

How is fish body size really distributed?

Freddie J. Heather Shane A. Richards Asta Audzijonyte

Abstract goes here

Table of contents

1	Background	1
2	Results & discussion	2
3	Methods	2
3.1	Body size data sources	2
3.2	Data filtering	4
3.3	Statistical analysis	4
4	Conclusion	5
5	References	5

1 Background

Body size is a key trait in determining how organisms interact with their environment, this is particularly true for marine organisms, where body size is a strong predictor of trophic position¹, growth rate, and predation mortality². Giometto et al.³, using the spherical diameter of unicellular protists, identified that the mean body size of a species, was sufficient to describe the entire body size distribution. The authors conjectured that this may hold for multicellular organisms. Here, we test this conjecture using 15 million+ individuals from 1064 fish species.

Body size is often assumed to be lognormally distributed due to the mult

2 Results & discussion

We found that the variance parameter of the lognormal distribution was constant across the fitted mean parameter of the lognormal distribution (Figure 1 A), further we found a consistent relationship between the standard deviation and the mean of the normal distribution fitted to observed body size (Figure 1 B). This is to say the mean size of the species is sufficient to describe the overall body size distribution. For both data sources, continuous and binned, we found the patterns between the mean and variance parameters of the lognormal distribution held, with $\text{sdlog} = 0.39 \pm 0.007$ (Reef Life Survey, binned data) and 0.3 ± 0.03 (Cryptobenthic fish, continuous). For normal distribution, we found a consistent relationship‘

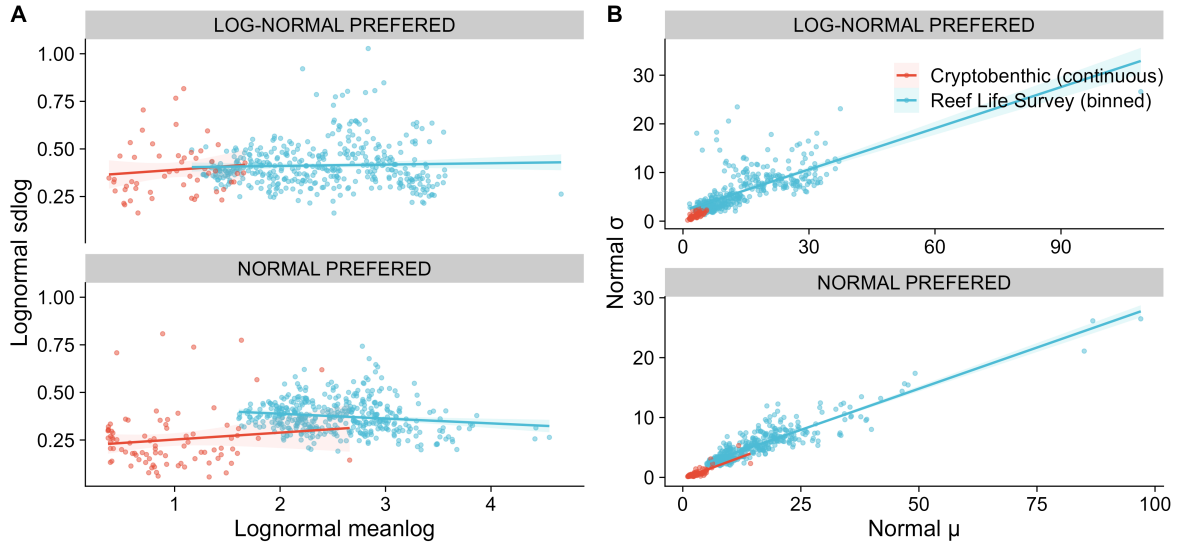


Figure 1: Mean size is sufficient to describe overall body size distribution.

3 Methods

3.1 Body size data sources

Fish body size data were obtained through two sources: Reef Life Survey (RLS; 15M+ individuals, binned body size)⁴, and Cryptobenthic reef fish (CBF; 8K+ individuals, continuous body size)⁵.

RLS surveys involve an underwater visual census method along 50m transect line, with a diver searching 5m either side of the transect line, the body size of the fish is estimated to the nearest body size bin (2.5, 5, 7.5, 10, 15, 20, 25, 30, ..., 400cm).

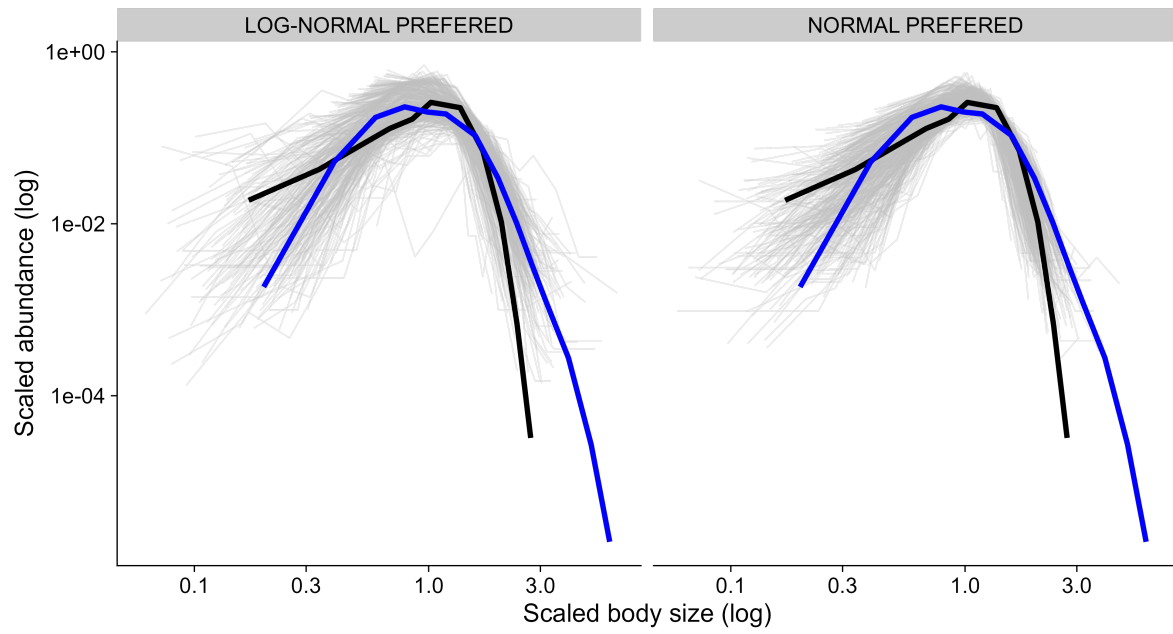


Figure 2: Some distributions are better described by Normal, some by lognormal. Blue is the lognormal fit, using the median lognormal parameters, and black is the normal fit with the median parameters.

The second data source was from field collections using enclosed clove-oil stations at six reefs; Moorea, GoO, AG, Lizard Island, Panama, and Belize. Reef outcrops are selected, measured, and covered with a bell-shaped fine mesh and tarpaulin, before being sprayed with a clove-oil:ethanol solution (1:5). Fish are collected with tweezers and placed in ziplock bags. See Brandl et al.⁵ for full methods.

3.2 Data filtering

For binned body size data, we excluded body size distributions that spanned fewer than four body size bins, we did not apply an exclusion for the continuous data based on the range of body sizes. For binned data we set a minimum count of 10 individuals per population to fit a distribution, for the binned data, this minimum count was 100. We performed a sensitivity analysis to these filtering parameters to show that across a range of filters for minimum count, the overall result did not substantially change.

Data were grouped at varying spatial scales to determine the appropriate spatial scale at which these distributions best fit. For the cryptobenthic fish data, populations were defined at two spatial scales; 1) species-level, and 2) location*species level. At the species level, all individuals belonging to the same species, irrespective of location were grouped in the same distribution. For RLS data, we defined three spatial scales: 1) species-level, 2) marine ecoregion⁶ level, and 3) 1x1 degree lat-lon gridcell level.

3.3 Statistical analysis

All statistical analyses were performed using the statistical programming language R⁷, and in combination with the statistical modelling language Stan⁸ for Bayesian analyses.

Cryptobenthic fish body size data, on a continuous scale, were fitted to one of two probability distributions, lognormal and normal, firstly at the species level (one distribution per species), and then at the localised reef scale (one distribution per species per reef).

$$L_p \sim \mathcal{N}(\mu_p, \sigma_p)$$

$$L_p \sim \mathcal{LN}(\nu_p, \tau_p)$$

Where L_p is the observed body length of population, p . Population body length, L_p , is assumed to come from either a normal, \mathcal{N} , or lognormal distribution, \mathcal{LN} . We have used ν and τ as the location and scale parameters of the lognormal distribution to avoid confusion with μ and σ , of the normal distribution. Each population has its own location (μ_p and ν_p) and scale (σ_p and τ_p) parameter for each of the two distributions, normal and lognormal, respectively.

Reef Life Survey body size data, in body size bins, were also fit using Stan, but using a method to account for the body size binning approach. The probability of being within each RLS body size bin (e.g. 2.5cm) was calculated by integrating the probability density function of the distribution between the upper and lower bounds of the size bin (e.g. between 3.75 and 1.25cm for the 2.5cm size bin).

$$Pr(B_b, K_k) = \frac{\int_{B_{bl}}^{B_{bu}} \mathcal{N}_{cdf}(\mu_k, \sigma_k)}{C}$$

Where C is a normalising constant, defined by the probability of being less than 1.25cm, that accounts for the body sizes not recorded in the RLS method. \mathcal{N}_{cdf} refers to the cumulative density function of the normal distribution.

$$C = \int_0^{1.25} \mathcal{N}_{cdf}(\mu_k, \sigma_k)$$

This method was run for both the normal and lognormal distributions. To account for misclassification error in the binning of body size an additional parameter (ϵ_p) was added that adjusts the bin probability by a small amount, which also has the benefit of avoiding zero probability likelihood estimates.

The likelihood of the model was defined as the summation of the probability of the data given the model parameters. To compare the log-likelihood values between species, with differing number of body size data per species, we calculated the log-likelihood per datum. This allowed for comparison between species, and a calculation of a log-likelihood per datum ratio (LLDR) between normal and lognormal fits. The posterior distribution of the LLDR could then be used as a measure of the normal vs lognormal fit for each species.

4 Conclusion

5 References

1. Jennings, S., Pinnegar, J. K., Polunin, N. V. & Boon, T. W. [Weak cross-species relationships between body size and trophic level belie powerful size-based trophic structuring in fish communities](#). *Journal of Animal Ecology* 934–944 (2001).
2. Goatley, C. H. R. & Bellwood, D. R. [Body size and mortality rates in coral reef fishes: A three-phase relationship](#). *Proceedings of the Royal Society B: Biological Sciences* **283**, 20161858 (2016).
3. Giometto, A., Altermatt, F., Carrara, F., Maritan, A. & Rinaldo, A. [Scaling body size fluctuations](#). *Proceedings of the National Academy of Sciences* **110**, 4646–4650 (2013).

4. Edgar, G. J. & Stuart-Smith, R. D. [Systematic global assessment of reef fish communities by the Reef Life Survey program](#). *Scientific Data* **1**, 140007 (2014).
5. Brandl, S. J. *et al.* [Demographic dynamics of the smallest marine vertebrates fuel coral reef ecosystem functioning](#). *Science* **364**, 1189–1192 (2019).
6. Spalding, M. D. *et al.* [Marine Ecoregions of the World: A Bioregionalization of Coastal and Shelf Areas](#). *BioScience* **57**, 573–583 (2007).
7. R Core Team. [R: A language and environment for statistical computing](#). (R Foundation for Statistical Computing, 2022).
8. Stan Development Team. [Stan modeling language users guide and reference manual, v2.21.0](#). (2023).