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-

mates of overall body size. In summary, the superiority of RMA analyses over OLS analyses for our data is not as clear cut as Green suggests.

Second, and more importantly, Green used the value of the correlation coefficient r to calculate the RMA slope (= OLS slope/r). As a glance at our Appendix 2 will show, 27 of the r-values for males and seven for females were not statistically significant. Dividing the OLS slope by a meaningless number can only give a meaningless result. One possible solution we considered was to use RMA analyses and to simply discard all traits with nonsignificant r-values. But 31 of the 34 nonsignificant values were for genitalic traits, so this would amount to throwing the baby out with the bathwater—dumping informative data to facilitate a statistical test. In fact, the imbalance in nonsignificant values between genitalia and nongenitalia (next paragraph) illustrates a major aspect of the news from our data, that there is often an especially poor correlation between the sizes of genitalic and nongenitalic body parts. Nor could we bring ourselves to accept the alternative of using meaningless RMA numbers as if they were meaningful (as was done by Green in both his RMA analyses and his calculations of the x on y values in his table 1, where he used b/r^2). A final possibility would be to assume that all RMA slopes calculated with nonsignificant r-values are zero, a solution that is probably inappropriate for RMA analyses (Sokal and Rohlf 1995, p. 544; doing this results in totals nearly identical to those we reported for OLS regressions: six RMA genitalic slopes higher and 62 lower than the median for the nongenitalic structures).

If one were to opt for throwing out babies and discarded all RMA values based on nonsignificant values of r and nonsignificant OLS slopes ($n = 4$), none of our original conclusions is altered. First consider the "baby." The trend for nonsignificant values of r to occur for genitalia rather than for nongenitalia is highly significant: whereas 24 of 63 r-values for male genitalia and seven of 12 for female genitalia were nonsignificant, only three of 59 r-values for nongenitalia in males and zero of 10 in females were nonsignificant ($\chi^2 = 19.4$, df = 1, P << 0.001 for males, P = 0.0053 with Fisher's exact test for females). All four nonsignificant OLS slopes were also for genitalia. What about the RMA slopes calculated with the data that are left? Considering species-by-species comparisons of male structures, only 15 species of the original 20 remain in the analysis. In 12 of these, the median RMA slope of nongenitalia is higher than the median RMA slope of genitalia ($\chi^2 = 5.4$, df = 1, P = 0.02), which is the same proportion as Green's 16 of 20. Removing the horn of *Onthophagus* (see below) does not affect this result. Comparing individual genitalic slopes with the median nongenitalic slope for each species, the proportion is 29 of 35, still highly significant ($\chi^2 = 15.1$, df = 1, P < 0.001). In females, which according to Green show no trend, only three of the five species remain in the analysis. In all three the mean genitalic RMA is lower than the mean nongenitalic RMA, and when individual genitalic slopes are compared with nongenitalic slopes, all six genitalic slopes are lower. The trend in females is thus similar to that in males. Finally, the lack of difference between species with rigid and soft female genitalia also remains. When the species with "?" in our Appendix 1 and with nonsignificant values for r and OLS

slope are removed, data remain for males of four rigid and six soft species; their median genitalic RMA slopes are 0.43, 0.70, 0.76, and 1.25 for the rigid species, and 0.46, 0.46, 0.83, 0.97, 1.01, and 1.02 for the soft species.

In summary, no matter whether the slopes are from OLS regressions, from RMA regressions in which all r-values are used, or from RMA regressions in which only statistically significant r and OLS values are used, the answer is the same: Genitalia have significantly lower slopes than nongenitalic body parts of the same animals. Our conclusions regarding the hypotheses we tested are not changed.

A second statistic which Green emphasizes is the correlation coefficient r. We avoided emphasizing r in our discussion because it does not allow one to distinguish between the tightness of the fit of data around the regression line (we used the "standard error of estimate" to measure this dispersion) and the slope of this line (we used the OLS regression slope to estimate this). Setting aside these statistical fine points, r has great biological importance for one of the questions we asked in our paper: Do male genitalia offer particularly reliable cues to females as indicators of male size? The low r-values for genitalia that were emphasized by Green ("This important difference in r-values indicates that genitalic measures covaried less well with the body size indicators . . .") give a clear negative answer to our question: No, male genitalia give particularly unreliable cues of male size to females.

Green's emphasis on statistics rather than biology shows up again in his discussion of whether or not genitalia tend to be negatively allometric (slopes less than 1.0). If one discards all RMA values based on nonsignificant r-values, 10 of 15 species have median slopes less than 1.0 (the corresponding number for OLS slopes was 20 of 20). But there are of course still other Model II regression techniques, and, as Green notes, different techniques have different biases and will tend to give higher or lower absolute values. The outcomes of comparisons with an absolute number like 1.0 will depend on the technique used. We therefore concluded that comparisons among slopes using the same statistical technique are more meaningful, and that is why we have emphasized such comparisons in both this reply and in our original paper.

It is worth noting that further calculations give reason to suppose that our decision to lump data from each species together in our original paper to avoid the possible pseudoreplication problem that was mentioned by Green (different measurements of an animal's genitalia may not be independent) was overly conservative. This is because in many species different measures of the genitalia show surprisingly weak correlations with each other, especially in comparison with nongenitalic structures. Of 84 Pearson correlation coefficients between different pairs of genitalic measures in the same species, 52.4% were not even statistically significant. These correlations were much weaker than those in pairs of nongenitalic structures in the same animals (only 11.1% of 117 were not significant; $\chi^2 = 33.6$, df = 1, P < 0.0001). In all 20 species, the median r value between genitalic traits was less than the median between nongenitalic traits. The biological meaning of this lack of strong correlations within genitalia is not clear; but in terms of the methodology of our

study, it means that pseudoreplication is less of a problem than both we and Green had anticipated.

Why Are Genitalia Different?

Now we turn to the question of why it should be that in all of the analyses just discussed genitalic structures tend to have lower slopes than nongenitalic structures when they are regressed on indicators of body size. Green evidently was not entirely convinced by his own claim that our evidence on this point was "not compelling", for he proposed three possible explanations for differences between genitalia and nongenitalia as alternatives to our "one size fits all" idea. We find these alternative hypotheses difficult to support for the following reasons.

Explanation 1. There was a systematic trend in our data toward higher measurement error in genitalia because genitalic structures were generally smaller. This objection has two problems. We found that the errors in measurements made under a microscope, as nearly all of ours were, depended more on problems regarding the presence of sharp, well-defined borders at sites where there were clear reference points ("measurability"), and on the difficulty in placing all specimens in the same orientation prior to measuring. As we noted in our paper, we carefully chose traits for which these problems were minimal. Problems regarding size could be (and were) often easily resolved by simply using a higher magnification. Secondly, we specifically tested for the effects of size on slopes (see entire last paragraph on p. 418 of our paper), because we had the same idea as Green that perhaps there was a size effect. We thought that perhaps increased measurement error might have occurred (though we suspected not), and also that animals may be less able to achieve precise developmental regulation of the, size of particularly small structures. Some of the genitalic structures we measured were relatively large, while some of the nongenitalic structures were relatively small. As we discussed in our paper, large genitalic structures (and small nongenitalic structures) did not have different slopes than did small genitalic structures (and large nongenitalic structures; the reader is invited to check the values, which we marked with "#" in our Appendix 2, for especially large genitalic traits and especially small nongenitalic traits).

Explanation 2. Green says that "little is known about the biology of most species. . ." in our study and suggests that, because of our ignorance of their sexual behavior, we inadvertently chose nongenitalic male characters that are under sexual selection and thus have especially high slopes. In fact, there is a large body of work on these species (much of which was cited in our paper) that describes which portions of the male's body are brought into play during sexual interactions and which thus enabled us to be careful in this regard. We and our coauthors have published descriptions (some in quite painful detail) of the sexual behavior of 11 of the 20 species: Huber and Eberhard (1997) on *Physocyclus globosus*; Huber (1997) on *Metagonia rica*; Eberhard (1993a) on *Macrodactylus sylphis*; Eberhard (1993b) on *Phyllophaga obsoleta*; Rodriguez (1994) on *Chelymorpha alternans*; Eberhard and Kariko (1996) on *Macrohaltica jamaicensis*; Rodriguez (1998a) on *Pseudoxychila tarsalis*; Eberhard and Pereira

(1996) and Eberhard (unpubl. ms.) on *Archisepsis diversiformis*; Briceño et al. (1996) and Briceño and Eberhard (1998) on *Ceratitis capitata* (the extensive literature on this species is reviewed by Eberhard, 1999); Rodriguez (1998b and 1999) on *Ozophora baranowskii*; and Briceño and Eberhard (1997) on *Paralabella dorsalis*. We also have made unpublished observations on a twelfth (*Ceratoma trifurca*) and have studied the sexual behavior of a congener of a thirteenth (Eberhard 1985 on *Hetaerina macropus*). Other workers have studied the sexual behavior of another species in our study (Robinson and Robinson 1980 on *Argiope trifasciata*) and still others have studied the sexual behavior of congeners of four additional species in our study (Robinson and Robinson 1980 on *Araneus*; Gerhardt 1921, 1923, and Bristowe 1929 on *Tetragnatha*; Alcock 1982 on *Hetaerina*, and C. Cordero [unpubl.] and W. Eberhard [unpubl.] on *Philoponella*). That leaves only three species for which direct observations of sexual behavior were not available to us. We were quite aware of the possible problem of sexual selection on nongenitalic body parts (among other reasons, because of Green 1992) and were careful to try to avoid including them. On p. 417 we stated that we specifically excluded those nongenitalic portions of the body that are specialized for contact with females in sexual contexts (for indications that such structures are often under sexual selection, see Eberhard 1985). Our only conscious inclusion of a sexually selected nongenitalic structure was that of the horn of the beetle *Onthophagus incensus* (for observations of its use as a weapon in male-male battles in a congeneric species, see Emlen 1995). We included this species and its horn because of interest in determining whether a species with dimorphic male weapons (see Eberhard and Gutierrez 1991, as well as footnote b in our Appendix 2) might have dimorphic genitalia (it did not). Deletion of the *O. incensus* horn data does not change the pattern for this species (all OLS and RMA genitalic slopes are lower than all nongenitalic slopes). Other than the horns, the behavioral observations available indicate that none of the other nongenitalic characters are specialized for use as weapons or as courtship devices.

Green's suggestion that we inadvertently included sexually selected nongenitalic body parts can also be tested independently by checking whether the traits we measured show sexual dimorphisms in slope. If Green were correct, the slopes for these characters would be consistently steeper in males than in females. There are only data to do this with seven traits in five species (three spiders, one beetle, and one fly). The pattern predicted by Green does not occur: in only two of the five species and three of the seven traits is the male RMA slope greater in a male versus female comparison: 1.22 versus 1.28 for femur I in *Philoponella*; 1.26 versus 0.99 for femur I in *Argiope*; 1.31 versus .1.45 and 0.76 versus 0.85 for tibia I and chelicerae in *Physocyclus*; 1.15 versus 1.04 and 1. 10 versus 0.91 for elytron length and total body length in *Chelymorpha*; 0.92 versus 1.16 for head width in *Ceratitis*. In short, Green's attempt to explain differences between genitalic and nongenitalic structures on the basis of our ignorance of the sexual biology of the species we studied is unfounded.

Explanation 3. Genitalic structures differ from nongenitalic structures due to a "closer ontogenetic or functional relationship" between nongenitalic structures and body size in-

dicators. This explanation resembles ours in implying that nongenitalic and genitalic structures evolve under different selective regimes. It may also include a mechanism (as opposed to a selective explanation) for the large pattern in our data. Natural selection might favor separation of the developmental pathways of genitalic traits from those of nongenitalic traits to enable males of varying body sizes to produce genitalic structures of relatively uniform size due to the "one size fits all" advantage we proposed in our paper. Green's proposal is so vague, however, that it seems difficult to test with empirical data. Other than using differences in allometric relations, which would of course not allow it to be distinguished from our "one size fits all" idea, how is one to establish these proposed relationships? We know of no data supporting any differences in ontogenetic or functional relationships in any of the species we studied, other than the fact that genitalia are specialized for use in sexual encounters and the other measured traits are not.

In summary, the RMA regression analyses which motivated Green to doubt our conclusions, may not be appropriate, and in any case are seriously flawed because many are based on statistically nonsignificant values. Even if Green's reanalyses are accepted, or if those slopes based on nonsignificant values are discarded or are set to zero, there are still significant differences between genitalic and nongenitalic allometry. The available data on our study species indicate that two of the three explanations Green offers as alternatives to our "one size fits all" idea to explain these differences are probably wrong. The third is related to our ideas, but lacks any direct supporting evidence.

Literature Cited

Alcock, J. 1982. Post-copulatory mate guarding by males of the damselfly *Hetaerina vulnerata* Selys (Odonata: Calopterygidae). Anim. Behav. 30:99-107.

Briceño, R. D., and W. G. Eberhard. 1995. The functional significance of male cercus morphology in 13 species of tropical earwigs (Dermaptera: Forficulidae, Labiidae, Carcinophoridae., Pygidicranidae). Smithson. Contrib. Zool. 555:1-63.

———. 1998. Medfly courtship duration: a sexually selected reaction norm changed by crowding. Ethol. Ecol. Evol. 10:369-382.

Briceño, R. D., D. Ramos, and W. G. Eberhard. 1996. Courtship behavior of male medflies (*Ceratitis capitata*; Diptera: Tephritidae) in captivity. Fla. Entomol. 79:1-15.

Bristowe, W. S. 1929. The mating habits of spiders, with special reference to the problems surrounding sexual dimorphism. Proc. Zool. Soc. Lond. 21:309-358.

Eberhard, W. G. 1985a. The behavioral ecology of the damselfly *Hetaerina macropus* (Calopterygidae). Odonatologica 15:51-60.

———. 1985b. Sexual selection and animal genitalia. Harvard Univ. Press, Cambridge, MA.

———. 1993a. Copulatory courtship and genital mechanics of three species of *Macrodactylus* (Coleoptera, Scarabaeidae, Melolonthinae). Ethol. Ecol. Evol. 5:19-63.

———. 1993b. Copulatory courtship and the morphology of genitalic coupling in seven *Phyllophaga* species (Coleoptera, Melolonthidae). J. Nat. Hist. 27:683-717.

———. 1999. Sexual behavior and sexual selection in the medfly, *Ceratitis capitata*. In M. Aluja and A. Norrbom, eds. Fruitflies (Tephritidae): phylogeny and evolution of behavior. CRC Press, Boca Raton, FL. In press.

Eberhard, W. G., and E. Gutierrez. 1991. Male dimorphisms in beetles and earwigs and the question of developmental constraints. Evolution 45:18-28.

Eberhard, W. G.. and S. Kariko. 1996. Copulation behavior inside and outside the beetle *Macrohaltica jamaicensis* (Coleoptera, Chrysomelidae). J. Ethol. 14:59-72.

Eberhard, W. G., and F. Pereira.. 1996. Functional morphology of male genitalic surstyli in the dungflies *Archisepsis diversiformis* and *A. ecalcarata* (Diptera: Sepsidae). J. Kans. Entomol. Soc. 69:43-60.

Eberhard, W. G., R. Achoy, J. Ugalde, and M. C. Marin. 1993. Natural history and behavior of two species of *Macrohaltica* beetles (Coleoptera: Chrysomelidae). Psyche 100:93-119.

Eberhard, W. G., B. A. Huber, R. L. Rodriguez S., 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.

Emlen, D. 1994. Evolution of male horn length dimorphism in the dung beetle *Onthophagus acuminatus* (Coleoptera: Scarabeidae). PhD. diss. Princeton University, Princeton, NJ.

Gerhardt, U. 1921. Vergleichende Studien über die Morphologie des männlichen Tasters und die Biologie der Kopulation der Spinnen. Arch. Naturgesc. 87(A,4):78-247.

———. 1923. Weitere Studien über die Biologie der Spinnen. Arch. Naturgesch. 90(A,5):85-192.

Green, A. J. 1992. Positive allometry is likely with mate choice, competitive display and other functions. Anim. Behav. 43:170-172.

———. 1999. Allometry of genitalia in insects and spiders: one size does not fit all. Evolution 53:1621-1624.

Huber, B. A. 1997. On American "*Micromerys*" and *Metagonia* (Pholcidae, Araneae), with notes on natural history and genital mechanics. Zool. Scripta 25:431-463.

Huber, B. A., and W. G. Eberhard. 1997. Courtship, copulation, and genital mechanics in *Physocyclus globosus* (Araneae, Pholcidae). Can. J. Zool. 74:905-918.

Robinson, M. H., and B. Robinson. 1980. Comparative studies of the courtship and mating behavior of tropical araneid spiders. Pac. Ins. Monogr. 36:1-218.

Rodriguez, V. 1994. Fuentes de variación en la precedencia de espermatozoides de *Chelymorpha alternans* Boheman 1854 (Coleoptera: Chrysomelidae: Cassidinae). MSc. thesis, Universidad de Costa Rica.

Rodriguez S., R. L. 1998a. Mating behavior of two *Pseudoxychila* beetles (Coleoptera: Cicindellidae). Can. Entomol. 130:735-750.

———. 1998b. Possible female choice during copulation in *Ozophora baranowskii* (Heteroptera: Lygaeidae): female behavior, multiple copulations, and sperm transfer. J. Ins. Behav. 11:725-741.

———. 1999. Mating behavior in two *Ozophora* bugs (Heteroptera: Lygaeidae). J. Kans. Entomol. Soc. *In press*.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry. The principles and practice of statistics in biological research. 3d ed. Freeman, New York.