

Fish intraspecies body size distributions.

Freddie J. Heather Shane A. Richards Simon Brandl
Asta Audzijonyte

Abstract goes here

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 under experimental lab conditions, 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, from reef ecosystems globally.

Estimations of unexploited biomass are necessary for the stock-assessments, yet these are often based on expert advice, which may lead to subjective or biased estimates of exploitation levels^{4,5}. Among fisheries scientists, there appears to be consensus on the body size distribution of an unexploited fish population, partly due to the majority of body size data coming from fishery-dependant sources with size selective biases (e.g., fishing net mesh size). Many modelling approaches assume the abundance of unexploited fish populations to monotonically decline with body size (e.g., exponential distribution)^{6,7}, whilst others

Here we test the likelihood of each of these distributions using two global scale body size distributions, spanning millimetre long cryptobenthic species, to metre long shark species. We use a combination of two data collection methods, an underwater survey method which involves body size estimation into pre-defined body size classes, and the second involving a destructive method allowing the continuous body size distribution to be calculated.

2 Results & discussion

Our observations from 607 unique fish species support the conjecture³ that a single body size metric (e.g. mean body size) is sufficient to approximate the body size distribution of

fish species. We show that for the coefficient of variation of both the normal and lognormal distributions, 95% of estimates fell within the range of 0.16 to 1.57, with a median coefficient of variation of 0.45 (Figure 1). Only one of the 139 cryptobenthic population body size distributions (continuous data) was better described by a normal distribution than a lognormal distribution (triangles in Figure 1 A), this population had also the smallest mean size of any population. On the other hand, 36% ($n = 1100$ populations) of the 3089 binned body size distributions from Reef Life Survey populations were better described by a lognormal distribution, over a normal distribution (64%, $n = 2989$ populations). The coefficient of variation of body size distributions that were better described by lognormal distributions were higher (median cov = 0.61) than the those better described by a normal distribution ((median cov = 0.32, Figure 1 B).

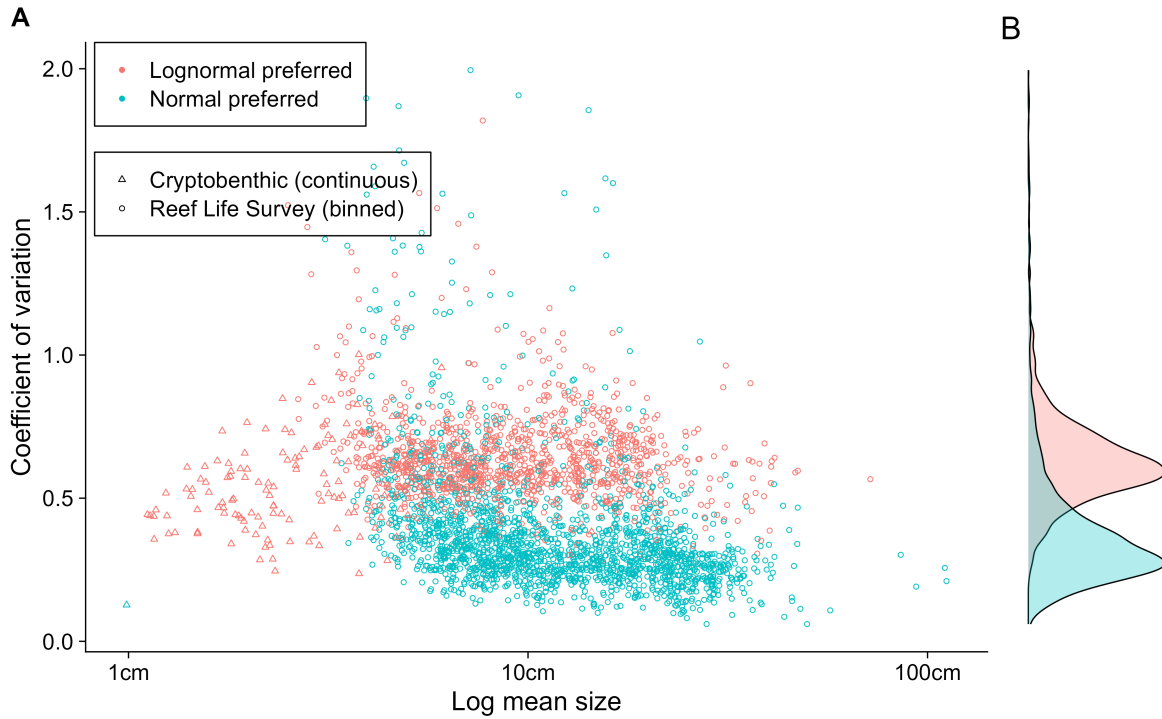


Figure 1: Mean size is sufficient to describe overall body size distribution.

Body size distributions once scaled by the total abundance of the population (i.e. abundance density, or probability-at-size),

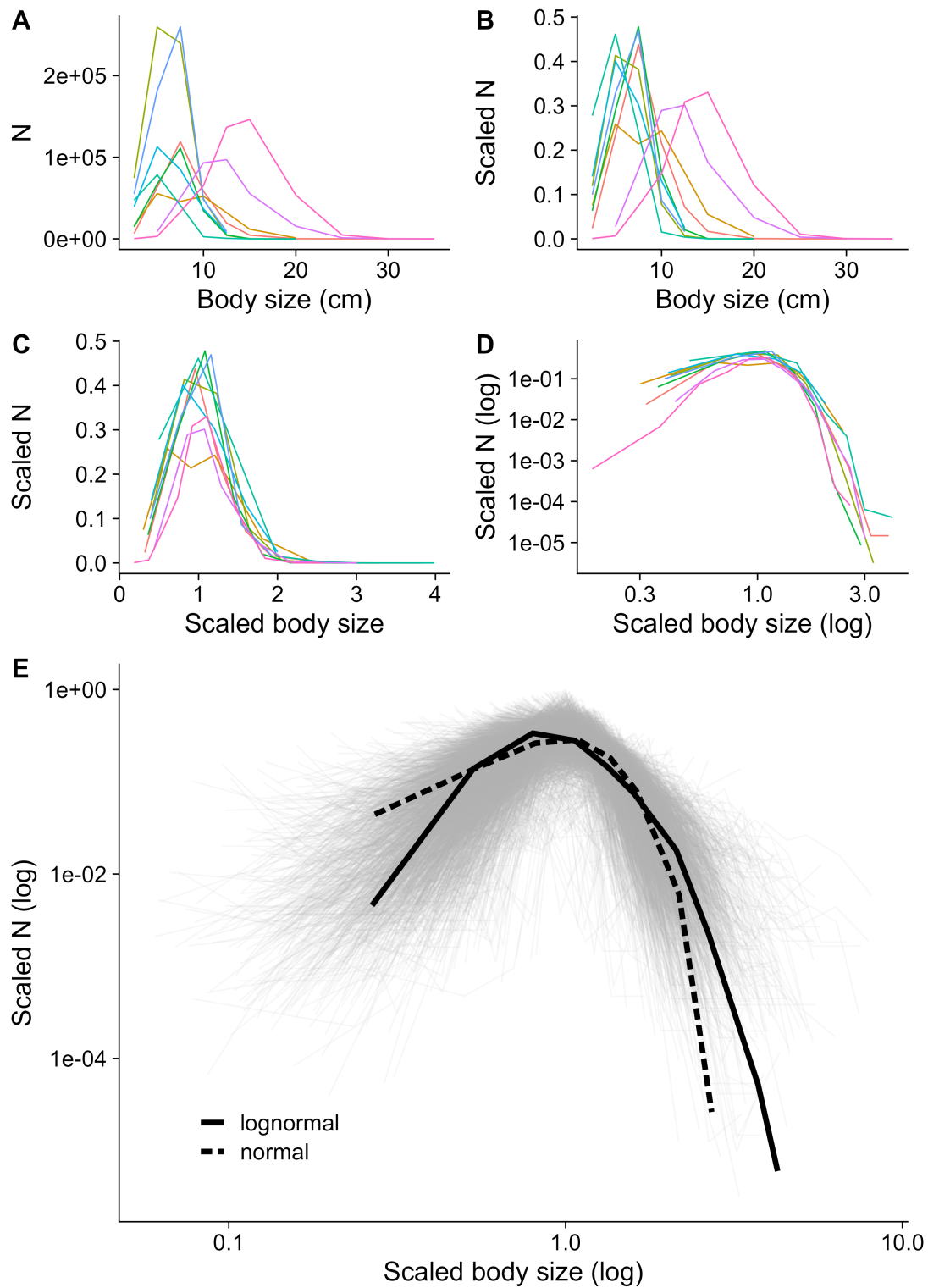


Figure 2: Observed body size distributions

3 Methods

3.1 Body size data sources

Fish body size data were obtained through two sources: 1) visual census data from Reef Life Survey (RLS; filtered to Australian populations only, over 12 million individuals, 497 species, binned body size estimates)⁸, and 2) lethal sampling of Cryptobenthic reef fishes (CBF; global spatial coverage, over 8000 individuals, 127 species, continuous body size measurements to the nearest 0.1mm)⁹.

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).

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.⁹ for full methods.

3.2 Data filtering

For binned body size data (visual census, Reef Life Survey), we excluded body size distributions that spanned fewer than four body size bins (e.g. 2.5cm, 5cm, and 7.5cm size bins only), we did not apply an exclusion for the continuous data (lethal sampling, cryptobenthic fishes) based on the range of body sizes. For the continuous 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.

A population was defined as a species within a location. For the Reef Life Survey dataset, to avoid biases associated with clumped data we determined a 'location' to be all surveys within a 1°x1° latitude-longitude gridcell ($n = 3089$ populations,). For the cryptobenthic fish data, a location was defined at the reef outcrop that was netted ($n = 139$ populations).

XX body size distributions were excluded from the final analysis due to model-fitting issues as a result of either a large proportion of individuals in a single body size bin ($n = XX$ populations) or due to bimodal body size distributions ($n = XX$ populations). See Supplementary Material XX for these distributions.

3.3 Statistical analysis

All statistical analyses were performed using the statistical programming language R¹⁰, and in combination with the bayesian statistical modelling language Stan¹¹.

A lognormal and normal distribution were fitted to each of the cryptobenthic population's continous body size distributions.

$$L_i \sim \mathcal{N}(\mu_i, \sigma_i)$$

$$L_i \sim \mathcal{LN}(\nu_i, \tau_i)$$

Where L_i is the observed body length of population, i . Population body length, L_i , is fitted to 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_i and ν_i) and scale (σ_i and τ_i) 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. 5cm) was calculated by integrating the probability density function of the distribution between the upper (b_u) and lower (b_l) bounds of the size bin, b (e.g. between 3.75cm and 6.25cm for the 5cm size bin). For example, the probability of an individual within the population i , being in size bin b , given body size is normally distributed is:

$$Pr(b, i) = \frac{\phi_i(b_u) - \phi_i(b_l)}{1 - \phi_i(1.25)}$$

Where ϕ_i refers to the cumulative density function of the normal (or lognormal) distribution, with population-level parameters. The demoninator is a normalising constant, defined by the probability of being less than 1.25cm; accounting for the body sizes not recorded in the RLS method.

This method was run for both the normal and lognormal distributions. Reasonable starting values were used in the MCMC process of fitting normal distributions to avoid numerical computational limits (i.e., floating point errors) of the integration of unreasonable distribution parameters.

The likelihood of the model was defined as the summation of the normalised probability of the abundance within each size bin given the model parameters. Since the lognormal and normal distributions have the same number of parameters, the log-likelihood of the fits of the distributions could be directly compared. For each population the preferred distribution (normal or lognormal) was determined as the distribution with the lowest median log-likelihood value of the posterior distribution.

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. Mangel, M. *et al.* [A perspective on steepness, reference points, and stock assessment.](#) *Canadian Journal of Fisheries and Aquatic Sciences* **70**, 930–940 (2013).
5. Froese, R. *et al.* [A new approach for estimating stock status from length frequency data.](#) *ICES Journal of Marine Science* **75**, 2004–2015 (2018).
6. Hordyk, A., Ono, K., Valencia, S., Loneragan, N. & Prince, J. [A novel length-based empirical estimation method of spawning potential ratio \(SPR\), and tests of its performance, for small-scale, data-poor fisheries.](#) *ICES Journal of Marine Science* **72**, 217–231 (2015).
7. Froese, R. *et al.* [A new approach for estimating stock status from length frequency data.](#) *ICES Journal of Marine Science* **75**, 2004–2015 (2018).
8. 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).
9. Brandl, S. J. *et al.* [Demographic dynamics of the smallest marine vertebrates fuel coral reef ecosystem functioning.](#) *Science* **364**, 1189–1192 (2019).
10. R Core Team. [R: A language and environment for statistical computing.](#) (R Foundation for Statistical Computing, 2022).
11. Stan Development Team. [Stan modeling language users guide and reference manual, v2.21.0.](#) (2023).