Don’t Forget the Biology: A Reply to Green
Author(s): W. G. Eberhard, B. A. Huber and R. L. Rodriguez
Reviewed work(s):
Published by: Society for the Study of Evolution
Stable URL: http://www.jstor.org/stable/2640910
Accessed: 25/07/2012 16:25

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of
content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms
of scholarship. For more information about JSTOR, please contact support@jstor.org.

Society for the Study of Evolution is collaborating with JSTOR to digitize, preserve and extend access to
Evolution.
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 performed 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.
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 (\( = \frac{b}{r} \)). 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 imbalance in nonsignificant values between genitalia and nongenitalic structures. Nor could we bring ourselves to accept the alternative of using meaningless RMA numbers as if they were significant OLS slopes, so this would be throwing the baby out with the bathwater. 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, 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 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 mean 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 genital 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 nongenitalic 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 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 genital 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-
Acknowledgments

We thank J. Lobo and J. Kelly for advice on statistics, and F. Mora and M. J. West-Eberhard for comments on a previous version of the manuscript.

Literature Cited


