Allometry of Genitalia in Insects and Spiders: One Size Does Not Fit All

Andy J. Green

COMMENTS

ALLOMETRY OF GENITALIA IN INSECTS AND SPIDERS: ONE SIZE DOES NOT FIT ALL

ANDY J. GREEN
Department of Applied Biology, Estación Biológica de Doñana, CSIC, Avda. María Luisa s/n, 41013 Sevilla, Spain
E-mail: andy@ebd.csic.es

Eberhard et al. (1998) reported that genitalia had lower allometric slopes than other body parts in all of 20 species of insects and spiders studied. This led them to suggest that selection on genitalia in both sexes favors intermediate, standard sizes appropriate for the most typical size of the opposite sex ("one size fits all") and that sexual selection has different effects on the scaling of genitalia and other body parts. Here I demonstrate that their results are dependent on the methodology used in their analyses and that reanalysis using more appropriate methods shows that the evidence that genitalia tend to be negatively allometric is not conclusive.

Eberhard et al. (1998) analyzed allometry by linear ordinary least squares (OLS) regression of log(length of body part) on log(length of body-size indicator). For each species, several different measures of genitalia and nongenitalia were all regressed on the "body size indicator" (another nongenitalic measure, usually of the thorax in insects and the cephalothorax in spiders). In males of *all* 20 species, Eberhard et al. found that the median OLS slope for genitalic body parts was lower than that for nongenitalian body parts. Combining all the genitalic slopes together (up to five for each species), in 60 of 63 cases the OLS slope was less than the median nongenitalic slope for the same species. (Note such combined analyses are not statistically sound because repeated measurements of the same individual's genitalia are not independent [as suggested by Eberhard et al., p. 418]. However, I include them in this article for the sake of comparison.)

These highly significant results obtained by Eberhard et al. are a consequence of their use of OLS regression, which is inappropriate because the assumptions of this model I regression are violated for studies of this type (for a fuller explanation and discussion of alternative methods than that presented below, see Ricker 1984; Rayner 1985; McArdle 1988; Pagel and Harvey 1988; LaBarbera 1989; Martin and Barbour 1989; Jolicoeur 1990; Sokal and Rohlf 1995; Mesplé et al. 1996).

The slope of the line of best fit to a bivariate plot is heavily dependent on the assumptions underlying the regression method used. The assumptions of the OLS method include: (1) that y is dependent on x in a strict, causal manner; and (2) that x is measured without sampling or measurement error and that only y is subject to such error. I emphasise that "error" as used here and below (with quotes) is not synonymous with measurement error, which is only one of its components.

In the study of Eberhard et al. (1998), y (body part I) is not directly dependent on x (body part II) in the manner required for OLS regression. The authors themselves state "our analyses are not meant to imply, as is usual in regression analyses, that one variable was dependent on another" (p. 417) and "we do not intend to make any claims regarding cause-effect relationships with body size" (p. 421–422). Likewise, the assumption that there is no "error" in x is also violated because x is subject to both natural variation and measurement error.

As a consequence, the OLS method used by Eberhard et al. systematically underestimates the slope best describing the functional relationship between x and y (hereafter described as the "true slope"). The correct way to place the best-fit line depends on the precise nature of the "error" in x and y (Rayner 1985), which is unknown for Eberhard et al.'s study. The true slope can be expected to lie somewhere between the OLS slope for y regressed on x (*b*) and the OLS slope for x regressed on y (equal to *b* divided by r^2; Ricker 1984; McArdle 1988; Martin and Barbour 1989).

Instead of presenting OLS slopes of y on x, Eberhard et al. would have had just as much statistical justification in presenting OLS slopes of x regressed on y. However, changing the slopes they presented to OLS regressions of x on y transforms their results, reversing the tendency for genitalic slopes to be shallower. Median x-on-y slopes for genitalia are higher than those for nongenitalia in 12 of 20 species (Table 1, sign test, $P > 0.8$). In 29 of 63 cases (sign test, $P > 0.6$), the genitalic slope is now higher than the median nongenitalic slope for the same species.

Eberhard et al. are either unaware of the problems of using OLS regression or they implicitly assume that, because they are comparing slopes of one character with another, the regression method is irrelevant (i.e., that the same character has the higher slope with any method). This is not the case because when the allometric slope for one character is higher with an OLS regression of y on x, it may easily be lower with other methods (Fig. 1).

In the presence of unmeasured variation in both x and y, no line-fitting method gives an unbiased result with real data. However, two model II regression methods, major axis (MA) and reduced major axis (RMA), provide valuable alternatives to OLS. Both MA and RMA are standard techniques in allometry studies (e.g., Boag 1984; LaBarbera 1989; Herrera 1992; Cane 1993; Simmons and Scheepers 1996; Silva 1998), and there is debate about their relative merits (Ricker 1984; McArdle 1988; Jolicoeur 1990; Sokal and Rohlf 1995).

The lower the value of the correlation coefficient *r*, the more the OLS method of regressing y on x tends to under-

1621

TABLE 1. Median slopes of log-log regressions of nongenitalic and genitalic characters in males of 20 species of spiders (top six species) and insects, recalculated from Eberhard et al. (1998). Slopes are from ordinary least squares regression of x on y and from reduced major axis regression.

Species	Nongenitalia		Genitalia	
	x on y	RMA	x on y	RMA
Philoponella vicina	0.802	0.778	3.182	1.699
Argiope trifasciata	1.339	1.259	1.096	0.757
Araneus expletus	1.550	1.364	0.875	0.700
Tetragnatha sp.	1.161	0.955	0.593	0.469
Physocyclus globosus	0.888	0.799	0.606	0.432
Metagonia rica	2.608	1.904	2.632	1.027
Macrodactylus sylphis	1.664	1.186	1.601	0.810
Phyllophaga obsoleta	2.050	1.284	2.284	0.933
Onthophagus incensus	0.938	0.904	4.617	1.510
Chelymorpha alternans	1.631	1.122	-5.920	-0.381
Macrohaltica jamaicensis	0.907	0.825	1.053	0.564
Ceratoma trifurca	1.434	1.133	1.812	0.905
E. kamerunicus	1.139	0.952	6.321	1.321
Pseudoxychila tarsalis	0.913	0.776	11.992	0.674
Archisepsis diversiformis	0.895	0.811	1.859	0.620
Ceratitis capitata	0.983	0.924	0.781	0.397
Camponotus sp.	1.785	0.767	4.662	0.660
Ozophora baranowskii	3.535	1.155	1.368	0.588
Paralabella dorsalis	3.440	1.273	4.400	1.008
Hetaerina fuscovittata	2.409	1.312	9.911	1.487

FIG. 1. Ordinary least squares (OLS) regression of y on x (dotted line) and reduced major axis regression (solid line) for $\log_e$ culmen length (mm) and $\log_e$ cranium length against $\log_e$ body weight (g) for male marbled teal *Marmaronetta angustirostris* (n = 26; Green, in press). The OLS slope for culmen length (0.30) is greater than that for cranium length (0.15), whereas the RMA slope for cranium length (0.79) is greater than that for culmen length (0.49). The corresponding OLS slopes for regressions of x on y (not shown) are 4.12 for cranium length and 0.79 for culmen length.

estimate the true slope. This leads to a positive correlation between r and b, which is readily observed in the appendices of Eberhard et al. (1998). The differences in OLS slopes they reported between genitalia and nongenitalia are largely a reflection of the difference in values of r between genitalian and nongenitalian body parts and the body-size indicators. The median r-value is lower for genitalia in *all* 20 species (sign test, $P < 0.001$). Thus, it is no surprise that the OLS slopes of y on x are consistently lower for genitalia. This important difference in r-values indicates that genitalic measures covaried less well with the body-size indicator than nongenitalic body parts. It may partly be a reflection of higher measurement error in genitalia (which were generally smaller) than in nongenitalian body parts. Although Eberhard et al. (1998) present repeated measurements (in their Table 1) for two of the 20 species studied, they present no estimates of measurement error, which is defined as that portion of the total sample variance made up by within-individual variation (Yezerinac et al. 1992). Therefore, it is unclear whether measurement error was higher for genitalic measures.

Another likely explanation for the lower r-values for genitalia is that the nongenitalic body parts (e.g., length of pronotum or width of head) have a closer ontogenetic or functional relationship with the measure chosen by Eberhard et al. as the body size indicator (e.g., width of pronotum or thorax). Both empirical and theoretical work suggest that developmentally or functionally related characters tend to covary (Fairbairn 1992; Cane 1993). For example, in birds different skeletal measures correlate more strongly with each other than with plumage measures and vice versa (Freeman and Jackson 1990; Gosler et al. 1998). If a different body-size indicator had been used by Eberhard et al. (e.g., total dry mass or PC1), the r-values for genitalic and nongenitalic body parts would have been more similar.
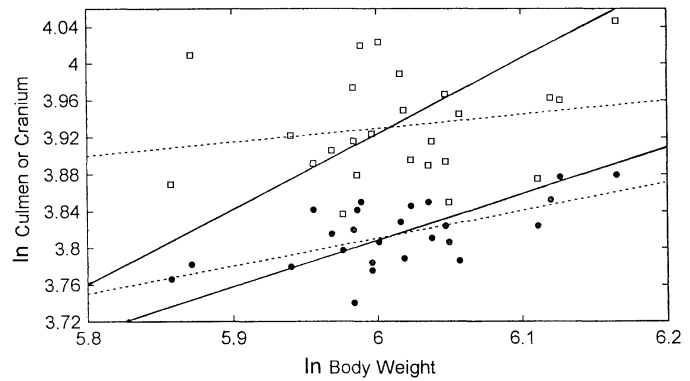
OLS slopes cannot be expected to reflect the functional relationship between the size of genitalia or other body parts and "body size." Eberhard et al. present sufficient data to allow calculation of the RMA regression slopes (equal to the OLS slope divided by r). The median genitalic RMA slope is lower than the median nongenitalic slope in 16 of 20 species (Table 1), a significant result (sign test, $P = 0.012$). Although a significant trend remains with RMA analyses, this is very different to the *universal* trend for lower genitalic slopes presented by Eberhard et al. (1998). In 45 of 63 cases, the genitalic slope was lower than the median nongenitalic slope for the same species (sign test, $P = 0.001$).

Results presented by Eberhard et al. for female genitalia undergo similar changes when reanalyzed. Eberhard et al. report that all 12 allometric slopes for female genitalia (in a total of five species) were lower than the median nongenitalic values for the same species ($P < 0.001$). With RMA analyses, only seven of 12 allometric slopes for female genitalia were lower than median nongenitalic values (sign test, $P > 0.7$).

For males of *all* 20 species studied and for females of the five species studied, the median OLS genitalic slopes presented by Eberhard et al. are less than one, thus suggesting negative allometry. However, with RMA methods median male genitalic slopes were below one for only 14 of 20 species (Table 1, sign test, $P > 0.1$) and female genitalic slopes were below one for only two of five species. Thus, RMA methods provide no compelling evidence for a general pattern of negative genitalic allometry in insects and spiders. The overall trend does not deviate from isometry, suggesting that there is no selection for a standard, nonvariable size of genitalia that fits all members of the opposite sex (contrary to Eberhard et al., p. 419).

As Eberhard et al. found with OLS regression slopes, the median RMA slopes for male genitalia in five species with rigid female genitalia (range -0.38–1.7) did not differ significantly from those in seven species with soft genitalia (range 0.6–1.0). Thus, there is no evidence that mechanical

lock-and-key considerations demand negative allometry in male genitalia.

Without information on the distribution of "error" in y or x, it is not possible to know whether RMA or MA methods are more appropriate for analyzing allometry in the data of Eberhard et al. The "error" distributions may vary considerably among the 20 species studied and between genitalic and nongenitalic y-values (e.g., in response to the different measurement methods and different human observers used; see Yezerinac et al. 1992; Palmeirim 1998). Therefore, different regression techniques may be more appropriate in different cases. Unless the trend for male genitalic slopes to be lower than nongenitalic slopes seen with RMA methods can be shown to be repeated with MA methods, it cannot be accepted with confidence.

In any case, such a comparison of genitalic and nongenitalic slopes is heavily influenced by the nature of the nongenitalic characters studied. Eberhard et al. aim (at least implicitly) to compare slopes for genitalia with those of other body parts not important in sexual selection. However, little is known about the biology of most species measured by Eberhard et al., and some of the characters included among the nongenitalia may have particularly high slopes because they are sexually selected. Current theory suggests that positively allometric characters are likely to be sexually selected, whether as weapons or visual stimuli (Shine 1990; Emerson and Voris 1992; Petrie 1992; Simmons and Scheepers 1996; Simmons and Tomkins 1996). Many of the 27 of 59 male nongenitalic characters included by Eberhard et al. whose RMA slopes were greater than one may be sexually selected (e.g., head horns or leg dimensions; Brown and Bartalon 1986; Zeh et al. 1992), and their inclusion may have biased the median nongenitalic slopes used in their analyses.

There are also problems with the analyses of coefficients of variation (CV) presented by Eberhard et al. They report that the CVs of male genitalic measurements were lower than the median CV for nongenitalic measurements of that species in 41 cases (a significant trend), whereas their table 1 shows this to be the case in only 38 cases (two others were equal; sign test, $P > 0.07$). They correctly report that the median CV in genitalic measurements was larger than the median nongenitalic CV in six species and smaller in 13. However, they give an erroneous significant χ^2 value of 4.26, when the actual value is 2.58 ($P > 0.1$). Thus, contrary to Eberhard et al. (1998), there appears to be no evidence for an overall difference in coefficient of variation between genitalia and other body parts.

In conclusion, a combination of statistical problems means that the evidence presented by Eberhard et al. (1998) that insect and spider genitalia tend to be negatively allometric is not compelling. RMA allometric slopes for genitalia do not differ from isometry. RMA genitalic slopes still tend to be lower than RMA nongenitalic slopes, although the consistent pattern reported by Eberhard et al. with OLS methods is no longer observed. Furthermore, this result may be a product of the inclusion of sexually selected nongenitalic characters in the analysis. The wide variation in allometric slopes observed for genitalia remains to be explained and suggests that the sexual and natural selection pressures acting on genitalian scaling vary widely between different characters and species. This is to be expected because the 20 species studied by Eberhard et al. represent 15 families of insects and spiders.

LITERATURE CITED

Boag, P. T. 1984. Growth and allometry of external morphology in Darwin's finches (*Geospiza*) on Isla Daphne Major, Galapagos. J. Zool. Lond. 204:413–441.

Brown, L., and J. Bartalon. 1986. Behavioral correlates of male morphology in a horned beetle. Am. Nat. 127:565–570.

Cane, W. 1993. The ontogeny of postcranial integration in the common tern, *Sterna hirundo*. Evolution 47:1138–1151.

Eberhard, W. G., B. A. Huber, R. L. Rodriguez, R. D. Briceño, I. Salas, and V. Rodriguez. 1998. One size fits all? Relationships between the size and degree of variation in genitalia and other body parts in twenty species of insects and spiders. Evolution 52:415–431.

Emerson, S. B., and H. Voris. 1992. Competing explanations for sexual dimorphism in a voiceless Bornean frog. Funct. Ecol. 6: 654–660.

Fairbairn, D. J. 1992. The origins of allometry: size and shape polymorphism in the common waterstrider, *Gerris remigis* Say (Heteroptera, Gerridae). Biol. J. Linn. Soc. 45:167–186.

Freeman, S., and W. M. Jackson. 1990. Univariate metrics are not adequate to measure avian body size. Auk 107:69–74.

Gosler, A. G., J. J. D. Greenwood, J. K. Baker, and N. C. Davidson. 1998. The field determination of body size and condition in passerines: a report to the British Ringing Committee. Bird Study 45:92–103.

Green, A. J. In press. The scaling and selection of sexually dimorphic characters: an example using the Marbled Teal. J. Avian Biol.

Herrera, C. M. 1992. Interspecific variation in fruit shape: allometry, phylogeny and adaptation to dispersal agents. Ecology 73: 1832–1841.

Jolicoeur, P. 1990. Bivariate allometry: interval estimation of the slopes of the ordinary and standardized normal major axes and structural relationship. J. Theor. Biol. 144:275–285.

LaBarbera, M. 1989. Analyzing body size as a factor in ecology and evolution. Annu. Rev. Ecol. Syst. 20:97–117.

Martin, R. D., and A. D. Barbour. 1989. Aspects of line-fitting in bivariate allometric analyses. Folia Primatol. 53:65–81.

McArdle, B. H. 1988. The structural relationship: regression in biology. Can. J. Zool. 66:2329–2339.

Mesplé, F., M. Troussellier, C. Casellas, and P. Legendre. 1996. Evaluation of simple statistical criteria to qualify a simulation. Ecol. Modell. 88:9–18.

Pagel, M. D., and P. H. Harvey. 1988. The taxon-level problem in the evolution of mammalian brain size: facts and artifacts. Am. Nat. 132:344–359.

Palmeirim, J. M. 1998. Analysis of skull measurements and measurers: can we use data obtained by various observers? J. Mamm. 79:1021–1028.

Petrie, M. 1992. Are all secondary sexual display structures positively allometric and, if so, why? Anim. Behav. 43:173–175.

Rayner, J. M. V. 1985. Linear relations in biomechanics: the statistics of scaling functions. J. Zool. Lond. 206:415–439.

Ricker, W. E. 1984. Computation and uses of central trend lines. Can. J. Zool. 62:1897–1905.

Shine, R. 1990. Function and evolution of the frill of the frillneck lizard, *Chlamydosaurus kingii* (Sauria: Agamidae). Biol. J. Linn. Soc. 40:11–20.

Silva, M. 1998. Allometric scaling of body length: elastic or geometric similarity in mammalian design. J. Mamm. 79:20–32.

Simmons, L. W., and J. L. Tomkins. 1996. Sexual selection and the allometry of earwig forceps. Evol. Ecol. 10:97–104.

Simmons, R. E., and L. Scheepers. 1996. Winning by a neck: sexual selection in the evolution of giraffe. Am. Nat. 148:771–786.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry. 3rd ed. W. H. Freeman and Co., New York.

Yezerinac, S. M., S. C. Lougheed, and P. Handford. 1992. Mea-

surement error and morphometric studies: statistical power and observer experience. Syst. Biol. 41:471–482.

Zeh, D. W., J. A. Zeh, and G. Tavakilian. 1992. Sexual selection and sexual dimorphism in the harlequin beetle *Acrocinus longimanus*. Biotropica 24:86–96.

DON'T FORGET THE BIOLOGY: A REPLY TO GREEN

W. G. EBERHARD,[1] B. A. HUBER,[2] AND R. L. RODRIGUEZ[3]

[1]*Escuela de Biología, Universidad de Costa Rica, Ciudad Universitaria, Costa Rica, and Smithsonian Tropical Research Institute*
E-mail: weberhar@cariari.ucr.ac.cr
[2]*Department of Entomology, American Museum of Natural History, New York, New York 10024*
[3]*Department of Entomology, Hawarth Hall, University of Kansas, Lawrence, Kansas 66045*

In our 1998 paper (Eberhard et al. 1998), we tested several hypotheses regarding the possible selective factors involved in the evolution of animal genitalia. We compared the slopes of log-log ordinary least squares (OLS) regressions of genitalia on indicators of body size with the slopes of other body parts on the same indicators in 20 species of insects and spiders. Our major conclusions regarding the rejection of male-female conflict, good viability genes, and lock and key arguments (see Abstract) are unaffected by the reanalysis proposed by Green (1999), even if it were more appropriate than ours, which we doubt (see below).

In neither our OLS regressions nor Green's reduced major axis (RMA) regressions are the slopes of genitalia greater than those of other body parts, as would be expected if genitalia were used as weapons in forceful intraspecific battles (the male-female conflict hypothesis) or as signals of male size (good viability genes hypothesis). Instead, the slopes show a statistically significant trend to be lower in both analyses. In addition, neither we nor Green found differences on comparing the slopes of the genitalia of the species in which lock and key considerations might be important (species in which male genitalia fit against rigid female genitalic structures) with the slopes in species in which lock and key can be ruled out because of the mechanical mesh of the male's genitalia with those of the female. So, conservatively, we conclude that the major conclusions of our paper are not affected under Green's reanalysis. It is important not to lose sight of the biological questions being tested in debates over statistical methods.

There are a number of reasons, however, to doubt several of Green's points. Green's claim that reanalysis is needed, the analysis he performed, and the additional explanations that he proposed in preference to those we mentioned in our original paper all have serious problems. We will discuss first the statistical questions, and then the more directly biological questions.

Is RMA Regression More Appropriate?

Many of Green's objections (see the first half of his final, summary paragraph) hinge on the different values he obtained when he used RMA regression analyses. We struggled with the question of whether we should use RMA rather than OLS regressions while we were preparing our original paper, and finally decided against RMA for two reasons. The first is that it is not obvious which of the two types of regression is more appropriate, despite the impression given by Green. For instance, Sokal and Rohlf (1995, p. 544) mention that the RMA regression technique has been the subject of "serious criticisms" (see also reservations expressed by McArdle 1988; Martin and Barbour 1989). LaBarbera (1989) states that "OLS regression is appropriate only when the goal is to . . . allow prediction of expected values given one of the two variables" (which is the "good genes" hypothesis we wanted to test). Sokal and Rohlf also note that if a causality relation exists between variables, one should employ OLS rather than RMA regressions (p. 545). While, as Green noted, we were careful in our paper to avoid claiming that we could establish cause-effect relations between the particular body size indicators and the other variables we used, we think that a complete lack of cause-effect between overall body size and the sizes of both genitalic and nongenitalic traits is highly improbable. Green's proposal that there is equal justification for presenting regression slopes of y regressed on x and x regressed on y does not make biological sense (the size of an animal's body does not seem likely to be determined by the size of its genitalia). Inasmuch as our indicator variables are indeed indicators of overall body size, then at least indirect cause effect relationships with other variables probably do exist. The analyses that we performed using alternative indicators of body size (p. 418, second paragraph of Results) generated nearly identical results, suggesting that our indicators may indeed be reasonable esti-