2018

On the Evolution of "Extreme" Morphology: Patterns of Cost and Benefit in Sexually Selected Weapons

Devin Mackenzie O'Brien

Let us know how access to this document benefits you.
Follow this and additional works at: https://scholarworks.umt.edu/etd

Recommended Citation
https://scholarworks.umt.edu/etd/11155

This Dissertation is brought to you for free and open access by the Graduate School at ScholarWorks at University of Montana. It has been accepted for inclusion in Graduate Student Theses, Dissertations, & Professional Papers by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact scholarworks@mso.umt.edu.
ON THE EVOLUTION OF "EXTREME" MORPHOLOGY: PATTERNS OF COST AND
BENEFIT IN SEXUALLY SELECTED WEAPONS

By

DEVIN MACKENZIE O'BRIEN

Bachelor of Sciences, University of Connecticut, Storrs, Connecticut, 2013

Dissertation

presented in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in Organismal Biology, Ecology, and Evolution

The University of Montana
Missoula, MT

May 2018

Approved by:

Scott Whittenburg, Dean of The Graduate School
Graduate School

Douglas J. Emlen, Chair
Division of Biological Sciences

Mark Briffa
School of Biological and Marine Sciences, Plymouth University

John P. McCutcheon
Division of Biological Sciences

Bret W. Tobalske
Division of Biological Sciences

H. Arthur Woods
Division of Biological Sciences
© COPYRIGHT

by

Devin Mackenzie O’Brien

2018

All Rights Reserved
On the Evolution of “Extreme” Morphology: Patterns of Cost and Benefit in Sexually Selected Weapons

ABSTRACT

Sexually selected weapons represent some of the most spectacular morphologies in the animal world. They grow out of proportion with body size or other, more typically proportioned structures, and are some of the largest traits in both absolute and relative size. It is therefore unsurprising that animal weapons are some of the most intensely studied structures in biology. Yet, despite this interest, surprisingly little is known about the expression and evolution of these traits. In particular, four questions remain unanswered: How does selection act on weapons in the wild? Do the costs of large weapons ever outweigh the benefits? How are these patterns of cost and benefit reflected in the morphology and development of modern weaponed species? Can we use these patterns to infer the strength and direction of selection when natural observation is unattainable? My dissertation aims to answer these questions by describing the costs and benefits surrounding sexually selected weapons in the wild. I use the frog legged leaf beetle (Sagra femorata) as my primary study system.

In Chapter 1, I provide the first description of S. femorata mating behavior in the wild and provide an explicit measure of selection acting on their hindleg weapons. In Chapters 2 and 3, I investigate factors that may shape patterns of selection observed in Chapter 1 – specifically, biomechanical and metabolic cost. In Chapter 4, I explore broad trends in morphological scaling that result from patterns of selection described in earlier chapters. I review the literature surrounding weapon evolution and propose a new method for characterizing selective history through measures of static morphological scaling. Collectively, this work provides a comprehensive analysis of weapons within and across taxa, expanding our understanding of sexually selected morphology and setting the stage for future studies of sexual selection and morphological evolution.
ACKNOWLEDGEMENTS

This dissertation would have been impossible without the help of many. First, I thank my committee – Doug Emlen, Mark Briffa, John McCutcheon, Bret Tobalske, and Art Woods – for their support and insight throughout the development of this project. In particular, I am grateful to my advisor, Doug Emlen, who, with his unbridled enthusiasm and constant encouragement, guided me through this process largely unscathed. Second, I thank my collaborators in Japan – Teruyuki Niimi, Hiroki Gotoh, Masako Katsuki, and Yasuzaka Okada – without whom I would have been lost during my time aboard. I also thank Ono-san, a kind man who visited my field site every day, bringing me food, water, and great conversation. As an overstressed and overheated graduate student, in a country where I had few friends and spoke little of the language, I valued his company much more than I think he knows. Third, I thank everyone who helped design experiments and collect, curate, and analyze the data presented here. Specifically, Erin McCullough, Christine Miller, Chrissie Painting, Sarah Solie, Jesse Webber, past and present OBEE graduate students, and the current Emlen Lab (Cerisse Allen, Romain Boisseau, Jill del Sol, and Jema Rushe).

Finally, I thank my close friends, Pat Burke, Matt Fuller, John and Liz Grohs, and Steve Lane, my parents, Sean and Mary O’Brien, my sister, Hannah, and especially my wife, Rachel, for their unconditional love and support.

This research was funded by the National Science Foundation (EAPSI, #1515085; Montana IOE), the Japanese Society for the Promotion of Science (EAPSI), Sigma Xi, the Drollinger-Dial Foundation, and the University of Montana.
# TABLE OF CONTENTS

Abstract iii

Acknowledgements iv

Table of Contents v

Overview 1

Chapter 1: Selection on an extreme weapon in the frog-legged leaf beetle (*Sagra femorata*) 4

Chapter 2: Overcoming mechanical adversity in extreme hindleg weapons 46

Chapter 3: Muscle mass drives cost in sexually selected insect weapons 78

Chapter 4: On the evolution of extreme structures: static scaling and the function of sexually selected signals 102
OVERVIEW

Sexually selected weapons are among the most extreme and diverse morphologies in the animal world. They are massive in size, and can grow drastically out of proportion with the body [e.g., 1–5]. It is therefore unsurprising that weapons have captured the attention of biologists for centuries. Darwin, for example, was obsessed with beetle horns [6], and even Aristotle touched on their peculiar form [7]. Yet, despite this intense interest, basic questions pertaining to the nature of sexual selection remain unanswered. For example, is selection on weapon size open-ended and directional? Or do the costs of producing and wielding these structures offset the benefits so that net selection on trait size is stabilizing? How are these patterns of selection reflected in modern populations, and to what extent can we use signatures of sexual selection to infer trait function when natural observation is unattainable? My dissertation aims to answer these questions by describing patterns of cost and benefit surrounding sexually selected weapons in the wild. I use the hindleg weapons of frog legged leaf beetles (*Sagra femorata*) as a focal system.

In Chapter 1 of my dissertation, I provide the first description of frog legged beetle reproductive biology in the wild. I found that male hindleg weapons display extreme sexual dimorphism compared to female hindlegs and scale hyperallometrically with body size. Males use these weapons to compete directly over access to females, where “intruding” males challenge rivals already *in copula*, and attempt to “steal” reproductive mates. I also present an explicit measure of selection acting on hindleg weapons. Previous study [8] and preliminary analyses suggested frog legged beetles may have experienced a history of strong selection for large weapon size. This was supported by the steep hyperallometric scaling relationship between weapon and body size and the critical role weapon size plays in fighting behavior. Upon further investigation, however, selection appears to have stabilized. Instead of directional selection for large weapons sizes, evidence suggests frog legged beetles experience stabilizing
selection for a particular relationship between weapon and body size. That is, selection appears to favor increases in relative weapon size, but this is limited by as-yet undiscovered costs associated with increasingly large weapons.

In Chapters 2 and 3, I explore two factors that may stabilize selection for large weapon sizes—biomechanical and metabolic costs. These were assessed in a variety of insect weapons, including frog legged beetle and leaf footed bug hindlegs, rhinocerous beetle horns, and stag beetle mandibles. I found that males with the largest weapons face intrinsic biomechanical limits to weapon strength, which likely hinder fighting success. This mechanical disadvantage can be overcome by disproportional muscle growth. However, this muscle growth is correlated with disproportional increase in metabolic strain. Overall, the necessity to maintain strong weapons but mitigate metabolic strain may limit selection for exceedingly large weapon sizes and help explain patterns of stabilizing selection observed in weapon bearing species.

In the final chapter of my dissertation, I explore broad trends in morphological scaling that result from the patterns of selection described in earlier chapters. I review the literature surrounding weapon evolution and propose a new method for characterizing a structure’s selective history through measures of static morphological scaling. I demonstrate this method by analyzing a suite of 29 extreme structures across taxa to show how one can reliably infer patterns of selection for static morphological measures.

Overall, this work expands our understanding of extreme sexually selected morphology. I provide one of the few direct measures of selection acting on weapons in the wild, describe the costs and benefits that shape these observed patterns of selection, and explore how these trends hold across animal taxa. This work provides a comprehensive analysis of sexually selected weapons within and across taxa, enriching our understating of extreme morphology in general, and setting the stage for future studies of sexual selection and morphological evolution.
Literature cited


CHAPTER 1

Selection on an Extreme Weapon in the Frog-legged Leaf Beetle (Sagra femorata)

Devin M. O'Brien¹, Masako Katsuki², Douglas J. Emlen¹

¹ Division of Biological Sciences, University of Montana, Missoula, MT, USA
² Department of Arts and Sciences, University of Tokyo, Tokyo, JP


Abstract

Biologists have been fascinated with the extreme products of sexual selection for decades. However, relatively few studies have characterized patterns of selection acting on ornaments and weapons in the wild. Here, we measure selection on a wild population of weapon-bearing beetles (frog legged leaf beetles: Sagra femorata) for two consecutive breeding seasons. We consider variation in both weapon size (hindleg length), and in relative weapon size (deviations from the population average scaling relationship between hindleg length and body size), and provide evidence for directional selection on weapon size per se and stabilizing selection on a particular scaling relationship in this population. We suggest that whenever growth in body size is sensitive to external circumstance such as nutrition, then considering deviations from population-level scaling relationships will better reflect patterns of selection relevant to evolution of the ornament or weapon than will variation in trait size per se. This is because trait size versus body size scaling relationships approximate underlying developmental reaction norms relating trait growth with body condition in these species. Heightened condition-sensitive expression is a hallmark of the exaggerated ornaments and weapons favored by sexual selection, yet this plasticity is rarely reflected in the way we think about – and measure – selection acting on these structures in the wild.
Keywords: Animal weapons, mating systems, reaction norms, sexual selection

Introduction

Since our earliest observations of the natural world, biologists have been obsessed with extremes. Elaborate and beautiful structures, like the antlers of ungulates, have inspired some of the most influential studies in evolutionary biology. Darwin was infatuated with beetle horns and the tails of peacocks, while Wallace spent years collecting extreme forms in the Malay Archipelago [1–3]. The majority of these “exaggerated” structures are products of sexual selection, either ornaments of female choice (intersexual selection) or weapons of male-male battle (intrasexual selection) [2,4]. However, despite considerable interest surrounding exaggerated traits and sexual selection [5,6], we know surprisingly little about how selection acts on ornaments and weapons in the wild.

Basic questions pertaining to the nature of selection acting on ornaments and weapons remain largely unanswered. For example, is selection on ornament and/or weapon size open-ended and directional, favoring larger or relatively larger structures? Or do the costs of producing and wielding these structures offset the benefits so that net selection on trait size is stabilizing in contemporary populations? Decades of theoretical work have addressed these questions [reviewed in 5–7]. Yet, to date, only a few dozen studies have quantified selection on ornament or weapon size in the wild (Table 1) – a surprising dearth considering the spectacular diversity of these structures, and the central role that they play in our understanding of sexual selection. As a result, consensus surrounding the patterns of selection expected for sexually selected structures is lacking [8]. We suggest that part of the problem may be that most researchers (ourselves included) have incorrectly approached sexually selected systems from the perspective of trait size per se, despite evidence suggesting weapon and ornament evolution is most accurately characterized by the heritable relationship between trait size and condition.
Selection on exaggerated ornaments and weapons: What should we expect?

Hypotheses I and II: Selection on trait size

Since Darwin’s initial proposal of sexual selection, the assumption has been that big traits are better than small ones [e.g., 2,9]. Ornaments and weapons look like they should be under strong directional selection for large trait size, and this long-standing view of directional selection has been supported by a large number of theoretical [e.g., 9–12] and empirical (Table 1) studies [4,13–24]. It represents a common, intuitive view of sexual selection and is the default expectation when analyzing selection on ornaments and weapons in the wild [but see 25].

An obvious alternative to directional selection for large trait size is that the reproductive benefits associated with large ornaments and weapons are balanced by natural selection costs associated with bearing large traits. Such costs have been documented in insects [e.g., 26–28], crustaceans [e.g., 29,30], ungulates [e.g., 31,32], and other taxa [33–35], suggesting stabilizing selection may be common. Given many of these traits have been present for millions of years, it is reasonable to assume contemporary populations have reached a point where the costs of bearing large structures now offset their reproductive benefits. If true, then the selective surface we should observe in the wild is one of stabilizing selection on ornament or weapon size, rather than directional selection.

Hypotheses III - V: Selection on reaction norms

Exaggerated weapons and ornaments often function as honest signals of overall quality [5,36,37]. The size, shape, or complexity of these traits amplifies subtle differences between competitors and facilitates the assessment of potential mates and rivals. Critical to signal function is a large degree of developmental plasticity. Almost without exception, the development of exaggerated ornaments and weapons is more plastic (more condition sensitive) than the growth of other body parts [36,38–41].
The condition-sensitive nature of ornaments and weapons has been recognized for decades [42–45], but the implications of this plasticity are often overlooked in studies quantifying sexual selection in the wild [37,43,44,46–48]. For example, the majority of studies to date measure selection on trait size per se – that is, they measure the relationship between performance and variation in ornament or weapon size irrespective of body size and/or condition [11,12,43,44,47,49]. However, as with any exquisitely phenotypically plastic trait, the genetic variation relevant to ornament and weapon evolution likely exists not as heritability of the trait itself, but as differences among genotypes in the way they couple trait growth with nutrition, body size, and/or condition [e.g., 50–52]. Thus, the most appropriate measures of phenotypic selection would consider performance in relation to individual variation in these underlying reaction norms.

Experimental evidence (e.g., rearing full siblings on high or low nutrition) suggests that individuals modulate expression of ornament or weapon size in a manner consistent with a developmental norm of reaction [36,41,51,53–55]. In addition, artificial selection experiments show that relationships between trait and body size are highly heritable and can evolve in response to selection [50,51,56]. Together, these studies support a view of exaggerated ornaments and weapons as plastic traits whose expression is governed by reaction norms. That is, the adaptive trait is not weapon/ornament size per se, but the shape of the underlying reaction norm. If true, then considering the ornament or weapon by itself would be misleading. Instead, variation in where animals fall relative to the scaling relationship of the population, or relative trait size, would be more appropriate for measures of selection - since for many sexually selected traits, the static scaling relationship between trait and body size approximates the reaction norm between trait size and condition [8,57] (Fig. 1).

Selection on the reaction norm relating ornament or weapon growth to condition can be proportional (i.e., selection to increase the intercept of trait size versus body size that would
change the proportion of that trait relative to other traits or body size, and it would do so for all individuals; Hypothesis III) or correlational (where the strength and direction of selection acting on the trait is correlated with body size, thereby altering the slope of the scaling relationship; Hypothesis IV; [8]). Selection on the reaction norm can also be stabilizing, if individuals on or close to the population average relationship have higher fitness than those that deviate above or below (Hypothesis V).

Proportional selection on reaction norms is expected when increases in trait size are similarly favored across the entire population. Trends like these have been demonstrated [50,58,59], but their presence in natural, sexually selected systems is likely rare. This is because selection within these systems is seldom uniform. For example, animals developing under ideal conditions (big individuals) benefit from growing large ornaments or weapons, while individuals developing under poor conditions (small individuals) do best by not investing in expensive structures that would prove useless due to their stature. In this situation, we expect selection to favor animals with ornaments or weapons that are relatively larger than those of their rivals when body sizes are big, and relatively small when body sizes are small [37,46,48,60]. As a result, theoretical treatments of ornaments and weapons suggest that selection for honest signaling will favor the evolution of increasingly steep scaling relationship slopes [37,46,48,61]. Steep slopes result from a combination of benefits to large traits in the best-condition individuals and disproportionate costs of those same traits in the poorest quality individuals [37,44,46,48,60,62].

Alternatively, populations may have reached a balance, where the costs of increased ornament/weapon size offset the benefits across the available range of body sizes (Hypothesis V). This pattern of stabilizing selection would be evident if individuals falling on the population average scaling relationship have higher fitness than those deviating in either direction. To our knowledge, only one study has directly measured mating success as a function of deviations
from a population-level scaling relationship [56]. This study found clear evidence for stabilizing selection, but focused on a trait (butterfly wing size) that is shaped primarily by natural selection (locomotion) and only secondarily by sexual selection (mate choice). So, strong stabilizing selection is expected. Whether proportional, correlational, or stabilizing selection is likely for exaggerated ornaments and weapons – traits functioning as signals in the context of sexual selection – is far from clear, and awaits studies that quantify selection on the reaction norms, rather than on trait size per se.

Here, we analyze these five hypotheses for how selection acts on an extreme, and extremely variable, sexually selected weapon: (I) directional selection on weapon size (Fig. 2 A), (II) stabilizing selection on weapon size (Fig. 2 B), (III) proportional selection on reaction norms (increased reaction norm intercept; Fig. 2 C), (IV) correlational selection on reaction norms (increased reaction norm slope; Fig 2 D), and (V) stabilizing selection on a particular reaction norm (Fig. 2 E). Using field measures of mating success from a natural population of frog-legged leaf beetles (Sagra femorata, family Chrysomelidae), we examine individual performance in the context of each of these alternatives and discuss our results as they pertain to agents of selection acting on weapons and ornaments in the wild.

Methods

Study site and population

All observations were conducted on a wild population of frog legged beetles along the Kushida River in Matsuzaka, Mie Prefecture, Japan. Data were collected over two breeding seasons, 2014 (July-August) and 2015 (June-August). This study is the first we are aware of to describe the natural history and reproductive behavior of frog legged beetles in the wild [22].
**Frog legged beetle behavior and natural history**

Frog legged beetles (*Sagra femorata*) are large, iridescent beetles found throughout Southeast Asia (Fig. 3 A and B). Populations form in large aggregations (typically 300-400 individuals), where most adults remain for their one-year life cycle. Populations appear to be consistent through time, with multiple generations aggregating on the same plants across multiple seasons. Aggregations form on small (3mx5m) patches of low vines (commonly kudzu; *Pueraria spp.*), where beetles feed, compete, and reproduce (Fig. 3 D). This clustering behavior allowed us to collect observations of behavior and mating success for the entire population across multiple seasons.

Frog legged beetles are not host specific and host plant varies throughout the beetle’s range (personal observation; personal comm., Dave Furth – October 2013). Adults chew wounds in thick, woody sections of vine and feed on oozing sap (Fig. 3 B). These “feeding sites” remain active from several days to several weeks and are abandoned once they stop producing fresh sap. The number of beetles feeding and turnover of beetles at feeding sites is highly variable. Beetles commonly feed alone and in small groups (4-10 animals) and feeding time varies from under one hour to several days. No guarding behavior of feeding sites has been observed.

During the breeding season, male frog legged beetles use their sexually dimorphic and exaggerated hindlegs (Fig. 3 A) as weapons to compete directly over access to females. Male weapons are only used in battle, and are not functional legs. In fact, during terrestrial locomotion, males drag their weapons behind them, using only their fore- and mid-limbs to walk (females use all six walking legs).

Competition between males takes place one-on-one and begins when one male (intruder) approaches a rival (resident) already *in copula* with a female (Appendix 1.1: male-male competition). Fights can progress in two ways. 1) The intruder uses his hindlegs to reach around the abdomen of the resident, administering a rapid succession of squeezes. The
resident, in turn, wraps his hindlegs around the weapons of his opponent, retaliating with similar behavior. 2) The intruding male attaches himself to the prothorax of the resident male using sexually dimorphic fore- and midlimb tarsi and wraps his legs under the vine upon which his rival and potential mate are attached [22]. The intruder then contracts his legs, as if he were squeezing the vine. This hinges the intruder, allowing him to lift his opponent off of the female (featured in Appendix 1.1: male-male competition). Throughout competition, females remain passive. The winning male mates with the female and guards her from subsequent mating for approximately one hour (Fig. 3 A). Losing males are either forcibly removed from the fighting area or retreat in search of another mate. Both males and females mate multiple times throughout the season.

Females lay their eggs in the stems of host plants. During development, larvae form galls in plant stems, where they feed and develop until mid-autumn and overwinter at the final larval stage (Fig. 3 C). Pupation begins in the spring, and adults emerge from galls from mid-June through early July. Newly emerged adults immediately begin feeding and competing over mates.

Behavioral observations:

At the beginning of each breeding season all visible beetles were captured, numbered, and rereleased. Each individual was assigned an identification number, drawn on the beetle’s right elytron with Uni® oil based paint markers. Markings did not appear to harm the animals or alter their behavior. As new, unmarked beetles emerged as adults, they were measured and numbered until the entire population was identified (Appendix 1.2: capture data for 2014 and 2015 field seasons). As the season progressed, markings were assessed and, if necessary, redrawn to limit degradation of markings and misidentification of beetles. Once individuals were marked and measured they were returned to the same branch from which they were collected.
At the time of initial capture, key morphological traits were measured. Measurements were based on preliminary behavioral observations collected on sourced beetles from LPS imports (Denver, CO) and Tropical Entomological House (Penang, Malaysia), and measurements collected in Katsuki et al. [22]. Measurements included elytra length (EL) as a metric of body size, and hindleg femur length (FL) as a metric of weapon size. Measurements were collected using digital or dial calipers and on the left leg when possible.

Observations of behavior were prioritized by reproductive activity. The population was surveyed for reproductive activity at 30 minute intervals. Once spotted, fighting or courting beetles were observed in close proximity until the behavior ended. Mating was considered complete when the mating male and female separated. Between surveys, focal females were followed and reproductive activity was recorded. This allowed the observation of reproductive males that may have been overlooked in surveys due to their small size and/or subtle behavior. Observations were conducted with minimal disturbance, and physical contact was avoided when possible. Observations were collected during hours of peak activity (typically from 5:30 – 13:00).

Statistical analyses:

All statistical analyses were performed in R 3.2.4 (R Core Development Team, 2016). Morphological data from the two field seasons were compared using two sample \( t \)-tests to compare means and Kolmogorov-Smirnov (KS) tests to compare trait distributions. No significant differences in morphology were found between data from the 2014 and 2015 field seasons (Appendix 1.3: comparison of 2014 and 2015 field seasons). Data from the two seasons were then combined for the remainder of analyses. In addition, a term representing year/season was incorporated into each of the models to control for the effect of year/season on patterns of selection, but these terms were non-significant for every model \( (p > 0.3 \text{ for all}) \) and were dropped from further analyses.
All morphological measures were log_{10} transformed before analysis. Major-axis (MA) regression was used to assess allometric relationships given the approximately equal error in measures of trait and body size ("smatr" package in R, Warton, D. [2005]). Using significance tests built into the "smatr" package in R, estimates of intercept were assessed using a Wald test and estimates of slope were compared using a Likelihood-ratio test. Residual values from MA models were collected as a estimate of relative weapon size.

All models were conducted using both measured values of weapon and body size and the principle components of weapon and body size. Analyses using principle components yielded similar results to those using measured values. Since measured values are more biologically intuitive than principle components, we only report those models constructed using measured values.

Table 2 shows all models used in analyses of selection. Ordinary least squares (OLS) regression was used to determine support for Hypotheses I-V in accordance with the methods described by Lande and Arnold [63] and Arnold and Wade [64]. Models only incorporating linear terms were used to assess directional selection. Models incorporating both linear and quadratic terms of the explanatory variable were used to assess patterns of stabilizing selection. In each analysis, relative mating success was used as a continuous response variable. Relative mating success was calculated for each male as the number of observed inseminations across the entire breeding season (ranging 0-5) divided by the average number of inseminations per male in the population across the entire breeding season. The Akaike information criterion (AIC) and significance of model coefficients were used to compare models.

Hypothesis I (directional selection on weapon size) and Hypothesis II (stabilizing selection on weapon size), were assessed by regressing mating success on weapon size using both linear and quadratic representations of weapon size. This allowed for the assessment of Hypotheses I and II simultaneously. Quadratic components of the model (representing Hypothesis II) were not significant, however, and were therefore removed from the model.
Hypothesis I was then assessed using a simple linear regression of mating success on weapon size.

Hypothesis III and IV were assessed by regressing relative mating success simultaneously on three parameters; weapon size, body size, and the interaction between weapon and body size. Hypothesis III, proportional selection on reaction norms, was assessed using the regression coefficients representing weapon size (after controlling for body size). Hypothesis IV, correlational selection on reaction norms, was assessed using regression coefficients representing the interaction between weapon and body size.

Hypothesis V, (stabilizing selection on reaction norms) was assessed by regressing relative mating success on relative weapon size using both linear and quadratic representations of relative weapon size (Table 2).

Because of the relatively low statistical power associated with analyses of stabilizing selection [e.g., see 25], we used two additional approaches to test Hypothesis V. First, we regressed mating success on absolute residual weapon size using only linear regression coefficients. This allowed us to assess the presence of stabilizing selection without incorporating quadratic coefficients into our analyses, increasing the statistical power of our test from 0.309 (2 coefficients, $R^2 = 0.0063$, $n = 446$, $\alpha = 0.05$) to 0.429 (1 coefficient, $R^2 = 0.0069$, $n = 446$, $\alpha = 0.05$). Second, we analyzed differences in variance in relative weapon size between mated and non-mated males in the population (treated as a binary response variable where males either successfully or unsuccessfully inseminated at least one female throughout the breeding season) using Levene’s test on the residual values from MA regression. As stabilizing selection is predicted to reduce variation within a population, we believe this complementary test will be generally useful in studies quantifying stabilizing selection in wild populations.

Finally, selection differentials and gradients were calculated on weapon size and relative weapon size to measure the strength and direction of selection associated with Hypotheses I and V. Selection differentials were calculated as the difference in mean weapon size/relative
weapon size of mated and non-mated animals. Selection gradients were calculated as the partial regression coefficients from OLS regression of relative mating success on weapon size/relative weapon size multiplied by two [63,65].

**Results**

*Morphology and scaling of frog legged beetles*

For male and female frog legged beetles femur length increased linearly with body size (Fig. 4). Wald Tests, Likelihood-ratio tests, and comparisons of confidence intervals confirm that these scaling relationships are significantly different between males and females (Table 2). For males, body size ranged from 9.0mm to 15.6mm and weapon size ranged from 5.5mm to 12.0mm. For females, body size ranged from 7.4mm to 14.3mm and femur length (a comparable measure to weapon size in males) ranged from 4.5mm to 9.9mm.

*Analyses of selection*

We found the greatest support for our first and fifth hypotheses, directional selection on weapon size (Hypothesis I) and stabilizing selection on reaction norms (Hypothesis V).

In support of Hypothesis V, the regression of relative mating success on absolute residual weapon size provided the best fit for the data (AIC = 1676.513) and showed a trend towards significance ($F_{1,444} = 3.088$, $p = 0.08$) (Fig. 5 A; Table 2). The regression of relative mating success on residual weapon size also fit the data well (AIC = 1678.785) and the quadratic regression coefficient in the model showed a trend towards significance ($t_{2,443} = -1.662$, $p = 0.097$) (Fig. 5 B; Table 2). In addition, mated animals showed reduced variation in relative weapon size compared to non-mated animals (variance of mated animals = 0.0005; variance of non-mated animals = 0.001; Levene's test $F_{1,455} = 4.17$, $p = 0.042$) (Fig. 6). For relative weapon size, the directional selection differential was 0.00037 and the quadratic selection gradient was $-5.2515 \pm 3.16028$. For the absolute value of relative weapon size, the
directional selection differential was -0.0029, and the directional selection gradient was -6.107 ± 3.4756.

In support of Hypothesis I, the regression of relative mating success on weapon size provided a similar fit to the data (AIC = 1676.908; Table 2) compared to the regression of relative mating success on absolute residual weapon size, and there was a trend toward significance (F_{1, 144} = 2.692, p = 0.102). The directional selection differential on weapon size was 0.0044, the directional selection gradient was 2.334 ± 1.422.

We found no support for our other hypotheses, stabilizing selection on weapon size and correlational and proportional selection on reaction norms. The model coefficients describing our second hypothesis did not approach significance (t_{2, 443} = -0.475, p = 0.635; Table 2). In models describing Hypothesis III and IV, neither the coefficient describing weapon size (t_{3, 442} = -0.264, p = 0.792) nor the interaction term between weapon and body size (t_{3, 442} = 0.312, p = 0.755) approached significance (Table 2).

**Discussion**

*Stabilizing selection on reaction norms*

We analyzed patterns of selection using raw and relative trait size approaches (Fig. 2). Our first two hypotheses, directional and stabilizing selection on weapon size, addressed the classic view of selection acting on trait size *per se*. Our remaining three hypotheses, proportional, correlational, and stabilizing selection on relative weapon size, considered sexually selected traits as reaction norms.

We found the greatest support for stabilizing selection on relative trait size (Hypothesis V; Fig. 5). In frog legged beetles, selection acts to stabilize the reaction norm between weapon size and body size. Mated animals show reduced variation in relative trait size. Steep scaling relationships in this population, relative to females (Fig. 4; Table 3) suggest a history of selection for steep reaction norms, probably resulting from differential costs/benefits to large
and small individuals [37,46,48,60] or persistent mating success of small males with small weapons [61]. However, we see no evidence that selection is acting to increase the intercept or steepness of these relationships in present day populations. In addition, we found similar support for directional selection on weapon size (Hypothesis 1), suggesting either a history of selection for increased weapon size, or a combination of directional selection on weapon size per se and stabilizing selection on the reaction norm relating weapon size to body size and/or condition.

Overall, we suggest frog legged beetles experienced a history of strong selection for increased weapon size, which led to the evolution of extreme trait size, heightened condition-sensitive (plastic) weapon growth, and a steep scaling relationship between weapon and body size in males. Over time, however, we suspect these animals reached a balance, where increasing costs to males with the largest weapons (i.e., costs associated with additional increases in the steepness of the scaling relationship) began to offset the reproductive advantages of increased relative weapon size. This stalled the evolution of trait reaction norms and stabilized the population around the present-day allometry. In modern populations, our data indicate that selection acts to reduce variation in residual weapon size so animals experience stabilizing selection for the existing scaling relationship between weapon and body size.

**Agents of selection**

The benefits of large hindleg weapons are clear. Beetles with the largest weapons perform best in combat (personal observation) and evidence suggests they have increased mating success as a result. However, selection for large weapons is not experienced equally across a population. The benefits of big weapons should be disproportionately higher for those animals with the largest body sizes and weapons [46,48,60], increasing the slope of reaction norms in addition to the intercept (Fig. 2; blue arrows).
The patterns of selection observed here suggest that costs (i.e., limits to fitness resulting from large traits) may offset the benefits of the biggest weapons (Fig. 2: red arrows). The costs of bearing sexually selected traits have been demonstrated repeatedly [66,67,29,34, but see 68], and theory predicts ornaments and weapons should be costly [37,69,70]. Costs have yet to be quantified in frog legged beetles, but two types of costs in particular are likely to be relevant.

First, animals may experience a cost to reproduction and survivorship resulting from the way their weapons influence performance outside of battle. Large weaponed animals perform well in fights, but large muscles associated with powerful weapons may impede critical processes such as predator avoidance and mate searching, and disproportionately increase resting and active metabolic rate. For example, large muscle mass negatively influences flight performance in stag beetles [28] and significantly increases metabolic rate in fiddler crabs [29].

Second, animals may experience a cost to performance during combat, a trend likely driven by mechanical limits to weapon strength. In many weapon systems (including frog legged beetles), weapon performance is directly related to fighting and reproductive success [13,14,18,20,71]. Large weaponed animals are predicted to experience a decrease in weapon strength through disproportional changes in the physical components of weapons as they become increasingly large [72–74]. This trend is grounded in simple lever physics where increases/decreases in one component of a lever require proportional change in another to maintain performance. For example, as the squeezing surface of a lever moves farther from the fulcrum of that lever system, as it would in an increasingly long hindleg weapon, the force exerted by that lever would decrease [75]. In principle, animals could compensate for this mechanical disadvantage by increasing the cross sectional area of the squeezing muscles or increasing the length of other components of the lever system [74,76]. However, space constraints within the animal or structure may place an upper bound on this growth, ultimately limiting the performance of the weapon system.
Under these conditions, net selection would act to push weapon growth reaction norms toward a relationship that balances the benefits and costs of relative weapon size. We suggest that for frog legged beetles, the shape of this reaction norm is approximated by the observed population static allometry. While the agents of selection outlined here are likely candidates for the drivers of our observed pattern of stabilizing selection, they do not represent a comprehensive analysis of costs and further work will be required to determine the specific agents of selection shaping weapon evolution in this system.

Going forward

Despite a long-standing interest in sexually selected traits, our understanding of how selection acts on exaggerated ornaments and weapons remains unclear. For example, if the genetic variation relevant to the evolution of ornaments and weapons exists primarily as variation in reaction norms relating trait growth with condition (as it should, given the notoriously low heritability of these traits and their exquisite condition sensitivity), then measures of selection focusing on variation in trait size per se may prove inaccurate [77]. Yet, the overwhelming majority of studies do just that. Despite over 35 years of research supporting a reaction norm view of sexually selected traits, only 17% of studies to date measure selection on anything other than ornament/weapon size, and none explicitly incorporate the reaction norm perspective (Table 1).

One probable reason for this is that accurately describing reaction norms responsible for ornament or weapon growth remains impractical for the majority of non-model systems. Performing the environmental manipulation experiments that are standard for estimating reaction norms in laboratory systems is impossible in most cases. Instead, researchers must depend on more laborious approaches to quantify these relationships, such as long-term pedigree analyses (T. Frankino, personal comm.)
Another significant hurdle is that for many animals there is no readily measurable proxy for condition [78–82]. In many insect systems, body size is overwhelmingly driven by larval or nymphal access to nutrition, a product of dominance in competitive interactions and maternal effects (e.g., eggs placed in the best possible locations at the best possible times [83,84]), which reflect variation in overall genetic quality [85–87]. Body size in these same species is often critically important for fighting success and dominance interactions. For these reasons, body size may serve as an easily quantified substitute for individual condition, and variation around the ornament/weapon size versus body size scaling relationship may be interpreted loosely as variation in the individual underlying reaction norms [8,50,51,57]. In other commonly studied systems, such as ungulates, the relationship between body size and condition will be confounded by age and/or high heritability [e.g., 88,89]. As a result, interpreting variation around a trait-size versus body size scaling relationship is more problematic. Body size [90] may still be the best available predictor of condition in these species, but future work will need to establish this.

Nevertheless, we maintain that a reaction norm view of ornaments and weapons is appropriate when studying the evolution of sexually selected traits. We use this logic here, and interpret selection on this scaling relationship as phenotypic selection for a particular form of conditional expression. The steep scaling relationship of male hindlegs (compared to females of the same species) is consistent with a history of directional selection for increased weapon size combined with disproportionate costs for small, poor-condition males [37,44,47,62]. However, our measures of mating success suggest that the benefits of increased weapon size are now balanced by as-yet-undescribed costs, such that males with hindleg sizes close to the population mean scaling had higher mating success than males with either relatively larger, or relatively smaller, weapon sizes. Even this is only a rough approximation, however, since it does not include metrics for estimating the shapes of the individual reaction norms, such as sibling or pedigree analyses, and measures of mating success remain an approximation of reproductive
success when true paternity is unknown. We suggest that future work be aimed at developing reliable and practical means of quantifying individual scaling relationships and paternity in non-model systems, allowing us to better connect performance in the wild with meaningful variation in sexually selected trait size.
**Author contributions:** DMO: Conception and design of the work, data collection, data analysis, drafting of article, final approval of article. MK: Significant assistance in data collection and analysis. DJE: Critical insight into execution of the work and data analysis, assistance in revision of article.

**Acknowledgements** We would like to thank the NSF and JSPS for funding, Sarah Solie for her assistance during field analyses, Teruyuki Niimi, Hiroki Gotoh, and Yasukazu Okada for their support in Japan, and Jesse Webber, Cerisse Allen, Steve Lane, Wenfei Tong, Jill del Sol, Romain Boisseau, and Jema Rushe for feedback on early drafts of the manuscript. The authors have no conflicts of interest to declare.
Literature cited

2. Darwin C. The Descent of Man: And Selection in Relation to Sex. Appleton; 1871.
3. Wallace AR. The Malay Archipelago: the land of the orang-utan and the bird of paradise; a narrative of travel, with studies of man and nature [Internet]. Courier Corporation; 1869.


25


83. Miller CW. Seasonal effects on offspring reproductive traits through maternal oviposition behavior. Behav Ecol. 2008;19: 1297–1304.


91. Falconer DS. Selection in different environments: effects on environmental sensitivity (reaction norm) and on mean performance. Genet Res. 1990;56: 57–70. doi:10.1017/S0016672300028883


Fig. 1: A) Relationship between environmental condition and trait size in a typical, condition-dependent sexually selected trait. Filled circle represents the realized phenotype of a single individual. Red line represents the cryptic genotype (reaction norm) of that same individual. Open circles depict alternative phenotypes that could have been produced by this genotype had the individual completed development at a different body condition [91]. B) Heritable variation relevant to the evolution of condition-sensitive ornaments and weapons exists as cryptic differences among genotypes in the precise relationship between trait size and condition. Realized phenotypes for each genotype indicated by filled circles [8]. C) The population allometry (grey line), estimated from realized phenotypes of individuals (filled circles), may serve as a proxy for the population average of the underlying developmental reaction norms. This estimation is especially relevant for systems where it is not possible/practical to quantify the shapes of the individual reaction norms.
Fig. 2: Five hypotheses for how selection acts on sexually selected traits in the wild. Blue arrows indicate positive selection. Red arrows indicate negative selection. A) Directional selection for increased weapon size. B) Stabilizing selection on weapon size. C) Proportional selection on the reaction norm of weapon and body size (increasing intercept). D) Correlational selection on the reaction norm (increasing slope). E) Stabilizing selection on the reaction norm. Note: at equilibrium, the strength of selection in each direction is likely similar for stabilizing selection.
Fig. 3: A) Mating pair of *Sagra femorata*. Male on top, female on bottom. B) 3 female *S. femorata* feeding on the stem of kudzu (*Pueraria spp*). C) *S. femorata* galls (red arrows) in stem of kudzu. D) Field location in Matsuzaka, Mie Prefecture, Japan. Entire population clustered within the frame. Photos: D. O’Brien.
Fig. 4: Static allometry of male (closed circles, solid line) and female (open circles, dotted line) *S. femorata* femur length (weapon in male beetles). Lines represent major axis regression.
Fig. 5: A) Relative mating success vs. absolute residual weapon size. Dotted line represents the ordinary least squares (OLS) linear regression of relative mating success on absolute residual weapon size. B) Relative mating success vs. residual weapon size. Dotted line represents the curvilinear OLS regression of relative mating success on residual weapon size. In both panels, males with the lowest residual weapon size tend to have the highest mating success in the population.
Fig. 6: Stabilizing selection on relative weapon size in male *S. femorata*. Mated males show reduced variation in relative weapon size compared to non-mated males. Points of male allometry plot color coded by mating success. Dotted line major axis (MA) regression between body size and weapon size.
Table 1: Examples of selection studies on wild, sexually selected systems. In each study listed, selection was estimated as a function of trait size. Surveyed studies are focused on structure size (e.g., tail length, horn length, etc.), rather than signals of color or overall body size. Only 17% of studies surveyed measured selection on anything other than trait size per se.

<table>
<thead>
<tr>
<th>Publication</th>
<th>Organism</th>
<th>Focal trait</th>
<th>Treatment of trait size</th>
<th>Pattern of selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andersson (1982)</td>
<td>Bird</td>
<td>Tail length</td>
<td>Absolute</td>
<td>Directional</td>
</tr>
<tr>
<td>Brodsky (1988)</td>
<td>Bird</td>
<td>Comb size</td>
<td>Absolute</td>
<td>Directional</td>
</tr>
<tr>
<td>Andersson (1989)</td>
<td>Bird</td>
<td>Tail length</td>
<td>Absolute</td>
<td>Directional</td>
</tr>
<tr>
<td>Barnard (1990)</td>
<td>Bird</td>
<td>Tail length</td>
<td>Absolute</td>
<td>Directional</td>
</tr>
<tr>
<td>Møller (1993)</td>
<td>Bird</td>
<td>Tail length</td>
<td>Absolute</td>
<td>Directional</td>
</tr>
<tr>
<td>Petrie et al. (1991)</td>
<td>Bird</td>
<td>Tail length</td>
<td>Absolute</td>
<td>Directional</td>
</tr>
<tr>
<td>Smith and Montgomery (1991)</td>
<td>Bird</td>
<td>Tail length</td>
<td>Absolute</td>
<td>Directional</td>
</tr>
<tr>
<td>Smith et al. (1991)</td>
<td>Bird</td>
<td>Tail length</td>
<td>Absolute</td>
<td>Stabilizing</td>
</tr>
<tr>
<td>Andersson (1992)</td>
<td>Bird</td>
<td>Tail length</td>
<td>Absolute</td>
<td>Directional</td>
</tr>
<tr>
<td>Evans and Hatchwell (1992)</td>
<td>Bird</td>
<td>Tail length</td>
<td>Absolute</td>
<td>Directional</td>
</tr>
<tr>
<td>Hannon and Eason (1995)</td>
<td>Bird</td>
<td>Comb area</td>
<td>Absolute</td>
<td>Directional</td>
</tr>
<tr>
<td>Møller (1995)</td>
<td>Bird</td>
<td>Tail length</td>
<td>Absolute</td>
<td>Directional</td>
</tr>
<tr>
<td>Yasmin and Yahya (1996)</td>
<td>Bird</td>
<td>Tail length</td>
<td>Absolute</td>
<td>Directional</td>
</tr>
<tr>
<td>Saino et al. (1997)</td>
<td>Bird</td>
<td>Tail length</td>
<td>Absolute</td>
<td>Directional</td>
</tr>
<tr>
<td>Ligon et al. (1998)</td>
<td>Bird</td>
<td>Comb size</td>
<td>Absolute</td>
<td>Directional</td>
</tr>
<tr>
<td>Jones and Hunter (1999)</td>
<td>Bird</td>
<td>Crest length</td>
<td>Absolute</td>
<td>Directional</td>
</tr>
<tr>
<td>Pryke et al. (2001)</td>
<td>Bird</td>
<td>Tail length</td>
<td>Absolute</td>
<td>Directional</td>
</tr>
<tr>
<td>Safran and McGraw (2004)</td>
<td>Bird</td>
<td>Tail length</td>
<td>Absolute</td>
<td>None</td>
</tr>
<tr>
<td>Takahashi et al. (2008)</td>
<td>Bird</td>
<td>Tail length</td>
<td>Absolute</td>
<td>None</td>
</tr>
<tr>
<td>Basolo (1990)</td>
<td>Fish</td>
<td>Tail length</td>
<td>Absolute</td>
<td>Directional</td>
</tr>
<tr>
<td>Haines and Gould (1994)</td>
<td>Fish</td>
<td>Tail length</td>
<td>Absolute</td>
<td>Directional</td>
</tr>
<tr>
<td>Rosenthal and Evans (1998)</td>
<td>Fish</td>
<td>Tail length</td>
<td>Absolute</td>
<td>None</td>
</tr>
<tr>
<td>McClintock and Uetz (1996)</td>
<td>Arachnid</td>
<td>Tuft size</td>
<td>Absolute</td>
<td>Directional</td>
</tr>
<tr>
<td>LeBas et al. (2003)</td>
<td>Insect</td>
<td>Pinnate area</td>
<td>Absolute</td>
<td>Directional</td>
</tr>
<tr>
<td>Wheeler et al. (2012)</td>
<td>Insect</td>
<td>Pinnate scales</td>
<td>Absolute</td>
<td>Non-linear</td>
</tr>
</tbody>
</table>

**Ornaments of female choice**

**Weapons and signals of male–male competition**
Table 1 (continued):

<table>
<thead>
<tr>
<th>Publication</th>
<th>Organism</th>
<th>Focal trait</th>
<th>Treatment of trait size</th>
<th>Pattern of selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Painting and Holwell (2014)</td>
<td>Insect</td>
<td>Rostrum length</td>
<td>Absolute</td>
<td>None</td>
</tr>
<tr>
<td>Katsuki et al. (2014)</td>
<td>Insect</td>
<td>Leg size</td>
<td>Absolute</td>
<td>Directional</td>
</tr>
<tr>
<td>Erict and Gwynne (2015)</td>
<td>Insect</td>
<td>Head size</td>
<td>Relative to body size</td>
<td>Directional</td>
</tr>
<tr>
<td><strong>Dual function traits (ornament of female choice and weapon/signals of male–male competition)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loyau et al. (2005)</td>
<td>Avian</td>
<td>Tail length</td>
<td>Absolute</td>
<td>Directional</td>
</tr>
<tr>
<td>Oliveira and Castódio (1998)</td>
<td>Crustacean</td>
<td>Claw size</td>
<td>Absolute</td>
<td>Directional</td>
</tr>
<tr>
<td>Latruffe et al. (1999)</td>
<td>Crustacean</td>
<td>Claw size</td>
<td>Absolute</td>
<td>Directional</td>
</tr>
<tr>
<td>Panhuis and Wilkinson (1999)</td>
<td>Insect</td>
<td>Eye-stalk length</td>
<td>Relative to body size</td>
<td>Directional</td>
</tr>
</tbody>
</table>
Table 2: Models constructed for the testing of Hypotheses I-V including the AIC for each model and parameter estimates, standard errors, and *p*-values. Model format *y*~*x*; *y* regressed on *x*.

 superscript 1 model constructed after non-significant quadratic term was dropped from the prior model.

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>Parameters</th>
<th>Estimate</th>
<th>Standard Error</th>
<th><em>p</em>-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>mating.success~weapon.size^2+weapon.size</td>
<td>Hypotheses 1 and 2</td>
<td>weapon.size</td>
<td>2.59009</td>
<td>1.57996</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td></td>
<td>weapon.size^2</td>
<td>-0.75056</td>
<td>1.57996</td>
<td>0.635</td>
</tr>
<tr>
<td></td>
<td></td>
<td>weapon.size</td>
<td>2.334</td>
<td>1.422</td>
<td>0.102</td>
</tr>
<tr>
<td>mating.success~weapon.size</td>
<td>Hypothesis 1</td>
<td>weapon.size</td>
<td>-7.223</td>
<td>27.36</td>
<td>0.792</td>
</tr>
<tr>
<td>mating.success~weapon.size+body.size+weapon.size*body.size</td>
<td>Hypotheses 3 and 4</td>
<td>weapon.size</td>
<td>5.739</td>
<td>24.599</td>
<td>0.816</td>
</tr>
<tr>
<td></td>
<td></td>
<td>weapon.size*body.size</td>
<td>7.814</td>
<td>25.032</td>
<td>0.755</td>
</tr>
<tr>
<td>mating.success~resid.weapon.size+resid.weapon.size^2</td>
<td>Hypothesis 5</td>
<td>resid.weapon.size</td>
<td>-0.34483</td>
<td>1.58014</td>
<td>0.8274</td>
</tr>
<tr>
<td></td>
<td></td>
<td>resid.weapon.size^2</td>
<td>-2.62576</td>
<td>1.58014</td>
<td>0.0973</td>
</tr>
<tr>
<td>mating.success~</td>
<td>resid.weapon.size</td>
<td>Hypothesis 5</td>
<td>resid.weapon.size</td>
<td>-6.1074</td>
<td>3.4756</td>
</tr>
</tbody>
</table>
Table 3: Scaling relationships of femur length on body size for male and female frog legged beetles. Estimated using major-axis (MA) regression. Reported values include estimate and 95% CI for male and female intercept and slope. Test statistics (Wald test for intercept, likelihood ratio test for slope) and p values reported for comparisons between male and female intercept and slope.

<table>
<thead>
<tr>
<th>Intercept:</th>
<th>Estimate</th>
<th>95% CI</th>
<th>Test statistic</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>-0.666</td>
<td>[-0.738, -0.595]</td>
<td>1633.94</td>
<td>1</td>
<td>&gt;0.0001</td>
</tr>
<tr>
<td>Female</td>
<td>-0.76</td>
<td>[-0.83, -0.69]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Slope:</th>
<th>Estimate</th>
<th>95% CI</th>
<th>Test statistic</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>1.566</td>
<td>[1.491, 1.646]</td>
<td>35.5669</td>
<td>1</td>
<td>&gt;0.0001</td>
</tr>
<tr>
<td>Female</td>
<td>1.147</td>
<td>[1.053, 1.25]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 1.1: Male-male competition in the frog legged beetle. Resident male and female in white, intruding male in grey. (a) Mating pair, (b) intruder tires to mount pair, (c) intruder grasps resident by hindlegs and swings his own body to remove resident male from the female, (d) resident male removes his genitalia from the female’s reproductive tract, (e) resident male steps back, and (f) resident male leaves. From [22], used with permission.
Appendix 1.2: Number of new beetles captured and marked per day during the A) 2014 and B) 2015 field seasons.
**Appendix 1.3: Mean trait values from 2014 and 2015 field seasons and statistical comparisons between seasons.**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>FL (male)</td>
<td>9.785</td>
<td>0.092</td>
<td>9.711</td>
<td>0.0614</td>
<td>283.43</td>
<td>0.508</td>
<td>0.548</td>
</tr>
<tr>
<td>FL (female)</td>
<td>7.084</td>
<td>0.083</td>
<td>7.126</td>
<td>0.04</td>
<td>172.42</td>
<td>0.648</td>
<td>0.261</td>
</tr>
<tr>
<td>EL (male)</td>
<td>13.427</td>
<td>0.092</td>
<td>13.429</td>
<td>0.058</td>
<td>268.57</td>
<td>0.987</td>
<td>0.205</td>
</tr>
<tr>
<td>EL (female)</td>
<td>12.459</td>
<td>0.108</td>
<td>12.549</td>
<td>0.069</td>
<td>197.51</td>
<td>0.471</td>
<td>0.347</td>
</tr>
</tbody>
</table>
CHAPTER 2

Overcoming Mechanical Adversity in Extreme Hindleg Weapons

Devin M. O'Brien¹ and Romain P. Biosseau¹

¹ Division of Biological Sciences, University of Montana, Missoula, MT, USA

Abstract

The size of sexually selected weapons and their performance in battle are both critical to reproductive success, yet these traits are often in opposition. Bigger weapons make better signals. However, due to the mechanical properties of weapons as lever systems, increases in size may inhibit other metrics of performance as different components of the weapon grow out of proportion with one another. Here, using direct force measurements, we investigated the relationship between weapon size and weapon force production in two hindleg weapon systems, frog legged beetles (*Sagra femorata*) and leaf footed cactus bugs (*Narnia femorata*), to test for performance tradeoffs associated with increased weapon size. In male *S. femorata*, relative force production decreased as weapon size increased. Yet, absolute force production was maintained across weapon sizes. Surprisingly, mechanical advantage was constant across weapon sizes and large weaponed males had disproportionately large leg muscles. In male *N. femorata*, on the other hand, there was no relationship between weapon size and force production, likely reflecting the importance of their hindlegs as signals rather than force-producing structures of male-male competition. Overall, our results suggest that when weapon force production is important for reproductive success, large weaponed animals may overcome mechanical challenges by maintaining proportional lever components and investing in (potentially costly) compensatory mechanisms.

**Keywords:** Animal weapons, sexual selection, biomechanics
Introduction:

Animal weapons have a history of strong selection for large size [1–17]. This, in part, results from their role as signals to potential mates [18–20] and rival males [19,21–25]. In both contexts, weapons typically function as honest signals of quality where the largest, most conspicuous traits make the best signals [17,26]. However, as selection pushes weapons toward larger sizes, they face intrinsic, mechanical challenges that impede their performance [19,27–29]. This is because animal weapons, like many other mechanical traits (e.g., jaws of fishes [30–32] or jumping legs in insects [33–35]), are lever systems, the components of which must appropriately interact to achieve high performance (e.g., Fig. 1C-E).

All lever systems are composed of a fulcrum (i.e., the pivot about which the lever turns), an “input” lever arm (L_in), an “output” lever arm (L_out), an input force (F_in), and an output force (F_out). (The relationships between these components are represented by Equation 1.)

\[ F_{\text{out}} = \frac{F_{\text{in}} L_{\text{in}}}{L_{\text{out}}} \]  

(1)

The components of lever systems must remain in proportion to maintain force output (F_out) [27,28,36]. While increased weapon size may be favored by selection acting toward more efficient signaling or increased reach during combat, variation in the strength of selection and/or constraint experienced by lever components may cause them to scale disproportionally with one another. If, for example, external structures (L_out – horns, antlers, etc.) are free to become large while internal structures (L_in and/or F_in – tendons, bone, muscle, etc.) are architecturally constrained in their growth, as selection acts to increase overall weapon size, L_out may scale with body size at a faster rate than F_in and/or L_in. When this occurs, the mechanical advantage of the lever system will decrease and weapon force output (F_out) will suffer [19,27,36].
The mechanical limits of lever systems should impede overall trait performance [19,27,28,36]. Large weapons may make compelling signals and even limit the frequency of combat [22,37,38]. However, the largest males in a population will still be tested by similarly armed opponents [22,37–42]. When this occurs, weapons need to perform well. If not, animals could sustain severe damage and/or death, thereby eliminating their reproductive potential [43,44]. Large traits that function only as signals or deterrents are not sustainable in the context of animal contests. For this reason, animal weapons are predicted to represent a selective balance between the need for large, conspicuous signals and strong, force-generating weapons [19,45].

To date, several studies have quantified the relationship between weapon size and mechanical performance [25,27,36,46–57]. Yet, the majority of these studies have focused on one of three ecological/evolutionary scenarios: the claws of crustaceans [46–48,54,56–58], jaws of lizards [25,49,50,52], or weapons that do not function as signals [53,55]. Since the relative importance of signaling and fighting may vary considerably depending on the ecology of the species, further work is necessary to understand how the relationship between weapon size and force production varies across taxa and context and how this variation influences the evolution of sexually selected weapons and signals.

Here, we evaluate weapon performance as a function of weapon size in two systems with sexually selected hindleg weapons, frog legged beetles (*Sagra femorata*: Fig. 1A) and leaf footed cactus bugs (*Narnia femorata*: Fig. 1B). Using a strain gauge force-transducer, we measured how weapon force production varies across the natural range of weapon sizes to better understand the balance between selection for increased weapon size and performance. In addition, we measured input lever arm length (\(L_{in}\)), output lever arm length (\(L_{out}\)), and muscle mass (estimate of \(F_{in}\)) in these weapons to evaluate patterns of constraint and compensation involved in maintaining weapon force output. We predicted that large weapons would have relatively (if not absolutely) lower force production than smaller ones (i.e., the “paradox of the
weakening combatant” [29]). This would result from decreasing mechanical advantage as weapons become large, which should in turn decrease relative force production ($F_{\text{out}}$) [19,27,36].

Materials and Methods

Study organisms

Male frog legged beetles (Coleoptera, Chrysomelidae, Sagra femorata, Dury) have large hindleg weapons, used in one-on-one battle over direct access to females (Fig. 1A). During combat, males attack one another, using their hindlegs to squeeze rival males, pry apart copulating pairs, and steal mates [59,60]. The hindlegs of frog legged beetles primarily function as weapons, but they also appear to function as signals in competitive assessment. Males “wave” or display their hindlimbs as deterents to encroaching males, suggesting hindleg size plays a role in competitive assessment [D. O’Brien, personal observation].

Leaf footed cactus bugs (Hemiptera, Coreidae, Narnia femorata, Stål) have enlarged hindleg weapons used in male-male competition over reproductive territories (Fig. 1B). Similar to frog legged beetles, rival males back up to one another and use their weapons to squeeze opponents and pull them away from potential mates [61–63]. Hindlegs appear to be honest indicators of male quality [63,64] and, like other sexually selected structures, are subject to selection for increased size and/or signal quality.

Adult frog legged beetles (100 females, 95 males) were collected from a wild population in Matsuzaka, Mie Prefecture, Japan. Upon capture, measurements of elytra length (body size) and femur length (weapon size [59,60]) were collected using digital calipers. Animals were housed in 150 ml plastic cups at 25°C and fed Kudzu (Pueraria spp.) leaves ad libitum. Juvenile leaf footed cactus bugs (53 females, 47 males) were initially collected from a wild population in Gainesville, Florida, USA. Nymphs were shipped to Missoula, Montana, USA where they were housed in 500 mL plastic cups at 28°C and fed cactus fruit and pads (Opuntia spp.) ad libitum.
Measurements of prothorax width (body size) and femur area (weapon size [61]) were collected for each adult using photographs and ImageJ 1.50i software (NIH, USA).

**Squeezing force**

Squeezing force of hindleg weapons was collected using a full bridge, strain gauge force transducer (Appendix 2.1). The transducer was composed of two needles, which were attached to parallel metal plates. These plates were constructed of flexible brass, which bent as the animal squeezed the needles. Bending of the brass plates (i.e., squeezing force) was recorded using attached strain gauges (model EA-06-062AQ-350, Vishay Measurements Group, NC USA) and was transmitted to a computer (Dell Vosro 220, Dell, TX USA) via amplifier (model 2160 Vishay Measurements Group, NC USA) and AD converter (PowerLab 8sp, ADInstruments, Sydney Australia). Raw values were collected as a change in voltage and converted to a measure of force (N).

The relationship between force and measured voltage was identified as non-linear during subsequent analyses, thereby overestimating squeezing force in the largest animals (particularly large weaponed *S. femorata*). The force transducer was therefore calibrated across a range of known weight (2g – 100g), a curvilinear ordinary least squares (OLS) regression was fit to the data, and the equation of the best fit curvilinear line (\(y = 93.362x - 10.239x^2 + 36\); \(F_{2,4} = 1646\); \(p < 0.001\)) was used to correct raw voltage output to accurate force measures. Corrected measures are reported here.

During squeezing trials, animals were held by an observer at the thorax and a single hindleg was placed on the force transducer. For both animals, closing force was measured at the most distal point of the true output lever (L<sub>out</sub>). In *S. femorata*, L<sub>out</sub> is equal to the linear distance from the center of the femur-tibia joint (fulcrum) to the distal spine of the tibia (Fig 1C, Appendix 2.2A). In *N. femorata*, L<sub>out</sub> is equal to the linear distance from the center of the femur-tibia joint (fulcrum) to the most distal point on the widened “leaf” of the tibia (Fig 1E, Appendix
2.2B). Leg placement during squeezing measures aimed to mimic leg position during male-male competition, estimated through personal observation and video recording [D. O'Brien; Miller Lab, University of Florida]. While the animal was squeezing, a second observer annotated each “squeeze”, sorting acceptable squeezes from inadequate ones (e.g., poor leg placement on the needles) and removing noise (e.g., insect leg bumping into the needle, rather than squeezing it). Even so, due to a lack of cooperation from the animals (especially *N. femorata*), there was appreciable variation in leg placement across trials. For each trial, animals were encouraged to squeeze the force transducer for 2-4 minutes or until they refused to squeeze. Maximum squeezing force was collected across two trials. All measures were recorded in Lab Chart v7.2 (ADinstruments, Sydney AUS).

**Dissections (muscle mass and measures of L\textsubscript{in} and L\textsubscript{out})**

Hindleg muscle mass was collected from a subset of *S. femorata* (n = 173) and all *N. femorata* (n = 100) used in squeezing analyses. Whole hindlegs (*S. femorata*) and femurs (*N. femorata*) were dissected, dried at 90°C, and weighed. After initial weighing, muscle was digested by fully submerging the leg in 10% KOH and incubating at 90°C for 12 (*S. femorata*) or 8 (*N. femorata*) hours to ensure total dissolution of soft tissues [65]. After digestion, hindlimbs were dried at 90°C and weighed a second time. The difference between the first and second weighing was taken as an estimate of dry muscle mass. Muscle mass was taken from a single leg (leg used in squeezing trial when available).

Hindlegs were dissected in a subset of *S. femorata* (n = 27) to determine the precise internal structure of the leg and to gain accurate measures of L\textsubscript{in} and L\textsubscript{out} (Fig. 1C, Fig S2A). L\textsubscript{in} was identified as the linear distance from the center of the femur-tibia joint to the muscle attachment sclerite of the tibia. L\textsubscript{out} was identified as the linear distance from the center of the femur-tibia joint to the distal spine of the tibia. Measurements of L\textsubscript{in} and L\textsubscript{out} were collected using photographs of dissected legs and ImageJ 1.50i software (NIH, USA). From these measures,
the relationships between \( L_{\text{out}} \) and tibia length and \( L_{\text{in}} \), and tibia length were calculated using ordinary least squares regression. There were no significant sex differences in these relationships (95% CI intercept \( L_{\text{in}} \) for males [-0.227, 0.972] and females [-0.294, 0.691], 95% CI slope \( L_{\text{in}} \) for males [-0.29, 0.11] and females [-0.038, 0.135], 95% CI intercept \( L_{\text{out}} \) for males [-2.144, 3.397] and females [-0.855, 1.11], 95% CI slope \( L_{\text{out}} \) for males [0.554, 1.197] and females [0.776, 1.086]). Thus, male and female data were combined into the two regressions reported here (\( L_{\text{in}} \): \( y = 0.079x + 0.03, F_{1,24} = 91.26, p < 0.0001 \); \( L_{\text{out}} \): \( y = 0.903x + 0.39, F_{1,24} = 795.8, p < 0.0001 \)). Equations from these regressions were then used to estimate \( L_{\text{in}} \) and \( L_{\text{out}} \) for every beetle using measures of tibia length described above.

Similarly, hindlegs \( N. \text{ femorata} \) were dissected to identify exact measures of \( L_{\text{in}} \) and \( L_{\text{out}} \). \( L_{\text{in}} \) was identified as the linear distance from the center of the femur-tibia joint to the attachment point of the flexor muscle on the tibia (Fig 1E, Appendix 2.2B). \( L_{\text{out}} \) was identified as the linear distance from the center of the femur-tibia joint to the most distal point on the widened “leaf” of the tibia. Both \( L_{\text{in}} \) and \( L_{\text{out}} \) were directly measured in all animals using photographs of dissected legs and ImageJ 1.50i software (NIH, USA).

Statistical analyses

All statistical analyses were performed in R 3.3.2 (R Core Development Team 2016). All data were log\(_{10}\) transformed prior to analysis. Ordinary least squares (OLS) regression was used to assess all relationships. For both species and in both sexes, weapon size (\( S. \text{ femorata} \), femur length; \( N. \text{ femorata} \), femur area), \( L_{\text{in}}, L_{\text{out}} \), and muscle mass were regressed on body size (\( S. \text{ femorata} \): elytra length, \( N. \text{ femorata} \): prothorax width) in separate models.

Maximum squeezing force was regressed on weapon size in both species and both sexes to assess overall weapon force output. For male \( S. \text{ femorata} \), linear models with interaction terms between weapon size and muscle mass were constructed to further explore
the effect of weapon size, muscle mass, and their interaction on squeezing force. Differences in maximum squeezing force between sexes were calculated using t-tests.

To determine whether the observed increase in muscle mass relative to body size represented a compensatory mechanism, 95% confidence intervals were generated from OLS regression and used to compare the observed scaling relationship between muscle mass and body size to the expected, isometric relationship (β₀ = 3 for volumetric measures). If the observed slope was greater than expected (i.e., β > 3), it was considered a compensatory mechanism [36].

Finally, since mechanical advantage is expected to decrease in the absence of compensation as weapons grow large [27,36], \( \log_{10} \) mechanical advantage ([L_in]/[L_out]) was regressed against weapon size.

Results

Results are summarized in Table 1.

Squeezing force

In male *S. femorata*, maximum squeezing force increased hypoallometrically with weapon size (Fig 2A; Table 1). There was no significant interaction between muscle mass and weapon size on maximum squeezing force (\( t_{84} = 0.669, p = 0.505 \)). In female *S. femorata*, there was no significant relationship between maximum squeezing force and weapon size (Fig. 2A; Table 1). In *S. femorata*, maximum squeezing force was higher in males than in females (mean\(_{\text{male}} = 0.338N\); mean\(_{\text{female}} = 0.109N\); \( t_{113.42} = 15.996, p < 0.0001 \)).

In male *N. femorata*, there was no significant relationship between maximum squeezing force and weapon size (Fig. 2B; Table 1). In females, maximum squeezing force increased isometrically with weapon size (Fig. 2B; Table 1). There was no significant difference in maximum squeezing force between sexes in *N. femorata* (\( t_{96.286} = -0.0396, p = 0.693 \)).
Morphological measures of lever components

A summary of all morphological measures is provided in Appendix 2.3. In *S. femorata*, weapon size increased hyperallometrically with body size in males and isometrically in females (Fig. 3A; Table 1). L_{in} increased isometrically with body size in males and hypoallometrically with body size in females (Fig. 4A; Table 1). L_{out} increased isometrically with body size in males and hypoallometrically with body size in females (Fig. 4B; Table 1). There was no significant relationship between mechanical advantage and weapon size in males or females (Fig 4C; Table 1).

In *N. femorata*, weapon size increased isometrically with body size in males and hypoallometrically with body size in females (Fig. 3B; Table 1). L_{in} increased isometrically with body size in males and increased slightly, but significantly, hyperallometrically with body size in females (Fig. 4D; Table 1). L_{out} increased hypoallometrically with body size in both males and females (Fig 4E; Table 1). Mechanical advantage increased hypoallometrically with weapon size in both males and females (Fig 4F; Table 1).

In male *S. femorata*, muscle mass increased hyperallometrically with body size, which is consistent with a compensatory mechanism (slope 95% CI: [3.11, 3.809]; Fig 5A; Table 1) [36]. In females, muscle mass increased isometrically with body size (Slope 95% CI: [2.778, 3.777]; Fig 5A; Table 1).

In both male and female *N. femorata*, muscle mass scaled isometrically with body size (slope 95% confidence intervals for males [0.914, 3.181] and females [2.311, 3.131]; Fig. 5B; Table 1).
Discussion

We measured weapon force output as a function of weapon size in two wild, sexually selected weapon systems, frog legged beetles (*S. femorata*) and leaf footed cactus bugs (*N. femorata*). In *S. femorata*, weapon force output increased hypoallometrically ($\beta = 0.630 \pm 0.135$) with weapon size, suggesting large males have relatively lower, but absolutely higher, force production than smaller rivals (Fig 2A; Table 1). As weapons grow large, mechanical advantage (and therefore weapon force output) is predicted to decrease in the absence of compensation and limit the relationship between weapon size and weapon force output (e.g., [27,36; Equation 1]). In *S. femorata*, however, mechanical advantage was maintained across all animals and absolute force production increased with weapon size (Fig. 2A; Fig. 4C; Table 1). This suggests *S. femorata* employs one or more compensatory mechanism, which partially mitigates the mechanical limits predicted to hinder large weapon sizes.

Here, we identified two potential compensatory mechanisms, proportional growth of weapon/hindleg lever components and disproportionate growth of femur muscle mass. Overall, male *S. femorata* do not experience mechanical disadvantage as weapons grow large, since they compensate for the increase in output lever length associated with increased in weapon size by similarly increasing input lever length. Male *S. femorata* displayed longer input and output levers than females, which resulted in constant mechanical advantage across weapon sizes and between sexes (Fig 4 A-C).

In addition, in male *S. femorata*, femur muscle mass ($F_{in}$) increased hyperallometrically with body size ($\beta > 3$; Fig. 5A; Table 1), which is consistent with compensatory mechanisms identified in other systems [e.g., 36]. It should be noted, however, that both absolute and relative weapon force output should increase with weapon size, given disproportionate muscle growth and the observed maintenance of mechanical advantage (Fig. 4C; Table 1). Clearly, there are as-yet undiscovered limits to weapon force production in this system (mechanical
and/or behavioral), and further work is necessary to uncover why exactly weapon force output scales hypoallometrically in the frog legged beetle.

Male *N. femorata* showed no significant relationship between weapon force output (*F*<sub>out</sub>) and weapon size (Fig. 2B; Table 1). This result was surprising given the observed increase in mechanical advantage with weapon size (Fig 4F; Table 1). However, leg muscles were relatively smaller in large weaponed animals (Fig 5B; Table 1), which may explain why weapon force output did not increase with weapon size in males of this species. This result was unexpected given the established role hindleg weapons play in male-male competition [e.g., 61], and the maintenance of mechanical advantage across weapon sizes. One explanation for this trend is that these hindlegs may be under relatively weak selection for increased force production in the context male-male combat. Instead, the hindlegs of leaf footed cactus bugs may serve a greater role as intersexual signals of male quality, a behavioral context in which weapon force output is not an important component of fitness and hindlimb area, rather than force production, is under strong selection for increased size. Indeed, previous work suggests hindleg area is an honest indicator of overall quality [63,64] and recent studies have detected directional selection for increased hindleg area in the wild [61]. If true, then focal animals may have been unwilling to perform at full capacity during squeezing trails (since their hindlegs function primarily as display signals rather than weapons).

Alternatively, the ability to squeeze an opponent between both femurs, rather than between the femur and tibia of a single leg (as measured here), may be the most relevant metric of fighting success in this system (personal observation; Miller lab, University of Florida). Either scenario would result in an underestimation of weapon force output and could explain the observed non-significant relationship between weapon size and weapon force output. While we maintain our measures of weapon size, *L*<sub>in</sub>, and *L*<sub>out</sub>, are relevant in this system and to understanding the forces produced by these weapons, further investigation is necessary to establish exactly how weapon length and force production influence the outcome of male-male
competition in the leaf footed cactus bug, and what role, if any, these traits play in overall reproductive success.

Compensatory muscle growth and honest signaling in the frog legged beetle

Sexually selected weapons act as signals of quality and weapons of male-male battle. In both contexts, honesty is essential. Weapon size must honestly display quality to potential mates [18–20] and fighting ability to rival males [19,21–25] and, when tested in combat by similarly armed opponents, large weapons must produce sufficient force [reviewed in 66]. If not, receivers are predicted to focus to other, more reliable indicators of quality/fighting ability and selection for large weapons/signals should relax. Honesty in sexually selected weapons can be maintained through several mechanisms, including exquisite sensitivity to stress [67], parasite load [68,69], environmental condition [70], and intrinsic cost associated large structures [71,72]. The latter is particularly relevant to weapon systems where large, conspicuous structures often hinder the animals that bear them [73–78]. When present, the costs of sexually selected weapons typically increase with trait size, so only the largest animals can develop and wield large weapons and high quality signals are restricted to high quality males [26,71,72,79].

We suggest the compensatory muscle growth identified in frog legged beetles comes at a cost and, through that cost, functions as mechanism of honesty. Muscle is notoriously expensive to develop [80–82] and maintain [73,75,83–87]. In preserving absolute weapon force output through compensatory muscle growth, frog legged beetles may experience added metabolic [73,87] and locomotor [73,75] strain. For example, fiddler crabs with large, muscular claws suffer from disproportionately high resting metabolic rates [73,87], while stag beetles with large mandibles experience decreased flight performance resulting from their heavy, muscular jaws [75]. Such costs are consistent with theoretical models of handicap and indicator traits, where cost helps maintain the honesty/integrity of sexually selected traits as signals [71,72,79,88–92]. We therefore suggest that compensation for mechanical disadvantage
through muscle growth may contribute to the integrity of weapon size as an honest indicator of quality and fighting ability in this system.

**Conclusion**

The size of sexually selected weapons is critical to their role as honest signals. Weapons signal overall quality to potential mates and display fighting prowess to rival males. In both contexts, large traits are favored. However, selection for large, conspicuous signals is likely balanced by the need for weapons to perform well during combat. Here, we analyzed the relationship between weapon size and weapon force production (i.e., performance) in two systems, frog legged beetles (*S. femorata*) and leaf footed cactus bugs (*N. femorata*). In male frog legged beetles, weapon force output scaled hypoallometrically with weapon size. This is partially consistent with lever theory, where both absolute and relative force output are predicted to decrease as weapons become large [e.g., 27,36]. However, absolute force output appears to be maintained in this system through the maintenance of mechanical advantage across all weapon sizes. In addition, we showed a disproportional increase in leg muscle mass, which may also help maintain overall weapon force output. Alternatively, male *N. femorata* showed no relationship between weapon size and force production, potentially reflecting the importance of hindleg area as an intersexual display of male quality rather than a force-producing weapon of male-male competition.

Overall, we suggest that when weapon force production is an important component of reproductive success, and animals experience mechanical limits to weapon force production, the evolution of compensatory mechanisms is likely [reviewed in 93]. We also suggest that some compensatory mechanisms, such as muscle growth in the frog legged beetle, could enhance signal honesty in the context of sexual selection, both by disproportionately increasing metabolic or other costs associated with the largest male weapons and by maintaining fight performance at even the largest weapon sizes. Clearly, more work is required to understand the
realized cost of heavily muscled weapons, how this influences individual fitness in the wild, and the ubiquity of the trends described here.
Author contributions

DMO: Conception of project, measurement of squeezing force for Sagra femorata and Narnia femorata, measures of morphology and muscle mass for Sagra femorata, primary contributor to writing and edits of manuscript. RPB: Measurement of squeezing force for Sagra femorata and Narnia femorata, measures of morphology and muscle mass for N. femorata, significant contribution to writing and edits of manuscript.

Acknowledgements

We thank Bret Tobalske for designing and building the force transducer and technical support, Sarah Solie for her assistance in animal capture and analyses for squeezing force in Sagra femorata, The Miller Lab (University of Florida) for collection of Narnia femorata and assistance in early planning of the project, Teruyuki Niimi, Hiroki Gotoh, Masako Katsuki, and Yasu Okada for their support in Japan, and Doug Emlen, Sheila Patek, Art Woods, and Steve Lane, for feedback on the project and manuscript. We thank the National Science Foundation in collaboration with the Japanese Society for the Promotion of Science (Award number: 1515085) for funding.
Literature cited


41. Maynard Smith J. Evolution and the Theory of Games: In situations characterized by conflict of interest, the best strategy to adopt depends on what others are doing. Am Sci. 1976;64: 41–45.


63. Sasson DA, Munoz PR, Gezan SA, Miller CW. Resource quality affects weapon and testis size and the ability of these traits to respond to selection in the leaf-footed cactus bug, Narnia femorata. Ecol Evol. 2016;


Fig. 1: A) Mating *Sagra femorata* (male on top, photo: D. O’Brien). B) Male *Narnia femorata* (photo: R. Boisseau). C-E) Illustrations of lever systems. C) *S. femorata* hindlimb. D) Simplified machine. E) *N. femorata* hindlimb. Components of the lever systems are color coded across all structures (L_in = input lever (dark red), L_out = output lever (dark blue), F_in = force in (orange), F_out = force out (light blue), fulcrum (light red)). All three systems are best described as 3rd order levers.
Fig. 2: Relationship between $\log_{10}$ weapon size and $\log_{10}$ maximum squeezing force (measured using a strain gauge force transducer). Males represented by closed circles and solid lines, females by open circles and dotted lines. Lines represent ordinary least squares regression between $\log_{10}$ tibia length and $\log_{10}$ maximum squeezing force. A) *Sagra femorata* (male: $n = 95, \beta = 0.630 \pm 0.135, F_{1,93} = 21.68, p < 0.001$; female: $n = 100, \beta = 0.153 \pm 0.116, F_{1,98} = 1.736, p < 0.191$) and B) *Narnia femorata* (male: $n = 38, \beta = 0.356 \pm 0.409, F_{1,36} = 0.756, p = 0.36$; female: $n = 43, \beta = 1.289 \pm 0.369, F_{1,41} = 12.21, p < 0.01$).
Fig. 3: Relationship between $\log_{10}$ weapon size and $\log_{10}$ body size for A) *Sagra femorata* (weapon size = femur length, male: n = 95, female: n = 99) and B) *Narnia femorata* (weapon size = femur area, male: n = 38, female: n = 43) measured with dial calipers and photographs/ImageJ (W. Rasband) respectively. Males represented by closed circles and solid lines, females by open circles and dotted lines. Lines represent ordinary least squares (OLS) regression of $\log_{10}$ weapon size on $\log_{10}$ body size. In *S. femorata*, $\log_{10}$ weapon size scales hyperallometrically with $\log_{10}$ body size in males ($\beta = 1.267 \pm 0.042$, $F_{1,93} = 903.6$, $p < 0.0001$) and isometrically with $\log_{10}$ body size in females ($\beta = 1.036 \pm 0.057$, $F_{1,97} = 327$, $p < 0.0001$). In *N. femorata*, $\log_{10}$ weapon size scales isometrically with $\log_{10}$ body size in males ($\beta = 2.009 \pm 0.17$, $F_{1,36} = 139.6$, $p < 0.0001$) and hypoallometrically with $\log_{10}$ body size in females ($\beta = 1.668 \pm 0.108$, $F_{1,41} = 236.69$, $p < 0.0001$).

**A. Sagra femorata**

![Graph showing the relationship between log10 weapon size and log10 body size for Sagra femorata.](image)

**B. Narnia femorata**

![Graph showing the relationship between log10 weapon size and log10 body size for Narnia femorata.](image)
Fig. 4: Relationships between log$_{10}$ lever components/mechanical advantage and log$_{10}$ body size for *S. femorata* (left) and *N. femorata* (right) measured with dial calipers and/or photographs/ImageJ (W. Rasband). Males represented by closed circles and solid lines, females by open circles and dotted lines. Lines represent ordinary least squares (OLS) regression of log$_{10}$ weapon size on log$_{10}$ body size. In *S. femorata*, A) input lever length ($L_{in}$) scaled isometrically with log$_{10}$ body size in males ($n = 95$, $\beta = 1.023 \pm 0.159$, $F_{1,93} = 41.56$, $p < 0.0001$) and hypoallometrically with log$_{10}$ body size in females ($n = 100$, $\beta = 0.166 \pm 0.066$, $F_{1,98} = 6.327$, $p = 0.014$) and B) output lever length ($L_{out}$) scaled isometrically with log$_{10}$ body size in males ($n = 95$, $\beta = 1.016 \pm 0.157$, $F_{1,93} = 41.7$, $p < 0.0001$) and hypoallometrically with log$_{10}$ body size in females ($n = 100$, $\beta = 0.164 \pm 0.065$, $F_{1,98} = 6.318$, $p = 0.014$). C) There was no relationship between log$_{10}$ mechanical advantage and log$_{10}$ body size for male ($n = 13$, $\beta = -0.001 \pm 0.003$, $F_{1,11} = 1.167$, $p = 0.69$) or female ($n = 13$, $\beta = -0.004 \pm 0.005$, $F_{1,11} = 0.8625$, $p = 0.373$) *S. femorata*. In *N. femorata*, D) input lever length ($L_{in}$) scaled isometrically with log$_{10}$ body size in males ($n = 38$, $\beta = 1.123 \pm 0.15$, $F_{1,36} = 53.93$, $p < 0.0001$) and slightly hyperallometrically with log$_{10}$ body size in females ($n = 43$, $\beta = 1.157 \pm 0.149$, $F_{1,41} = 60.56$, $p < 0.0001$) and E) output lever length ($L_{out}$) scaled hypoallometrically with log$_{10}$ body size in both males ($n = 38$, $\beta = 0.503 \pm 0.083$, $F_{1,36} = 36.29$, $p < 0.0001$) and females ($n = 43$, $\beta = 0.763 \pm 0.113$, $F_{1,41} = 46.36$, $p < 0.0001$). F) log$_{10}$ mechanical advantage scaled hypoallometrically with log$_{10}$ body size in both male ($n = 38$, $\beta = 0.227 \pm 0.07$, $F_{1,36} = 10.6$, $p = 0.002$) and female ($n = 43$, $\beta = 0.157 \pm 0.093$, $F_{1,41} = 2.855$, $p = 0.01$) *N. femorata*. 
Fig. 5: Relationship between log$_{10}$ hindlimb muscle mass (measured using by KOH muscle digestion) and log$_{10}$ body size for A) Sagra femorata (male: n = 88, female: n = 85) and B) Narnia femorata (male: n = 47, female: n = 53). Males represented by closed circles and solid lines females by open circles and dotted lines. Lines represent ordinary least squares regression between log$_{10}$ body size and log$_{10}$ hindlimb muscle mass. In S. femorata, log$_{10}$ hindlimb muscle mass scaled hyperallometrically with log$_{10}$ body size in males ($\beta = 3.406 \pm 0.176$, $F_{1,86} = 387.9$, $p < 0.0001$) and females ($\beta = 3.278 \pm 0.251$, $F_{1,83} = 170.4$, $p < 0.0001$). In N. femorata, log$_{10}$ hindlimb muscle mass scaled hypoallometrically with log$_{10}$ body size in both males ($\beta = 2.408 \pm 0.563$, $F_{1,45} = 13.24$, $p < 0.001$) and females ($\beta = 2.721 \pm 0.123$, $F_{1,51} = 177.7$, $p < 0.0001$)
Table 1: Models constructed for analyses in $y \sim x$ format. SE = standard error. For *S. femorata*, weapon size = tibia length, body size = elytra length. For *N. femorata*, weapon size = tibia length, body size = prothorax width.

<table>
<thead>
<tr>
<th>Weapon size</th>
<th>Model</th>
<th>Intercept</th>
<th>SE</th>
<th>Slope</th>
<th>SE</th>
<th>n</th>
<th>$R^2$</th>
<th>$F$</th>
<th>d.f.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. femorata</em> (male)</td>
<td>weapon size ~ body size</td>
<td>-0.443</td>
<td>0.048</td>
<td>1.267</td>
<td>0.042</td>
<td>95</td>
<td>0.906</td>
<td>903.3</td>
<td>1,95</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><em>S. femorata</em> (female)</td>
<td>weapon size ~ body size</td>
<td>-0.29</td>
<td>0.063</td>
<td>1.036</td>
<td>0.057</td>
<td>99</td>
<td>0.709</td>
<td>327.5</td>
<td>1,98</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><em>N. femorata</em> (male)</td>
<td>weapon size ~ body size</td>
<td>-0.574</td>
<td>0.095</td>
<td>2.009</td>
<td>0.17</td>
<td>38</td>
<td>0.789</td>
<td>139.6</td>
<td>1,39</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><em>N. femorata</em> (female)</td>
<td>weapon size ~ body size</td>
<td>-0.272</td>
<td>0.065</td>
<td>1.668</td>
<td>0.108</td>
<td>43</td>
<td>0.849</td>
<td>236.6</td>
<td>1,44</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Muscle mass</th>
<th>Model</th>
<th>Intercept</th>
<th>SE</th>
<th>Slope</th>
<th>SE</th>
<th>n</th>
<th>$R^2$</th>
<th>$F$</th>
<th>d.f.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. femorata</em> (male)</td>
<td>muscle mass ~ body size</td>
<td>-2.912</td>
<td>0.199</td>
<td>3.460</td>
<td>0.178</td>
<td>88</td>
<td>0.819</td>
<td>387.9</td>
<td>1,87</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><em>S. femorata</em> (female)</td>
<td>muscle mass ~ body size</td>
<td>-2.975</td>
<td>0.278</td>
<td>3.278</td>
<td>0.251</td>
<td>85</td>
<td>0.672</td>
<td>170.4</td>
<td>1,86</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><em>N. femorata</em> (male)</td>
<td>muscle mass ~ body size</td>
<td>1.545</td>
<td>0.315</td>
<td>2.048</td>
<td>0.553</td>
<td>47</td>
<td>0.227</td>
<td>13.24</td>
<td>1,46</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><em>N. femorata</em> (female)</td>
<td>muscle mass ~ body size</td>
<td>1.073</td>
<td>0.123</td>
<td>2.721</td>
<td>0.204</td>
<td>53</td>
<td>0.777</td>
<td>177.7</td>
<td>1,47</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>L_1</th>
<th>Model</th>
<th>Intercept</th>
<th>SE</th>
<th>Slope</th>
<th>SE</th>
<th>n</th>
<th>$R^2$</th>
<th>$F$</th>
<th>d.f.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. femorata</em> (male)</td>
<td>$L_1$ ~ body size</td>
<td>-1.316</td>
<td>0.179</td>
<td>1.023</td>
<td>0.159</td>
<td>95</td>
<td>0.301</td>
<td>41.56</td>
<td>1,95</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><em>S. femorata</em> (female)</td>
<td>$L_1$ ~ body size</td>
<td>-0.481</td>
<td>0.278</td>
<td>0.166</td>
<td>0.066</td>
<td>100</td>
<td>0.051</td>
<td>6.223</td>
<td>1,96</td>
<td>0.014</td>
</tr>
<tr>
<td><em>N. femorata</em> (male)</td>
<td>$L_1$ ~ body size</td>
<td>-1.08</td>
<td>0.085</td>
<td>1.123</td>
<td>0.112</td>
<td>38</td>
<td>0.599</td>
<td>53.93</td>
<td>1,39</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><em>N. femorata</em> (female)</td>
<td>$L_1$ ~ body size</td>
<td>-1.181</td>
<td>0.099</td>
<td>1.157</td>
<td>0.149</td>
<td>43</td>
<td>0.507</td>
<td>60.56</td>
<td>1,40</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>L_2</th>
<th>Model</th>
<th>Intercept</th>
<th>SE</th>
<th>Slope</th>
<th>SE</th>
<th>n</th>
<th>$R^2$</th>
<th>$F$</th>
<th>d.f.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. femorata</em> (male)</td>
<td>$L_2$ ~ body size</td>
<td>-0.248</td>
<td>0.178</td>
<td>1.014</td>
<td>0.157</td>
<td>95</td>
<td>0.302</td>
<td>41.19</td>
<td>1,95</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><em>S. femorata</em> (female)</td>
<td>$L_2$ ~ body size</td>
<td>0.582</td>
<td>0.278</td>
<td>0.164</td>
<td>0.065</td>
<td>100</td>
<td>0.051</td>
<td>6.241</td>
<td>1,96</td>
<td>0.014</td>
</tr>
<tr>
<td><em>N. femorata</em> (male)</td>
<td>$L_2$ ~ body size</td>
<td>0.364</td>
<td>0.047</td>
<td>0.503</td>
<td>0.083</td>
<td>38</td>
<td>0.502</td>
<td>36.72</td>
<td>1,39</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><em>N. femorata</em> (female)</td>
<td>$L_2$ ~ body size</td>
<td>0.183</td>
<td>0.068</td>
<td>0.703</td>
<td>0.113</td>
<td>43</td>
<td>0.514</td>
<td>66.33</td>
<td>1,40</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mechanical Advantage</th>
<th>Model</th>
<th>Intercept</th>
<th>SE</th>
<th>Slope</th>
<th>SE</th>
<th>n</th>
<th>$R^2$</th>
<th>$F$</th>
<th>d.f.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. femorata</em> (male)</td>
<td>mechanical adv. ~ weapon size</td>
<td>0.102</td>
<td>0.034</td>
<td>-0.001</td>
<td>0.003</td>
<td>13</td>
<td>-0.074</td>
<td>1.166</td>
<td>1,13</td>
<td>0.30</td>
</tr>
<tr>
<td><em>S. femorata</em> (female)</td>
<td>mechanical adv. ~ weapon size</td>
<td>-1.268</td>
<td>0.052</td>
<td>0.227</td>
<td>0.07</td>
<td>38</td>
<td>0.206</td>
<td>16.01</td>
<td>1,39</td>
<td>0.002</td>
</tr>
<tr>
<td><em>N. femorata</em> (male)</td>
<td>mechanical adv. ~ weapon size</td>
<td>-1.221</td>
<td>0.068</td>
<td>0.157</td>
<td>0.093</td>
<td>43</td>
<td>0.402</td>
<td>2.853</td>
<td>1,40</td>
<td>0.11</td>
</tr>
<tr>
<td><em>N. femorata</em> (female)</td>
<td>mechanical adv. ~ weapon size</td>
<td>0.133</td>
<td>0.133</td>
<td>0.630</td>
<td>0.133</td>
<td>55</td>
<td>0.190</td>
<td>21.68</td>
<td>1,56</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Squeezing force</th>
<th>Model</th>
<th>Intercept</th>
<th>SE</th>
<th>Slope</th>
<th>SE</th>
<th>n</th>
<th>$R^2$</th>
<th>$F$</th>
<th>d.f.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. femorata</em> (male)</td>
<td>maximum force ~ weapon size</td>
<td>1.183</td>
<td>0.134</td>
<td>0.630</td>
<td>0.133</td>
<td>55</td>
<td>0.190</td>
<td>21.68</td>
<td>1,56</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><em>S. femorata</em> (female)</td>
<td>maximum force ~ weapon size</td>
<td>1.315</td>
<td>0.010</td>
<td>0.153</td>
<td>0.116</td>
<td>100</td>
<td>0.007</td>
<td>1.759</td>
<td>1,101</td>
<td>0.191</td>
</tr>
<tr>
<td><em>N. femorata</em> (male)</td>
<td>maximum force ~ weapon size</td>
<td>-1.592</td>
<td>0.306</td>
<td>0.536</td>
<td>0.409</td>
<td>38</td>
<td>-0.007</td>
<td>0.756</td>
<td>1,39</td>
<td>0.39</td>
</tr>
<tr>
<td><em>N. femorata</em> (female)</td>
<td>maximum force ~ weapon size</td>
<td>-2.241</td>
<td>0.272</td>
<td>1.289</td>
<td>0.509</td>
<td>43</td>
<td>0.211</td>
<td>12.21</td>
<td>1,40</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>
Appendix 2.1: Schematic of constructed force transducer. A) Rigid metal bar used to stabilize the transducer stationary during trials. B) Flexible, brass arms that bend during squeezing trials. C) Needles that the animals squeeze during trials. Squeezing force (red) causes deformation in brass arms (B). Deformation is recorded by strain gauges (blue) in a full bridge configuration, as they are placed under tension (T1 and T2) and compression (C1 and C2).
Appendix 2.2: Lever components of A) Sagra femorata and B) Narnia femorata hindlimbs.
Appendix 2.3: Summary of morphological measurements.

<table>
<thead>
<tr>
<th>Species</th>
<th>Sex</th>
<th>Trait</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. femorata</td>
<td>male</td>
<td>Weapon size (femur length)</td>
<td>13 mm</td>
<td>6.1 mm</td>
<td>9.84 mm</td>
<td>0.123</td>
</tr>
<tr>
<td>S. femorata</td>
<td>female</td>
<td>Weapon size (femur length)</td>
<td>5 mm</td>
<td>8.3 mm</td>
<td>7.191 mm</td>
<td>0.058</td>
</tr>
<tr>
<td>S. femorata</td>
<td>male</td>
<td>Body size (elytra length)</td>
<td>9.7 mm</td>
<td>15.8 mm</td>
<td>13.569 mm</td>
<td>0.128</td>
</tr>
<tr>
<td>S. femorata</td>
<td>female</td>
<td>Body size (elytra length)</td>
<td>10.5 mm</td>
<td>14.3 mm</td>
<td>12.82 mm</td>
<td>0.906</td>
</tr>
<tr>
<td>S. femorata</td>
<td>male</td>
<td>Input lever (L\text{in})</td>
<td>0.251 mm</td>
<td>0.852 mm</td>
<td>0.7 mm</td>
<td>0.012</td>
</tr>
<tr>
<td>S. femorata</td>
<td>female</td>
<td>Input lever (L\text{in})</td>
<td>0.378 mm</td>
<td>0.733 mm</td>
<td>0.505 mm</td>
<td>0.005</td>
</tr>
<tr>
<td>S. femorata</td>
<td>male</td>
<td>Output lever (L\text{out})</td>
<td>2.918 mm</td>
<td>9.780 mm</td>
<td>8.053 mm</td>
<td>0.121</td>
</tr>
<tr>
<td>S. femorata</td>
<td>female</td>
<td>Output lever (L\text{out})</td>
<td>4.363 mm</td>
<td>8.426 mm</td>
<td>5.817 mm</td>
<td>0.055</td>
</tr>
<tr>
<td>N. femorata</td>
<td>male</td>
<td>Weapon size (femur area)</td>
<td>3.742 mm</td>
<td>8.55 mm</td>
<td>5.677 mm</td>
<td>0.160</td>
</tr>
<tr>
<td>N. femorata</td>
<td>female</td>
<td>Weapon size (femur area)</td>
<td>3.912 mm</td>
<td>7.556 mm</td>
<td>5.418 mm</td>
<td>0.133</td>
</tr>
<tr>
<td>N. femorata</td>
<td>male</td>
<td>Body size (prothorax width)</td>
<td>3.137 mm</td>
<td>4.624 mm</td>
<td>3.627 mm</td>
<td>0.045</td>
</tr>
<tr>
<td>N. femorata</td>
<td>female</td>
<td>Body size (prothorax width)</td>
<td>3.358 mm</td>
<td>4.935 mm</td>
<td>3.993 mm</td>
<td>0.050</td>
</tr>
<tr>
<td>N. femorata</td>
<td>male</td>
<td>Input lever (L\text{in})</td>
<td>0.278 mm</td>
<td>0.475 mm</td>
<td>0.354 mm</td>
<td>0.006</td>
</tr>
<tr>
<td>N. femorata</td>
<td>female</td>
<td>Input lever (L\text{in})</td>
<td>0.246 mm</td>
<td>0.426 mm</td>
<td>0.348 mm</td>
<td>0.007</td>
</tr>
<tr>
<td>N. femorata</td>
<td>male</td>
<td>Output lever (L\text{out})</td>
<td>3.962 mm</td>
<td>4.937 mm</td>
<td>4.407 mm</td>
<td>0.038</td>
</tr>
<tr>
<td>N. femorata</td>
<td>female</td>
<td>Output lever (L\text{out})</td>
<td>3.178 mm</td>
<td>5.079 mm</td>
<td>4.410 mm</td>
<td>0.06</td>
</tr>
</tbody>
</table>
CHAPTER 3

Muscle Mass Drives Cost in Sexually Selected Insect Weapons

Devin M. O’Brien¹, Romain P. Boisseau¹, Ummat Somjee², Meghan Duell³, Erin McCullough⁴, Sarah Solie¹,⁵, Douglas J. Emlen¹

¹ Division of Biological Sciences, University of Montana, Missoula, MT
² Entomology and Nematology Department, University of Florida, Gainesville, FL
³ School of Life Sciences, Arizona State University, Tempe, AZ
⁴ Centre for Evolutionary Biology, University of Western Australia, Crawley, WA 6009, AUS
⁵ Biology Department, Duke University, Durham, NC, USA

Abstract

Sexually selected weapons function as signals of quality where weapon size displays condition to potential mates and fighting ability to competitive rivals. Honesty is essential to this process. If weapons misrepresent quality, and poor quality animals produce high quality signals, receivers should focus on other, more reliable structures. Cost is one way honesty is maintained. As weapons grow large, cost increases, thereby restricting large weapons to high quality animals that can bear the strain. However, while rich in historical precedent, and commonly observed across taxa, the idea of costly weapons remains controversial. This is in part because the cost of weapons appears unpredictably variable in both type and severity.

We suggest this variation may be the result of variation in the types of weapons studied. In particular, variation in the muscle mass directly associated with weapon movement. Here, we measure the maintenance metabolic cost of sexually selected weapons in five insect species - frog legged leaf beetles (Sagra femorata), Indonesian stag beetles (Cyclommatus metallifer), Japanese rhinoceros beetles (Trypoxylus dichotomus), heliconia bugs (Leptoscelis tricolor), leaf footed cactus bugs (Narnia femorata) - and directly relate these measures to weapon muscle mass. We show that animals with large weapon muscles have high resting metabolic rates.
compared to animals with small muscles, and provide evidence that this trend is driven by
weapon muscle mass *per se*.

**Keywords:** Animal weapons, sexual selection, cost, metabolic rate

**Introduction**

Sexually selected weapons are some of the largest and most diverse structures in the
animal world. They grow out of proportion with body size and other, more typically proportioned
structures [1–6] and, when viewed across clades (e.g., beetle horns, *Scarabaeidae* [7–9],
Cervid antlers, *Cervidae* [10–12]), are unparalleled in morphological diversity [13].

Weapons typically function as tools of intrasexual competition [reviewed in 13–18].
Animals use these structures to compete with same sex rivals over direct access to mates [19–
23], or over resources that otherwise grant access to mates (e.g., feeding territories [24–29]).
Evidence also suggests weapons function as intra- and intersexual signals. Weapon size
typically correlates with overall body size [1–6], and overall body size often reflects individual
quality/condition [30–34, but see 35,36]. Through this connection, weapon size provides an
effective signal of the resource holding potential (RHP) of an opponent [e.g., 28,39–43], and
members of the opposite sex may use weapon size to assess the condition of a potential mate
[e.g., 28,37,38].

Honesty is essential to signal function, and weapons are no exception [44–51]. If poor
quality animals cheat and produce high quality signals, receivers should shift focus to other,
more reliable cues. One way signal honesty is maintained is through cost, particularly when
costs are steepest for poor condition males [3,44,46–49,51–55]. Costs tend to increase as
structures get big. Thus, big structures are both more conspicuous, and more difficult to fake,
helping explain why sexual selection so often favors increases in weapon size [44–51]. Indeed,
costly weapons have been identified in a variety of species (Cervids [56], Bovids (*Bovidae*) [57],
dung beetles [58,59], stag beetles (*Lucanidae*) [60–62], fiddler crabs (*Uca*) [63–66]). However, the type (metabolic [65], locomotor [61,62], developmental [59], etc.) and severity of cost experienced by weaponed species is highly variable, even across closely related taxa [e.g., 58,59,67–69]. To date, the cause of this variation remains unknown.

Here, we focus on metabolic costs of sexually selected weapons. Muscle is consistently expensive to maintain [61,65,66,70–73], and we predict that variation in weapon muscle mass (resulting from variation in the way weapon strength is generated) will help explain species differences in the relative metabolic costs of sexually selected weapons. We report the metabolic cost of sexually selected weapons in five insect species (Fig. 1) - frog legged leaf beetles (*Sagra femorata*), Indonesian stag beetles (*Cyclommatus metallifer*), Japanese rhinoceros beetles (*Trypoxylus dichotomus*), heliconia bugs (*Leptoscelis tricolor*), leaf footed cactus bugs (*Narnia femorata*). For each species, we used flow through respirometry to measure resting metabolic rate (RMR) as an estimate of maintenance metabolic cost, and directly measured weapon muscle mass using potassium hydroxide (KOH) digestion. We show that animals with large weapon muscles have higher RMR than animals with small muscles, and provide evidence that this trend is indeed driven by muscle content. We discuss our results in the context of honest signaling and costly weapons, and show that observed variation in weapon cost is likely driven by variation in the properties of weapons studied.

**Materials and methods**

**Study Species**

Three criteria were used in choosing focal species. First, all have sexually selected weapons used in male-male competition over access to females and/or reproductive territories. Frog legged leaf beetles (Coleoptera, Chrysomelidae, *Sagra femorata*, Dury) [23,74], leaf footed cactus bugs (Hemiptera, Coreidae, *Narnia femorata*, Stål) [29,75], and heliconia bugs (Hemiptera, Coreidae, *Leptoscelis tricolor*, Westwood) [76] have enlarged hindleg weapons
Indonesian stag beetles (Coleoptera, Lucanidae, *Cyclommatus metallifer*, Boisduval) have enlarged mandibles [77,78]. Japanese rhinoceros beetles (Coleoptera, Scarabaeidae, *Trypoxylus dichotomus*, Linnaeus) have a bifurcated head horn and a smaller, bifurcated thoracic horn [26]. Second, all species have been previously studied in the context of costly sexually selected weapons (or, in the case of *L. tricolor*, studies exist in a closely related species [79]), providing multiple measures of cost beyond those surveyed here (*S. femorata* [79], *C. metallifer* [61,62,80], *T. dichotomus* [67–69], *N. femorata* and *L. tricolor* [79]). Third, all species were either a) easily collected in the wild or b) available through commercial breeders.

Adult *S. femorata* were collected from a wild population in Matsuzaka, Mie Prefecture, Japan. Animals were communally housed in 20L plastic containers and fed Kudzu (*Pueraria spp.*) leaves *ad libitum*. Specimens were transported to the National Institute for Basic Biology (NIBB; Okazaki, Aichi Prefecture, JP) where RMR was measured as described below. Adult *C. metallifer* were sourced from a breeding population at the University of Nagoya (Nagoya, Aichi Prefecture, JP) and shipped to the University of Montana (Missoula MT, USA). Animals were housed individually in 500mL plastic containers and fed organic apples *ad libitum* until RMR was measured. Larval *T. dichotomus* were sourced from a commercial insect distributor (Yasaka Kobuto Kuwagata World, Hamada City, JP) and shipped to the University of Montana. Larvae were reared to adulthood at 28°C in individual 500mL plastic containers filled with a combination of compost and decomposed wood shavings. Adults were housed individually in 2L plastic containers and fed organic apples *ad libitum* until RMR was measured. Juvenile *N. femorata* were collected from a wild population in Gainesville, Florida, USA and transported to the University of Montana in 500mL plastic containers. Animals were individually reared to adulthood at 28°C, and fed cactus (*Opuntia spp.*) fruit and pads *ad libitum* until RMR was measured as described below. Adult *L. tricolor* were collected from a wild population near Gamboa, Panama. Animals were housed individually in 500mL plastic containers at 28°C and fed Heliconia flower (*H. platystachys* and *H. mariae*) *ad libitum* until RMR was measured.
Resting Metabolic Rate (RMR)

Flow through respirometry was used to measure CO$_2$ emission at rest (estimate of RMR, see below). Sample sizes were as follows: S. femorata, n = 38; C. metallifer, n = 10; T. dichotomus, n = 16; N. femorata, n = 44; L tricolor, n = 52. The entire system was calibrated using pure N$_2$ and 2000ppm CO$_2$ (NorLab$^\text{®}$) every other day of analysis.

For each animal, rate of CO$_2$ emission was measured for one hour using a two-cell infrared analyzer (Licor LI-7000, Licor, Lincoln, NE, USA) in differential mode. Dry, CO$_2$-free compressed air was directed through the reference cell (Cell A), which measured the fractional CO$_2$ concentration of incident air, then through a cylindrical glass chamber containing the focal animal (14 mL for S. femorata, N. femorata, and C. metallifer, L. tricolor, 500 mL for T. dichotomus) and into the measurement cell (Cell B), which measured fractional CO$_2$ concentration of excurrent air. All gas circulated in 3mm inner diameter plastic tubing (Bevaline-IV, Cole Parmer, Vernon Hills, IL, USA). Air flow was controlled by a mass-flow controller (Unit instruments, Yorba linda, CA, USA; 0 – 500cm$^3$/min, calibrated with air), connected to controlling electronics (MFC-4, Sable Systems International, Las Vegas, NV, USA). Flow rates were selected based on the body size of each species to balance detectability and temporal resolution and were as follows: S. femorata, C. metallifer, T. dichotomus = 500mL/min, N. femorata, L. tricolor = 250mL/min [81]. Temperature was monitored using T-type thermocouples connected to a thermocouple meter (TC-1000, Sable Systems International). Activity was monitored visually or, when possible, using an activity detector (AD-1, Sable Systems International). Before and after each trial, baseline CO$_2$ in the system was measured for two minutes with the experimental chamber empty. These measures were used to correct for baseline drift by modeling a line between CO$_2$ levels at the beginning and end of the trial and subtracting it from each CO$_2$ measurement.

For S. femorata, data were collected using Lab Chart (v7.2 ADinstruments, Sydney AUS) receiving signals from an AD converter (PowerLab 8sp, ADinstruments). For all other
species, data were collected using ExpeData software (v. 1.1.9, Sable Systems International) receiving signals from an AD converter (UI2, Sable Systems International). AD converters received analog signals from the two-cell infrared analyzer, mass-flow controller, controlling electronics, thermocouple meter, and activity detector. The traces collected showed relative concentration of CO₂ (ppm) according to time (sampling frequency: 1Hz). Raw measures were converted to molar rates of CO₂ (Ṁ CO₂) production using known flow rate and the Ideal Gas Law. O₂ consumption was not measured and respiratory exchange ratio (Ṁ CO₂/Ṁ O₂) was not calculated.

Finally, a continuous period of at least 20 minutes during which the animal was completely inactive was isolated, and mean Ṁ CO₂ production during this time was calculated. This measure was taken as an estimate of RMR. The first 10 minutes of each trial were excluded to avoid effects of handling stress.

Morphological measures and muscle digestion

All morphological measures were collected after measuring RMR. For S. femorata, C. metallifer, and T. dichotomus, measures of body size (S. femorata, C. metallifer = elytra length; T. dichotomus = prothorax width) and weapon size (S. femorata = femur length, femur width, tibia length; C. metallifer = mandible length; T. dichotomus = horn length, horn width) were measured on live animals using digital calipers. For N. femorata and L. tricolor, measures of body size (prothorax width) and weapon size (femur length) were collected from photographs of preserved specimens using ImageJ 1.50i software (NIH, USA).

Weapon muscle mass was measured using KOH digestion. Weapons and associated muscle were dissected (S. femorata, N. femorata, L. tricolor = whole hindlimbs; C. metallifer = mandibles, head; T. dichotomus = horn, head, prothorax), dried at 90°C, and weighed. After initial weighing, weapons were completely submerged in 10% KOH and incubated at 90°C to digest soft tissue, primarily muscle (S. femorata, C. metallifer, T. dichotomus = 12 hours; N.
femorata, L. tricolor = 8 hours). After digestion, weapons were rinsed with water and dried at 90°C. Once dry, weapons were weighed a second time. The difference between first and second weighing was taken as an estimate of dry muscle mass.

Statistical analyses

All statistical analyses were performed in R 3.4.3 (R Core Development Team, 2017). All data were log_{10} transformed prior to analysis. For S. femorata and T. dichotomus, a principle components analysis (PCA) using different measures of weapon size was conducted. For S. femorata, the PCA included femur length, femur width, and tibia length. PC1 explained 87.83% of the variation in weapon size and was used as a measure of overall weapon size in subsequent analyses. For T. dichotomus, the PCA included horn length and horn width. PC1 explained 100% of the variation in weapon size and was used as a measure of overall weapon size in subsequent analyses. For all other species, only one measure of weapon size was collected and PCA were not performed.

Ordinary least squares regression (OLS) was used to assess the relationship between weapon size and body size, and muscle mass and body size for all species. These relationships were compared to isometry (slope (\(\beta\)) = 1 for linear measures, slope (\(\beta\)) = 3 for volumetric measures) to look for patterns of hyperallometry (\(\beta >\) isometry) or hypoallometry (\(\beta <\) isometry). RMR was regressed on weapon size, body size, and muscle mass within the same model to determine the effect of interactions on RMR. All models were assessed and compared using the Akaike information criterion (AIC).

Results

Weapon size increased hyperallometrically with body size in S. femorata (\(\beta = 39.778, \ F_{1,36} = 449, p < 0.0001\)), C. metallifer (\(\beta = 3.126, \ F_{1,8} = 16.8, p < 0.01\)), T. dichotomus (\(\beta = 32.035, \ F_{1,14} = 92, p < 0.0001\)), and L. tricolor (\(\beta = 2.128, \ F_{1,7} = 9.388, p = 0.018\)) (Table 1).
Weapon size increased hypoallometrically with body size in *N. femorata* ($\beta = 0.727, F_{1,42} = 0.727, p < 0.0001$) (Fig. 1; Table 1).

In *S. femorata*, muscle mass increased hyperallometrically (i.e., $\beta > 3$; [80]) with body size ($\beta = 3.625, F_{1,35} = 508, p < 0.0001$). In *C. metallifer*, muscle mass also increased hyperallometrically with body size ($\beta = 5.065, F_{1,14} = 82.42, p < 0.001$). In *T. dichotomus*, muscle mass increases hypoallometrically with body size ($\beta = 2.505, F_{1,55} = 78.36, p < 0.0001$). In *L. tricolor*, there was no significant relationship between muscle mass and body size. In *N. femorata* muscle mass increased hypoallometrically with body size ($\beta = 2.043, F_{1,42} = 13.44, p < 0.001$) (Fig. 1; Table 3).

In *S. femorata*, RMR increased hyperallometrically with body size ($\beta = 2.755, F_{1,36} = 13.94, p < 0.001$; Fig. 1; Table 2). The model containing body size and muscle mass provided the best fit for the data (AIC = -22.935, $F_{3,34} = 7.462, p < 0.001$) and included a significant, positive interaction between body size and muscle mass ($p = 0.035$; Table 2). In *C. metallifer*, RMR increased hyperallometrically with body size ($\beta = 2.63, F_{1,6} = 16.61, p < 0.01$). The model containing body size and muscle mass provided the best fit for the data (AIC = -30.585, $F_{3,6} = 16.61, p < 0.01$; Table 2), but did not contain significant interaction term between body size and muscle mass. The model containing weapon size and body size also fit the data well (AIC = -28.289, $F_{3,6} = 12.79, p < 0.01$) and contained a significant negative interaction between weapon size and body size ($p = 0.031$; Table 2). There was no significant relationship between RMR and body size or RMR and weapon size in *T. dichotomus, L. tricolor, or N. femorata* (Fig. 1; Table 2).
Discussion

Sexually selected weapons often function as signals of quality, where weapon size displays RHP and/or overall condition to potential mates [28,37,38] and competitive rivals [28,39–43]. Honesty is essential to this process [44–51], and costs - either of producing or maintaining a weapon – can play a critical role ensuring the long term integrity of these signaling systems [3,44,46–49,51–55]. As weapons become large, costs are predicted to increase, restricting large weapons to high quality individuals with sufficient resources to bear the strain. Yet, while rich in historical precedent [44–51], and commonly observed across taxa [56–63,65,82], the idea of costly weapons remains controversial [46,49,67–69,83]. This is in part because the cost of weapons appears unpredictably variable in both type and severity [59,61,62,65,66,68,69].

We suggest that much of the observed variation in costs of sexually selected weapons stems from variation in the types of weapons studied. Here, we surveyed resting metabolic rate (RMR) as a metric of metabolic cost in five insect species, and related these measures to weapon muscle mass. Species with high weapon muscle mass consistently showed high RMR for their body size, while those with low weapon muscle mass showed no significant relationship between RMR and body size (Fig. 1; Table 2). In *S. femorata* and *C. metallifer*, two species where muscle mass increased hyperallometrically with body size, RMR also increased steeply with body size (Fig. 1; Table 2; Table 3). *S. femorata* also showed a significant, positive interaction between body size and muscle mass in explaining RMR (Table 2), but this interaction was not significant in *C. metallifer*, possibly due to relatively low sample size (n = 10). In contrast, all other species showed either hypoallometric or non-significant relationships between muscle mass and body size (Fig. 1; Table 3), and none showed a significant relationship between RMR and body size or RMR and weapon size (Fig. 1; Table 2).
Our results suggest that when weapon muscle mass increases hyperallometrically with body size, animals experience steep metabolic cost. This type of muscle growth should be especially prevalent in weapons where strength depends on muscle content (rather than strength-amplifying levers), and hyperallometric scaling of weapon muscle mass is required to overcome mechanical disadvantage in large weapons [79,80,84]. Indeed, this is the case for S. femorata [79] and C. metallifer [80] and, in both systems, RMR increased steeply with body and/or weapon size (Fig. 1; Table 2). By extension, we suggest that the inconsistency in cost observed in the weapon literature may, in part, result from interspecific variation in the way weapon strength is generated and the associated variation in weapon muscle mass. For example, T. dichotomus experiences no locomotor, immune, developmental, or metabolic cost associated with large horns (Fig. 1; Table 2) [67–69]. Yet, their horns are hollow [69], and the prothoracic muscles regulating weapon movement are relatively small (Fig. 1; Table 3).

Similarly, N. femorata shows no strength [79] or metabolic cost associated with hindeg weapons, and they too have small weapon muscles (Fig. 1; Table 3). In these species, weapon strength may be generated through alterations to the weapon lever system, rather than hyperallometric increases in weapon muscle [e.g., 78,84], and the resulting low muscle mass may help explain why costs have never been observed in these species.

It should be noted, however, that weapon honesty is still expected in species where metabolic cost was not observed. These weapons still function as signals, and should reliably display RHP and/or quality. Rather than metabolic cost driven by large muscles, species with small muscles may experience developmental costs resulting from differential allocation during weapon development [56,59,85], or locomotor costs, not from heavy, muscular weapons, but from bulky, otherwise lightweight structures [58,86]. Honesty may also be maintained through heightened-condition dependent development [33,52,54,87–95]. Sexually selected weapons are famously sensitive to nutrition [94,95], environment [96,97], parasite load [98], and stress [99]. When growth of the weapon is sensitive to these factors, only the highest quality individuals can
produce large weapons. Similar to costs, condition dependence can effectively restrict the
biggest weapons to the largest, highest quality individuals, ensuring that weapon size persists
as a reliable signal through time.

Overall, we suggest that much of the controversy surrounding the presence/absence of
cost in weapon systems can be resolved, in part, by recognizing that both the type and severity
of cost should be dependent on the kind of weapon studied. Notably, variation in the way
weapon force is generated, and associated variation in weapon muscle mass. We therefore
encourage those exploring the costs surrounding sexually selected weapons to incorporate a
priori knowledge of the biomechanical mode of action of the structure itself, and the behavioral
ecology of the focal species, to direct their study, keeping in mind that in some (perhaps most)
weapon systems, cost and condition dependence may be working in tandem to maintain
honesty. We suggest future work be aimed at better connecting the properties of weapons to
measures of cost, and exploring how connections vary across species and weapon type.
Acknowledgments

We would like to thank the NSF and JSPS for funding (D.M.O.: award #1515085; D.J.E. NSF IOS-1456133), Art Woods (U. of Montana) for use of lab space, respirometry equipment, and assistance in data analysis, Teruyuki Niimi (National Institute for Basic Biology, Okazaki Japan) for use of lab space, Christine Miller for collection of Narnia femorata and assistance in early planning of the project, Hiroki Gotoh for supplying Cylcomattus metallifer, and Masako Katsuki, and Yasu Okada for assistance in collecting Sagra femorata. Figure 1 image credit: S. femorata, R. Niese, used with permission; C. metallifer, H. Gotoh, used with permission; T. dichotomus, Lsandonkey (https://creativecommons.org/licenses/by-sa/4.0/deed.en); L. tricolor; U. Somjee; N. femorata, R. Boisseau.

Author contributions

Literature cited


Figure 1: Scaling relationship between weapon size and body size, resting metabolic rate (RMR) and body size, and muscle mass and body size for all species. Each row contains information for one species, indicated by name and photograph in the left-most column. From top to bottom: In frog legged leaf beetles (*Sagra femorata*, *n* = 38), weapon size increased hyperallometrically with body size, RMR increased hyperallometrically with body size, and weapon muscle mass increased hyperallometrically with body size. In Indonesian stag beetles (*C. metallifer*, *n* = 10), weapon size increased hyperallometrically with body size, RMR increased hyperallometrically with body size, and weapon muscle mass increased hyperallometrically with body size. In Japanese rhinoceros beetles (*Trypoxylus dichotomus*, *n* = 16), weapon size increased hyperallometrically with body size, but there was no significant relationship between RMR and body size, and weapon muscle mass increased hypoallometrically with body size. In heliconia bugs (*Leptoscelis tricolor*, *n* = 52), weapon size increased hyperallometrically with body size, but there was no significant relationship between RMR and body size or weapon muscle mass and body size. In leaf footed cactus bugs (*Narnia femorata*, *n* = 44), weapon size increased hypoallometrically with body size, and there was no significant relationship between RMR and body size or weapon muscle mass and body size. All data were log_{10} transformed prior to analysis. Results summarized in Tables 1, 2, and 3. Image credits in acknowledgements.
Table 1: Scaling relationship between weapon size and body size for all species.

<table>
<thead>
<tr>
<th>Species</th>
<th>Model</th>
<th>AIC</th>
<th>F_{df}</th>
<th>p (model)</th>
<th>Coefficient</th>
<th>Estimate (coef.)</th>
<th>Std error (coef.)</th>
<th>p (coef.)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Frog legged beetle (S. femorata)</strong>*</td>
<td>WS ~ BS</td>
<td>50.818</td>
<td>449_{1.36}</td>
<td>&lt; 0.0001</td>
<td>Slope</td>
<td>39.778</td>
<td>1.877</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td><strong>Indonesian stag beetle (C. metallifer)</strong></td>
<td>WS ~ BS</td>
<td>-20.165</td>
<td>16.8_{1.8}</td>
<td>&lt; 0.01</td>
<td>Slope</td>
<td>3.126</td>
<td>0.763</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td><strong>Japanese rhinoceros beetle (T. dichotomous)</strong></td>
<td>WS ~ BS</td>
<td>29.073</td>
<td>92_{1.14}</td>
<td>&lt; 0.0001</td>
<td>Slope</td>
<td>32.035</td>
<td>3.34</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td><strong>Heliconia bug (L. tricolor)</strong></td>
<td>WS ~ BS</td>
<td>-30.935</td>
<td>9.388_{1.7}</td>
<td>0.018</td>
<td>Slope</td>
<td>2.128</td>
<td>0.695</td>
<td>0.018</td>
</tr>
<tr>
<td><strong>Leaf footed cactus bug (N. femorata)</strong></td>
<td>WS ~ BS</td>
<td>-210.799</td>
<td>65.57_{1.42}</td>
<td>&lt; 0.0001</td>
<td>Slope</td>
<td>0.727</td>
<td>0.089</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

* weapon size estimated through principle components analysis (PCA)

All data were log_{10} transformed prior to analysis
Table 2: Scaling relationship between resting metabolic rate (RMR), body size, weapon size, and muscle mass for all species.

Frog legged beetle (S. femorato*)

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>Fstat</th>
<th>p (model)</th>
<th>Coefficient</th>
<th>Estimate (coef.)</th>
<th>Std error (coef.)</th>
<th>p (coef.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMR ~ BS</td>
<td>-20.145</td>
<td>13.94</td>
<td>&lt; 0.001</td>
<td>BS</td>
<td>2.755</td>
<td>0.738</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>RMR ~ WS * BS</td>
<td>-20.318</td>
<td>6.212</td>
<td>&lt; 0.01</td>
<td>BS</td>
<td>-0.756</td>
<td>0.431</td>
<td>0.088</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BS</td>
<td>1.808</td>
<td>2.667</td>
<td>0.502</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>WS * BS</td>
<td>0.712</td>
<td>0.38</td>
<td>0.07</td>
</tr>
<tr>
<td>RMR ~ BS * MM</td>
<td>-22.934</td>
<td>7.462</td>
<td>&lt; 0.001</td>
<td>BS</td>
<td>-6.206</td>
<td>3.574</td>
<td>0.092</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MM</td>
<td>-8.8</td>
<td>4.275</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BS * MM</td>
<td>8.255</td>
<td>3.755</td>
<td>0.035</td>
</tr>
</tbody>
</table>

Indonesian stag beetle (C. metallifer)

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>Fstat</th>
<th>p (model)</th>
<th>Coefficient</th>
<th>Estimate (coef.)</th>
<th>Std error (coef.)</th>
<th>p (coef.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMR ~ BS</td>
<td>-23.521</td>
<td>16.61</td>
<td>&lt; 0.01</td>
<td>BS</td>
<td>2.63</td>
<td>0.645</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>RMR ~ WS * BS</td>
<td>-28.289</td>
<td>12.79</td>
<td>&lt; 0.01</td>
<td>BS</td>
<td>14.839</td>
<td>5.333</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BS</td>
<td>16.785</td>
<td>4.928</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BS * WS</td>
<td>-12.072</td>
<td>4.289</td>
<td>0.031</td>
</tr>
<tr>
<td>RMR ~ BS * MM</td>
<td>-30.585</td>
<td>16.61</td>
<td>&lt; 0.01</td>
<td>BS</td>
<td>8.322</td>
<td>1.767</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MM</td>
<td>0.049</td>
<td>0.047</td>
<td>0.339</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BS * MM</td>
<td>-0.046</td>
<td>0.037</td>
<td>0.259</td>
</tr>
</tbody>
</table>

Japanese rhinoceros beetle (T. dichotomous*)

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>Fstat</th>
<th>p (model)</th>
<th>Coefficient</th>
<th>Estimate (coef.)</th>
<th>Std error (coef.)</th>
<th>p (coef.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMR ~ BS</td>
<td>-4.697</td>
<td>0.426</td>
<td>0.524</td>
<td>BS</td>
<td>0.759</td>
<td>1.163</td>
<td>0.524</td>
</tr>
</tbody>
</table>

Helliconia bug (L. tricolor)

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>Fstat</th>
<th>p (model)</th>
<th>Coefficient</th>
<th>Estimate (coef.)</th>
<th>Std error (coef.)</th>
<th>p (coef.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMR ~ BS</td>
<td>-7.421</td>
<td>0.019</td>
<td>0.894</td>
<td>BS</td>
<td>0.351</td>
<td>2.557</td>
<td>0.894</td>
</tr>
</tbody>
</table>

Leaf footed cacuts bug (N. femorato)

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>Fstat</th>
<th>p (model)</th>
<th>Coefficient</th>
<th>Estimate (coef.)</th>
<th>Std error (coef.)</th>
<th>p (coef.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMR ~ BS</td>
<td>2.035</td>
<td>0.621</td>
<td>0.435</td>
<td>BS</td>
<td>0.7946</td>
<td>0.788</td>
<td>0.435</td>
</tr>
</tbody>
</table>

* weapon size estimated through principle components analysis (PCA)

All data were log$_{10}$ transformed prior to analysis
Table 3: Scaling relationship between muscle mass and body size for all species.

**Frog legged beetle (S. femorata)**

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>$F_{df}$</th>
<th>$p$ (model)</th>
<th>Coefficient</th>
<th>Estimate (coef.)</th>
<th>Std error (coef.)</th>
<th>$p$ (coef.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MM ~ WS</td>
<td>-131.439</td>
<td>488.5 $^{1,35}$</td>
<td>&lt; 0.0001</td>
<td>Slope</td>
<td>0.091</td>
<td>0.004</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>MM ~ BS</td>
<td>-132.795</td>
<td>508 $^{1,35}$</td>
<td>&lt; 0.0001</td>
<td>Slope</td>
<td>3.625</td>
<td>0.161</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>

**Indonesian stag beetle (C. metallifer)**

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>$F_{df}$</th>
<th>$p$ (model)</th>
<th>Coefficient</th>
<th>Estimate (coef.)</th>
<th>Std error (coef.)</th>
<th>$p$ (coef.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MM ~ WS</td>
<td>-23.63</td>
<td>56.76 $^{1,14}$</td>
<td>&lt; 0.001</td>
<td>Slope</td>
<td>1.22</td>
<td>0.162</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>MM ~ BS</td>
<td>-28.6</td>
<td>82.42 $^{1,14}$</td>
<td>&lt; 0.0001</td>
<td>Slope</td>
<td>5.065</td>
<td>0.558</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>

**Japanese rhinoceros beetle (T. dichotomous)**

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>$F_{df}$</th>
<th>$p$ (model)</th>
<th>Coefficient</th>
<th>Estimate (coef.)</th>
<th>Std error (coef.)</th>
<th>$p$ (coef.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MM ~ WS</td>
<td>-140.801</td>
<td>93.15 $^{1,55}$</td>
<td>&lt; 0.0001</td>
<td>Slope</td>
<td>0.063</td>
<td>0.006</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>MM ~ BS</td>
<td>-172.633</td>
<td>204 $^{1,55}$</td>
<td>&lt; 0.0001</td>
<td>Slope</td>
<td>2.505</td>
<td>0.175</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>

**Helliconia bug (L. tricolor)**

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>$F_{df}$</th>
<th>$p$ (model)</th>
<th>Coefficient</th>
<th>Estimate (coef.)</th>
<th>Std error (coef.)</th>
<th>$p$ (coef.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MM ~ WS</td>
<td>16.68</td>
<td>0.002 $^{1,7}$</td>
<td>0.968</td>
<td>Slope</td>
<td>0.144</td>
<td>3.481</td>
<td>0.968</td>
</tr>
<tr>
<td>MM ~ BS</td>
<td>15.934</td>
<td>0.612 $^{1,7}$</td>
<td>0.46</td>
<td>Slope</td>
<td>7.344</td>
<td>9.387</td>
<td>0.46</td>
</tr>
</tbody>
</table>

**Leaf footed cacuts bug (N. femorata)**

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>$F_{df}$</th>
<th>$p$ (model)</th>
<th>Coefficient</th>
<th>Estimate (coef.)</th>
<th>Std error (coef.)</th>
<th>$p$ (coef.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MM ~ WS</td>
<td>-54.668</td>
<td>19.42 $^{1,42}$</td>
<td>&lt; 0.0001</td>
<td>Slope</td>
<td>2.505</td>
<td>0.5685</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>MM ~ BS</td>
<td>-50.159</td>
<td>13.44 $^{1,42}$</td>
<td>&lt;0.001</td>
<td>Slope</td>
<td>2.043</td>
<td>0.557</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

* weapon size estimated through principle components analysis (PCA)

All data were log$_{10}$ transformed prior to analysis.
CHAPTER 4:

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

Devin M. O’Brien¹, Cerisse E. Allen¹, Melissa J. Van Kleeck², David Hone³, Robert Knell³, Andrew Knapp³, Stuart Christiansen⁴, Douglas J. Emlen¹

¹ Division of Biological Sciences, University of Montana, Missoula, MT, USA
² Department of Biology, University of Hawai’i at Mānoa, Honolulu, HI, USA
³ School of Biological and Chemical Sciences, Queen Mary University of London, UK
⁴ Sentinel High School, Missoula, MT, USA

Abstract:

The ‘positive allometry hypothesis’ predicts that ornaments and weapons of sexual selection will scale steeply when among-individual variation in trait size is compared with variation in overall body size. Intuitive and striking, this idea has been explored in hundreds of contemporary animal species and sparked controversy in paleobiology over the function of exaggerated structures in dinosaurs and other extinct lineages. Recently, however, challenges to this idea have raised question regarding the validity of the hypothesis.

We address this controversy in two ways. First, we suggest the positive allometry hypothesis be applied only to morphological traits that function as visual signals of individual body size. Second, because steep scaling slopes make traits better signals than other body parts, we propose that tests of the positive allometry hypothesis compare the steepness of the scaling relationships of focal, putative signal traits, to those of other body parts in the same organism (rather than to an arbitrary slope of 1).

We provide data for a suite of 29 extreme structures and show that steep scaling relationships are common when structures function as signals of relative body size, but not for comparably extreme structures that function in other contexts. We discuss these results in the context of animal signaling and sexual selection, and conclude that patterns of static scaling
offer powerful insight into the evolution and function of disproportionately large, or extreme, animal structures. Finally, using data from a ceratopsid dinosaur and a pterosaur, we show that our revised test can be applied to fossil assemblages, making this an exciting and powerful method for gleaning insight into the function of structures in extinct taxa.

**Keywords:** Sexual selection, scaling, animal signals, fossils

**Introduction**

Understanding how morphology scales with body size is one of the most pervasive topics in organismal biology [1–10]. 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 [11]; Rubner’s Surface Rule [12,13]; Cope’s Rule [14]; Rensch’s Rule [15–17]). Others account for transformations in shape arising during ontogeny (e.g., brain/body weight [5,18,19]; Dyar’s Law [20]). Here we focus on “static” allometry, scaling that occurs among individuals of the same age sampled from within populations [sensu 18,21,22].

Perhaps the most striking pattern in the study of static scaling is the observation that extreme products of sexual selection – ornaments of choice and weapons of battle –scale steeply with body size [4,23–38]. Specifically, when examined on a log scale, the relationship between the size of these structures and body size is greater than one (“positive allometry”) [3,39–41]. These steep scaling relationships cause ornaments and weapons to attain extraordinary proportions in the largest individuals, inspiring descriptions such as “extreme”, “exaggerated” [42] and “bizarre” [4] (Fig. 1).

Early studies of static scaling often focused on extreme products of sexual selection, including cervid antlers [1,2,43], fiddler crab (*Uca*) chelae [1], and beetle (Scarabaeidae) horns [44,45]. Since then, hundreds of extreme sexually selected structures have been examined, and
the overwhelming majority scale steeply with body size [e.g., 31,46–51]. 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 [52]; frog (Anura) forelimbs [53]; earwig forceps [25]), 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 [54]).

The positive allometry hypothesis has, however, been met with resistance. Bonduriansky [55] noted that the near universality of this pattern may be an artifact 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 [31,49,50]. Some extreme structures known to function as sexually selected ornaments, such as elaborate plumage in birds, do not scale positively with body size [56], nor do many genitalic traits, despite the fact that some experience strong selection for increased size [57,58]. 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 [55].

Furthermore, at least a few naturally selected structures, such as long bones in large mammals [59,60] and cranial horns in lizards [61], also scale positively with body size. Clearly, sexual selection need not lead to the evolution of steep scaling, and other agents of selection, such as locomotion and predator defense, 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, presence/absence of sexual dimorphism). 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 signaling 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 signaling 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-signaling contexts.

Methods

Specimen/structure selection and morphological measures

Species with putatively “extreme” structures – hereafter referred to as “focal structures” (see Appendix 4.4 for our classification of “extreme”) – were chosen from available taxa at the Phillip L. Wright Zoological Museum at the University of Montana (MT, USA), the Museum of
Comparative Zoology at Harvard (MA, USA), the Natural History Museum, London (UK), and the Emlen Lab Entomological Collection (MT, USA). Based on literature from these or closely related taxa, each focal structure was categorized as a “sexually selected signal” or “non-signal” structure (Table 1). Three 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 within the same organism, and ceratopsids (*Protoceratops andrewsi*) and pterosaurs (*Rhamphorhynchus muensteri*) to test the described methods on fossil datasets.

Reference structures were then chosen for each species as structures that could be consistently measured across all samples and lacked an obvious functional connection with the focal structure (minimizing the likelihood of correlational selection). 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 not available, 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 are provided in Table 1.

Dung beetles (*Sulcophanaeus menelas*), earwigs, large bee flies (*Bombylus major*), mantidflies (*Climaciella brunnea*), sabre wasps (*Rhyssa persuasoria*), and wildebeest (*Connochaetes taurinus*) were measured using photographs and ImageJ 1.50i software (NIH, USA). Ceratopsians (*Protoceratops andrewsi*) and pterosaurs (*Rhamphorhynchus muensteri*) were measured using ImageJ software, digital models, and digital calipers. All other species were measured using digital calipers.
**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 [62–65]. For every species, focal structure size and reference structure size were regressed on body size in separate models. Slope estimates ($\beta_{\text{focal}}$ and $\beta_{\text{reference}}$) were collected and 95% confidence intervals constructed. The 95% confidence intervals were then compared between focal and reference structures within the same species.

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

Coefficients of variation were calculated for every structure. Mean coefficient of variance was calculated across all signal structures and compared to the mean coefficient of variance calculated 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 ($\beta_{\text{focal}}$ and $\beta_{\text{reference}}$), differences between slopes ($\Delta \beta_{\text{focal-reference}}$), 95% confidence intervals surrounding $\beta_{\text{focal}}$, $\beta_{\text{reference}}$, and $\Delta \beta_{\text{focal-reference}}$, and coefficients of variation. For the majority of species with sexually selected signal structures, $\beta_{\text{focal}}$ was significantly greater than $\beta_{\text{reference}}$ (Table 1; Appendix 4.1). For three of these species, whitetail deer, wildebeest and pronghorn, $\beta_{\text{focal}}$ was greater than $\beta_{\text{reference}}$, but 95% confidence intervals surrounding these estimates were overlapping.

For all the species with exaggerated, non-signal structures, $\beta_{\text{focal}}$ and $\beta_{\text{reference}}$ were not significantly different (Table 1; Appendix 4.2). Unlike extreme sexually selected signal structures, extreme non-signal structures appear to scale similarly to reference structures within the same organism. Mean slope ($\beta_{\text{focal}}$) of all exaggerated sexually selected signal structures was greater than the mean slope ($\beta_{\text{focal}}$) of all exaggerated, non-signal structures ($t_{13.543} = -3.835$, $p < 0.01$) and 95% confidence intervals were non-overlapping (95% CI mean $\beta_{\text{focal}}$ for sexually selected signal structures [1.709, 4.56]; 95% CI mean $\beta_{\text{focal}}$ for non-signal structures [0.374, 0.783]). Mean $\Delta \beta_{\text{focal-reference}}$ for sexually selected signal structures was greater than mean $\Delta \beta_{\text{focal-reference}}$ for non-signal structures ($t_{14.164} = 4.079$, $p = 0.001$; SI 3) and 95% confidence intervals did not overlap (95% CI mean $\Delta \beta_{\text{focal-reference}}$ for sexually selected signal structures [1.072, 3.831]; 95% CI mean $\Delta \beta_{\text{focal-reference}}$ for non-signal structures [-0.501, 0.078]; Appendix 4.3).

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

Within species, sexually selected signal structures scaled steeply with body size (Table 1; Appendix 4.1). In the majority of sexually selected species surveyed here, the scaling relationship of the signal ($\beta_{\text{focal}}$) was significantly steeper than that of the reference structure ($\beta_{\text{reference}}$). Surprisingly, this pattern did not hold for whitetail deer (*Odocoileus virginianus*), wildebeest, or pronghorn (*Antilocapra americana*). In these species, $\beta_{\text{focal}}$ was greater than $\beta_{\text{reference}}$, but 95% confidence intervals surrounding these estimates were overlapping (Table 1; Appendix 4.1). This was likely either an artifact of relatively small sample size (n < 18 for whitetail deer and pronghorn) or biased sampling (e.g., hunters favoring largest antlered males in sampled populations), since previous work has shown positive allometry and/or strong selection for these, and similar, weapons [e.g., 66–68]. 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 [25,28,31,47,48,50,69,70].

Every exaggerated non-signal structure measured scaled with a slope that was not significantly different than that of the reference structure (Table 1; Appendix 4.2). In addition, across species, the scaling relationship ($\beta_{\text{focal}}$) of sexually selected signal structures was significantly steeper than that of non-signal structures ($t_{11.902} = -3.23, p < 0.01$; SI 3). 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 71–73]. In the context of sexual selection, receivers are primarily 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 [71,74,75]. In both cases, information encoded in the signal pertains to the overall genetic quality and/or condition of the bearer [reviewed in 76].

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 [71]. 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 [25,48,56,77–83]. Hypervariability in trait size amplifies associated variation in male quality, making these otherwise subtle differences easier to see [77,83,84].

Effective signals must also be honest. If poor quality males can cheat by producing effective signals, then receivers should focus on other traits. One form of honesty arises when the growth of signal traits is condition-sensitive [28,31,85–92]. Condition-sensitive growth of signal structures may “capture” genetic variation underlying overall quality, making these signals virtually impossible to fake [80,93,94]. Indeed, sexually selected signal structures are notoriously sensitive to stress, parasite load, and nutrition [66,95–101].

Both hypervariability and heightened condition sensitivity cause structures to be reliable and informative as signals of quality [71,74,102,103], and these basic characteristics are shared by a wealth of sexually selected signals [reviewed in 71]. 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 [31,47,69,70,92]. 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 [41,82,104,105]. 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 [106] and birds [107]. 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., 28,81,107,108]. Similarly, signals that vary in other ways besides size (e.g., color, behavior, 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 (Saturnia pyri) antennae, measured here; Table 1), are not predicted to scale steeper than reference structures, since hypervariation and/or condition sensitivity may decrease performance. 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 gabonica*) fangs, elephant shrew (*Elephantatus fuscus*) snouts, and mantidfly forelegs all scaled with relatively steep slopes (i.e., $\beta > 1$), but the reference structures were hyperallometric too (Appendix 4.2; 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 [e.g., 109–111]. 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., 106, 112, 113]. 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 abundance, 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 [74, 114–116]. Similarly, antagonistic coevolutionary arms races arising from conflict between males and females can drive rapid evolution of genitalia [117–121]. 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., 26, 122, 123].

Exaggerated size can also arise through natural selection as, for example, in some locomotor, prey capture, and feeding structures [reviewed in 104]. 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” [124–127]. However, similar to 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 non-signaling structures (Table 1; Appendix 4.2). 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 [128], 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 behavior of extinct taxa cannot be observed. Even so, lines of evidence can be drawn from static, morphological data to provide testable hypotheses of behavior [129]. 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 [54,e.g., 130]. 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 [43,54,131]. However, such inferences remain controversial [e.g., 132–138]. 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., 137]. As a result, detection of even fundamental patterns in morphology, such as sexual dimorphism, remains elusive [138, but see 139]. 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; 140,141].

Despite these limitations, we suggest behavior 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 131], 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 signaling 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 these methods 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 sexual dimorphism and trait specific coefficients of variation.
Author contributions


Acknowledgements

We would like to thank the NSF for funding (IOS – 1456133), Libby Beckman, Sandra Chapman, Judith Chupasko, Natalie Cooper, Tannis Davidson, Amy Henrici, Rick Hunter, Julian Kiely, Matthew Lemanna, Carl Mehling, Hana Merchant, Robert Portela Miguez, Robert Niese, José Rosado, Daniela Schwartz-Wings, Justyna Slowiak, Zhao Qi, Stig Walsh for access to specimens and data. Thank you to Brenden Holland, Emy Roberge, and Kellie Sorenson for help with collection of Jackson’s Chameleon data. We are also grateful to Erica McAlister, Alessandro Giusti and Gavin Broad for allowing access to the Diptera, Lepidoptera and Hymenoptera collections at the Natural History Museum, London, UK (NHMUK).

Photo credit Fig 1: Sheep, Jeremy Weber; T. jacsonii, Bengimint444; Mantid, Oliver Koemmerling (creativecommons.org/licenses/by-sa/3.0/deed.en); Wasp, Seney Natural History association; Viper Brimac the 2nd; Scarabaeidae, Bernanrd DUPONT. Fig 3: P. andrewsi, FunkMonk, Rhamphorhynchus, M0tty (gnu.org/licenses/fdl-1.3.en.html; creativecommons.org/licenses/by-sa/3.0/deed.en). Licensed under creativecommons.org/licenses/by/2.0/deed.en unless otherwise specified.
Literature cited


2. Thompson DW. On growth and form. Growth Form. 1917


42. Darwin C. The Descent of Man: And Selection in Relation to Sex. Appleton; 1871.


47. Petrie M. Intraspecific variation in structures that display competitive ability: large animals invest relatively more. Anim Behav. 1988;36: 1174–1179.


73. Searcy WA, Nowicki S. Signal interception and the use of soft song in aggressive interactions. Ethology. 2006;112: 865–872.


186. Szucsich NU, Krenn HW. Morphology and function of the proboscis in Bombyliidae (Diptera, Brachycera) and implications for proboscis evolution in Brachycera. Zoomorphology. 2000;120: 79–90. doi:10.1007/s004350000025


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. gabonica*; ns), and dung beetle horns (Scarabaeidae, ss). Photo credits in Acknowledgments.
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].
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.
<table>
<thead>
<tr>
<th>Common name</th>
<th>Scientific name</th>
<th>n</th>
<th>Sexual trait</th>
<th>Reference trait</th>
<th>Body-size estimate</th>
<th>Reference trait estimate</th>
<th>Statistical test</th>
<th>Reference (f)</th>
<th>Reference (r)</th>
<th>Reference for assigned trait function (selection/signal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amatitlana</td>
<td><em>Mammuthus pacificus</em></td>
<td>13</td>
<td>ear length</td>
<td>cranial length</td>
<td><em>Mammuthus pacificus</em></td>
<td>n</td>
<td>0.366</td>
<td>(0.28, 0.45)</td>
<td>0.851</td>
<td>0.301</td>
</tr>
<tr>
<td>Ameloceras</td>
<td>Phaneroceras rubrum</td>
<td>10</td>
<td>trunk length</td>
<td>snout-vent length</td>
<td><em>Phaneroceras rubrum</em></td>
<td>n</td>
<td>2.652</td>
<td>0.851</td>
<td>0.301</td>
<td>Redmond and Bradley (2006)</td>
</tr>
<tr>
<td>Ammonites</td>
<td><em>Phaneroceras rubrum</em></td>
<td>10</td>
<td>trunk length</td>
<td>snout-vent length</td>
<td><em>Phaneroceras rubrum</em></td>
<td>n</td>
<td>2.652</td>
<td>0.851</td>
<td>0.301</td>
<td>Redmond and Bradley (2006)</td>
</tr>
<tr>
<td>Anemone</td>
<td><em>Phaneroceras rubrum</em></td>
<td>10</td>
<td>trunk length</td>
<td>snout-vent length</td>
<td><em>Phaneroceras rubrum</em></td>
<td>n</td>
<td>2.652</td>
<td>0.851</td>
<td>0.301</td>
<td>Redmond and Bradley (2006)</td>
</tr>
<tr>
<td>Ancyloceras</td>
<td><em>Phaneroceras rubrum</em></td>
<td>10</td>
<td>trunk length</td>
<td>snout-vent length</td>
<td><em>Phaneroceras rubrum</em></td>
<td>n</td>
<td>2.652</td>
<td>0.851</td>
<td>0.301</td>
<td>Redmond and Bradley (2006)</td>
</tr>
<tr>
<td>Anomaloceras</td>
<td><em>Phaneroceras rubrum</em></td>
<td>10</td>
<td>trunk length</td>
<td>snout-vent length</td>
<td><em>Phaneroceras rubrum</em></td>
<td>n</td>
<td>2.652</td>
<td>0.851</td>
<td>0.301</td>
<td>Redmond and Bradley (2006)</td>
</tr>
</tbody>
</table>

Table 1: Summary of study species and results. (f) = focal trait, (r) = reference trait, CV = coefficient of variation, β = slope of scaling relationship between trait size and body size, Δβ = difference between β(f) and β(r). * = sexual dimorphism may be impossible to detect [see 137].
Appendix 4.1: 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 4.2: Scaling relationships for extreme naturally selected/non-signal structures. Lines represent ordinary least squares regression of log₁₀ standardized structure size on log₁₀ 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.

Standardized log₁₀ body size
Appendix 4.3: Comparison of $\Delta\beta_{\text{focal-reference}}$ (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 ($n = 15$). $\Delta\beta_{\text{focal-reference}}$ of extreme sexually selected signal structures is significantly greater than $\Delta\beta_{\text{focal-reference}}$ of extreme non-signal structures ($t_{15.616} = 4.153 \ p < 0.001$).
Appendix 4.4: 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 [142]. 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 [reviewed in 42,50,74]. 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 [142]. 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 \textit{(Acrocinus longimanus)} forelegs are relatively larger than midlegs or hindlegs [143]) 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 [143]). 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.

\textbf{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 [144,Dipodidae; 145] and the raptorial forelimbs of mantidflies [mantispidae; 146]. 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., 145]; and/or c) by mapping changes in trait size onto a phylogeny and testing for lineage specific changes in relative trait size [145].