Is the body size-GRS relationship a statistical artifact?

Null models [1] in macroecology have not been widely considered (but see [2–4]). Here we raise the possibility that the classic macroecological relationship between body size and GRS is the result of a statistical artifact, as opposed to a functional relationship arising from the energetics of body size and its ecological consequences. First, we randomize the data to derive a null expectation. Then we show how the empirical relationship deviates systematically from the null.

1. A null-expectation: the effect of multiplying uni-modal skewed distributions

The triangular shape of the body size-GRS relationship has been recognized for 30+ years and generally has been interpreted as the result of functional constraints on GRS arising from metabolic energetics. To our knowledge, a consideration of the pattern expected by chance and whether the empirical pattern deviates from this expectation has not been put forth in the literature.

The distributions of mammal mass and GRS are both uni-modal and skewed, even on a logarithmic scale. Further, the skews are opposite, such that log_{10} mass is skewed right and log_{10} GRS is skewed left (figure S2a). What is the effect of multiplying two random, uni-modal, oppositely-skewed distributions? Figure S1 illustrates this effect for the extreme case of two mirror-opposite, highly skewed distributions. We generated random variables X and Y with log-normal distribution and length n = 3,628 (the same sample size as the empirical data set) and multiplied the distribution of Y by -1 (figure S1a). The correlation between X and Y produces a roughly triangular pattern, with a dearth of large values of X combined with small values of Y (figure S1b). Thus, the expected pattern when the two random, uni-modal, oppositely-skewed distributions X and Y are combined is a near right triangle with a systematically increasing lower bound. The pattern is similar to the observed macroecological relationship between body size and GRS in mammals (and birds), which has a systematically increasing lower bound that has traditionally been interpreted as functional.

To derive a specific null-expectation for the body size-GRS relationship in mammals, we resampled all the data without replacement for each variable. This produced two new random variables (mass_{rand} and GRS_{rand}), but with the same respective distributions as the original variables (figure S2a). Visually, the observed pattern (figure S2b) is similar to the random pattern (figure S2c). We offer several lines of statistical and other evidence that the empirical relationship is in fact not well-described by the null.

2. Does the empirical relationship deviate from the null?

(i) The null-model fails to predict the internal structure of the trait-space. Figure S3a shows GRS_{rand} versus mass_{rand} for a single permutation, with regressions on either side of the modal...
mass $\log_{10}(1.6)$. This figure is analogous to that shown for the empirical data in figure 2c of the manuscript. Compared with the empirical relationship, which is significantly negative for small species ($< \log_{10}(1.6)$) and positive for large species ($> \log_{10}(1.6)$), the regressions on either side of the mode in the permuted data are flat (not significant). The results of 1000 permutations (figure S3b) show that, compared to random, the probability of the observed slopes and regression coefficients for both small and large species are extremely low (all p’s < 0.0001). In essentially no permutations did the random null-model predict a significant relationship on either side of the mode, or that the relationship flips sign on either side of the mode. This result can also be seen by comparing the shape of a local smoothing function passed through the empirical (figure S2b) versus randomized data (figure S2c). Empirically, we observe a dip in the smoothing function as we pass through the modal body size moving from smaller to larger body size. No such dip is observed in the null-model.

(ii) The null-model over predicts the number of large-bodied species with small GRS. Although the null-model fails to predict the relationship on either side of the mode, it does predict the shape of the lower bounds on minimum GRS (compare figure S4a with figure 2d in the manuscript). Following identical procedures described in the manuscript (see section Material and methods), we binned the mass data and extracted the minimum observed GRS in each body size class for each permutation. For large species, the results of 1000 permutations (figure S4b) show that the empirical regression line has a significantly steeper slope ($p = 0.01$) and higher $R^2$ ($p = 0.023$) than expected by chance alone. This indicates that the null-model over predicts the number of large-bodied species with small GRS and under predicts the rate at which minimum GRS increases as a function of body size.

(iii) The empirical relationship is indistinguishable from the null with respect to the negative lower bound. For small species, the results of 1000 permutations (figure S4b) show that the empirical regression line has a slope ($p = 0.193$) and $R^2$ ($p = 0.148$) that are not significantly different from random. Thus, this aspect of the trait space can be said to be well-described by the null-model. Does this mean that the observed negative lower bound on minimum GRS is a statistical artifact? No, but the possibility cannot be ruled out [3]. However, we argue that the cumulative statistical evidence indicates that the body size–GRS trait-space is not simply the result of multiplying two random variables.

3. Additional lines of evidence that the body size–GRS relationship is real

Finally, we reiterate two important points made in the manuscript that amplify the idea that the body size–GRS relationship is functional. We also raise a third point that, while not addressed in this study, is relevant to this discussion.

(iv) The non-linear relationship between body size and GRS and the position of the breakpoint around the modal body size was predicted independently based on a model of energetics and optimal body size in mammals [5]. Our finding of evidence for this prediction lends more support to the idea that the observed relationship is functional.

(v) The non-linear relationship between body size and GRS and the position of the breakpoint around the modal body size also has been found for home range size [6,7]. This is despite the facts that there are many fewer data available for home range size than GRS and the relationships are otherwise very different (body size–home range size is log-linear, while body size–GRS is polygonal).
(vi) The same general pattern of a triangular trait-space and the same positive lower bound on minimum GRS as observed for body size also has been identified in the dimension of mass-independent basal metabolic rate (MIBMR). These results will be presented elsewhere [8], but are worth noting here. MIBMR is normally distributed (not skewed) and the expected random correlation with GRS does not include a positive lower bound. Yet, we observe this lower bound [8], which indicates that, above and beyond the energetic demand that accrues with increases in body size, species with supra-allometric BMR are restricted to large GRS.

References

Figure S1. The effect of multiplying two random, uni-modal, oppositely-skewed distributions. (a) Two random log-normally distributed variables, $X$ and $Y$, with opposite skew. (b) The correlation between $X$ and $Y$. 
Figure S2. Comparison of the empirical body size–GRS relationship to that expected from random. (a) Empirical distributions of mass and GRS. Green lines are density functions. (b) Empirical relationship between mass and GRS. (c) Randomized relationship between mass and GRS for three permutations. Red lines in (b,c) are local smoothing functions.
Figure S3. Randomized relationship between GRS and mass on either side of the modal mass. (a) Results from a single permutation. Red lines are lines-of-best-fit from ordinary least squares regression. (b) Results of 1000 permutations and associated regression parameters. Red lines indicate the observed empirical value. P-values are the number of times the observed value was exceeded by a randomly generated value, divided by 1000.
**Figure S4.** Randomized relationship between minimum GRS and mass on either side of the modal mass. (a) Results from a single permutation. Lines are lines-of-best-fit with 95% confidence intervals about the fit. Blue = “small” species; Red = “large” species. (b) Results of 1000 permutations and associated regression parameters. Red lines indicate the observed empirical value. P-values are the number of times the observed value was exceeded by a randomly generated value, divided by 1000.