

Supporting Information Methods

Data collection

Home range data were collected as a mean, range or maximum linear distance. Data for the same species from different locations were separated. Most values were means with an associated estimate of standard deviation (SD). For values given as ranges, we calculated means and SDs based on the specified minimum and maximum home range. For values given as maxima, we set the SD to 0. For values given as means but without associated SDs, we used a generic estimate of SD based on the median observed coefficient of variation (approximately 1, i.e. the SD was assumed to be equal to the mean).

Fish survey data were collected based on visual census counts along belt transects of varying dimensions: 10×2 m (20 m^2), 15×2 m (30 m^2), 50×2 m (100 m^2), and 50×3 m (150 m^2). Very small fishes in Belize were counted within 5×5 m quadrats (25 m^2). Transect counts were combined as follows to calculate fish densities: $N_{\text{fish}} \times \text{Area}_{\text{transect}}^{-1} \times N_{\text{transects}}^{-1}$. The resulting data were then averaged across sites to yield means \pm SDs.

Ecological information on our final set of species was downloaded from Fishbase (Froese and Pauly 2016) using R (Boettiger et al. 2012). Functional groups were assigned based on diet and habitat as: (1) Browser, (2) Corallivore, (3) Excavator/Bioeroder, (4) Grazer/Detritivore, (5) Invertivore, (6) Invertivore/Omnivore, (7) Piscivore, (8) Piscivore/Invertivore, (9) Planktivore, and (10) Scraper/Small excavator. Fishery values were assigned based on a simple score for Fishbase records under “Fisheries”, “Gamefish”, and “Aquarium”. Highest values of 7 represented the maximum sum of scores for “Fisheries” (0 for “of no interest”, 1 for “subsistence fisheries”, 2 for “minor commercial”, 3 for “commercial”, and 4 for “highly commercial”), “Gamefish” (1 for “yes”), and “Aquarium” (1 for “public aquariums” and 2 for “commercial”). The lengths of fishes are maximum lengths as specified in the Fishbase dataset. Most values were given as total lengths (TL). To approximate TL for values given as standard length (SL), we used a generic conversion: $TL = 1.1 * SL$. Overall, our final set of 66 species covered 17 common families of coral reef fishes, including a diverse range of lengths, functional roles, and fishery values; from very small and sedentary butterflyfishes to large and highly mobile snappers and sharks (see Table S1 and Fig. S1).

Data analysis

To highlight predictions for standard reserves around the world, we assumed a median global reserve area of 4 km² (Halpern 2003), which indicates that the median reserve diameter is 2 km (quadratic reserve) or 2.26 km (circular reserve). Reserves in the Philippines were used as an example of a particularly small diameter: 387 m (quadratic reserve) or 437 m (circular reserve) indicated by a median reserve area of 0.15 km² (Weeks et al. 2010).

To determine whether mean home ranges (HR), mean densities (D) or maximum fish lengths (L) alone could be used to approximate simulation-derived predictions of effective reserve sizes (ERS) required to achieve 50%, 75% and 95% protection, we used least-square linear regressions on log-transformed data: (1) $ERS+1 \sim I + b1 (HR) + b2 (D) + b3 (HR \times D)$; (2) $ERS+1 \sim I + b (HR)$; and (3) $ERS+1 \sim I + b (L)$. Here, I is the intercept, and the parameters b represent the slopes of regression lines. The goodness of fit of regressions was assessed based on the coefficient of determination (R^2) and by visual examination of residuals.

Modelling procedure

Modelling step 1 (counting fish): Simulations were started by firstly determining the size of the seascape in which a reserve was enforced (e.g. 1 km). The model then implemented multiple, hypothetical fish surveys (N = 100 per reserve size and species), which sampled fish counts at random from the negative binomial distribution. The negative binomial distribution ensured that hypothetical fish counts could not be negative while still allowing for parameterizing the sampling such that simulations matched precisely the mean and natural variability of densities observed in the field.

Modelling step 2 (locating fish): Following the sampling of fish numbers, individuals were assigned to specific locations throughout the reserve. For this, we used two alternative methods: (1) randomly, and (2) evenly. The first method assumed that intraspecific interactions do not prevent individuals from occupying the same or nearby locations as conspecifics. In contrast, the second method assumed that individuals are territorial, choosing the maximum possible distance between themselves and conspecifics. For species observed to form pairs or schools with conspecifics, we aggregated individuals in schools whose sizes matched the mean and natural variability of schools observed in the field. As described

above, this was done by sampling at random from the negative binomial distribution. If school sizes could not be calculated at species-level, we used data at the genus-level or family-level (rare). Two different thresholds for schooling behavior were initially considered: (1) at least 20% of observations representing pairing or schooling individuals; (2) at least 50% of observations representing pairing or schooling individuals. However, since outcomes were very similar, and since the 20% threshold matched closely the information on schooling behavior available on Fishbase, we do not present outcomes for the 50% threshold. Species-specific information on assigned modes of distribution and school sizes is given in Table S1.

Modelling step 3 (assessing protection): Having determined the locations of individuals, each of them was assigned its individual home range value. For individuals aggregating in schools, we assumed the same home range value. The location of an individual or group of individuals was assumed to represent the center of its home range, such that movements were assumed to be confined to $0.5 \times$ the assigned home range value on either side of the assigned location. Home range values were sampled at random again from the negative binomial distribution, ensuring that sampled values could not be negative while still matching the mean and natural variability of field measurements. In the final step, our model then calculated whether the movements of individuals exceeded reserve boundaries. Conservatively, we assumed that this situation would lead to eventual mortality from fishing. That is, only individuals whose entire home range was contained within reserve boundaries were assumed to be protected.

References

- Boettiger C., Lang D.T., Wainwright P.C. (2012) rfishbase: exploring, manipulating and visualizing FishBase data from R. *J Fish Biol* **81**, 2030-2039.
- Froese R., Pauly D. (2016) FishBase (www.fishbase.org).
- Halpern B.S. (2003) The impact of marine reserves: Do reserves work and does reserve size matter? *Ecol Appl* **13**, S117-S137.
- Weeks R., Russ G.R., Alcala A.C., White A.T. (2010) Effectiveness of Marine Protected Areas in the Philippines for Biodiversity Conservation. *Conserv Biol* **24**, 531-540.

Supporting Information Figures

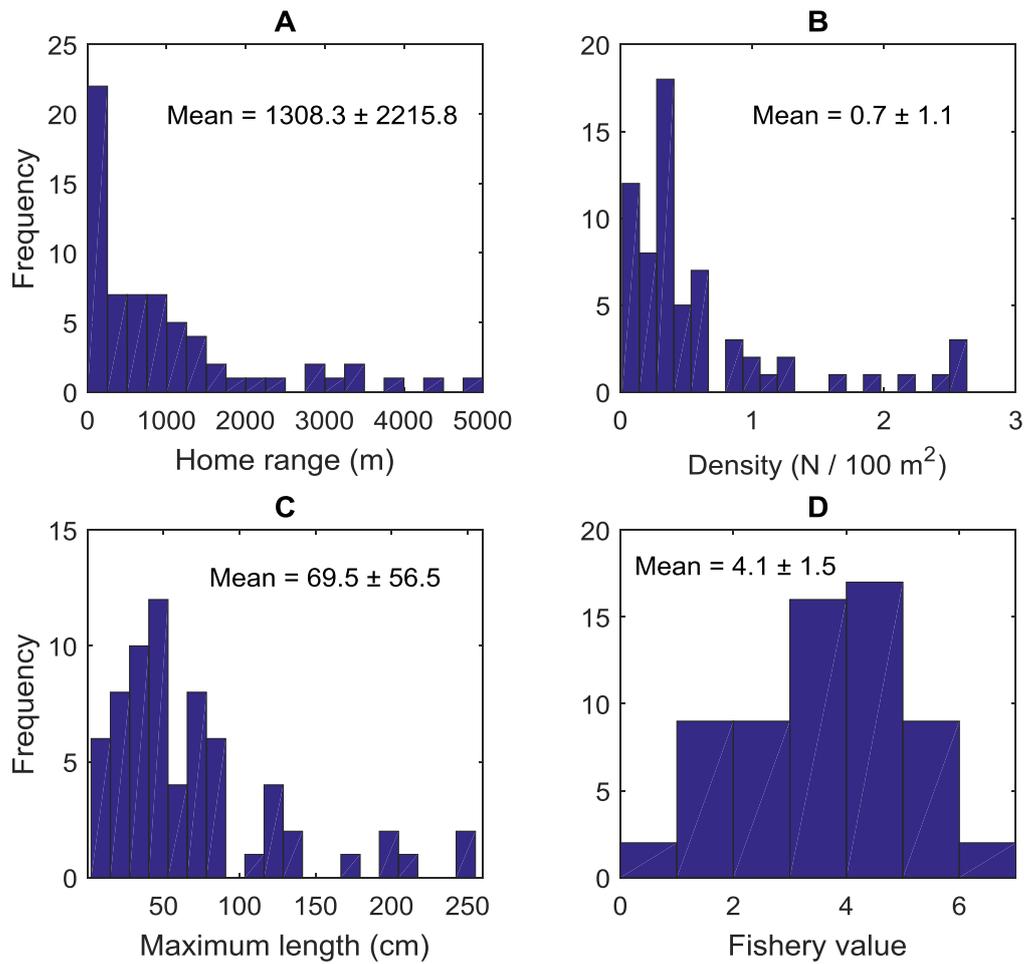


Figure S1: Histograms of the primary data for 66 coral reef fish species used for simulations in this study. Values in plots are means \pm SDs. Two outlying species with mean home ranges of 12 km in A (*Aprion virescens* and *Sphyraena barracuda*) and with densities of 4.5 and 6.5 fish per 100 m² in B (*Haemulon plumierii* and *Kyphosus sectatrix*, respectively) were excluded to represent the remaining data adequately. See text and Table S1 for details.

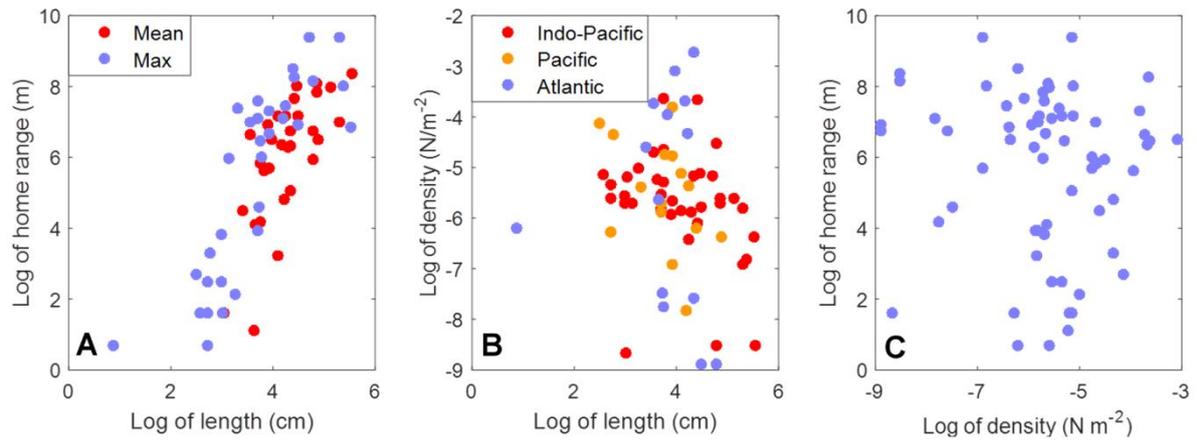


Figure S2: Relationships between the home range, density and maximum body length in 66 coral reef fishes. Correlations between length and home range were strongly positive (A, Pearson's $r = 0.77$, $p = 1.73 \times 10^{-14}$). The relationship between length and density was negative, albeit marginally significant (B, Pearson's $r = -0.22$, $p = 0.07$). The relationship between density and home range was unclear (C, Pearson's $r = -0.06$, $p = 0.60$).

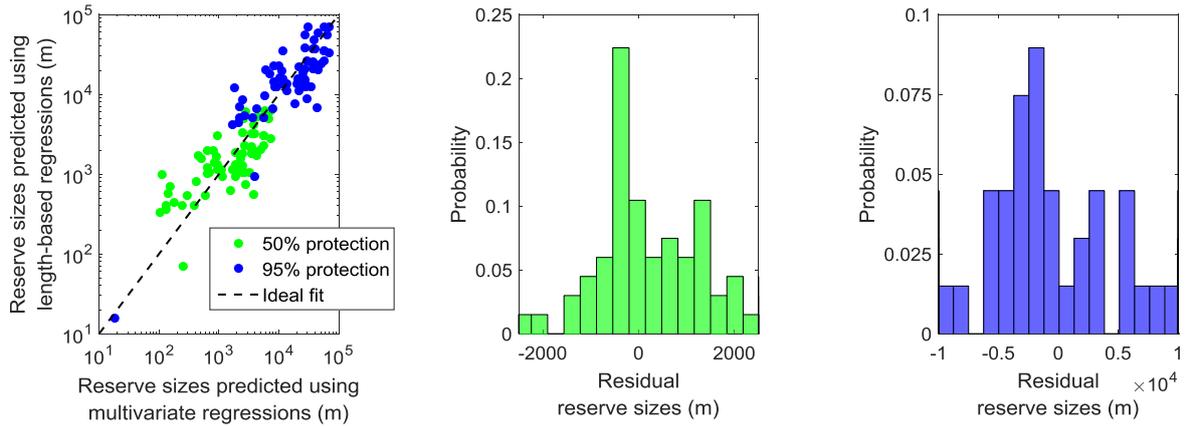


Figure S3: Reserve sizes predicted based on multivariate vs univariate regression models. The multivariate model included home range and density information, while the univariate model used maximum fish lengths only. Histograms in B (50% protection) and C (95% protection) specify probabilities of positive and negative deviations in model predictions (multivariate – univariate). To represent most data points adequately, histograms are capped at deviations of 2.5 km (B) and 10 km (C), which captured 88% and 51%, respectively, of all data points.

Supporting Information Tables (partly in separate Excel files)

Table S1: The data set used for simulations in this study (Excel).

Table S2: Taxonomic information and maximum length of coral reef fish species recorded in Kimbe Bay, Papua New Guinea (Excel).

Table S3: Mean number of protected individuals for all simulated species and reserve sizes (Excel).

Table S4: Mean percentage of expected maximum protection for all simulated species and reserve sizes (Excel).

Table S5: Summary of univariate linear regressions based on mean home ranges > 200 m fitted to simulation-based predictions of effective reserve sizes. Mean home ranges > 200 m were the single most important predictor across all protection levels ($p < 0.0001$), explaining at least 64% of the variation in simulation-based reserve size predictions (Fig. 3A). Values in brackets specify lower and upper 95% confidence intervals. For species with home ranges ≤ 200 m, reserve sizes of 1 km, 2 km and 5 km can be estimated to protect 50%, 75% and 95% of all individuals, respectively (see Fig. 3A).

Protection level	R^2	Intercept	B (Home range)
50%	0.764	1.305 (0.172; 2.438)	0.914 (0.756; 1.072)
75%	0.794	2.340 (1.340; 3.340)	0.880 (0.740; 1.019)
95%	0.636	4.056 (2.535; 5.577)	0.872 (0.656; 1.089)