an extreme sexually selected signal, horn length, and an extreme non-signal prey capture structure, tongue length, were surveyed, horn size scaled at a much steeper rate compared to the reference structure than did tongue size (Table 1; Fig. 2).

Why signals should scale more steeply than other body parts

Many studies have considered what makes a good signal (reviewed in Bradbury & Vehrencamp, 1998; Maynard Smith & Harper, 2003; Searcy & Nowicki, 2006). In the context of sexual selection, receivers are often females who use variation in signal expression as a basis for mate choice, or males who use these signals to determine the resource holding potential (i.e., fighting ability) of rival males (M. B. Andersson, 1994; Bradbury & Vehrencamp, 1998; Hardy & Briffa, 2013). In both cases, information encoded in the signal pertains to the overall genetic quality and/or condition of the bearer (reviewed in Neff & Pitcher, 2005).

Although any phenotype could, in principle, be used as a signal (provided it is detectable and variable across individuals), some make more effective signals than others. The best signals are conspicuous – bigger or brighter than other body parts (Bradbury & Vehrencamp, 1998). However, it is not just the structure that must be conspicuous. *Variation* in the expression of that structure is key to mate and rival assessment, and the more pronounced the differences, the better. For this reason, signal structures are often selected to be more variable in their expression than other, surrounding, non-signal structures (Alatalo, Höglund, & Lundberg, 1988; José Javier Cuervo & Møller, 2001; Emlen, Warren, Johns, Dworkin, & Lavine, 2012; Fitzpatrick, 1997; Petrie, 1992; Pomiankowski & Moller, 1995; Rowe & Houle, 1996; L. W. Simmons & Tomkins, 1996; Tazzyman, Iwasa, & Pomiankowski, 2014; Wallace, 1987).

Hypervariability in trait size amplifies associated variation in male quality, making these otherwise subtle differences easier to see (Hasson, 1991; Tazzyman et al., 2014; Wallace, 1987).

Effective signals must also be honest. If poor quality males can cheat by producing effective signals, then reliability of the signal plummets and receivers should focus on other traits. One form of honesty arises when the growth of signal traits is condition-sensitive (Biernaskie, Grafen, & Perry, 2014; Bonduriansky, 2006; Bonduriansky & Day, 2003; Grafen, 1990; Iwasa, Pomiankowski, & Nee, 1991; Johnstone, 1997; Kodric-Brown et al., 2006; Nur & Hasson, 1984; Pomiankowski, 1987; Zeh & Zeh, 1988). Condition-sensitive growth of signal structures may ‘capture’ genetic or environmental variation underlying overall quality, making these signals virtually impossible to fake (Miller & Moore, 2007; Rowe & Houle, 1996; Wilkinson & Taper, 1999). Indeed, sexually selected signal structures are notoriously sensitive to stress, parasite load, and nutrition (Cotton, Fowler, & Pomiankowski, 2004; Ezenwa & Jolles, 2008; Gosden & Chenoweth, 2011; Hamilton & Zuk, 1982; Izzo & Tibbetts, 2015; Knell & Simmons, 2010; Kruuk et al., 2002; Skarstein & Folstad, 1996).

Hypervariability through heightened condition sensitivity causes structures to be reliable and informative as signals of quality (M. B. Andersson, 1994; M. Andersson & Iwasa, 1996; M. Andersson & Simmons, 2006; Bradbury & Vehrencamp, 1998), and these basic characteristics are shared by a wealth of sexually selected signals (reviewed in Bradbury & Vehrencamp, 1998). When information contained in a sexually selected signal involves individual differences in the size of a structure, and when among-individual variation in condition or genetic quality manifests as differences in overall body size, then selection for increasingly effective signals should lead to the evolution of not just higher trait-specific coefficients of variation, but also to a relatively steeper scaling relationship slope (Biernaskie et al., 2014; Green, 1992; Kodric-Brown & Brown, 1984; Kodric-Brown et al., 2006; Petrie, 1988). The steeper the slope, the more variable the focal structure will be relative to surrounding body parts. Mechanistically, when variation in condition is driven by differential access to nutrition, then the evolution of heightened condition-sensitive growth in a particular structure, relative to others, will also manifest as an increase in the steepness of the slope for that structure (Emlen et al., 2012; Lavine, Gotoh, Brent, Dworkin, & Emlen, 2015; Mirth, Frankino, & Shingleton, 2016; Shingleton & Frankino, 2013). Thus, for this particular subset of signal structures, the positive allometry hypothesis should hold. Indeed, the steeper the scaling relationship slope, the better the signal will be, leading to the evolution of larger and larger structures with steeper and steeper patterns of static scaling.

A few exceptions should be noted, however. First, body size is not always correlated with overall genetic quality or condition, as is the case for many fishes (Bolger & Connolly, 1989) and birds (José J. Cuervo & Møller, 2009). In these species, signals are still expected to be condition-sensitive and hypervariable. However, because condition is not correlated with body size, differences in the relative sizes of signal structures may not covary with body size (e.g., Bonduriansky & Day, 2003; José J. Cuervo & Møller, 2009; Fitzpatrick, 1997; Pomfret & Knell, 2006). (This was true for several focal non-signal traits, and several reference traits surveyed here (indicated by low adjusted R^2 values; Table 1). Indeed, future analyses may benefit from choosing reference structures that more tightly covary with body size.) Similarly, signals that vary in other ways besides size (e.g., colour, behaviour, chemical signals) are also not expected to scale with body size. Finally, sexually selected traits that do not function as signals (e.g., peacock moth antennae, measured here; Table 1; Appendix 5), are not predicted to scale steeper than reference structures, since hypervariation and/or condition sensitivity may actually decrease performance. This includes sexually selected weapons that function only as tools of battle and not as signals of quality, condition, or resource holding potential (McCullough et al., 2016). For these structures, trait expression should be proportional across the entire population, even when selection favours large relative trait sizes. Large structures may display especially high scaling relationship intercepts compared to other traits in the body, but since there is no hypervariation and/or heightened condition sensitivity, the slope should not differ from that of a reference structure. Consequently, we suggest much of the confusion regarding the link between positive allometry and sexual selection can be resolved by recognizing that the positive allometry hypothesis applies only to those structures that act as visual signals of among-individual variation in condition or genetic quality and, in fact, it applies only to a subset of these, signals whose information involves differences in signal size in species where quality is approximated by variation in overall size. For these structures, sexual selection is predicted to drive the evolution of extreme trait size and unusually steep scaling.

Testing the positive allometry hypothesis against reference structures, rather than isometry

We suggest three reasons for testing the positive allometry hypothesis in comparison with reference structures, rather than with isometry. First, inferring signal function for a structure that scales steeply only makes sense if that structure scales more steeply than other body parts. Steep scaling relationship slopes are relevant because they cause structures to be better signals than other, surrounding body parts. The properties that make them effective signals are relative: they are more variable and more condition-sensitive in their growth than other body parts. Sexual selection favours receivers who pay attention to these structures because, by doing so, individuals make more informed decisions than they would if they focused on other body parts. Consequently, the pattern that matters for inferring a sexually selected signal function is the difference in slope between the putative signal and other, non-signal, structures.

Second, detecting hyperallometry in a focal structure without comparing the slope to a control can be misleading. It is possible for non-signal structures to scale steeply. Indeed, in our sample of non-signal exaggerated structures, gaboon viper (*Bitis gaboncia*) fangs, elephant shrew (*Elephantulus fuscus*) snouts, and mantidfly forelegs all scaled with relatively slopes (i.e., $\beta > 1$), but the reference structures were hyperallometric too (Appendix 5; Table 1). Had we focused only on the absolute value of the scaling relationship slope we would have erroneously inferred a signal function for these structures when, in fact, their scaling relationship slopes were no different from those of surrounding body parts. These structures lack the critical properties of an informative signal despite being hyperallometric.

Finally, comparing measured slopes with isometry places undue emphasis on the estimated slope *per se*. Isometry may be intuitive in principle, but actually detecting it, or rejecting it, depends a lot on the particular landmarks selected, the units of measurement involved, and the chosen measure of body size (Bookstein, 1989; Jungers, Falsetti, &

Wall, 1995; e.g., Mosimann & James, 1979). For this reason, focusing tests of the positive allometry hypothesis exclusively on rejection of a slope of one may be misleading, especially in the context of interspecific comparisons where landmarks and measures of body size/condition often differ (e.g., Bolger & Connolly, 1989; Jakob, Marshall, & Uetz, 1996, p. d; Peig & Green, 2010). Focusing instead on the slopes of focal structures compared to those of reference structures delivers an internally controlled assay for the properties of a structure's expression that matter. Significant increases in the slope of a focal structure relative to other body parts means that the focal structure has the predicted properties of a signal, and we suggest this constitutes evidence in favour of a function for that structure as a sexually selected signal.

Diversity of exaggerated morphology

Not all sexually selected structures are signals, but many experience strong selection for increased size. In arthropods with low population density, for example, males search for receptive females and selection can lead to the evolution of elaborate antennae and/or enlarged eyes (e.g., peacock moth antennae, measured here; Table 1). This results in pronounced sexual dimorphism in relative trait size and, in some species, exaggerated male sensory structures (M. B. Andersson, 1994; Bertin & Cezilly, 2003; Lefebvre, 2000; Thornhill, 1981). Similarly, antagonistic coevolutionary arms races arising from conflict between males and females can drive rapid evolution of genitalia (Arnqvist & Rowe, 2002, 2005; Brennan, Clark, & Prum, 2009; Parker, 1979; Leigh W. Simmons, 2014). In both contexts, sexual selection drives the evolution of extreme size, but these structures do not function as signals. There is little covariance between trait *variation* and fitness and, thus, no benefit in traits being hypervariable or extra condition sensitive. For these traits, steep scaling slopes are not expected (e.g., Eberhard, 1998, 2010; Hosken & Stockley, 2004).

Exaggerated size can also arise through natural selection as, for example, in some locomotor, prey capture, and feeding structures (reviewed in Lavine et al., 2015). Appendages such as praying mantis forelimbs and antlion mandibles function like levers, snapping closed to grasp prey. For these species, longer forelimbs or mandibles perform better than shorter ones both because they move faster at their tips, and because they sweep through a larger 'kill zone' (Loxton & Nicholls, 1979; Maldonado, Levin, & Pita, 1967). However, like sensory and genitalic structures of sexual selection, large size in these naturally selected structures is not related to a signal function. There is no benefit to hypervariability or heightened condition sensitivity, and steep scaling relationship slopes are not expected.

Here, we provide measures of static allometry for 15 extreme non-signalling structures (Table 1; Appendix 5). None are sexually dimorphic, and none scaled more steeply than other, typically proportioned, body parts. Jackson's chameleons provide perhaps the best example of all, since males in this species have both types of extreme structure: three horns on the head that function as a signal of competitive ability (Bustard, 1958), and an elongated tongue used to capture prey. Even though the tongue is relatively larger than the horns, tongues scaled with a slope that was shallower than the reference structure. Horns, in contrast, scaled disproportionately steeply (Fig. 2). Clearly, the evolution of extreme structures need not entail relative increases in static allometry slope, and steep slopes, when they occur, can provide valuable clues to a sexually selected signal function.

Inferring function for extreme structures in extinct taxa

Unlike most organisms described above, the behaviour of extinct taxa cannot be observed. Even so, lines of evidence can be drawn from static, morphological data to provide testable hypotheses of behaviour (D. W. E. Hone & Faulkes, 2014). For example, hypotheses surrounding mechanical function, such as those involving anchors for musculature or levers that increase moment arms, can be assessed (and potentially rejected) using data from fossils (e.g., D. W. Hone, Naish, & Cuthill, 2012; Knell & Fortey, 2005). Similarly, we maintain the use of static scaling relationship slopes and coefficients of variation may provide a means for inferring a sexually selected signal function for extreme morphology in the fossil record.

Static scaling relationships have been used already to infer function in the fossil record (Gould, 1973; D. W. Hone et al., 2016; Knell & Fortey, 2005). However, such inferences remain controversial (e.g., Padian & Horner, 2011, 2013, 2014; Knell & Sampson, 2011; Knell et al., 2013a; D. W. Hone & Mallon, 2017; Mallon, 2017). One issue is that collecting multiple individuals from the same fossil locality and horizon (i.e., a single population) is difficult. Sample sizes are often small or gathered from animals separated in space and/or time, and animals are rarely sexed (e.g., D. W. Hone & Mallon, 2017). As a result, detection of even fundamental patterns in morphology, such as sexual dimorphism, remains elusive (Mallon, 2017; but see Sengupta, Ezcurra, & Bandyopadhyay, 2017). Another issue is that distinguishing between different signal functions is often difficult. Social dominance and sexually selected signals, for example, are often confluent and distinguishing between them is complex. In addition, the cooption of extreme structures to multiple functions, thereby exposing them to multiple patterns of selection, may further confound these data (e.g., dugong tusks; Anderson, 1979; Domning & Beatty, 2007).

Despite these limitations, we suggest behaviour can be inferred from the fossil record using the methods and logic described above. We predict that when focal structures act as signals of overall body size, both the slope of the static scaling relationship and the coefficient of variation will be steeper/greater in the putative signal structure than in reference structures used as controls. As ‘proof of concept’ for this approach, we included two putative sexually selected signal structures from the fossil record in our analyses, the enlarged cephalic frill of the ceratopsian dinosaur *Protoceratops andrewsi* (adapted and expanded from D. W. Hone et al., 2016), and the tail vane of the pterosaur, *Rhamphorhynchus*. In both cases, the focal structure scaled more steeply with body size and had a higher coefficient of variation than reference structures measured in the same individual (Fig. 3; Table 1), implying a signalling function.

Overall, we believe this method useful for inferring extreme structure function in the fossil record (perhaps even more useful when analysed in conjunction with other patterns in morphology - e.g., changes in complexity during ontogeny, high variation in trait shape and size between species lineages). Both morphological scaling relationships and coefficients of variation can be reliably measured in fossil specimens, even when sample size is small. We recommend the use of these methods in subsequent analyses of extreme or 'bizarre' morphology in the fossil record, and are hopeful that they might provide insight into the ongoing debate regarding sexual selection in non-avian dinosaurs.

Overall, we suggest that when applied specifically and exclusively to disproportionately large animal structures that function as signals of overall body size, and when assessed through comparison with surrounding, non-signal structures rather than through detection of an estimated slope greater than 1, the positive allometry hypothesis holds. Sexually selected signal structures are predicted to – and, in fact, appear to – scale more steeply with body size than non-signal structures. For this reason, we suggest that relative patterns of trait scaling offer powerful clues to trait function, particularly when combined with other measures of trait expression such as trait specific coefficients of variation.

Data archive

Datasets supporting this article will be uploaded to Dryad

Competing interests

The authors have no competing interests to report

References

Abouheif, E., & Fairbairn, D. J. (1997). A comparative analysis of allometry for sexual size dimorphism: assessing Rensch's rule. *The American Naturalist*, *149*(3), 540–562.

Alatalo, R. V., Höglund, J., & Lundberg, A. (1988). Patterns of variation in tail ornament size in birds. *Biological Journal of the Linnean Society*, *34*(4), 363–374.

Anderson, P. K. (1979). Dugong behavior: on being a marine mammalian grazer. *Biologist*, *61*(4), 113–144.

Andersson, M. B. (1994). *Sexual selection*. Princeton University Press.

Andersson, M., & Iwasa, Y. (1996). Sexual selection. *Trends in Ecology & Evolution*, *11*(2), 53–58. https://doi.org/10.1016/0169-5347(96)81042-1

Andersson, M., & Simmons, L. W. (2006). Sexual selection and mate choice. *Trends in Ecology & Evolution*, *21*(6), 296–302.

Arnqvist, G., & Rowe, L. (2002). Correlated evolution of male and female morphologies in water striders. *Evolution*, *56*(5), 936–947.

Arnqvist, G., & Rowe, L. (2005). Sexual conflict. *Prince—ton University Press, Princeton, NI*.

Baker, R. H., & Wilkinson, G. S. (2001). Phylogenetic analysis of sexual dimorphism and eye-span allometry in stalk-eyed flies (Diopsidae). *Evolution*, *55*(7), 1373–1385.

Bateson, W., & Brindley, H. H. (1892). 3. On some cases of Variation in Secondary Sexual Characters, statistically examined. *Journal of Zoology*, *60*(4), 585–596.

Bels, V. L., Aerts, P., Chardon, M., Vandewalle, P., Berkhoudt, H., Crompton, A. W., … Frazzetta, T. H. (2012). *Biomechanics of feeding in vertebrates* (Vol. 18). Springer Science & Business Media.

Bergmann, P. J., & Berk, C. P. (2012). The evolution of positive allometry of weaponry in horned lizards (Phrynosoma). *Evolutionary Biology*, *39*(3), 311–323.

Bertin, A., & Cezilly, F. (2003). Sexual selection, antennae length and the mating advantage of large males in Asellus aquaticus. *Journal of Evolutionary Biology*, *16*(4), 698–707.

Bertin, A., & Fairbairn, D. J. (2007). The Form of Sexual Selection on Male Genitalia Cannot Be Inferred from Within-Population Variance and Allometry—a Case Study in Aquarius Remigis. *Evolution*, *61*(4), 825–837. https://doi.org/10.1111/j.1558-5646.2007.00074.x

Bertram, J. E., & Biewener, A. A. (1990). Differential scaling of the long bones in the terrestrial Carnivora and other mammals. *Journal of Morphology*, *204*(2), 157–169.

Biernaskie, J. M., Grafen, A., & Perry, J. C. (2014). The evolution of index signals to avoid the cost of dishonesty. *Proceedings of the Royal Society of London B: Biological Sciences*, *281*(1790), 20140876.

Blanckenhorn, W. U., Kraushaar, U. R. S., Teuschl, Y., & Reim, C. (2004). Sexual selection on morphological and physiological traits and fluctuating asymmetry in the black scavenger fly Sepsis cynipsea. *Journal of Evolutionary Biology*, *17*(3), 629–641. https://doi.org/10.1111/j.1420-9101.2004.00693.x

Blanckenhorn, W. U., Meier, R., & Teder, T. (2007). Rensch’s rule in insects: patterns among and within species. *Sex, Size, and Gender Roles: Evolutionary Studies of Sexual Size Dimorphism*, 60–70.

Bolger, T., & Connolly, P. L. (1989). The selection of suitable indices for the measurement and analysis of fish condition. *Journal of Fish Biology*, *34*(2), 171–182.

Bonduriansky, R. (2006). The evolution of condition-dependent sexual dimorphism. *The American Naturalist*, *169*(1), 9–19.

Bonduriansky, R. (2007). Sexual selection and allometry: a critical reappraisal of the evidence and ideas. *Evolution*, *61*(4), 838–849.

Bonduriansky, R., & Day, T. (2003). The Evolution of Static Allometry in Sexually Selected Traits. *Evolution*, *57*(11), 2450–2458. https://doi.org/10.1111/j.0014-3820.2003.tb01490.x

Bookstein, F. L. (1989). “ Size and shape”: a comment on semantics. *Systematic Zoology*, *38*(2), 173–180.

Bradbury, J. W., & Vehrencamp, S. L. (1998). Principles of animal communication. *Behavioral Ecology*, *12*, 283–286.

Brennan, P. L., Clark, C. J., & Prum, R. O. (2009). Explosive eversion and functional morphology of the duck penis supports sexual conflict in waterfowl genitalia. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20092139.

Bustard, H. R. (1958). Use of horns by Chamaeleo jacksoni. *Brit J Herpetol*, *2*, 105–107.

Cheverud, J. M. (1982). Relationships among ontogenetic, static, and evolutionary allometry. *American Journal of Physical Anthropology*, *59*(2), 139–149.

Christiansen, P. (1999). Scaling of the limb long bones to body mass in terrestrial mammals. *Journal of Morphology*, *239*(2), 167–190. https://doi.org/10.1002/(SICI)1097-4687(199902)239:2<167::AID-JMOR5>3.0.CO;2-8

Cock, A. G. (1966). Genetical aspects of metrical growth and form in animals. *The Quarterly Review of Biology*, *41*(2), 131–190.

Cotton, S., Fowler, K., & Pomiankowski, A. (2004). Condition dependence of sexual ornament size and variation in the stalk-eyed fly Cyrtodiopsis dalmanni (Diptera: Diopsidae). *Evolution*, *58*(5), 1038–1046.

Cuervo, J. J., & Møller, A. P. (2001). Components of phenotypic variation in avian ornamental and non-ornamental feathers. *Evolutionary Ecology*, *15*(1), 53–72.

Cuervo, J. J., & Møller, A. P. (2009). The allometric pattern of sexually size dimorphic feather ornaments and factors affecting allometry. *Journal of Evolutionary Biology*, *22*(7), 1503–1515.

Darwin, C. (1871). *The Descent of Man: And Selection in Relation to Sex* (Vol. 2). Appleton.

Dial, K. P., Greene, E., & Irschick, D. J. (2008). Allometry of behavior. *Trends in Ecology & Evolution*, *23*(7), 394–401.

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

Domning, D. P., & Beatty, B. L. (2007). Use of tusks in feeding by dugongid sirenians: Observations and tests of hypotheses. *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology*, *290*(6), 523–538. https://doi.org/10.1002/ar.20540

Dyar, H. G. (1890). The number of molts of lepidopterous larvae. *Psyche: A Journal of Entomology*, *5*(175–176), 420–422.

Eberhard, W. G. (1980). Horned beetles. *Scientific American*, *242*(3), 166–183.

Eberhard, W. G. (1998). Sexual behavior of Acanthocephala declivis guatemalana (Hemiptera: Coreidae) and the allometric scaling of their modified hind legs. *Annals of the Entomological Society of America*, *91*(6), 863–871.

Eberhard, W. G. (2010). Evolution of genitalia: theories, evidence, and new directions. *Genetica*, *138*(1), 5–18.

Egset, C. K., Hansen, T. F., Le Rouzic, A., Bolstad, G. H., Rosenqvist, G., & Pélabon, C. (2012). Artificial selection on allometry: change in elevation but not slope. *Journal of Evolutionary Biology*, *25*(5), 938–948. https://doi.org/10.1111/j.1420-9101.2012.02487.x

Emlen, D. J. (1996). Artificial Selection on Horn Length-Body Size Allometry in the Horned Beetle Onthophagus acuminatus (Coleoptera: Scarabaeidae). *Evolution*, *50*(3), 1219–1230. https://doi.org/10.2307/2410662

Emlen, D. J. (2008). The evolution of animal weapons. *Annual Review of Ecology, Evolution, and Systematics*, *39*, 387–413.

Emlen, D. J., & Allen, C. E. (2003). Genotype to Phenotype: Physiological Control of Trait Size and Scaling in Insects. *Integrative and Comparative Biology*, *43*(5), 617–634. https://doi.org/10.1093/icb/43.5.617

Emlen, D. J., Marangelo, J., Ball, B., & Cunningham, C. W. (2005). Diversity in the Weapons of Sexual Selection: Horn Evolution in the Beetle Genus Onthophagus (coleoptera: Scarabaeidae). *Evolution*, *59*(5), 1060–1084. https://doi.org/10.1111/j.0014-3820.2005.tb01044.x

Emlen, D. J., & Nijhout, H. F. (2000). The development and evolution of exaggerated morphologies in insects. *Annual Review of Entomology*, *45*(1), 661–708.

Emlen, D. J., Warren, I. A., Johns, A., Dworkin, I., & Lavine, L. C. (2012). A Mechanism of Extreme Growth and Reliable Signaling in Sexually Selected Ornaments and Weapons. *Science*, *337*(6096), 860–864. https://doi.org/10.1126/science.1224286

Ezenwa, V. O., & Jolles, A. E. (2008). Horns honestly advertise parasite infection in male and female African buffalo. *Animal Behaviour*, *75*(6), 2013–2021.

Fairbairn, D. J. (1997). Allometry for sexual size dimorphism: pattern and process in the coevolution of body size in males and females. *Annual Review of Ecology and Systematics*, *28*(1), 659–687.

Fastovsky, D. E., Weishampel, D. B., Watabe, M., Barsbold, R., Tsogtbaatar, K. H., & Narmandakh, P. (2011). A nest of Protoceratops andrewsi (Dinosauria, Ornithischia). *Journal of Paleontology*, *85*(6), 1035–1041.

Fitzpatrick, S. (1997). Patterns of morphometric variation in birds' tails: length, shape and variability. *Biological Journal of the Linnean Society*, *62*(1), 145–162.

Fromhage, L., & Kokko, H. (2014). Sexually selected traits evolve positive allometry when some matings occur irrespective of the trait. *Evolution*, *68*(5), 1332–1338.

Gosden, T. P., & Chenoweth, S. F. (2011). On the evolution of heightened condition dependence of male sexual displays. *Journal of Evolutionary Biology*, *24*(3), 685–692.

Gould, S. J. (1966). Allometry and Size in Ontogeny and Phylogeny. *Biological Reviews*, *41*(4), 587–638. https://doi.org/10.1111/j.1469-185X.1966.tb01624.x

Gould, S. J. (1973). Positive allometry of antlers in the "Irish elk", Megaloceros giganteus. *Nature*, *244*(5415), 375–376.

Gould, S. J. (1974a). Allometry in primates, with emphasis on scaling and the evolution of the brain. *Contributions to Primatology*, *5*, 244–292.

Gould, S. J. (1974b). The origin and function of'bizarre'structures: antler size and skull size in the'Irish Elk,'Megaloceros giganteus. *Evolution*, 191–220.

Gould, S. J. (1977). *Ontogeny and phylogeny*. Harvard University Press.

Grafen, A. (1990). Biological signals as handicaps. *Journal of Theoretical Biology*, *144*(4), 517–546.

Green, A. J. (1992). Positive allometry is likely with mate choice, competitive display and other functions. *Animal Behaviour*, *43*(1), 170–172. https://doi.org/10.1016/S0003-3472(05)80086-7

Hamilton, W. D., & Zuk, M. (1982). Heritable true fitness and bright birds: a role for parasites? *Science*, *218*(4570), 384–387.

Handa, N., Watabe, M., & Tsogtbaatar, K. (2012). New specimens of Protoceratops (Dinosauria: Neoceratopsia) from the Upper Cretaceous in Udyn Sayr, southern Gobi area, Mongolia. *Paleontological Research*, *16*(3), 179–198.

Hardy, I. C., & Briffa, M. (2013). *Animal contests*. Cambridge University Press.

Hasson, O. (1991). Sexual displays as amplifiers: practical examples with an emphasis on feather decorations. *Behavioral Ecology*, *2*(3), 189–197.

Hogue, C. L. (1984). Observations on the plant hosts and possible mimicry models of Lantern Bugs (Fulgora spp.)(Homoptera: Fulgoridae). *Rev. Biol. Trop*, *32*(1), 145–150.

Hone, D. W. (2012). A new specimen of the pterosaur Rhamphorhynchus. *Historical Biology*, *24*(6), 581–585.

Hone, D. W. E., & Faulkes, C. G. (2014). A proposed framework for establishing and evaluating hypotheses about the behaviour of extinct organisms. *Journal of Zoology*, *292*(4), 260–267.

Hone, D. W., Farke, A. A., Watabe, M., Shigeru, S., & Tsogtbaatar, K. (2014). A new mass mortality of juvenile Protoceratops and size-segregated aggregation behaviour in juvenile non-avian dinosaurs. *PloS One*, *9*(11), e113306.

Hone, D. W., & Mallon, J. C. (2017). Protracted growth impedes the detection of sexual dimorphism in non-avian dinosaurs. *Palaeontology*, *60*(4), 535–545.

Hone, D. W., Naish, D., & Cuthill, I. C. (2012). Does mutual sexual selection explain the evolution of head crests in pterosaurs and dinosaurs? *Lethaia*, *45*(2), 139–156.

Hone, D. W., Wood, D., & Knell, R. J. (2016). Positive allometry for exaggerated structures in the ceratopsian dinosaur Protoceratops andrewsi supports socio-sexual signaling. *Palaeontologia Electronica*, *19*(1), 1–13.

Hongo, Y. (2007). Evolution of male dimorphic allometry in a population of the Japanese horned beetle Trypoxylus dichotomus septentrionalis. *Behavioral Ecology and Sociobiology*, *62*(2), 245–253. https://doi.org/10.1007/s00265-007-0459-2

Hosken, D. J., & Stockley, P. (2004). Sexual selection and genital evolution. *Trends in Ecology & Evolution*, *19*(2), 87–93.

Huxley, J. (1932). Problems of relative growth. Retrieved from http://agris.fao.org/agris-search/search.do?recordID=US201300287973

Huxley, J. S., & Teissier, G. (1936). Terminology of Relative Growth. *Nature*, *137*(3471), 780–781. https://doi.org/10.1038/137780b0

Iwasa, Y., Pomiankowski, A., & Nee, S. (1991). The evolution of costly mate preferences II. The'handicap'principle. *Evolution*, 1431–1442.

Izzo, A., & Tibbetts, E. A. (2015). Heightened condition dependence of a sexually selected signal in male Polistes dominulus paper wasps. *Ethology*, *121*(6), 586–592.

Jakob, E. M., Marshall, S. D., & Uetz, G. W. (1996). Estimating Fitness: A Comparison of Body Condition Indices. *Oikos*, *77*(1), 61–67. https://doi.org/10.2307/3545585

Johnstone, R. A. (1997). The evolution of animal signals. *Behavioural Ecology: An Evolutionary Approach*, 155–78.

Jungers, W. L., Falsetti, A. B., & Wall, C. E. (1995). Shape, relative size, and size-adjustments in morphometrics. *American Journal of Physical Anthropology*, *38*(S21), 137–161.

Katsuki, M., Yokoi, T., Funakoshi, K., & Oota, N. (2014). Enlarged Hind Legs and Sexual Behavior with Male-Male Interaction in Sagra femorata (Coleoptera: Chrysomelidae). *Entomological News*, *124*(3), 211–220. https://doi.org/10.3157/021.124.0306

Kelly, C. D. (2005). Allometry and sexual selection of male weaponry in Wellington tree weta, Hemideina crassidens. *Behavioral Ecology*, *16*(1), 145–152.

Kennedy, M., Taylor, S. A., Nádvorník, P., & Spencer, H. G. (2013). The phylogenetic relationships of the extant pelicans inferred from DNA sequence data. *Molecular Phylogenetics and Evolution*, *66*(1), 215–222.

Kerkhoff, A. J., & Enquist, B. J. (2009). Multiplicative by nature: why logarithmic transformation is necessary in allometry. *Journal of Theoretical Biology*, *257*(3), 519–521.

Kilmer, J. T., & Rodríguez, R. L. (2016). Ordinary least squares regression is indicated for studies of allometry. *Journal of Evolutionary Biology*, n/a-n/a. https://doi.org/10.1111/jeb.12986

Kingdon, J. (1974). *East African Mammals; an Atlas of Evolution in Africa. Volume II Part A (Insectivores and Bats).* Academic Press.

Kleiber, M. (1932). Body size and metabolism. *ENE*, *1*(9). Retrieved from http://biology.unm.edu/jhbrown/Miami/Kleiber1932.pdf

Knell, R. J., & Fortey, R. A. (2005). Trilobite spines and beetle horns: sexual selection in the Palaeozoic? *Biology Letters*, *1*(2), 196–199.

Knell, R. J., Naish, D., Tomkins, J. L., & Hone, D. W. (2013a). Is sexual selection defined by dimorphism alone? A reply to Padian and Horner. In *Proc. Zool. Soc. Lond* (Vol. 35, pp. 491–562). Citeseer.

Knell, R. J., Naish, D., Tomkins, J. L., & Hone, D. W. (2013b). Sexual selection in prehistoric animals: detection and implications. *Trends in Ecology & Evolution*, *28*(1), 38–47.

Knell, R. J., & Sampson, S. (2011). Bizarre structures in dinosaurs: species recognition or sexual selection? A response to Padian and Horner. *Journal of Zoology*, *283*(1), 18–22.

Knell, R. J., & Simmons, L. W. (2010). Mating tactics determine patterns of condition dependence in a dimorphic horned beetle. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1692), 2347–2353.

Kodric-Brown, A., & Brown, J. H. (1984). Truth in advertising: the kinds of traits favored by sexual selection. *American Naturalist*, 309–323.

Kodric-Brown, A., Sibly, R. M., & Brown, J. H. (2006). The allometry of ornaments and weapons. *Proceedings of the National Academy of Sciences*, *103*(23), 8733–8738. https://doi.org/10.1073/pnas.0602994103

Kratzing, J. E., & Woodall, P. F. (1988). The rostral nasal anatomy of two elephant shrews. *Journal of Anatomy*, *157*, 135–143.

Kruuk, L. E. B., Slate, J., Pemberton, J. M., Brotherstone, S., Guinness, F., & Clutton-Brock, T. (2002). Antler Size in Red Deer: Heritability and Selection but No

Evolution. *Evolution*, *56*(8), 1683–1695. https://doi.org/10.1111/j.0014-3820.2002.tb01480.x

Lavine, L., Gotoh, H., Brent, C. S., Dworkin, I., & Emlen, D. J. (2015). Exaggerated Trait Growth in Insects. *Annual Review of Entomology*, *60*, 453–472.

Lefebvre, H. (2000). *Everyday life in the modern world*. A&C Black.

Loxton, R. G., & Nicholls, I. (1979). The functional morphology of the praying mantis forelimb (Dictyoptera: Mantodea). *Zoological Journal of the Linnean Society*, *66*(2), 185–203.

Lundrigan, B. (1996). Morphology of horns and fighting behavior in the family Bovidae. *Journal of Mammalogy*, *77*(2), 462–475.

Maldonado, H., Levin, L., & Pita, J. C. B. (1967). Hit distance and the predatory strike of the praying mantis. *Zeitschrift Für Vergleichende Physiologie*, *56*(3), 237–257. https://doi.org/10.1007/BF00333669

Mallon, J. C. (2017). Recognizing sexual dimorphism in the fossil record: lessons from nonavian dinosaurs. *Paleobiology*, 1–13.

Maynard Smith, J., & Harper, D. (2003). *Animal Signals*. OUP Oxford.

McCullough, E. L., Miller, C. W., & Emlen, D. J. (2016). Why Sexually Selected Weapons Are Not Ornaments. *Trends in Ecology & Evolution*.

Melnycky, N. A., Weladji, R. B., Holand, Ø., & Nieminen, M. (2013). Scaling of antler size in reindeer (Rangifer tarandus): sexual dimorphism and variability in resource allocation. *Journal of Mammalogy*, *94*(6), 1371–1379. https://doi.org/10.1644/12-MAMM-A-282.1

Miljutin, A. (2008). Trends of Specialisation in Rodents: the Five-toed Jerboas, Subfamily Allactaginae (Dipodoidea, Rodentia). *Acta Zoologica Lituanica*, *18*(4), 228–239. https://doi.org/10.2478/v10043-008-0033-9

Miller, C. W., & Emlen, D. J. (2010). Dynamic effects of oviposition site on offspring sexually-selected traits and scaling relationships. *Evolutionary Ecology*, *24*(2), 375–390. https://doi.org/10.1007/s10682-009-9312-6

Miller, C. W., & Moore, A. J. (2007). A potential resolution to the lek paradox through indirect genetic effects. *Proceedings of the Royal Society of London B: Biological Sciences*, *274*(1615), 1279–1286.

Mirth, C. K., Frankino, W. A., & Shingleton, A. W. (2016). Allometry and size control: what can studies of body size regulation teach us about the evolution of morphological scaling relationships? *Current Opinion in Insect Science*, *13*, 93–98.

Moczek, A. P., & Emlen, D. J. (2000). Male horn dimorphism in the scarab beetle, Onthophagus taurus: do alternative reproductive tactics favour alternative phenotypes? *Animal Behaviour*, *59*(2), 459–466.

Mosimann, J. E., & James, F. C. (1979). New statistical methods for allometry with application to Florida red-winged blackbirds. *Evolution*, *33*(1Part2), 444–459.

Neff, B. D., & Pitcher, T. E. (2005). Genetic quality and sexual selection: an integrated framework for good genes and compatible genes. *Molecular Ecology*, *14*(1), 19–38.

Nijhout, H. F., & Emlen, D. J. (1998). Competition among body parts in the development and evolution of insect morphology. *Proceedings of the National Academy of Sciences*, *95*(7), 3685–3689.

Nur, N., & Hasson, O. (1984). Phenotypic plasticity and the handicap principle. *Journal of Theoretical Biology*, *110*(2), 275–297.

O'Brien, D. M., Katsuki, M., & Emlen, D. J. (2017). Selection on an extreme weapon in the frog-legged leaf beetle (Sagra femorata). *Evolution*, *71*(11), 2584–2598.

Ohl, M., Barkalov, A. V., & Xin-Yue, C. (2004). *Annotated catalog of the Mantispidae of the world (Neuroptera)*. Associated Publishers.

Orians, G. H. (1969). Age and hunting success in the brown pelican (Pelecanus occidentalis). *Animal Behaviour*, *17*, 316–319. https://doi.org/10.1016/0003-3472(69)90016-5

Otte, D., & Stayman, K. (1979). BEETLE HORNS. *Sexual Selection and Reproductive Competition in Insects*, 259.

Padian, K., & Horner, J. R. (2011). The evolution of "bizarre structures" in dinosaurs: biomechanics, sexual selection, social selection or species recognition? *Journal of Zoology*, *283*(1), 3–17.

Padian, K., & Horner, J. R. (2013). Misconceptions of sexual selection and species recognition: a response to Knell et al. and to Mendelson and Shaw. *Trends in Ecology & Evolution*, *28*(5), 249–250.

Padian, K., & Horner, J. R. (2014). The species recognition hypothesis explains exaggerated structures in non-avialan dinosaurs better than sexual selection does. *Comptes Rendus Palevol*, *13*(2), 97–107.

Painting, C. J., & Holwell, G. I. (2013). Exaggerated Trait Allometry, Compensation and Trade-Offs in the New Zealand Giraffe Weevil (Lasiorhynchus barbicornis). *PLOS ONE*, *8*(11), e82467. https://doi.org/10.1371/journal.pone.0082467

Parker, G. A. (1979). Sexual selection and sexual conflict. *Sexual Selection and Reproductive Competition in Insects*, 123–166.

Paulian, R. (1935). polymorphisme des males de coléoptères. Retrieved from http://agris.fao.org/agris-search/search.do?recordID=US201300400362

Peig, J., & Green, A. J. (2010). The paradigm of body condition: a critical reappraisal of current methods based on mass and length. *Functional Ecology*, *24*(6), 1323–1332. https://doi.org/10.1111/j.1365-2435.2010.01751.x

Pélabon, C., Bolstad, G. H., Egset, C. K., Cheverud, J. M., Pavlicev, M., & Rosenqvist, G. (2013). On the relationship between ontogenetic and static allometry. *The American Naturalist*, *181*(2), 195–212.

Petrie, M. (1988). Intraspecific variation in structures that display competitive ability: large animals invest relatively more. *Animal Behaviour*, *36*(4), 1174–1179.

Petrie, M. (1992). Are all secondary sexual display structures positively allometric and, if so, why? *Animal Behaviour*, *43*(1), 173–175.

Pomfret, J. C., & Knell, R. J. (2006). Sexual selection and horn allometry in the dung beetle Euoniticellus intermedius. *Animal Behaviour*, *71*(3), 567–576.

Pomiankowski, A. (1987). Sexual selection: The handicap principle does work–sometimes. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *231*(1262), 123–145.

Pomiankowski, A., & Moller, A. P. (1995). A resolution of the lek paradox. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *260*(1357), 21–29.

Rowe, L., & Houle, D. (1996). The lek paradox and the capture of genetic variance by condition dependent traits. *Proceedings of the Royal Society of London B: Biological Sciences*, *263*(1375), 1415–1421.

Rubner, M. (1883). Ueber den einfluss der korpergrosse auf stoffund kaftwechsel. *Zeitschrift Fur Biologie*, *19*, 535–562.

Schmidt-Nielsen, K. (1984). *Scaling: Why is Animal Size So Important?* Cambridge University Press.

Schreiber, R. W., Woolfenden, G. E., & Curtsinger, W. E. (1975). Prey Capture by the Brown Pelican. *The Auk*, *92*(4), 649–654. https://doi.org/10.2307/4084778

Schulte-Hostedde, A., Kuula, S., Martin, C., Schank, C. C. M., & Lesbarrères, D. (2011). Allometry and sexually dimorphic traits in male anurans. *Journal of Evolutionary Biology*, *24*(5), 1154–1159.

Searcy, W. A., & Nowicki, S. (2006). Signal interception and the use of soft song in aggressive interactions. *Ethology*, *112*(9), 865–872.

Sengupta, S., Ezcurra, M. D., & Bandyopadhyay, S. (2017). A new horned and long-necked herbivorous stem-archosaur from the Middle Triassic of India. *Scientific Reports*, *7*(1), 8366. https://doi.org/10.1038/s41598-017-08658-8

Shingleton, A. W., & Frankino, W. A. (2013). New perspectives on the evolution of exaggerated traits. *Bioessays*, *35*(2), 100–107.

Shingleton, A. W., Frankino, W. A., Flatt, T., Nijhout, H. F., & Emlen, D. J. (2007). Size and shape: the developmental regulation of static allometry in insects. *BioEssays*, *29*(6), 536–548. https://doi.org/10.1002/bies.20584

Shingleton, A. W., Mirth, C. K., & Bates, P. W. (2008). Developmental model of static allometry in holometabolous insects. *Proceedings of the Royal Society of London B: Biological Sciences*, *275*(1645), 1875–1885. https://doi.org/10.1098/rspb.2008.0227

Simmons, L. W. (2014). Sexual selection and genital evolution. *Austral Entomology*, *53*(1), 1–17.

Simmons, L. W., & Tomkins, J. L. (1996). Sexual selection and the allometry of earwig forceps. *Evolutionary Ecology*, *10*(1), 97–104.

Skarstein, F., & Folstad, I. (1996). Sexual dichromatism and the immunocompetence handicap: an observational approach using Arctic charr. *Oikos*, 359–367.

Smith, R. J. (2009). Use and misuse of the reduced major axis for line-fitting. *American Journal of Physical Anthropology*, *140*(3), 476–486.

Stanley, S. M. (1973). An explanation for Cope's rule. *Evolution*, *27*(1), 1–26.

Stern, D. L., & Emlen, D. J. (1999). The developmental basis for allometry in insects. *Development*, *126*(6), 1091–1101.

Tazzyman, S. J., Iwasa, Y., & Pomiankowski, A. (2014). Signaling Efficacy Drives the Evolution of Larger Sexual Ornaments by Sexual Selection. *Evolution*, *68*(1), 216–229. https://doi.org/10.1111/evo.12255

Templeton, C. N., Greene, E., & Davis, K. (2005). Allometry of alarm calls: black-capped chickadees encode information about predator size. *Science*, *308*(5730), 1934–1937.

Thompson, D. W. (1917). On growth and form. *On Growth and Form.* Retrieved from https://www.cabdirect.org/cabdirect/abstract/19431401837

Thornhill, R. (1981). Panorpa (Mecoptera: Panorpidae) Scorpionflies: Systems for Understanding Resource-Defense Polygyny and Alternative Male Reproductive Efforts. *Annual Review of Ecology and Systematics*, *12*, 355–386.

Urban, J. M., & Cryan, J. R. (2009). Entomologically famous, evolutionarily unexplored: The first phylogeny of the lanternfly family Fulgoridae (Insecta: Hemiptera: Fulgoroidea). *Molecular Phylogenetics and Evolution*, *50*(3), 471–484. https://doi.org/10.1016/j.ympev.2008.12.004

Voje, K. L. (2016). Scaling of morphological characters across trait type, sex, and environment: a meta-analysis of static allometries. *The American Naturalist*, *187*(1), 89–98.

Von Bertalanffy, L. (1957). Quantitative laws in metabolism and growth. *The Quarterly Review of Biology*, *32*(3), 217–231.

Wallace, B. (1987). Ritualistic combat and allometry. *The American Naturalist*, *129*(5), 775–776.

Warton, D. I., Duursma, R. A., Falster, D. S., & Taskinen, S. (2012). smatr 3–an R package for estimation and inference about allometric lines. *Methods in Ecology and Evolution*, *3*(2), 257–259.

Warton, D. I., Wright, I. J., Falster, D. S., & Westoby, M. (2006). Bivariate line-fitting methods for allometry. *Biological Reviews*, *81*(2), 259–291.

Wellnhofer, P. (1975). Die Rhamphorhynchoidea (Pterosauria) der Oberjura-Plattenkalke S\c ddeutschlands, Teil III: Palˆkologie und Stammesgeschichte. *Palaeontographica Abteilung A*, 1–30.

West, G. B., & Brown, J. H. (2005). The origin of allometric scaling laws in biology from genomes to ecosystems: towards a quantitative unifying theory of biological structure and organization. *Journal of Experimental Biology*, *208*(9), 1575–1592.

West, G. B., Brown, J. H., & Enquist, B. J. (1997). A General Model for the Origin of Allometric Scaling Laws in Biology. *Science*, *276*(5309), 122–126. https://doi.org/10.1126/science.276.5309.122

Wilkinson, G. S. (1993). Artificial sexual selection alters allometry in the stalk-eyed fly Cyrtodiopsis dalmanni (Diptera: Diopsidae). *Genetics Research*, *62*(3), 213–222. https://doi.org/10.1017/S001667230003192X

Wilkinson, G. S., & Taper, M. (1999). Evolution of genetic variation for condition-dependent traits in stalk-eyed flies. *Proceedings of the Royal Society of London B: Biological Sciences*, *266*(1429), 1685–1690.

Wu, S., Zhang, F., Edwards, S. V., Wu, W., Ye, J., Bi, S., … Organ, C. L. (2014). The evolution of bipedalism in jerboas (Rodentia: Dipodoidea): origin in humid and forested environments. *Evolution*, *68*(7), 2108–2118.

Zeh, D. W., & Zeh, J. A. (1988). Condition-dependent sex ornaments and field tests of sexual-selection theory. *The American Naturalist*, *132*(3), 454–459.

Zeh, D. W., Zeh, J. A., & Tavakilian, G. (1992). Sexual Selection and Sexual Dimorphism in the Harlequin Beetle Acrocinus longimanus. *Biotropica*, *24*(1), 86–96. https://doi.org/10.2307/2388476

Figure 1: Extreme non-signal (ns) and sexually selected (ss) signal structures. Clockwise from top right; bighorn sheep horns (*O. canadensis*; ss), Jackson's chameleon horns (*T. jacsonii*; ss), praying mantis forelimbs (Mantodea; ns), ichneumon wasp ovipositor (Ichneumonoidea; ss non-signal), gaboon viper fangs (*B. gaboncia*; ns), and dung beetle horns (Scarabaeidae, ss). Photos credited in Acknowledgments.

Fig. 1

Figure 2: Static scaling relationships for an extreme sexually selected signal structure (horns; red; left; n = 40) and an extreme, non-signal naturally selected structure (tongue; blue; right; n = 25) in Jackson's chameleons (*T. jacksonii*). Red and blue indicate focal structures. Grey indicates the reference structures. Lines represent ordinary least squares regression of standardized $\log_{10}$ structure size on standardized $\log_{10}$ body size. In Jackson's chameleon, the extreme sexually selected signal (horn length) scales at a significantly steeper rate than the reference structure (hindfoot length). The extreme non-signal structure (tongue length) does not. 95% CI for horn length [3.358, 5.159], tongue length [0.251, 0.949], and hindlimb length [1.13, 1.979].

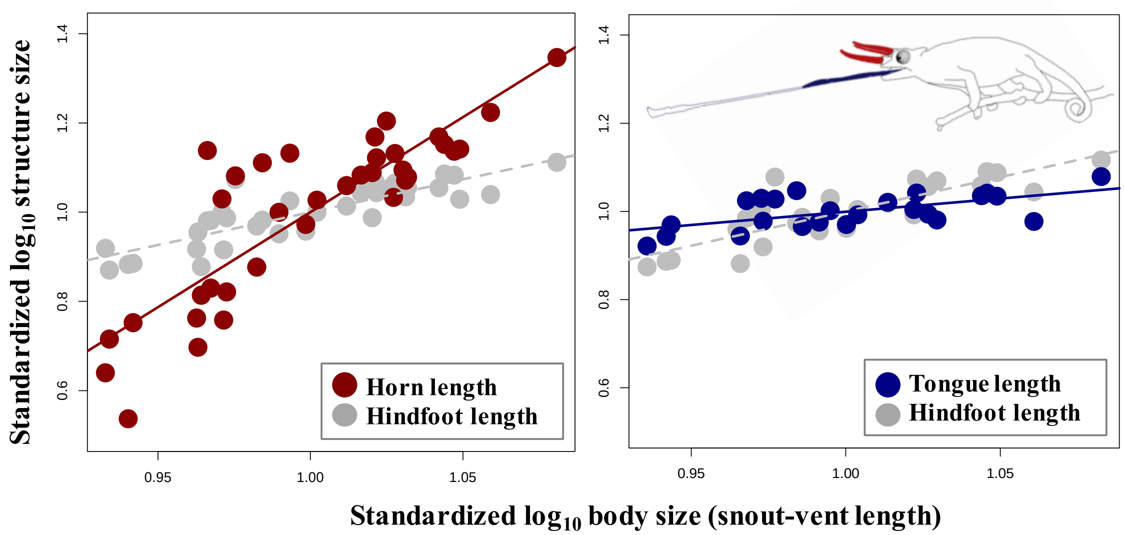


Fig. 2

Figure 3: Static scaling relationships for extreme putative sexually selected signal structures in ceratopsians (*Protoceratops andrewsi*; left; n = 38) and pterosaurs (*Rhamphorhynchus muensteri*; right; n = 10). Red indicates putative signal structures. Grey indicates reference structure. Lines represent the ordinary least squares regression of standardized $\log_{10}$ structure size on standardized $\log_{10}$ body size. In both species, the scaling relationship of the putative signal trait is steeper than that of the reference trait (*P. andrewsi*: 95% CI for slope of focal structure [1.173, 1.353], 95% CI for slope of reference structure [0.925,1.039]; *R. muensteri*: 95% CI for slope of focal structure [1.332, 2.930], 95% CI for slope of reference structure [0.871, 1.262]), consistent with a history of selection for a hypervariable sexually selected signal. Inlaid photographs display study species with focal structures highlighted in red. Photos credited in Acknowledgments.

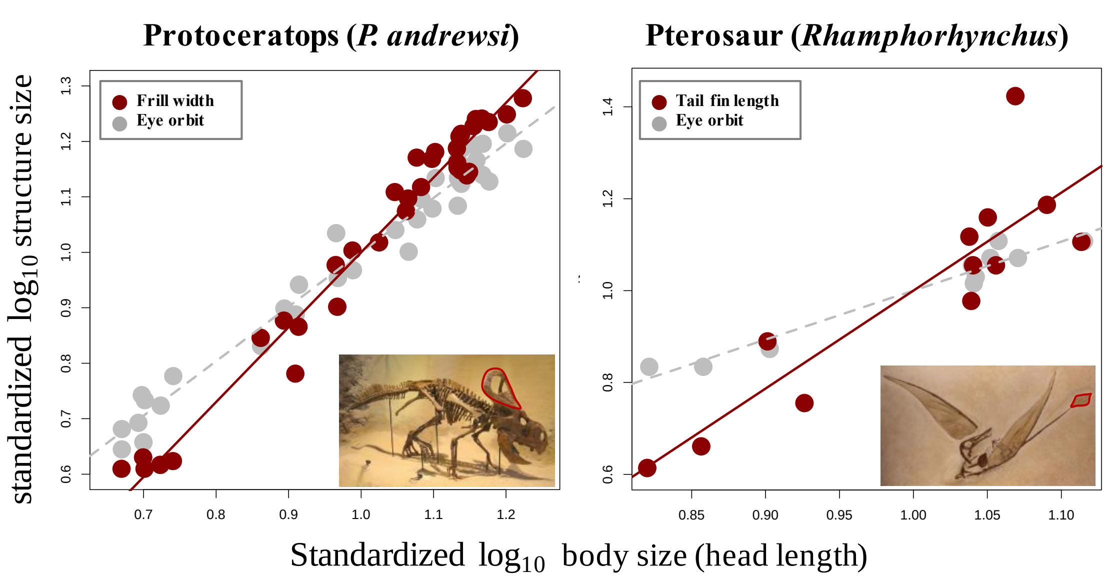


Fig. 3

Table 1: Summary of study species and results.

INCLUDED SEPARATELY AS EXCEL TABLE

Table 1 footnotes

(f) = focal trait, (r) = reference trait, CV = coefficient of variation, β= slope of scaling relationship between trait size and body size, Δβ = difference between $\beta_{(f)}$ and $\beta_{(r)}$, ✝= extinct species, $R^2_{(f)}$ = adjusted R^2 of scaling relationship between focal trait size and body size, $R^2_{(r)}$ = adjusted R^2 of scaling relationship between reference trait and body size, * = sexual dimorphism may be impossible to detect (see D. W. Hone & Mallon, 2017). F and *p* value from ANCOVA comparing $\beta_{(f)}$ and $\beta_{(r)}$ within the same species.

Appendix 1: Identifying extreme morphology

Many extreme structures appear self-evident. Some, like beetle horns, are massive in absolute and relative size and few would contest their designation as extreme. Others are more ambiguous. Butterfly wings, for example, rarely earn the title of extreme yet, their ontogenetic growth and relative size are more akin to that of beetle horns than other insect wings (Nijhout & Emlen, 1998). Examples like this highlight the ambiguity surrounding extreme morphology and the subjective nature of categorizing structures as extreme. This uncertainty, in part, stems from the lack of established criteria for designating a structure as extreme. For over a century, researches have explored the evolution of extreme morphology (M. B. Andersson, 1994; reviewed in Darwin, 1871; Emlen, 2008). Yet, to our knowledge, not once has the term 'extreme' been defined.

Recognizing and limiting bias is a vital component of biological research and, given the large body of work dedicated toward putatively extreme structures, we believe a consistent method for identifying these structures is needed. Here we suggest three (potentially overlapping) categories of extreme - ontogenetically, statically, and evolutionarily extreme – and provide guidelines for assigning structures to each category.

Ontogenetically Extreme**:** Ontogenetically extreme structures are those displaying rates of growth, often occurring in bursts close to reproductive maturity, that outpace other surrounding structures. Examples include the horns of beetles and the wings of lepidopterans, both of which grow to drastic proportions during the same timeframe as other, more typically proportioned structures (Nijhout & Emlen, 1998). Ontogenetically extreme should be distinguished by rates of growth that are faster than those of reference structures within the same organism.

***Statically Extreme*:** Statically extreme structures are disproportionately larger than other structures when sampled across same stage (generally adult) individuals within a population. Relative size of a focal trait can be assessed by comparing the size of the focal trait to other, analogous traits in the same sex (e.g., harlequin beetle (*Acrocinus longimanus*) forelegs are relatively larger than midlegs or hindlegs (Zeh, Zeh, & Tavakilian, 1992)) or by comparing the size of the same trait across sexes (e.g., harlequin beetle forelegs are disproportionately larger in males than they are in females (Zeh et al., 1992)). Statically extreme structures should be distinguished by comparing slopes and/or intercepts of the static scaling relationships (trait size versus body size) of the focal and reference traits.

***Evolutionarily Extreme*:** Evolutionarily extreme structures are extreme when compared with homologous structures in closely related organisms. Examples include the hindlegs of jerboas, which are relatively longer than the hindlegs of their quadrupedal ancestors (Miljutin, 2008; Dipodidae; Wu et al., 2014) and the raptorial forelimbs of mantidflies (mantispidae; Ohl, Barkalov, & Xin-Yue, 2004). Evolutionarily extreme structures can be distinguished by a) comparing static scaling relationships (slopes and/or intercepts) of individuals sampled from populations of ancestral and derived species; b) comparing mean relative trait size of ancestral and derived species (e.g., Wu et al., 2014); and/or c) by mapping changes in trait size onto a phylogeny and testing for lineage specific changes in relative trait size (Wu et al., 2014).

Appendix 2: Sources for *Protoceratops andrewsi* data. AMNH = American Museum of Natural History (New York, USA); MPC = Mongolian Palaeontological Centre (Ulaanbaatar, MN); IVPP = Institute of Vertebrate Palaeontology and Palaeoanthropology (Beijing, CN); ZPAL = Zoological Institute of Paleobiology, Polish Academy of Sciences (Warsaw, PL); CMNH/CM = Carnegie Museum of Natural History (Pittsburgh, USA); NHM = Natural History Museum (London, UK).

Protoceratops andrewsi

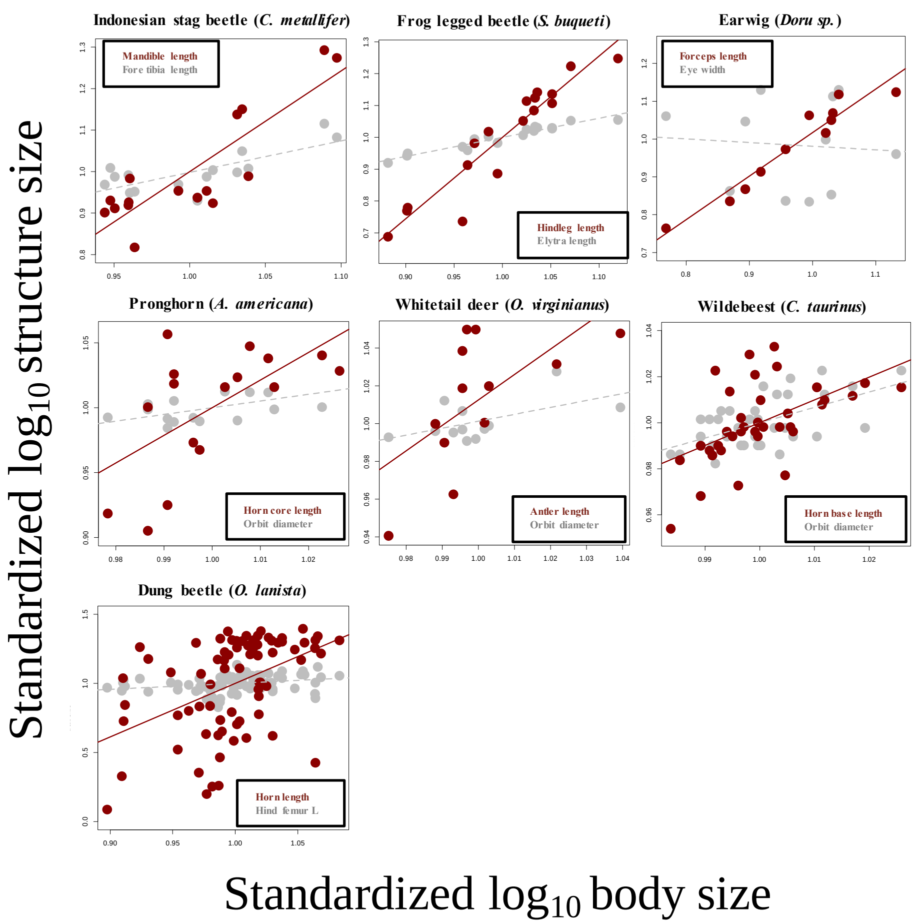
Source	Specimen	Number
Dodson (1976)	AMNH	6419
Dodson (1976)	AMNH	6434
Dodson (1976)	AMNH	6430
Dodson (1976)	AMNH	6251
Dodson (1976)	AMNH	6431
Dodson (1976)	AMNH	6486
Dodson (1976)	AMNH	6432
Dodson (1976)	AMNH	6428
Dodson (1976)	AMNH	6409
Dodson (1976)	AMNH	6480
Dodson (1976)	AMNH	6444
Dodson (1976)	AMNH	6485
Dodson (1976)	AMNH	6408
Dodson (1976)	AMNH	6433
Dodson (1976)	AMNH	6429
Dodson (1976)	AMNH	6439
Dodson (1976)	AMNH	6441
Dodson (1976)	AMNH	6477
Dodson (1976)	AMNH	6417
Dodson (1976)	AMNH	6425
Dodson (1976)	AMNH	6413
Dodson (1976)	AMNH	6414
Dodson (1976)	AMNH	6438
Dodson (1976)	AMNH	6466
Dodson (1976)	AMNH	6467
Handa et al. (2012)	MPC	100/539
Hone et al. (2014)	MPC	100/534
Hone et al. (2014)	MPC	100/526 B
Hone et al. (2014)	MPC	100/526 C
Fastovsky et al. (2011)	MPC	100/530 A
Fastovsky et al. (2011)	MPC	100/530 B
Fastovsky et al. (2011)	MPC	100/530 C
Fastovsky et al. (2011)	MPC	100/530 D
Fastovsky et al. (2011)	MPC	100/530 E
Fastovsky et al. (2011)	MPC	100/530 F
Fastovsky et al. (2011)	MPC	100/530 G
Fastovsky et al. (2011)	MPC	100/530 H
Unpublished photos	IVPP	23899
Unpublished photos	IVPP	Unnumbered Medium
Unpublished photos	IVPP	Unnumbered Small
Unpublished photos	ZPAL	MgD-II/2b
Unpublished photos	ZPAL	MgD-II/5
Unpublished photos	CMNH	9185
Unpublished photos	NHM	5134
Unpublished photos	NHM	6442
Unpublished photos	NHM	6440
Unpublished photos	AMNH	6418
Unpublished photos	AMNH	6637
Unpublished photos	AMNH	6422
Unpublished photos	AMNH	6485

Appendix 3: Sources for *Rhamphorhynchus muensteri* data. BSP = Palaeontological Museum, Munich (Munich, DE); YPM = Yale Peabody Museum (CT, USA) ;SMF = Forschungsinstitut und Naturmuseum Senckenberg (Frankfourt, DE); CMNH/CM = Carnegie Museum of Natural History (Pittsburgh, USA); SOS = Jura Museum (Eichstätt DE); NHM = Natural History Museum (London, UK); TMP = Royal Tyrell Museum of Palaeontology (Alberta, CA); MBR = Museo Argention de Ciencias Naturales (Buenos Aires, AR); BMNS = Brazoport Museum of Natural Science (TX, USA); NMS = National Museums of Scotland (Edinburgh, UK); TPI = Thanksgiving Point Institute (North American Museum of Ancient Life, UT, USA).

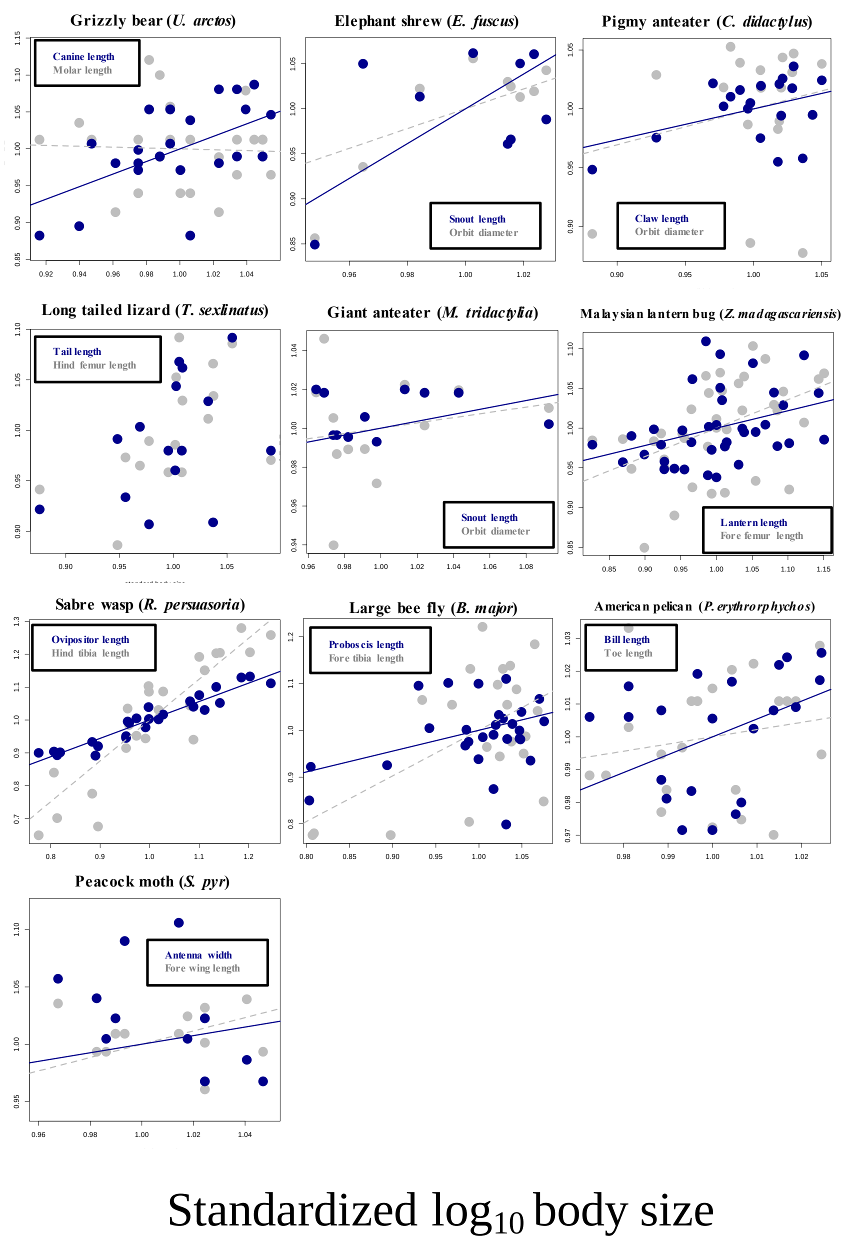
Rhamphorhynchus muensteri

Specimen	Specimen	Number	Wellnhofer number
Wellnhofer (1975)	BSP	1960.I.470	9
Wellnhofer (1975)	BSP	1938.I.503	11
Wellnhofer (1975)	Eichstaett		28
Wellnhofer (1975)	YPM	1778	33
Wellnhofer (1975)	SMF	R 4128	43
Wellnhofer (1975)	CM	11429	53
Wellnhofer (1975)	BSP	1907 I 37	60
Wellnhofer (1975)	SOS	3558	77
Wellnhofer (1975)	NA	NA	102
Hone (2012)	NHM	W1198z0077/0001	
Direct Measurement	TMP	2008.041.0001	
Direct Measurement	MBR	3650.3	
Direct Measurement	BMNS	21	
Measured from photo	NMS	G.1994.13.1.	
Measured from Photo	TPI	1012	

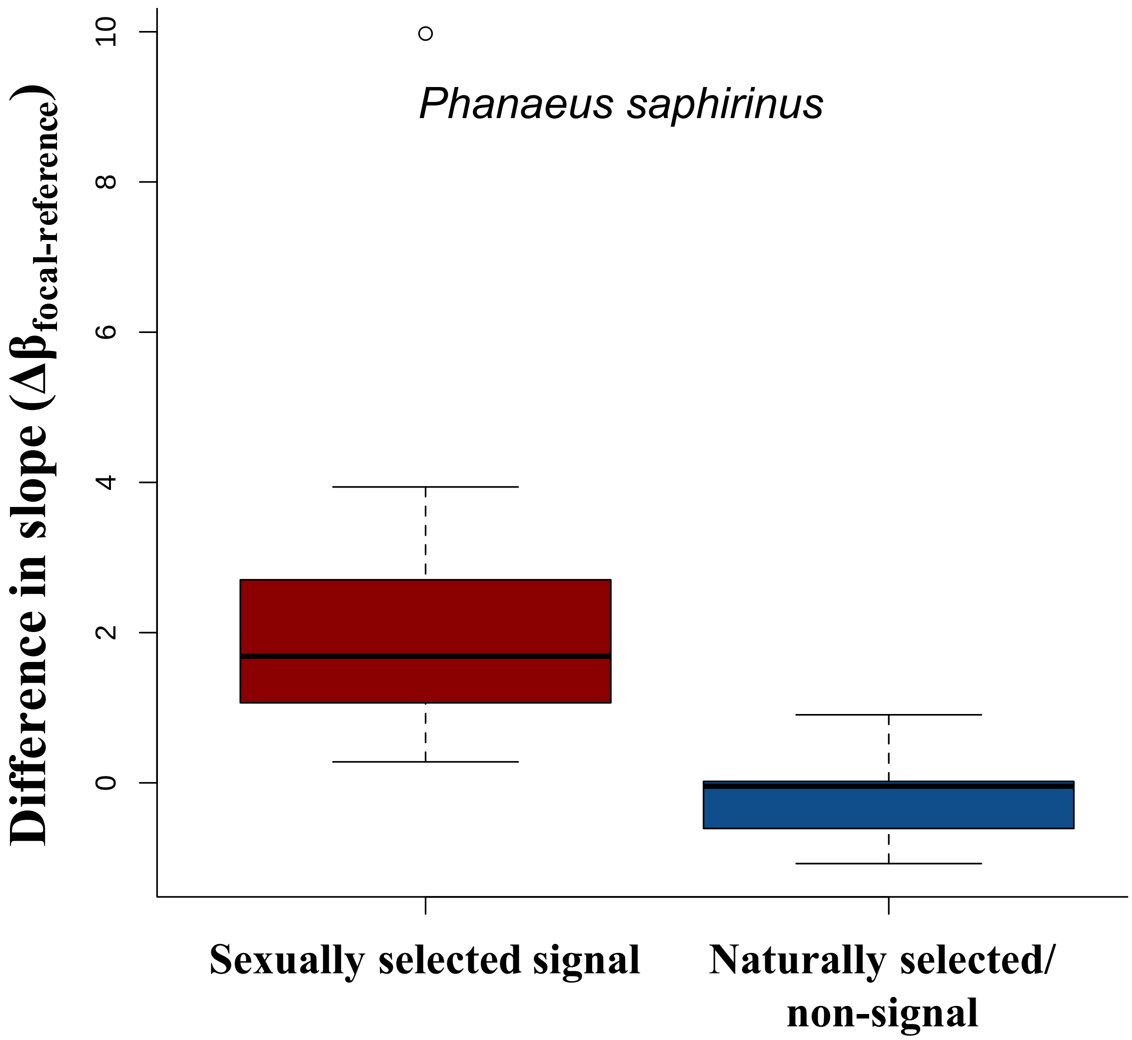
Appendix 4: Scaling relationships for extreme sexually selected signal structures. Lines represent ordinary least squares regression of $\log_{10}$ standardized structure size on $\log_{10}$ standardized body size (slope estimates and sample sizes reported in Table 1). Red points and lines represent focal traits. Grey points and lines represent reference traits.



Appendix 5: Scaling relationships for extreme naturally selected/non-signal structures. Lines represent ordinary least squares regression of $\log_{10}$ standardized structure size on $\log_{10}$ standardized body size (slope estimates and sample sizes reported in Table 1). Blue points and lines represent focal traits. Grey points and lines represent reference traits.



Appendix 6: Comparison of $\Delta\beta_{focal\text{-}reference}$ (difference between the scaling relationship slope of focal traits and reference traits) between extreme sexually selected signal traits (n = 14) and extreme non-signal selected traits (n = 15). $\Delta\beta_{focal\text{-}reference}$ of extreme sexually selected signal structures is significantly greater than $\Delta\beta_{focal\text{-}reference}$ of extreme non-signal structures ($t_{15.616} = 4.153$ $p < 0.001$).



Appendix 7: Comparison of coefficients of variation (CV) between extreme sexually selected signal traits (n = 14) and extreme non-signal selected traits (n = 15). CVs of extreme sexually selected signal structures is significantly greater than CVs of extreme non-signal structures ($t_{16.043} = 3.37$, $p < 0.01$).