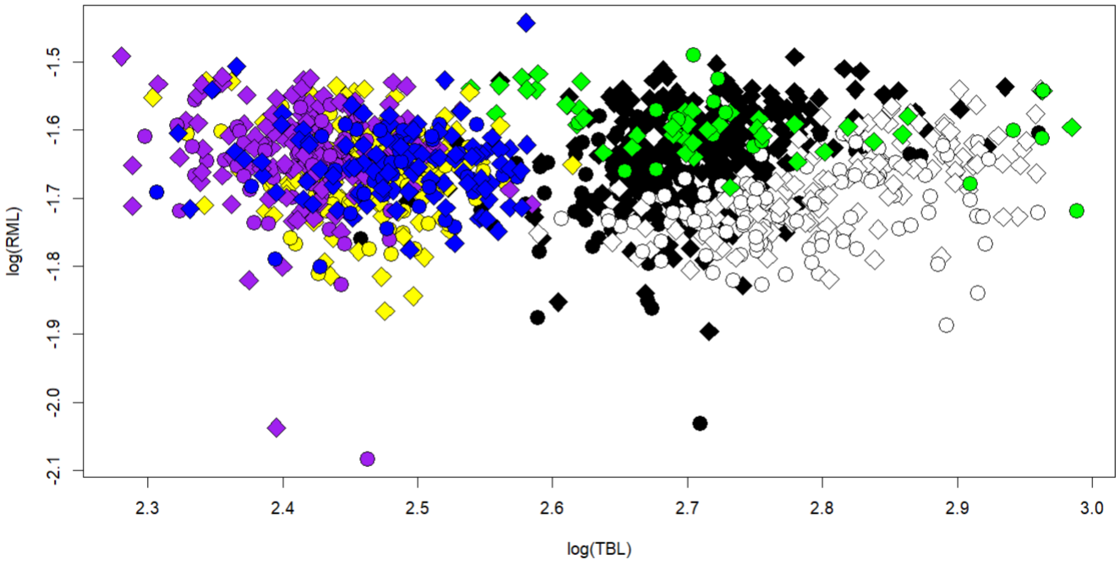
In this paper, the authors evaluated the sexual dimorphism in six semi-arboreal species of tiger beetles. Even though they used several morphological characteristics, it seems that the only ones that have most effect are body size (males are bigger than females) and elytra width (females have wider elytra than males). The authors then conclude that the main differences are related to fecundity selection (wider elytra = more room to make eggs) and male-male competition (larger body size = higher selective pressure on fighting).

While I enjoyed reading the paper because it is an interesting group and the data is amazing (I acknowledge the authors' hard work on this), I feel that the aim and how the data was analyzed lackluster. The authors start off by saying that sexual dimorphism is driven by sexual selection. However, that is not true as there are several cases of sexual dimorphism driven by differences in feeding niche. Thus, there is a whole section of ideas and putative mechanisms that are completely ignored – even though there is evidence that males have longer mandibles than female (L58-63), and that can suggest differences in feeding niche. After that, the authors add a whole lot of morphological measurements and discuss those that are significant by suggesting a function or something like that.

Another downside is using six species to do the exact same analysis. While I understand that is valuable information, it makes the Results section hard to read. Essentially, the Results section becomes "this species differed in X, while the other species only differed in Y". And the Discussion becomes a mix of ideas of why that pattern emerged in one species and another pattern on the other species. That is really hard to follow as a reader.

Another issue stemming from six species is that you will inherently have two types of patterns: one within species and another among species. Both are valid and valuable directions to talk about in a paper, but they are entirely different processes. The theory and mechanisms that work within species are different from the ones that work among species, and not having that clear in the text makes it extremely difficult to follow. To illustrate this example, I used the Supplementary data provided by the authors.

To do this, I got all the body size data (TBL, x-axis on the graph below) and plotted it against mandible length (RML, y-axis). Both axes are logged, but you remove the log the same pattern emerges. Circles are females and diamonds are males, and each color is a different species (doesn't matter which one because this is simply an example). Among species, you will have the investment patterns each species has (the slopes, which I did not calculate), but you will also have the overall morphospace occupied by the characteristic. If you look at the y-axis only, you will note that species don't differ as much in RML. But there is a clear gap around 2.6 TBL: there is a clear gap between species that is not occupied by any male or female of these groups. Why? This is the sort of pattern you get when you're looking among species: you have the areas of morphospace species occupy and areas where no species occupy. You could also use linear regressions to calculate a slope value within each species and compare them among species: which species has disproportionally larger mandibles? Does that correlate with levels of sexual dimorphism within species?



The notion behind using linear regressions for this type of allometric work is also why I don't think the authors should be using standardizations. I understand standardizing variables for the PCA, but you will also lose the information on the proportionalities that come with using linear regressions. Further, if you look at Figure 4, the WE boxplots, you will note that males are almost all of the same size, females are the ones varying the median. There is also a high number of outliers in that graph, which begs the question of why.

This has become longer that I anticipated, but my main message is: the authors need to think about their paper harder. My suggestion is to focus on either within species or among species patterns and focus on relevant morphological measures. Do not use all measures you have just because you measured them: think about what they mean for the animal and compare them accordingly. This is similar to predicting function from morphology, but much more robust (i.e., if X morphology does this, the pattern I expect is Y). By simply comparing everything you are bound to a find a significant difference.

Sorry for being overly negative, but I think the authors have an amazing dataset. By adding more about the biology of the group to the hypothesis and the statistics they are doing, this paper can shine.

I truly hope my revision helps.

Note on supplementary data: the data is accessible. The only thing I would add is a description of what the acronyms in the columns are and their units. I know they are in the paper, but it helps a lot to have them in the datasheet.