Consistency in body size frequency among hundreds of marine fishes with diverse life histories

Freddie J. Heather[1], Shane A. Richards[2], Nils C. Krueck[1], Rick D. Stuart-Smith[1], Simon J. Brandl[3], Jordan M. Casey[3], Graham J. Edgar[1], Neville Barrett[1], Valeriano Parravicini[4,5], Asta Audzijonyte[1,6]

1. Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Australia

2. School of Natural Sciences, University of Tasmania, Hobart, Australia

3. Department of Marine Science, The University of Texas at Austin, Marine Science Institute, Port Aransas, TX 78373, USA

4. PSL Université Paris: EPHE-UPVD-CNRS, Université de Perpignan, Perpignan, France

5. Laboratoire d'Excellence "CORAIL", Perpignan, France

6. Centre for Marine Socioecology, Tasmania, Australia

Abstract

Animal body size distributions are shaped by complex interactions among growth, mortality and recruitment. Theoretical expectations of fish body size distributions, derived from growth (k) and mortality (M) rates, are extensively used in fish stock assessments worldwide, yet rely on two life-history parameters (M and k) that are difficult to estimate. Using survey data for 822 species (3,228 populations) of shallow water marine fishes, we show that observed body length distributions can be reconstructed in the absence of these two parameters, and instead using only a single body length metric of a population (e.g., the mean or asymptotic body length). When scaled to the mean of the population, the frequency of body lengths from temporally pooled samples showed a consistent unimodal shape, which could be approximated by the truncated normal or lognormal distribution with a coefficient of variation value of around 0.3. This observation was evident for species with diverse life histories spanning two orders of magnitude in maximum body length. The consistency in population and species-level body length distributions suggests evolutionary convergence on a narrow range of viable outcomes even though multiple intrinsic and extrinsic factors are expected to influence body sizes. The reconstruction of fish body length distributions without knowledge of growth and mortality rates has important implications for the assessment and management of data-poor coastal fisheries worldwide.

Significance

Scientists have long debated whether size distributions of fish (and other animals with indeterminate growth) follow general rules or are highly variable. Using continental scale observations from hundreds of coastal fishes, we demonstrate remarkable consistency in the shape of length frequency distributions across populations and species with vastly different body sizes and life-histories. This similarity of species and population level body size distributions supports life-history predictions that variation in growth and mortality is constrained to optimize.

Main

The distribution of individual body sizes within animal populations depends on interacting rates of reproduction, recruitment, growth, mortality, and energy allocation[1,2]. Although the large number of physiological, demographic, and ecological processes influencing body size could potentially lead to high variation in the shape of body size distributions among populations and species, the viable set of growth and mortality characteristics in natural populations is limited[3,4]. For example, slow growth combined with high mortality may lead to extinction, whereas fast growth and low mortality may be impossible, given that fast resource acquisition demands elevated metabolism and increased predation risk[5–7]. Many life-history parameters, such as maturation or maximum size, growth, natural mortality and maximum age, show similar and predictable relationships across a wide range of animal taxa[6,8–12]. Thus, body size distributions, which emerge from these processes, may also show consistent shapes in natural populations. A strong consistency in body size distributions has been suggested for 13 species of unicellular organisms grown in experimental conditions. Despite an order of magnitude difference in the overall species-level mean cell diameter, scaled distributions of protists were remarkably similar and could be described by a single parameter, mean population body size[13]. Similar general patterns in intra-specific body size distributions might emerge across a range of other organisms. For example, in the case of fishes, predictions about body size distributions have been formalized from life-history theory rules, suggesting that for a given set of reproduction, mortality and growth parameters, emergent intra-specific size frequencies will look similar[2,14–16]. However, reproduction, mortality and growth parameters are unknown for most species and it remains debated whether general patterns in these parameters, and consequently in the shapes of body size frequency distributions, can be expected[17].

Understanding and predicting natural body size distributions of fishes is important from both theoretical and applied perspectives. Fishes are the most diverse group of vertebrates, encompassing species of different sizes and life-histories. Critically, they also represent the largest source of non-farmed animal protein, essential nutrition, and income for humans worldwide[18]. Fisheries are generally size-selective, often targeting the largest-bodied individuals within a species or population first. The relative abundance of individuals across the size spectrum of a species can be analyzed to infer fishery-induced population depletion, representing a commonly applied basis for stock assessments worldwide. Yet, estimating depletion requires knowledge about expected unfished size frequencies, which remain unknown for the vast majority of fish species. To assess whether population and species level size distributions of natural fish populations show predictable patterns, we used data collected from shallow water reef ecosystems through two divergent methods: underwater visual census observations and exhaustive sampling of, generally smaller-bodied, cryptobenthic fishes. The two data sources comprise 3,228 populations of 822, mostly unfished, temperate and tropical reef fish species that span a broad range of ecological and life-history traits[19] as well as maximum body lengths ranging from 1.1 cm to 2.5 meters.

Results and discussion

Despite the multitude of processes expected to influence individuals of different body sizes within a population, the observed length distributions of all 3,228 populations were remarkably similar in shape (after scaling individual body lengths by the population mean length to allow direct comparison among species of varying sizes; Fig. 1). Across three spatial scales and assumed organizational levels (population-level, metapopulation-level, and whole species-level; see Methods), and both sampling methods, most observed length distributions (98%) were approximately unimodal or hump-shaped.

For these 98% of populations (n = 3,166) we fit two common statistical distributions, the normal (truncated at the lowest observable size classes) and lognormal (median distribution fits in Fig. 1A), to the length-frequency data. Bayesian methods were used to assess which of the two distributions better described the observations, and to estimate their mean and coefficient of variation (CV). Most populations (93% of the cryptobenthic fishes, n = 127 populations, and 89% of larger bodies species in visual census surveys, n = 2,571 populations) were better described by a truncated normal distribution (Extended Data Fig. 1). Yet, regardless of the best fitting distribution, the relative spread (i.e., CV) around the mean body length was similar across populations and species (Fig. 2). For larger fish, analyzed through visual census, 80% of CV estimates were between 0.22 and 0.52 for the truncated normal distribution and between 0.27 and 0.51 for the lognormal distribution (95% of estimates from 0.19 to 0.55 and 0.24 to 0.58, respectively) (Fig. 2). The median CV values for the normal and lognormal distributions were very similar at 0.34 (SE = 0.002) and 0.37 (SE = 0.006), respectively. For the smaller, cryptobenthic fish populations and species, variation in body size was slightly lower (e.g. see Figs. 1B and 2), with the median value of the normal distribution CV at 0.23 (only 9 populations were better described by the lognormal distribution). The difference in CVs between small cryptobenthic and larger visual census species was partially explained by different observation methods and the binning of visual census data (versus individual measurements to the nearest mm for the cryptobenthic species). When cryptobenthic fish data were placed into bins similar to the visual census data, their CVs increased from the median CV value of 0.23 to 0.28 (Extended Data Fig. 2).

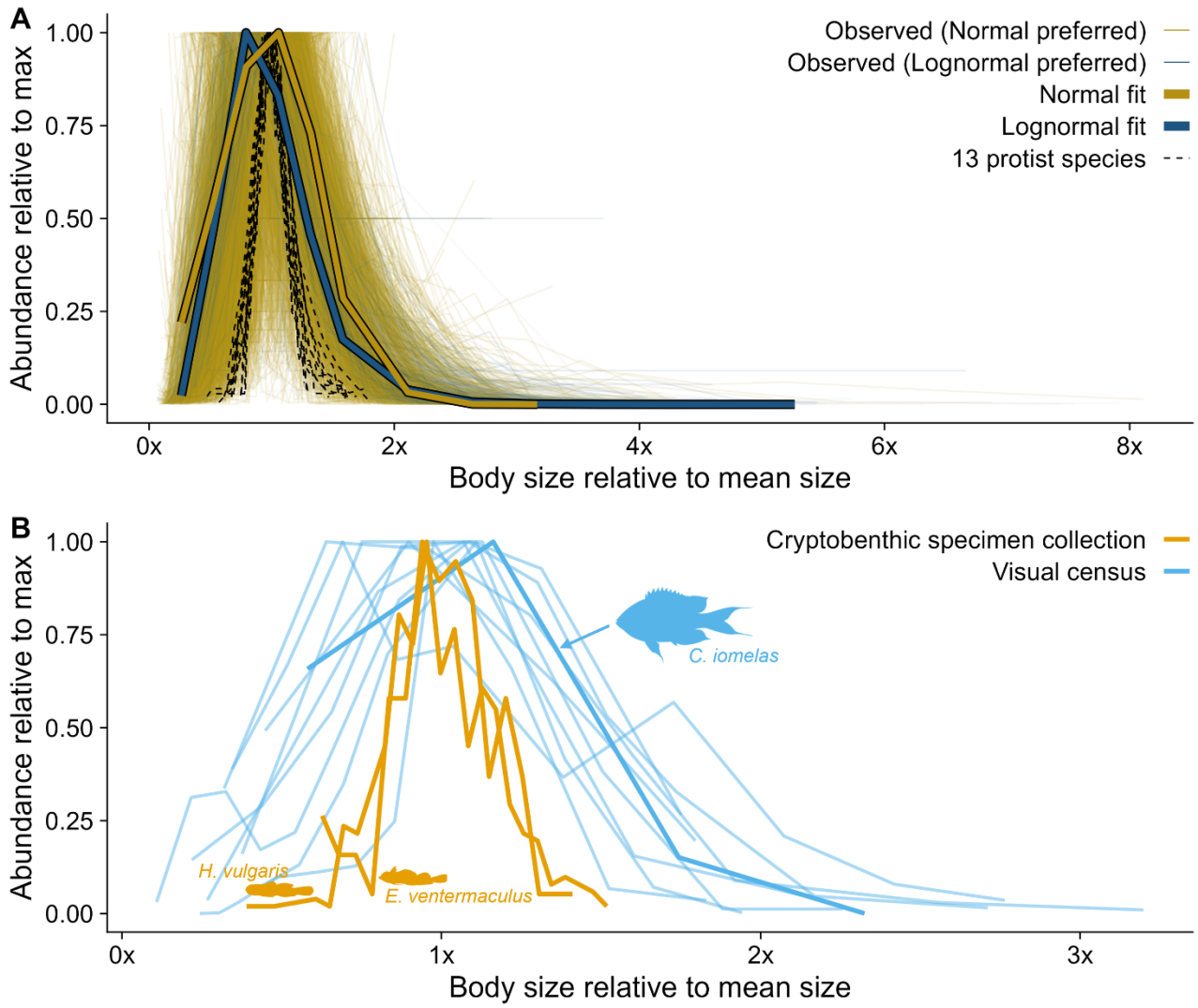


Figure 1: Observed body length distributions across 3,228 fish populations show a similar unimodal shape. *A) Observed individual body length distributions shown in thin solid lines with the median parameter estimates from normal-preferred and lognormal-preferred model fits shown as thick yellow and dark blue lines, respectively. Dashed lines in (A) show scaled body size (cell diameter) distributions of 13 freshwater protist species extracted from Giometto et al.* (2013)*. B) A random selection of 12 populations for visualisation purposes; three species have been highlighted (thicker lines) to indicate that each line represents a population-level body length distribution. Binned body length data were normalized to the bin width to allow comparison with continuous body length data. Abundance density (y-axis) is scaled to the maximum abundance density of the population, and body size (x-axis) is scaled to the mean body size of the population.*

The estimated CV values for fish length were consistent across three different spatial scales: population-level (Fig. 2), metapopulation or ecoregion-level[20] (Extended Data Fig. 3), and the entire species-level (Fig. 2). The results were also similar when species that may be targeted by recreational or commercial fisheries (12% of total, n = 97 species) were removed from the analysis (Extended Data Fig. 4). Naturally, over short temporal and fine spatial scales, local recruitment and mortality events might increase the variability of size distributions and produce multimodal size frequencies, such as sometimes seen in fisheries samples, especially from highly seasonal species[21,22]. Multi-modality is likely to be rarer when body size data encompasses multiple sampling

events and when seasonality is weaker, as is the case for our data. Only about 2% of coral reef and temperate rocky reef populations in our dataset revealed multiple peaks in body length frequencies (Extended Data Figs. 5).

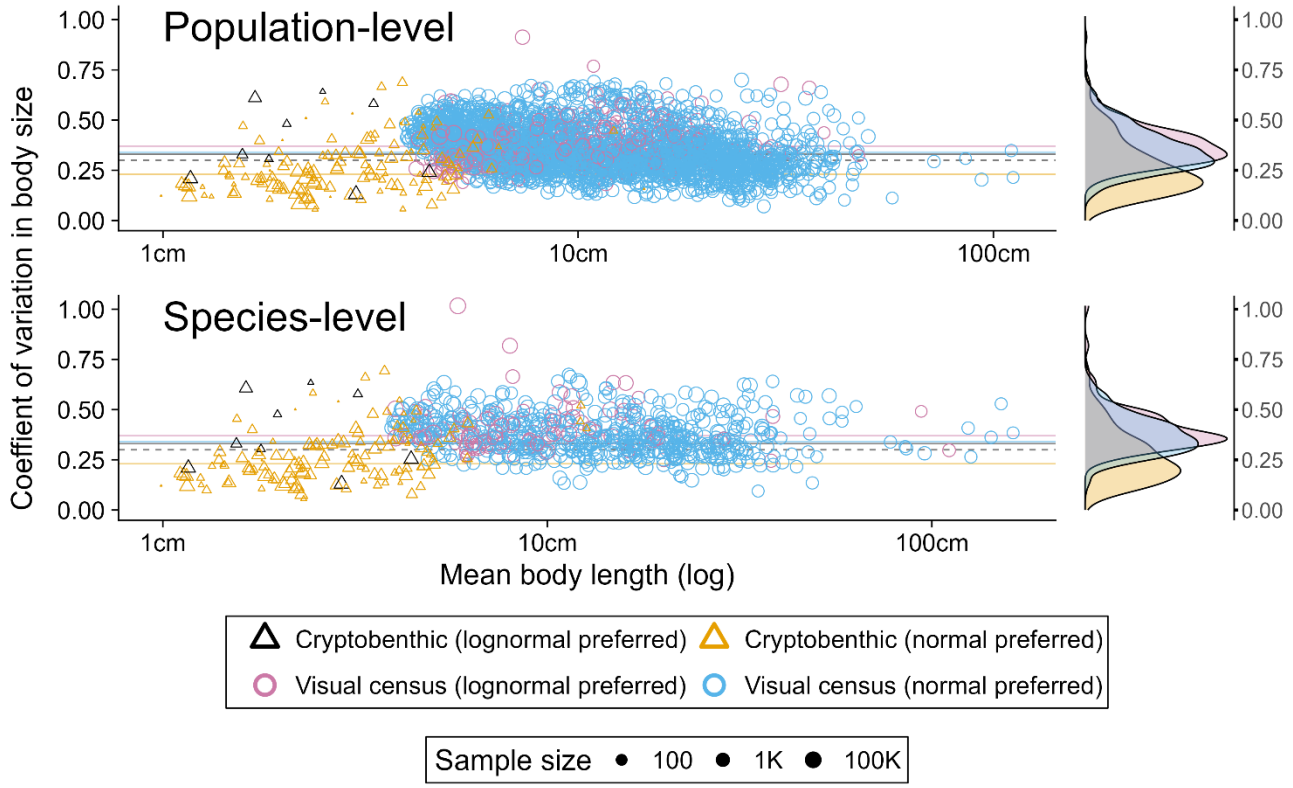


Figure 2: Coefficient of variation (CV) of body length distributions of ca 3000 fish populations of larger bodied species (visual census and largest cryptobenthic fish) is approximately 0.34, regardless of the mean body length of the population and the assumed statistical distribution. *Each symbol shows the CV value of a single fish population (grid cell, see Methods) or species. The color and shape of the symbol indicates the better-fitting statistical distribution and the sampling method. CV (y-axis) is calculated as the median of the posterior parameters associated with the better-fit distribution. Colored horizonal lines represent the median CV values, dashed line shows the estimated CV value of protists from Giometto et al.[13]. Margin density plots show the distribution of CV values across populations; for cryptobenthic lognormal preferred (black triangles), marginal plots are not shown due to small sample sizes (n = 9 populations). Symbol size is relative to the sample size (log scale).*

Across nearly two orders of magnitude in mean body length, the relative variation around the mean in the size distribution of fish species was largely independent of the mean body size (Fig. 2). The smallest cryptobenthic species revealed slightly lower variability in body sizes, while populations of the larger cryptobenthic species (mean body length > 5 cm) were mostly consistent with the level of variability observed in the visual census dataset (even before accounting for the binning). Tiny cryptobenthic species, such as *Eviota* spp.[23] (Extended Data Fig. 6), represent an extreme range of fish life-histories, with high mortality rates, rapid growth throughout their lives, and unique demographic dynamics[24]. Physiological and ecological constraints may eliminate larger individuals of cryptobenthic species and reduce body length variation within this group.

Interestingly, body size variability in 13 experimentally grown protists species[13] was also similar to those of the fish studied here, with the CV of cell diameter at ca 0.3 (dashed lines in Figs. 1A and 2). These findings suggest that body size distributions of many other organisms might be well-approximated with normal or lognormal distributions and a relatively narrow range of associated CV values. This universality has important theoretical and ecological implications and should be further investigated with a broader dataset of natural population body size data across taxa, sampling methods, and habitats.

Assuming a CV value other than 0.34 produced a worse fitting distribution. We tested this claim using calculations of dissimilarity (Kolmogrov-Smirnov test) by comparing the predictions of the simple single-parameter model using a range of CV values with the observed body length distributions. Assuming other CV values quickly reduced the predictive power (Fig. 3D). Importantly, when a CV of 0.34 was used, predictive power was similar regardless of which underlying statistical distribution (truncated normal or lognormal) was used to approximate the observed size distribution (Fig. 3). This is a useful property because the best suited statistical distribution would generally be unknown.

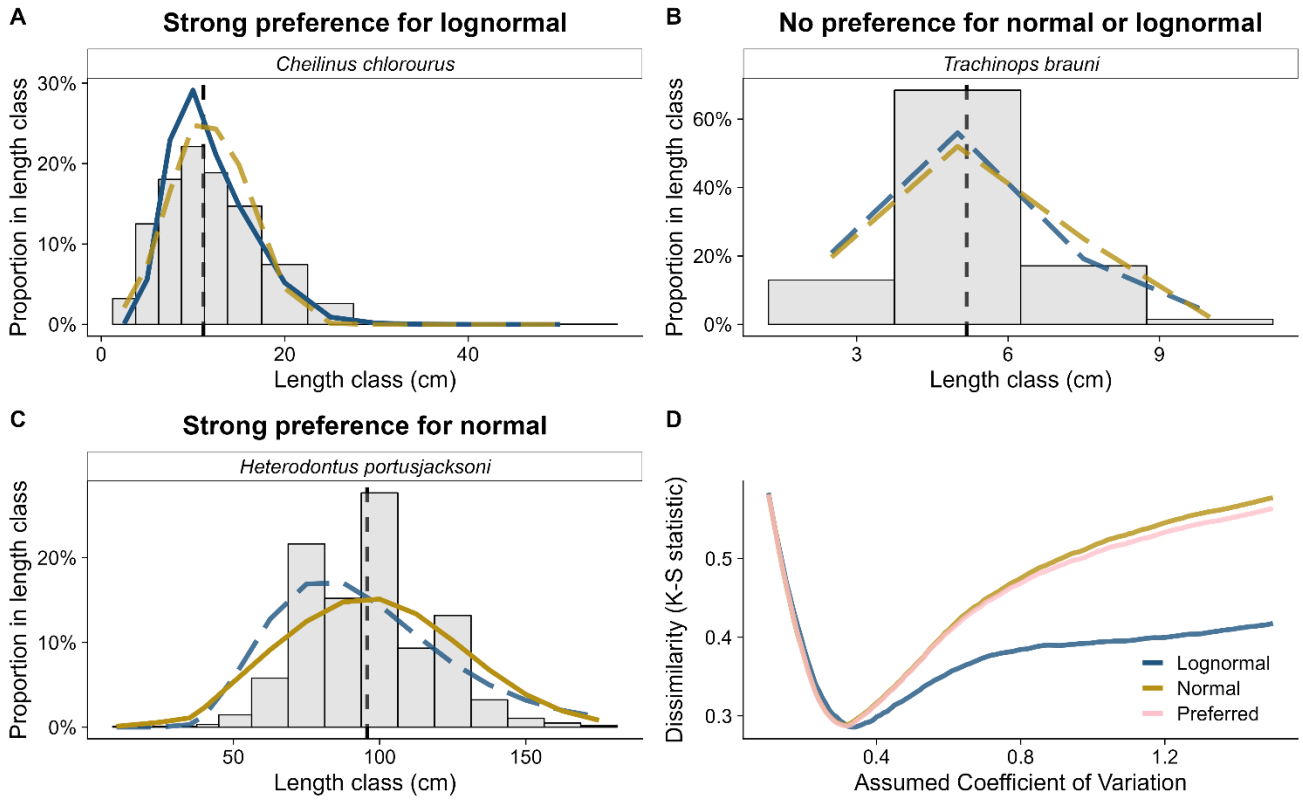


Figure 3: Empirical body size distributions (grey bars) can be reconstructed from the mean body size of the population (vertical black dashed line) and assumed CV of 0.34. *Panels A-C show randomly selected populations from one of three population groups for which body size distributions were best described by lognormal, normal or either of the two statistical distributions. Grey bars show observed data and lines show predicted distributions assuming either truncated normal or lognormal statistical distribution and CV of 0.34. Panel D shows that assuming CV of 0.34 provides the best fitting distributions with lowest median Kolmogorov-Smirnov (K-S) statistic across all populations regardless of the assumed statistical distribution (lognormal for all 3000 populations, truncated normal for all, or estimated using the best fitting distribution). Lower or higher CV values quickly reduced the predictive power, shown by higher K-S values.*

Fish population dynamics theory and expectations about body size distributions under different growth and mortality rates were originally formulated by Beverton and Holt[14] in the mid-20th century and later operationalized in size-based stock status assessment models by Hordyk *et al.*[2,15] and Froese *et al.*[16]. According to classic theory, the shape of a scaled body size distribution is flexible and determined by the relationship between mortality (M) and growth (k)[15,16]. Empirically estimated M/k ratios across species and phylogenetic groups vary between at least 0.5 and 3.0[2,12,17,25], although estimates of both M and k are notoriously hard to obtain. In fact, both life-history parameters are available for only around 17% of 7000 exploited species of fishes globally[26], and even for these species parameter values remain uncertain[17]. This long-standing challenge undermines the robustness of stock assessments, including applications of size-based approaches for data-poor conditions, hindering the effective management of world fisheries.

The similarity of shapes in body size distributions observed here (i.e., a narrow range of CV values) suggests that M/k ratios in natural populations might be constrained. In fact, such constraints have been proposed earlier, including estimates of an evolutionary optimal M/k ratio of 1.5, which is often used as a default assumption for stock assessment applications investigating poorly researched fish species[11,26,27]. However, whether an M/k ratio of 1.5 is optimal or broadly representative is under debate[28,29], and empirical tests are limited because traditional fish body size datasets have been confined to economically important fishery species, for which unfished body size distributions are generally unknown. With more widespread application of underwater videos and diver-based survey methods, the availability of datasets from all fish species in a community opens new possibilities to test and expand fish population dynamics theory. Overall, our findings suggest that across a broad range of environmental conditions, from tropical to temperate reefs, environmental and demographic parameters shaping growth and mortality are likely constrained, resulting in similar relative abundances of small, medium, and large individuals in unharvested populations. Yet the direct comparison between consistent CV values reported here and estimated M/k ratios in previous work is not straightforward. For example, even species with very low estimated M/k ratios (e.g. long-lived and relatively early maturing sea sweep, *Scorpis aequipinus*, or banded morwong, *Chirodactylus spectabilis*[25,30]) had CV values close to the median value in our dataset (Extended Data Fig. 6).

More importantly, the relatively consistent CV values observed in our study across multiple populations and species, and across a range of body sizes, suggests that even without prior information on life-history parameters, a single parameter – mean body length (or its approximation from the asymptotic length, see Methods and Extended Data Fig. 7) – allows for a sufficiently accurate assessment of observed (i.e., post-recruitment) body length distributions for many fish populations and species. To assess the performance of our simple method against a more complex population model we compared visual survey observations (grey bars in Fig. 4) to expected distributions generated using fish population dynamics theory[2,14,15], as implemented in a widely used length-based population assessment model[15] (the length-based spawning potential ratio, LBSPR, estimation approach, red line in Fig. 4). The theoretical prediction of length-frequency distributions based on LBSPR requires knowledge of three critical life-history parameters - asymptotic body length (L_{inf}) and the ratio between mortality rate (M) and the von Bertalanffy growth coefficient (k). The method further assumes continuous recruitment and constant length variability around mean size at age (default CV value is 0.1). The alternative approach derived from findings in this study requires only a single parameter - asymptotic body length L_{inf} (same as for the LBSPR based predictions) - which is then used to estimate mean body length (Extended Data Fig. 7) and assumes a truncated normal size distributions with a CV of body length of 0.34.

The life-history parameters required for LBSPR analyses were available for 68 (out of our 822) species that are unlikely to be intensively fished, all from the visual census data. We found that for these 68 species the LBSPR model approximated the observed body size distributions reasonably well (grey bars and red line in Fig. 4 A-G),

largely returning unimodal, hump shaped body size distributions as observed. However, length distributions reconstructed from just the asymptotic body length and an assumed CV value of 0.34 (made on a continuous scale, then clustered into bins corresponding to empirical data) provided a similarly good, or even better, fit. The dissimilarity between predictions and observations (based on Kolmogrov-Smirnov test) was generally lower for the simple method (in all cases assumed asymptotic length was the same between the two methods). This was true regardless of whether comparisons were done on all fish size classes or whether individuals <10cm were excluded due to possible observation bias in the data (Fig. 4H). Using a generic M/k ratio of 1.5 (see Froese et al.[16]), rather than species-specific growth and mortality parameters[17], did not noticeably change the LBSPR model predictions (Fig. 4H).

As expected, the LBSPR model predicted slightly higher abundances for the smallest size classes (<10cm) than observed, which could either indicate incorrect modelling assumptions or, more likely, sampling selectivity in our data (Extended Data Fig. 8). Visual census is inevitably subject to observation bias because divers are unlikely to reliably detect the smallest individuals[31]. Sampling bias is likely to be considerably smaller in the exhaustively sampled cryptobenthic fishes, although this assumption could not be assessed here, since life-history parameters required for theoretical predictions remain unknown for most cryptobenthic species. Yet, regardless of the sampling method, the smallest body size range, represented by eggs and larvae, often occur in locations that are difficult to study (i.e., the open ocean), and thus cannot be accurately observed. As such, generalizations about body length distributions presented here refer to the observable, post-recruitment life stages of reef fish populations, which are also of the highest relevance for reef ecosystems and associated fisheries.

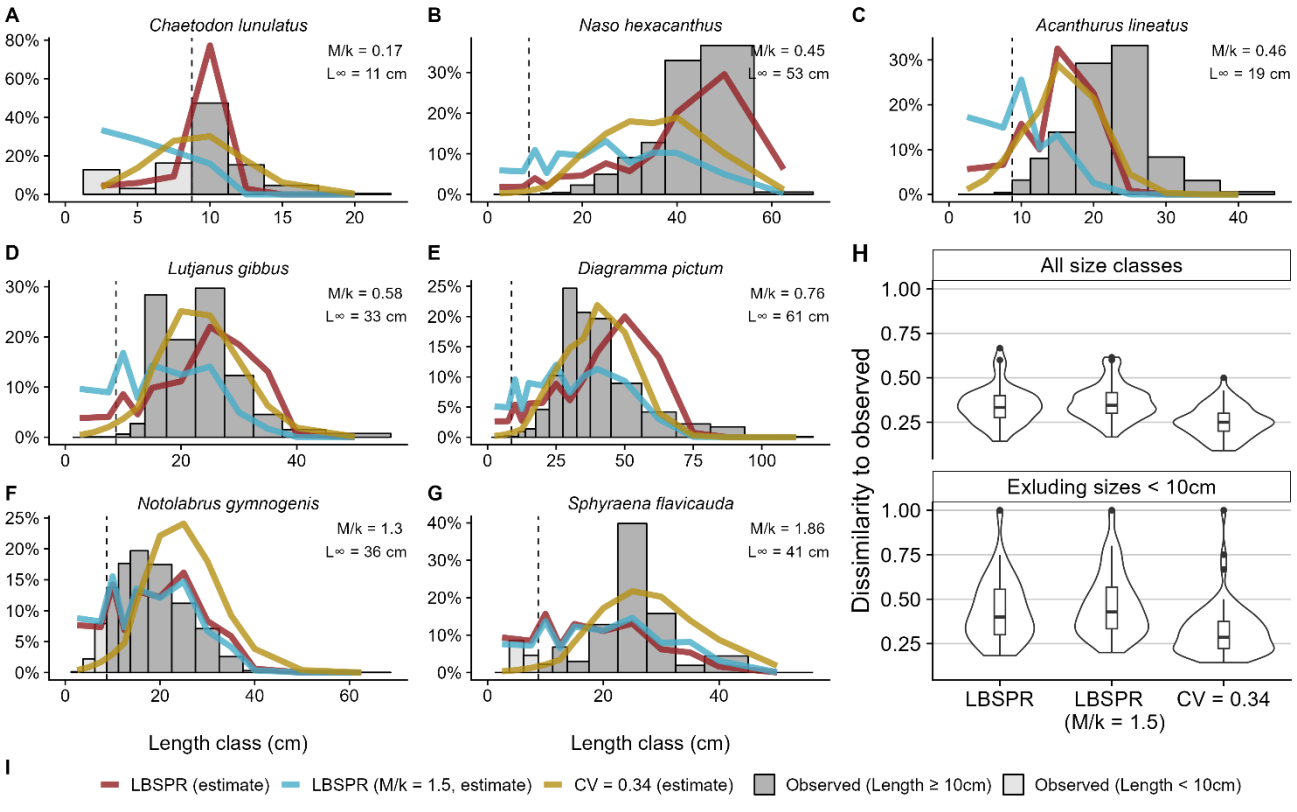


Figure 4: Simple one parameter model that assumes truncated normal distribution and CV = 0.34 matches observations similarly well or better than commonly used size-based methods with more parameters (LBSPR). *Panels A-G shows observations (grey bars) and predictions (solid lines) for seven randomly selected (from 68 species with available life-history parameters). Predictions are shown for a simple method using truncated normal distribution (yellow) with a CV = 0.34 and from LBSPR using either empirically estimate M and k values from Prince et al.*[17] *(red) or assuming a generic M/k ratio of 1.5*[16]*. Panel H shows the dissimilarity (Kolmogorov-Smirnov statistic) between the prediction and the observed body size distribution for all 68 species using all size classes or removing individuals that may have observational sampling bias (less than 10cm in body length).*

Findings from this study are likely to be of particular interest for practical applications in fisheries. Moreover, the similarity of scaled body length distributions contribute to the advancement of life-history theory and fitness optimization models[4,6] which predict a narrow range of optimal body size distributions. Indeed, our data suggest that in natural populations, over broad temporal and spatial scales, individual growth and life-histories are optimized to maximize fitness.

Our study calls for concerted efforts to bring together diverse fish body size datasets to test predictions across an even wider range of environmental conditions, as well as spatial and temporal scales. Further work is also needed to test the sensitivity of the approach to the misspecification of statistical distribution (e.g., lognormal or normal), sample sizes and conditions required to fit observations to the theoretical expectations with sufficient accuracy. From an ecological perspective, it is important to understand whether, and how, the observed variability in CV values (Fig. 2) and preferred distribution, especially smaller values in cryptobenthic fishes, relates to environmental or phylogenetic factors, and whether they remain consistent across temporal and spatial gradients within a species.

For data-poor and unassessed fishery species, which account for approximately 80% of all fishery species worldwide[32], size-based stock depletion assessments offer a key opportunity to improve our understanding of the health of fish populations and associated requirements for sustainable fisheries management[16,33,34]. This opportunity is greatly upscaled by the now widespread availability of increasingly automated collections of fish body size data. Findings from our study can support the use of these data by helping to quantitatively estimate baseline (unfished) size frequency distributions without the need to parameterize highly uncertain rates of growth and natural mortality[17,35–37]. Observed size frequency distributions can then be contrasted to these empirically-grounded expectations of unfished size frequency distributions to estimate population depletion. Helping to better understand the status and management needs of unassessed fishery stocks addresses a problem of global ecological and socio-economic significance, given that unassessed fish populations worldwide are more likely to be in poor condition[18,32,38–40].

Methods

Body size data sources

We analyzed two datasets. The first dataset was obtained from underwater visual surveys around the Australian continent (over 12 million individuals, 3,089 fish populations of 710 species) by divers in the Reef Life Survey program (RLS)[41,42] and the Australian Temperate Reef Collaboration (ATRC)[43]. Visual census surveys involved a standardized protocol of a diver searching 5m either side of a 50m transect line, recording individual fish lengths to the nearest size bin (2.5, 5, 7.5, 10, 15, 20, 25, 30, …, 400cm) and recording taxonomic ID. Full details of the standardized methods are available online[44].

The second dataset was obtained from local-scale community collections of cryptobenthic fishes from locations around the world (over 8000 individuals, 139 populations of 132 species)[45,24,46]. Extractive sampling of cryptobenthic fishes (CRF data) was performed in the field using enclosed clove-oil stations at six locations: Mo'orea (French Polynesia), Fujairah (Gulf of Oman), Abu Dhabi (Arabian/Persian Gulf), Lizard Island (Australia), Bocas del Toro (Panamá), and the Mesoamerican Barrier Reef in Belize. Detailed descriptions of the collections across these locations are found in Brandl et al.[45,24,46]. Reef outcrops were selected, measured, and covered with a bell-shaped fine mesh and tarpaulin, before being sprayed with a clove-oil:ethanol solution (1:5). Fish were collected with tweezers and placed in ziplock bags. In the laboratory, all fishes were measured (nearest 0.1mm, and binned to 1mm bins for statistical distribution fitting – see *Statistical analysis*), weighed (nearest 0.001g), photographed, and identified. See Brandl et al. [24] for full methods.

The majority (88%) of the data come from fish species that are unlikely to be targeted by either commercial or recreational fishing, providing opportunities to investigate complete body sizes in close to natural unfished populations. A population was defined as the collection of all individuals of a species within a location. For the visual census datasets, to avoid biases associated with clumped survey-locations, we determined a 'location' to be all surveys within a 1°x1° latitude-longitude grid cell (n = 3,089 populations). For the cryptobenthic fish data, a location was defined as each sampled reef outcrop (n = 139 populations). When analyzing the influence of spatial scale, for visual census data we also included a meta-population spatial scale, defined as all samples within an ecoregion based on the Marine Ecoregions of the World[20]. For both data sets we also included a species level scale, where all data from a single species were pooled.

Data filtering

Before analyses data were filtered for errors by exploring unusual body sizes and out of area observations. For visual census data, we further excluded populations where body size distributions spanned fewer than four body size bins (e.g., spanning 2.5cm, 5cm, and 7.5cm size bins only) or where more than half of individuals were in the smallest observable body size bin (i.e., 2.5cm bin). For the CRF data, where fish were measured individually, we set a minimum number of 10 individuals per population to fit a distribution, for the visual census data, where fish are counted in size bins, the minimum was set to 100 individuals. We performed a sensitivity analysis to show that these filtering parameters had no significant influence on the overall CV estimates (Extended Data Fig. 9).

Of the 3,089 visual census populations, 215 populations were not included in analyses when calculating the coefficient of variation: 59 populations were removed prior to distribution-fitting due to bimodal distributions, and 152 populations with greater than 50% of individuals within a single body size class. The removal of bimodal distributions did not significantly influence the results (Extended Data Fig. 10).

Statistical analysis

The goal of our statistical analysis was to evaluate the suitability of the positive-normal (PN, i.e., truncated normal) and lognormal (LN) probability density functions in describing the length-binned visual data for each fish population, and the more accurately measured cryptobenthic species. For each population we identified which of these two distributions better described the observed fish lengths and we also assessed whether the mean of the better distribution informed its coefficient of variation in a consistent manner across species or populations.

Consider a fish population of interest and let n_i denote the number of individuals observed within length-bin i, which is bounded by lengths l_i and l_{i+1} (cm). When assessing fish populations using data collected via cryptobenthic sampling length data were associated with 1mm size bins. Let $F_m(l|\mu,\sigma)$ denote the cumulative probability density associated with length distribution m = PN or LN. Here we have parameterized both length distributions according to a central tendency parameter μ and a variance parameter σ. The probability of observing a randomly selected fish being in length-bin i according to length distribution m is:

$$P_i^m = k\big(F_m(l_{i+1}|\mu,\sigma) - F_m(l_i|\mu,\sigma)\big),$$

where k is the normalization constant, $k = 1/(1 - F_m(l_1|\mu,\sigma))$. The fish observations are not independent as they have been collected at multiple locations and times. This non-independence will result in variation in counts across length-bins being greater than expected according to the multinomial distribution. A common way to account for non-independence and accommodate overdispersion is to assume that the counts distributed across bins are consistent with the Dirichlet-multinomial distribution. According to this assumption the likelihood of observing all the $n = \sum_i n_i$ fish counts is:

$$L(\mu,\sigma,\theta_0) = \frac{\Gamma(\theta_0)\Gamma(n+1)}{\Gamma(n+\theta_0)} \prod_{i=1}^{I} \frac{\Gamma(n_i+\theta_i)}{\Gamma(\theta_i)\Gamma(n_i+1)},$$

where θ_0 is a positive constant describing the degree of overdispersion, and $\theta_i = \theta_0 P_i^m$.

For each population and both length distributions, the three parameters $\{\mu,\sigma,\theta_0\}$ were estimated using Bayesian methods. We considered uninformative priors for all parameters and coded the likelihood using the stan programming language (implemented via the package 'rstan'). Convergence and estimation of the posterior parameter distributions was performed using three chains of length 2,000 iterations (including a 1,000-iteration warmup). As both size-distribution models were described by the same number of parameters (i.e. $\{\mu,\sigma,\theta_0\}$), we identified the better fitting size-distribution according to the highest median of the marginal posterior log-likelihood. Of the 3,068 fitted populations, 9 did not converge for the lognormal distribution but did for the normal, in which case the normal distribution was determined as the preferred distribution, for all other populations convergence was achieved for both distributions. The posterior median parameters: μ and σ, from the better fitting model, were then used to calculate the mean and the coefficient of variation of fish lengths for the whole fish population.

All statistical analyses were performed using the statistical programming language R[47] in combination with the Bayesian statistical modelling language Stan[48].

Reconstructing the body size distribution from a single parameter

Assuming a coefficient of variation of 0.34, and the mean body length of the population ($\bar{L}$), we can calculate the two parameters of each of the lognormal and normal distribution (truncated at the smallest body size observable; 1.25cm for visual census data). For the normal distribution, the mean of the distribution is equal to

the observed mean of the population, ($\bar{L}$), the standard deviation σ of the distribution is calculated equal to 0.34$\bar{L}$. For the lognormal distribution, if parameterized as $LN(\mu, \sigma^2)$, we can calculate σ as $\sqrt{log(0.34^2 + 1)}$, and μ as $log(\bar{L} - \frac{\sigma^2}{2})$.

To reconstruct the natural population body length distribution from just asymptotic length (L_{inf}) for the comparisons of performances between our simple model and a commonly used size-based population assessment models LBSPR we used Prince et al. estimates of L_{inf} that were available for 68 mostly unfished species in our dataset. The two size measures were correlated (Extended Data Fig. 10) (R^2 = 74%) and mean body sizes $\bar{L}$ for the 68 species were estimated using empirically derived regression $log(\bar{L}) = 0.37 + 0.78 \cdot log(L_{inf})$.

Reconstructing body size distributions using LBSPR model

To compare how our simple method of reconstructing body size distributions performed compared to traditional fisheries population dynamics predictions, we used the commonly applied size based population assessment method LBSPR (length-based spawning potential ratio)[15], as implemented in the R package LBSPR[33]. The LBSPR method reconstructs expected size distributions using Beverton and Holt fisheries population dynamics predictions[14], assuming that fish growth follows the Von Bertalanffy growth model, mortality at ages recruited to the fishery is constant, recruitment is continuous and the coefficient of variation around length at age is 10%[15,16]. The function 'LBSPRsim' from the LBSPR package was used to reconstruct unfished length distribution using mortality and growth ratio (M/k) and asymptotic length (L_{inf}).

Species-level M, k, and L_{inf} were obtained from Prince et al.[17]. If multiple values of M and k were given for a single species, the M and k combination with the closest M/k ratio to the median of the species was selected. Only species that appeared in both the visual census dataset, had published M/k ratio and L_{inf} values from Prince et al.[17] and were deemed to be not intensively fished were selected (n = 68 species, including species with some level of fishing in part of their geographic distribution). These 68 species were used to compare how the model presented here (normal truncated size distribution with the CV around mean size of 0.34) and LBSPR predict observed length frequencies using the same L_{inf} parameter (Fig. 4).

Observed versus expected dissimilarity

To compare between empirically observed and predicted body length distributions (e.g., normal versus LBSPR expected in Fig.3, or normal versus lognormal in Fig. 4), we used the Kolmogorov-Smirnov (K-S) test. The K-S test is a nonparametric test of equality between two distributions, in our case an observed and predicted body length distribution. Here we use the K-S statistic (using the 'ks.test' function within the 'stats' package of R) as a relative measure of dissimilarity between two samples (observed, and predicted, proportion in body length bin).

Code availability statement

Data and code (R and stan) to recreate the figures in this study are available at github.com/FreddieJH/size_dist_fitting (and will be archived on Zenodo before publication).

Acknowledgements

We thank the many Reef Life Survey divers and members of the Australian Temperate Reef Collaboration who collected the visual census data used here. These data used in the analyses are managed through, and were

sourced from, Australia's Integrated Marine Observing System (IMOS) – IMOS is enabled by the National Collaborative Research Infrastructure Strategy (NCRIS). We also thank the many divers that assisted with cryptobenthic fish collections. We also thank Rainer Froese, Ken Haste Andersen and Gustav Delius for comments on the manuscript drafts, Michael Spence for efforts in improving computation speeds in Bayesian modelling and Wanwan Kurniawan for help with the LBSPR model. This study was supported by the Australian Research Council (ARC) Discovery Project (DP220102446) to AA, SAR and NCK, ARC Future Fellowship (FT190100599) to RD S-S, and Pew Fellowships in Marine Conservation to AA and RD S-S.

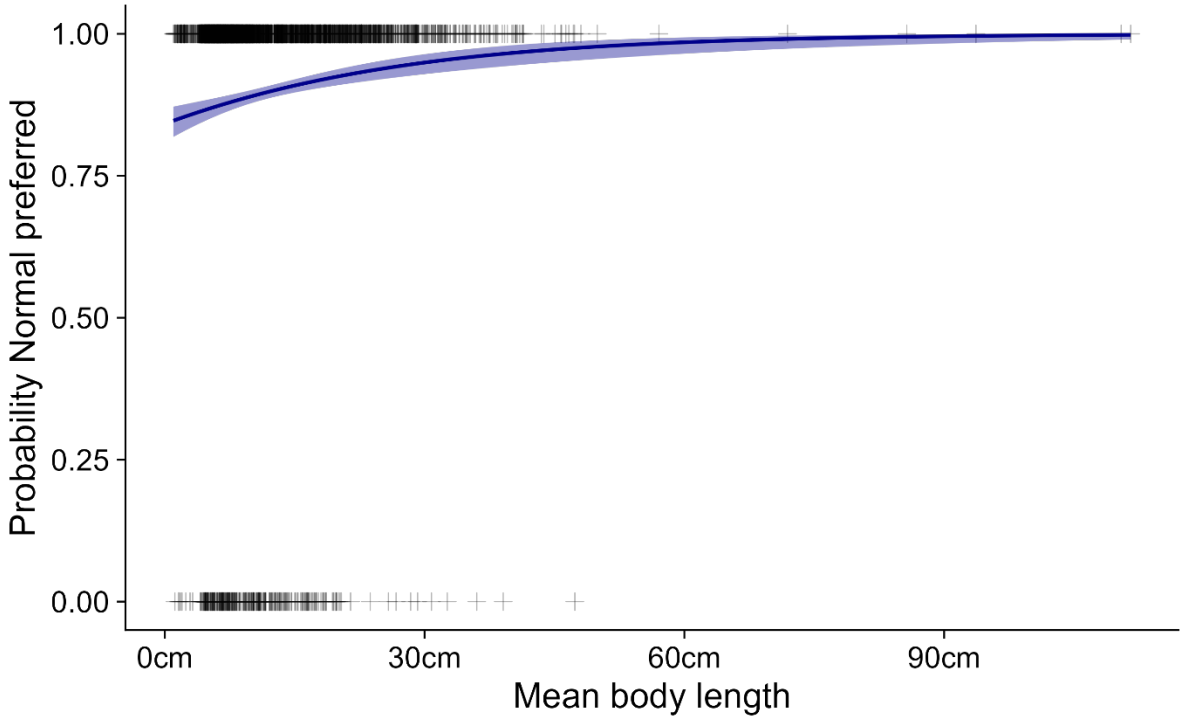
References

1. Andersen, K. H. *et al.* Characteristic Sizes of Life in the Oceans, from Bacteria to Whales. *Annu. Rev. Mar. Sci.* **8**, 217–241 (2016).
2. Hordyk, A., Ono, K., Sainsbury, K., Loneragan, N. & Prince, J. Some explorations of the life history ratios to describe length composition, spawning-per-recruit, and the spawning potential ratio. *ICES Journal of Marine Science* **72**, 204–216 (2015).
3. Charnov, E. L., Turner, T. F. & Winemiller, K. O. Reproductive constraints and the evolution of life histories with indeterminate growth. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9460–9464 (2001).
4. Jensen, A. L. Beverton and Holt life history invariants result from optimal trade-off of reproduction and survival. (1996).
5. Beverton, R. & Holt, S. *A Review of the Lifespans and Mortality Rates of Fish in Nature, and Their Relation to Growth and Other Physiological Characteristics*. vol. 5 (Wiley Online Library, 1959).
6. Charnov, E. L. *Life History Invariants: Some Explorations of Symmetry in Evolutionary Ecology*. (Oxford University Press, 1993).
7. Jørgensen, C. & Holt, R. E. Natural mortality: Its ecology, how it shapes fish life histories, and why it may be increased by fishing. *Journal of Sea Research* **75**, 8–18 (2013).
8. Dureuil, M. & Froese, R. A natural constant predicts survival to maximum age. *Commun Biol* **4**, 641 (2021).
9. Gislason, H., Daan, N., Rice, J. C. & Pope, J. G. Size, growth, temperature and the natural mortality of marine fish: Natural mortality and size. *Fish and Fisheries* **11**, 149–158 (2010).
10. Healy, K., Ezard, T. H. G., Jones, O. R., Salguero-Gómez, R. & Buckley, Y. M. Animal life history is shaped by the pace of life and the distribution of age-specific mortality and reproduction. *Nat Ecol Evol* **3**, 1217–1224 (2019).
11. Pauly, D. On the interrelationships between natural mortality, growth parameters, and mean environmental temperature in 175 fish stocks. *ICES Journal of Marine Science* **39**, 175–192 (1980).
12. Prince, J., Hordyk, A., Valencia, S. R., Loneragan, N. & Sainsbury, K. Revisiting the concept of Beverton-–Holt life-history invariants with the aim of informing data-poor fisheries assessment. *ICES Journal of Marine Science* **72**, 194–203 (2015).
13. Giometto, A., Altermatt, F., Carrara, F., Maritan, A. & Rinaldo, A. Scaling body size fluctuations. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 4646–4650 (2013).
14. Beverton, R. & Holt, S. *On the Dynamics of Exploited Fish Populations.* (Her Majesty's Stationary Office, 1957).

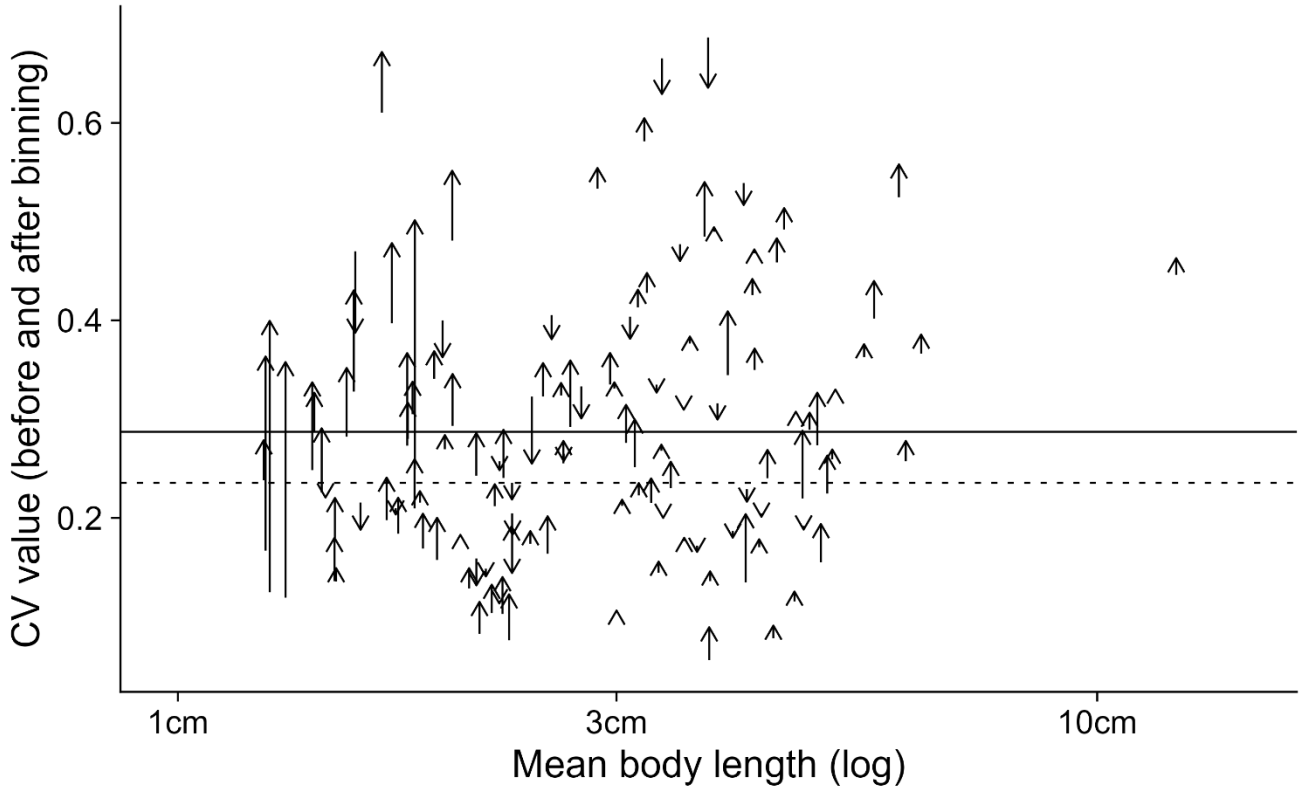
15. Hordyk, A., Ono, K., Valencia, S., Loneragan, N. & Prince, J. A novel length-based empirical estimation method of spawning potential ratio (SPR), and tests of its performance, for small-scale, data-poor fisheries. *ICES Journal of Marine Science* **72**, 217–231 (2015).
16. Froese, R. *et al.* A new approach for estimating stock status from length frequency data. *ICES Journal of Marine Science* **75**, 2004–2015 (2018).
17. Prince, J. D., Wilcox, C. & Hall, N. How to estimate life history ratios to simplify data-poor fisheries assessment. *ICES Journal of Marine Science* **80**, 2619–2629 (2023).
18. FAO. *The State of World Fisheries and Aquaculture 2022*. http://www.fao.org/documents/card/en/c/cc0461en (2022) doi:10.4060/cc0461en.
19. Brandl, S. J., Goatley, C. H. R., Bellwood, D. R. & Tornabene, L. The hidden half: ecology and evolution of cryptobenthic fishes on coral reefs. *Biol Rev* **93**, 1846–1873 (2018).
20. Spalding, M. D. *et al.* Marine Ecoregions of the World: A Bioregionalization of Coastal and Shelf Areas. *BioScience* **57**, 573–583 (2007).
21. Huston, M. A. & DeAngelis, D. L. Size Bimodality in Monospecific Populations: A Critical Review of Potential Mechanisms. *The American Naturalist* **129**, 678–707 (1987).
22. Macdonald, P. D. M. & Pitcher, T. J. Age-Groups from Size-Frequency Data: A Versatile and Efficient Method of Analyzing Distribution Mixtures. *J. Fish. Res. Bd. Can.* **36**, 987–1001 (1979).
23. Depczynski, M. & Bellwood, D. R. Extremes, plasticity, and invariance in vertebrate life history traits: insights from coral reef fishes. *Ecology* **87**, 3119–3127 (2006).
24. Brandl, S. J. *et al.* Demographic dynamics of the smallest marine vertebrates fuel coral reef ecosystem functioning. *Science* **364**, 1189–1192 (2019).
25. Thorson, J. T., Munch, S. B., Cope, J. M. & Gao, J. Predicting life history parameters for all fishes worldwide. *Ecological Applications* **27**, 2262–2276 (2017).
26. Froese, R. & Binohlan, C. Empirical relationships to estimate asymptotic length, length at first maturity and length at maximum yield per recruit in fishes, with a simple method to evaluate length frequency data. *Journal of Fish Biology* **56**, 758–773 (2000).
27. Beverton, R. J. H. Patterns of reproductive strategy parameters in some marine teleost fishes. *Journal of Fish Biology* **41**, 137–160 (1992).
28. Mangel, M. Invariant ratios vs. dimensionless ratios. *Science* **310**, 1426–1427 (2005).
29. Nee, S., Colegrave, N., West, S. A. & Grafen, A. The Illusion of Invariant Quantities in Life Histories. *Science* **309**, 1236–1239 (2005).
30. Coulson, T., Gaillard, J. & Festa-Bianchet, M. Decomposing the variation in population growth into contributions from multiple demographic rates. *Journal of Animal Ecology* **74**, 789–801 (2005).
31. Ackerman, J. & Bellwood, D. Reef fish assemblages: a re-evaluation using enclosed rotenone stations. *Mar. Ecol. Prog. Ser.* **206**, 227–237 (2000).
32. Costello, C. *et al.* Status and Solutions for the World's Unassessed Fisheries. *Science* **338**, 517–520 (2012).
33. Hordyk, A. Package LBSPR. (2019).
34. Rudd, M. B. & Thorson, J. T. Accounting for variable recruitment and fishing mortality in length-based stock assessments for data-limited fisheries. *Can. J. Fish. Aquat. Sci.* **75**, 1019–1035 (2018).
35. Barnett, L. A. K., Branch, T. A., Ranasinghe, R. A. & Essington, T. E. Old-Growth Fishes Become Scarce under Fishing. *Current Biology* **27**, 2843-2848.e2 (2017).

36. Froese, R. *et al.* On the pile-up effect and priors for Linf and M/K: response to a comment by Hordyk et al. on “A new approach for estimating stock status from length frequency data”. *ICES Journal of Marine Science* **76**, 461–465 (2019).
37. Hordyk, A. R., Prince, J. D., Carruthers, T. R. & Walters, C. J. Comment on “A new approach for estimating stock status from length frequency data” by Froese et al. (2018). *ICES Journal of Marine Science* **76**, 457–460 (2019).
38. Hilborn, R. *et al.* Effective fisheries management instrumental in improving fish stock status. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 2218–2224 (2020).
39. Hilborn, R. *et al.* State of the World’s Fisheries. *Annu. Rev. Environ. Resour.* **28**, 359–399 (2003).
40. Worm, B. *et al.* Rebuilding Global Fisheries. *Science* **325**, 578–585 (2009).
41. Edgar, G. J. *et al.* Establishing the ecological basis for conservation of shallow marine life using Reef Life Survey. *Biological Conservation* **252**, 108855 (2020).
42. Edgar, G. J. & Stuart-Smith, R. D. Systematic global assessment of reef fish communities by the Reef Life Survey program. *Sci Data* **1**, 140007 (2014).
43. Edgar, G. J. & Barrett, N. S. Effects of the declaration of marine reserves on Tasmanian reef fishes, invertebrates and plants. *Journal of Experimental Marine Biology and Ecology* **242**, 107–144 (1999).
44. Reef Life Survey Foundation. Standardised Survey Procedures for Monitoring Rocky & Coral Reef Ecological Communities. https://reeflifesurvey.com/wp-content/uploads/2019/02/ NEW-Methods-Manual_150815.pdf (2019).
45. Brandl, S. J., Casey, J. M., Knowlton, N. & Duffy, J. E. Marine dock pilings foster diverse, native cryptobenthic fish assemblages across bioregions. *Ecology and Evolution* **7**, 7069–7079 (2017).
46. Brandl, S. J. *et al.* Extreme environmental conditions reduce coral reef fish biodiversity and productivity. *Nat Commun* **11**, 3832 (2020).
47. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, Vienna, Austria, 2022).
48. Stan Development Team. Stan Modeling Language Users Guide and Reference Manual, v2.21.0. (2023).

Extended Data



Extended Data Figure 1. Populations with larger mean body lengths are usually better described by a truncated normal than a lognormal distribution. The blue line shows the logistic regression fit to the data. The populations include both visual census and cryptobenthic data.

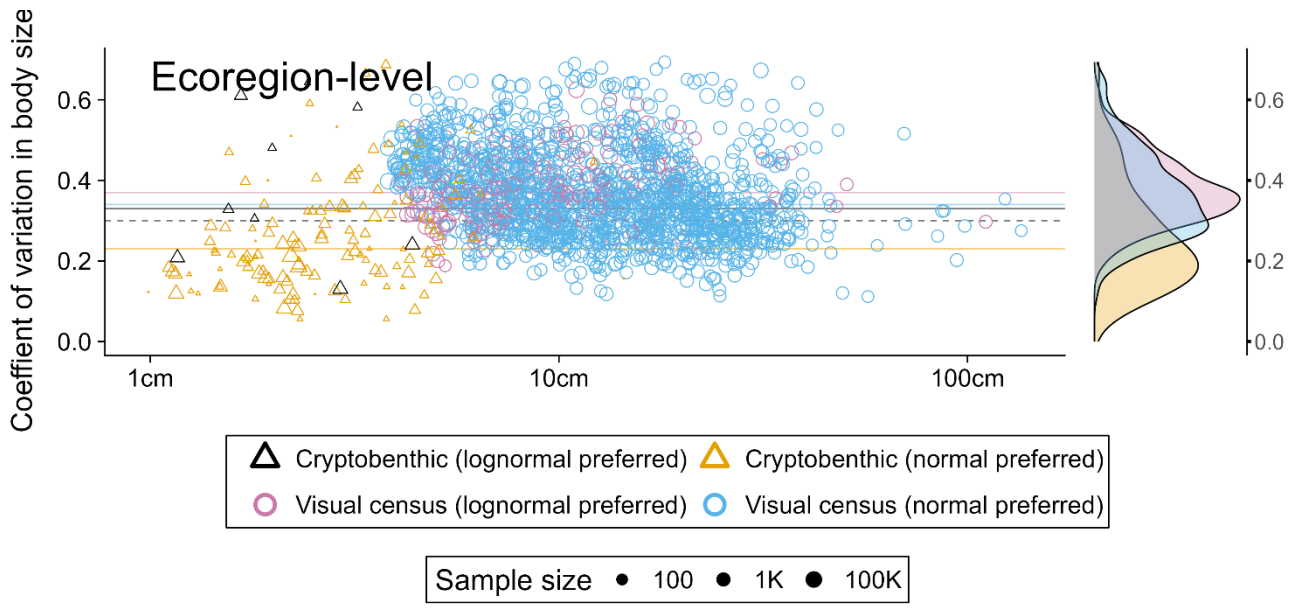


490

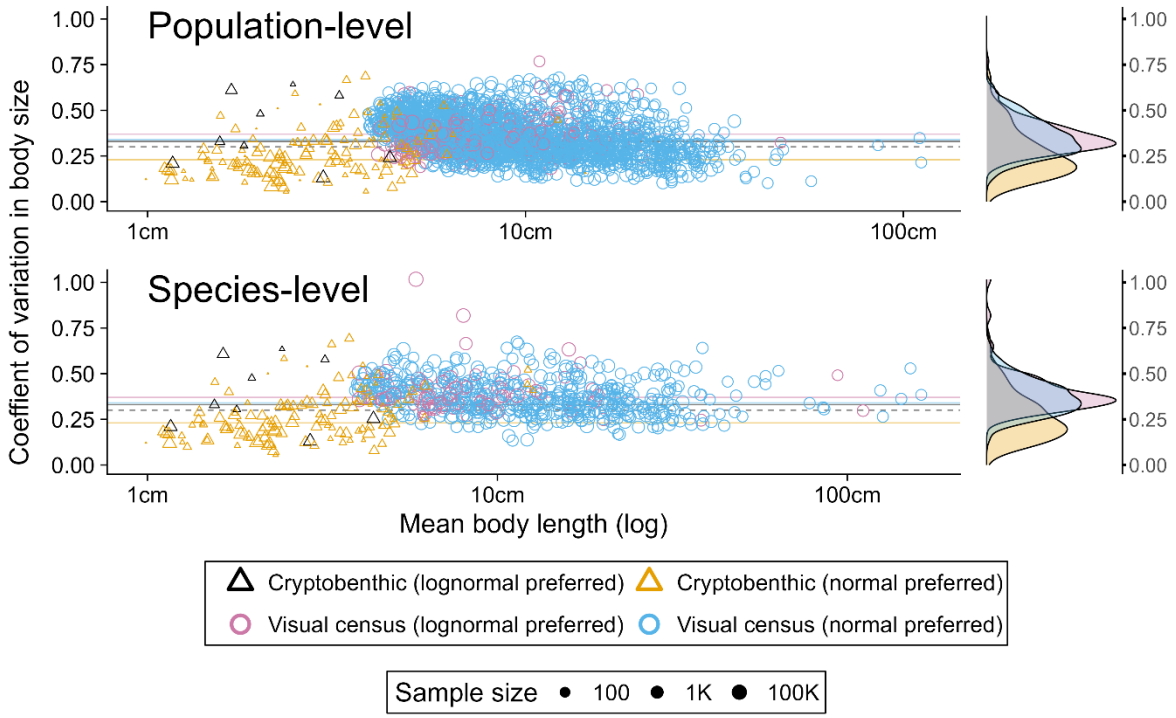
491 **Extended Data Figure 2.** Binning the cryptobenthic data (median CV = 0.23, dotted horizontal line) into visual
492 census bins results in a greater median CV value (median CV = 0.28, solid horizontal line). Arrows represent the
493 change from the CV value (either normal or lognormal) of the continuous data to the CV value of the binned
494 data. Note: To be binned into the same visual census bins (2.5cm, 5cm, 7.5cm etc.), cryptobenthic body
495 lengths were multiplied by a constant (constant = 3) before binning.

496

497



Extended Data Figure 3. Relationship between population mean body length and CV at the ecoregion-level in the visual census data and location level in the cryptobenthic data. Fisheries-targeted species and distributions deemed to be bimodal have been excluded. The median CV value is 0.35, 80% of CV values fall in the range of 0.23 to 0.51 and 95% of the CV values fall in the range of 0.19 and 0.56.



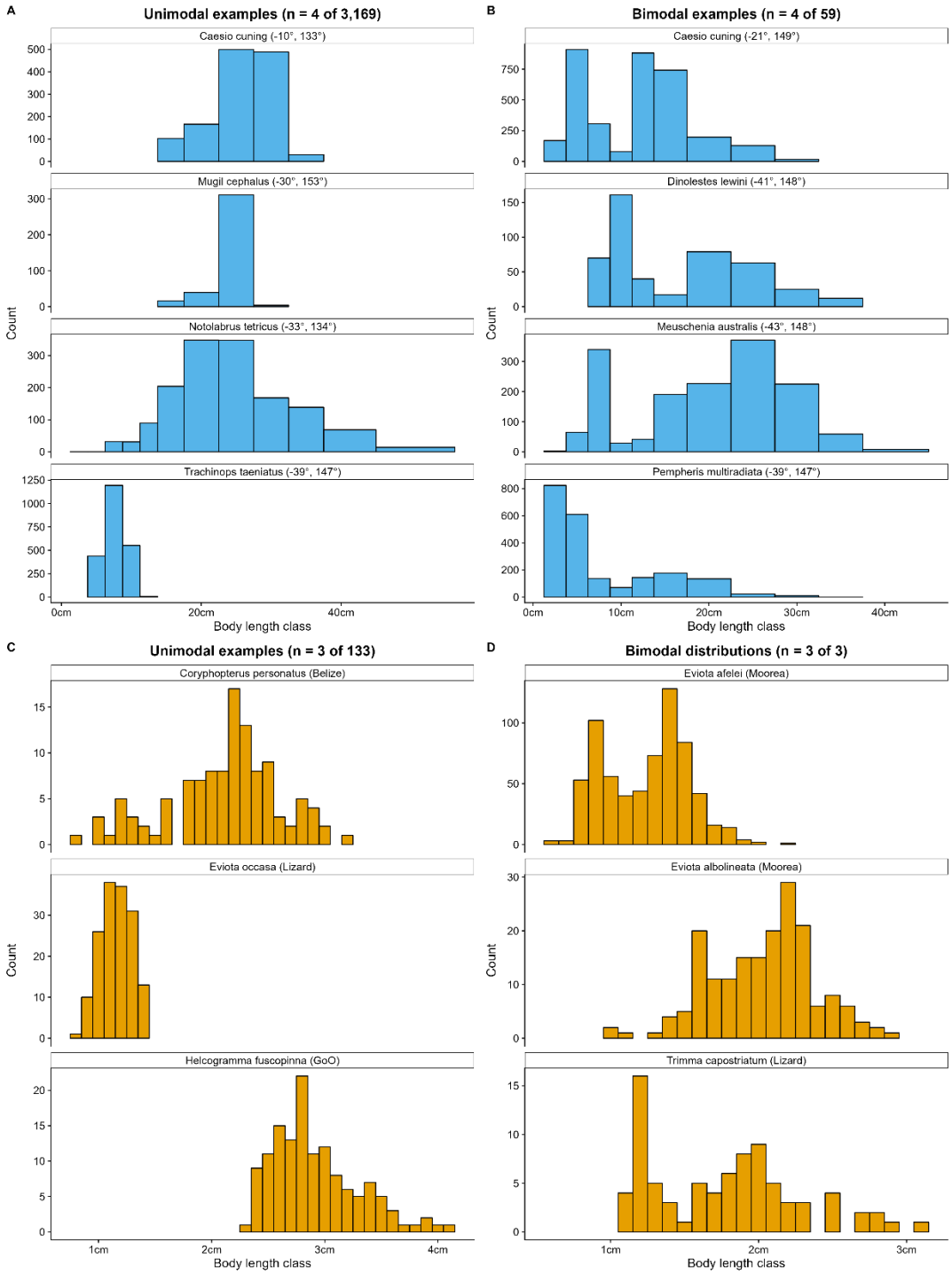
507

508 **Extended Data Figure 4.** Reconstruction of Fig. 2 of the main text but excluding species and genera that are
509 potentially targeted by fishing. Excluding these species has almost no influence on the estimate of CV. At the
510 population level the median CV value is 0.34, 80% of CV values fall in the range of 0.22 to 0.51 and 95% of the
511 CV values fall in the range of 0.19 and 0.55. At the species-level the median CV value is 0.35, 80% of CV values
512 fall in the range of 0.22 to 0.50 and 95% of the CV values fall in the range of 0.17 and 0.56.

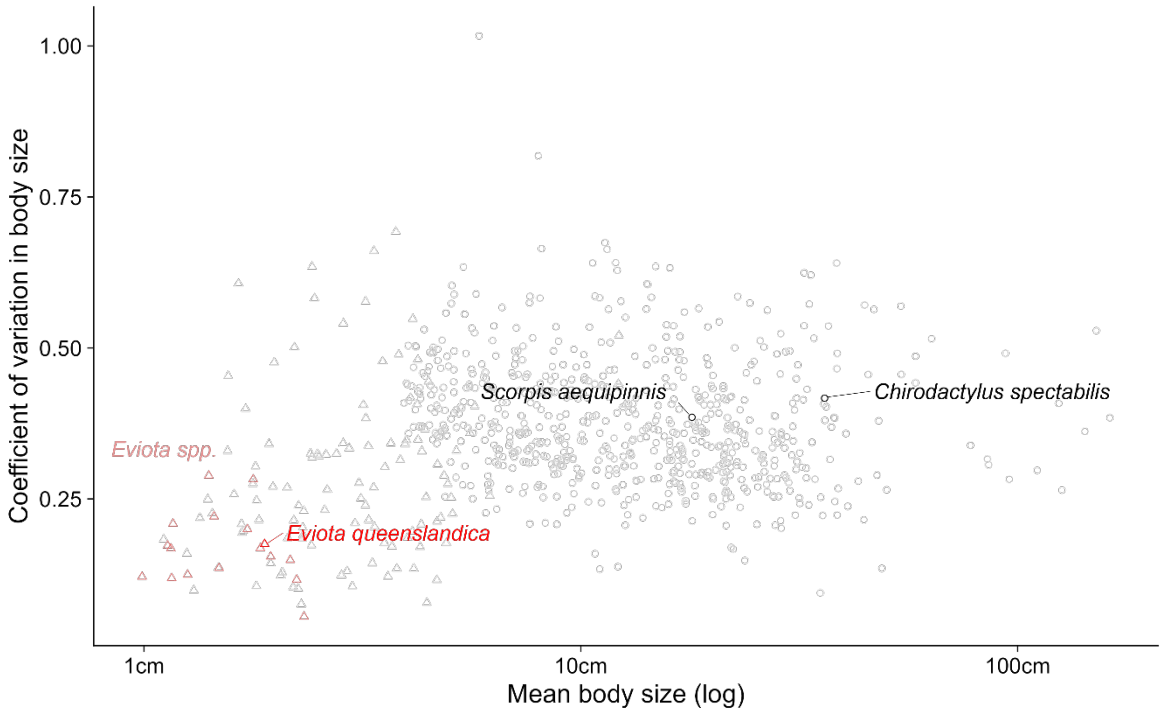
513

514

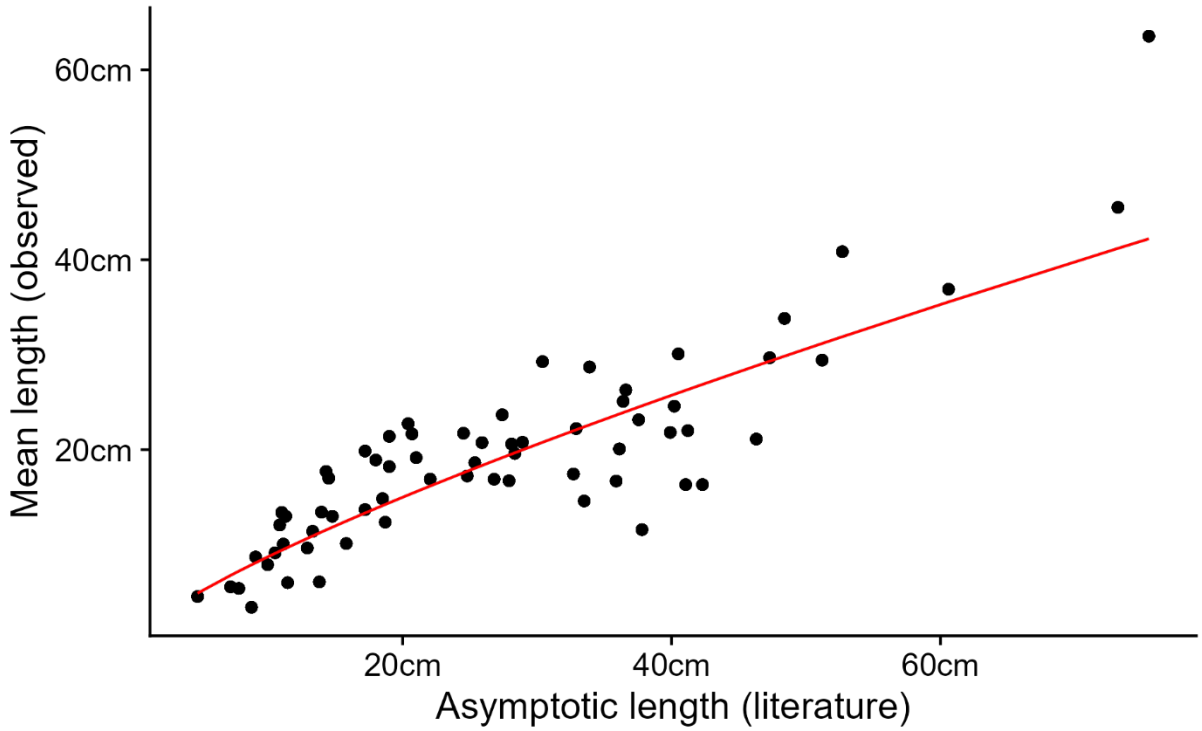
515



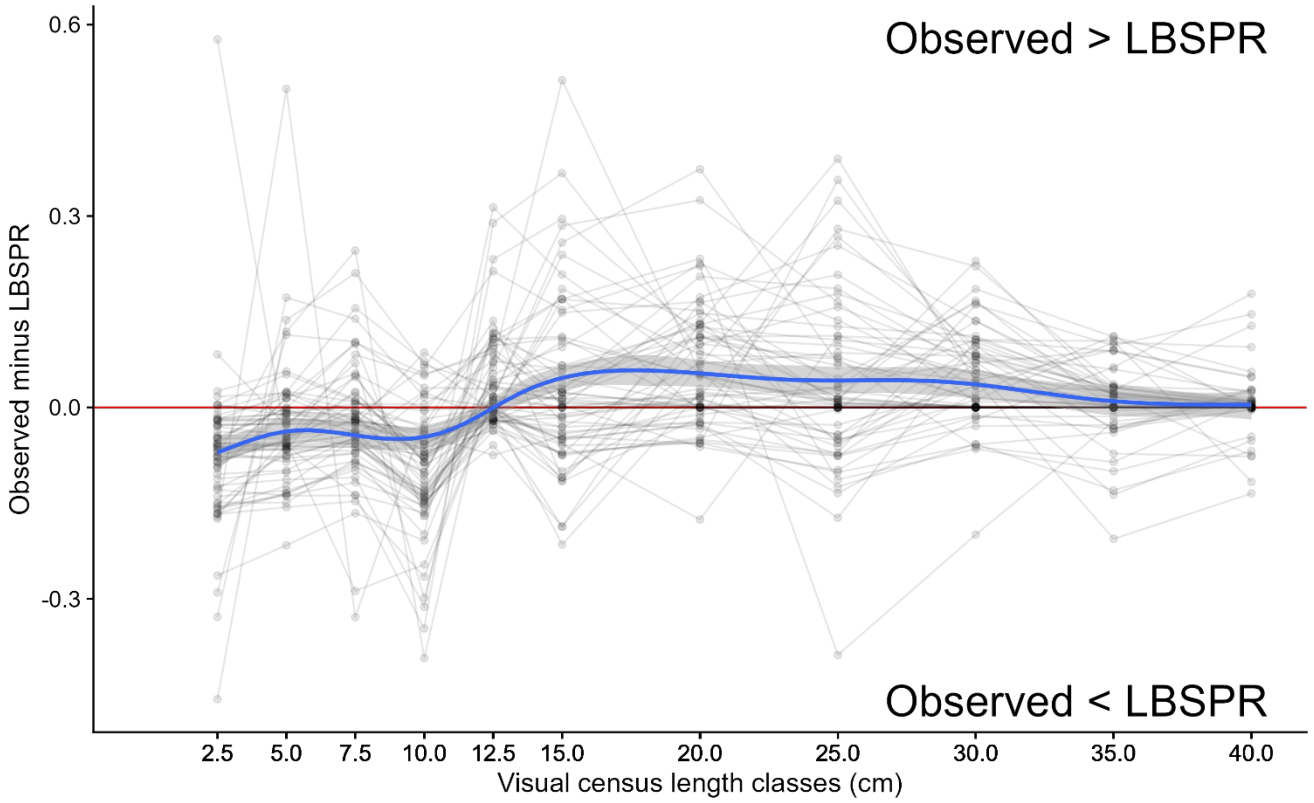
Extended Data Figure 5. Randomly selected examples of approximately unimodal (A, C) and bimodal (B, D) observed body length distributions in visual census sampling (blue, A, B) and cryptobenthic sampling (orange, C, D). Approximately bimodal shapes were found in 59 out of 3,089 (2%) of visual census, and 3 of 136 (2%) cryptobenthic population-level datasets. Bimodal distributions were removed from distribution-fitting analyses.



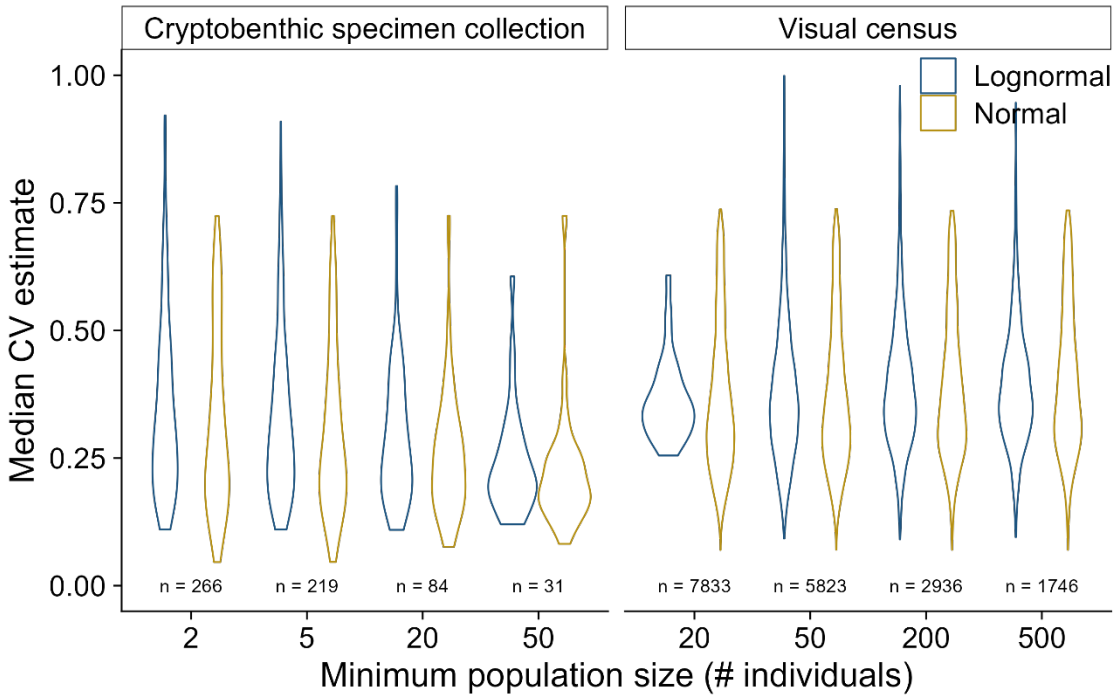
Extended Data Figure 6. Relationship between mean body length and coefficient of variation (CV), highlighting three species with known extreme life-histories. The two species in black text are known to be long-lived, slow growing species (*Scorpis aequipinus* and *Chirodactylus spectabilis)*, yet their CV values fall in the center of the of the distribution of CV values. On the other hand, *Eviota spp.* and in particular, *E. queenslandica,* are known to be fast-growing and short-lived. *Eviota* spp. are on the lower end of the range of CV estimates.



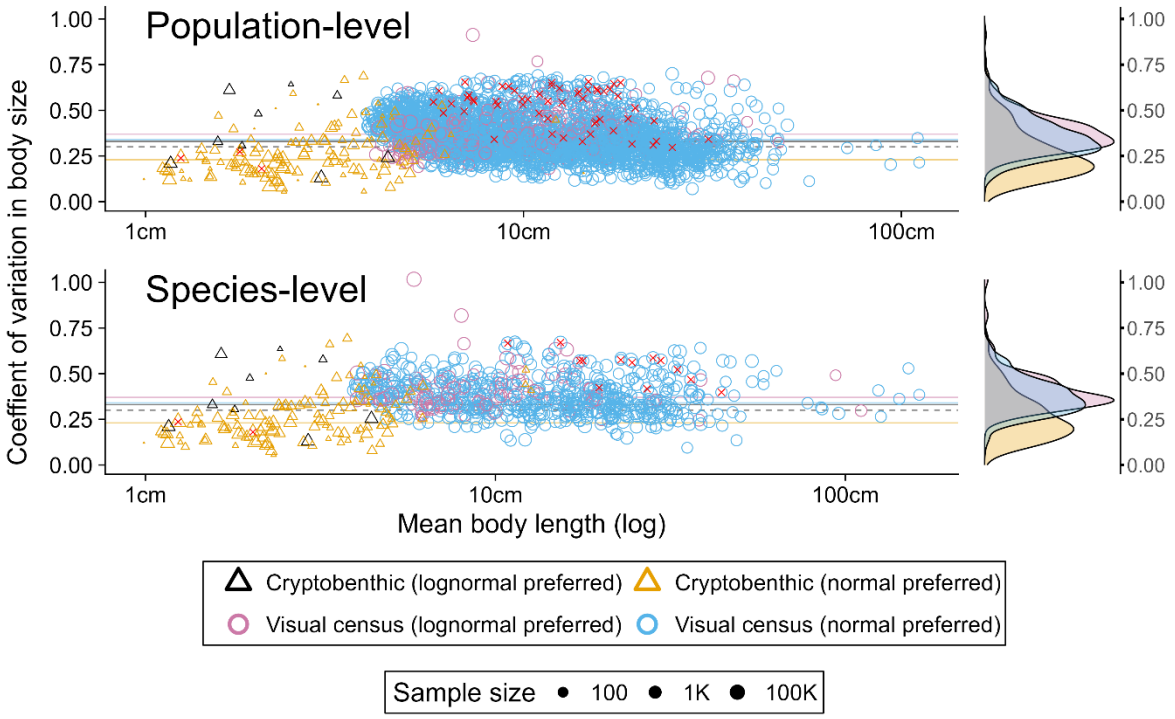
Extended Data Figure 7. *A fitted linear model to predict mean population body length (from empirical body length distributions) from asymptotic body length (available from Prince et al.[17]). This fitted model was used to reconstruct the body length distribution when only asymptotic body length is known, which is often the case for many species or populations. Fitted model (red line): log(mean_length) = 0.37 + 0.78*log(asymptotic_length).*



Extended Data Figure 8. *Observed proportion within a length bin minus expected proportion as predicted by the theoretical model LBSPR. LBSPR is parameterized by asymptotic length, mortality (M) and growth rate (k). Points below the red horizontal line indicate there is a lower observed proportion of individuals in the size bin than predicted by LBSPR. The x-axis has been truncated to 40cm to emphasize this region where observational sampling bias is most likely to occur.*



Extended Data Figure 9. Changing the minimum sample size (individuals within a population) required for distribution fitting has minimal influence of the median estimate of the coefficient of variation (CV). CV values greater than one have been excluded from this figure.



554

555 **Extended Data Figure 10.** Distributions deemed to be bimodal (red crosses, 2% of total populations) were
556 excluded in Fig. 2 in the main text. Including these distributions has minor influence on the estimate of CV. At
557 the population level the median CV value is 0.34, 80% of CV values fall in the range of 0.22 to 0.51 and 95% of
558 the CV values fall in the range of 0.19 and 0.56. At the species-level the median CV value is 0.36, 80% of CV
559 values fall in the range of 0.22 to 0.51 and 95% of the CV values fall in the range of 0.17 and 0.57.

560