# Global drivers and vulnerabilities of coral reef fish functions: Methods

## 1. Underwater visual census database

We used a published global database of reef fish abundance and sizes collected along belt transects1. This database encompasses 9118 transects across 585 sites (98 localities) in the Central Indo-Pacific, Central Pacific, Eastern Pacific, Western Indian, Eastern Atlantic, Western Atlantic. The database only includes sites at the outer reef slope and with a hard reef bottom. Transects were carried out at a constant depth, parallel to the reef crest. We selected the species inside families for which we have body stoichiometric data, that were at least 7cm to minimize the bias related to the identification of small individuals, and finally we discarded rare species, for which less than 20 individuals were ever recorded across all transects. The dataset then included 1110 species that belong to 25 families (Acanthuridae, Balistidae, Bothidae, Chaetodontidae, Cirrhitidae, Fistulariidae, Haemulidae, Holocentridae, Kyphosidae, Labridae, Lethrinidae, Lutjanidae, Monacanthidae, Mugilidae, Mullidae, Ostraciidae, Pempheridae, Pomacanthidae, Pomacentridae, Sciaenidae, Scorpaenidae, Serranidae, Siganidae, Tetraodontidae, Zanclidae).

## 2. Bioenergetic modeling

Here, we focused on 5 key processes mediated by fish: N excretion rate (gN/day/m2), P excretion rate (gP/day/m2), production of body mass through growth (gC/day/m2), herbivory, i.e. ingestion rate of macrophytes (gC/day/m2), and piscivory, i.e. ingestion rate of fishes (g/day/m2). These 5 processes were estimated in each transect using individual-based bioenergetic models that predicts elemental fluxes, including ingestion rate, excretion rates of N and P, and growth rate. The bioenergetic model framework integrates elements of metabolic theory, stoichiometry, and flexible elemental limitation2. We quantified the input parameters, including elements of metabolism, growth, and diet and body stoichiometry, for all 1110 species through the integration of empirical data, data synthesis, and Bayesian phylogenetic models (See supplementary methods). We ran the model for each combination of species identity, body size, and sea surface temperature (n = 30668) to get the contribution of each individual to each process in each transect and the cumulated estimates for the fish community per surface area. Each process is thus expressed in dry mass per day per square meter. We note that N excretion, P excretion, and biomass production include contributions of all fishes, whereas herbivory and piscivory are carried out by a subset of the community, with respect to their trophic guild3. To reduce the occurrence of misclassification of herbivores and piscivores, we categorized a species as a herbivore or piscivore if it had both the highest probability to be classified in that trophic group and this probability was more than 0.5, based on the probability scores of trophic guilds for a global fish species database that defines trophic guilds based on empirical data using a quantitative, unbiased, and fully reproducible framework3. Further, as a comparison, we quantified herbivory and piscivory rates using two alternative trophic guild classifications based on expert opinion3,4 (Extended data figure 5). Both the herbivory and piscivory rates are congruent with the expert opinion trophic guild classifications.

## 3. Relationship between functions and biomass

The standing stock biomass of communities is inevitably related to all functions because of the additive nature of the quantification and general metabolic theory. Furthermore, because of the known relationship between temperature and parameters related to growth and respiration, all functions are also positively correlated with temperature. To model the effect of biomass and sea surface temperature (sst), independent of other factors, we performed a Bayesian mixed effect regression of each log-transformed function for community-level observations ():

We then assessed the covariation between functions, independent of biomass and sst. To do so, we first extracted the median residuals for each function per transect. In some transects, there were no piscivores or herbivores observed. In those cases, we did not include these transects in the analysis. We then quantified the correlations that exist among the different functions using these median residuals. Finally, for the purpose of visualizing the residual variation of functions per locality on a world map, we ran a supplemental model, similar to the model described above but including random effects both per site and locality. We then extracted and plotted the location effects, which can be interpreted as the average variation per locality.

## 4. Effect of community structure on ecosystem functions

To investigate the effect of the community structure while still accounting for the effects of standing biomass and sea surface temperature, we quantified a set of variables that characterize the community. These variables describe the size, age, and trophic distribution of the community, as these may all affect functions2. Specifically, we calculated the 2.5%, 50% and 97.5% quantiles of the total length, immaturity, and trophic level of all individuals per transect. The total length is based on the visual estimation by divers. The immaturity is quantified using the following formula:

where is the species-specific growth rate parameter and is the species-specific asymptotic adult length, and is the total length of individual i. Essentially, this is the derivative of the Von Bertalanffy growth model for a certain length, and the higher this value is, the younger the individual. Finally, the trophic level was extracted from fishbase5. Additionally, we quantified the transect-level species richness. For each log-transformed function we then fitted a Bayesian mixed-effect model with all 12 above-mentioned variables, after verifying that there are no strong correlations between variables (the highest correlation coefficient was 0.5, and 50% of the variable pair correlations varied between -0.1 and 0.2).

To compare effects across functions and assess the relative importance of each variable, we standardized all variables prior to model fitting. We fitted all 5 models by using 4 cores, that each had 2000 iterations with a warm-up of 1000 iterations, and used weakly-informative priors6.

## 5. Species dominance and contributions to functions

We quantified the relative contribution of each species to each function for all transects as followed:

where i is a certain species, j is a transect, F is the value of function f.

Then, we quantified the degree of species dominance per function for each transect. We did this by first ranking species according to their contribution to function, followed by quantifying the cumulative contributions of species to functions. Then, we used the area under the species accumulation curve as a measure for the degree of dominance. Specifically, the degree of dominance (DD) was calculated as followed:

where is the area under the curve, is the theoretical area under the curve where each species has an equal contribution to a certain function, is the theoretical area under the curve where one species performs the entire function. They are quantified as:

where is the contribution of a certain species and R is the number of species contributing to a certain function. The degree of dominance thus ranges between 0 and 1, where 0 means that each species contributes equally and 1 means that a single species performs the entire function. In the case of N excretion, P excretion, and production, R equals the species richness, while for herbivory and piscivory R represents the number of herbivores and piscivores, respectively.

Finally, to know how often species are contributing more than average for a certain function, we quantified the frequency of dominance, i.e. the number of times a species is dominant divided by the total number of transects in which that species is observed. A species is considered dominant for a certain function in a given transect if their contribution is higher than 1/R, i.e. they contribute more than the situation in which each species contributes equally to a certain function.

## 6. Vulnerability to fishing and climate change

For each species, we quantified two measures of vulnerability: vulnerability to climate change and vulnerability to fishing pressure7. For species’ vulnerability to climate change, we solely focus on their vulnerability to the loss of live corals. Vulnerability to climate change induced coral loss is related to diet specialization, habitat specialization for live coral and body size7. Graham et al. (2011)7 developed a score for climate change vulnerability for 134 species. We used these scores to fit a Bayesian mixed effect predictive model that relates the vulnerability with the log-transformed maximum size of fish (extracted from Fishbase 5), the dependence on coral for food (3 categories: not dependent, facultative corallivore, and obligate corallivore), and dependence on coral for habitat (2 categories: dependent vs. not dependent)8,9. We also included a random effect for family. To verify the fit of the model we inspected the posterior predictive plot, which indicated a good fit. Further, the model had a Bayesian R2 of 0.97. We thus used this model to extrapolate the vulnerability measure to all 1110 species in our dataset. For species’ vulnerability to fishing, we extracted the index from Cheung et al. (2005)10. Next, we calculated vulnerability scores per function on the community level by averaging the species-level scores weighted by the contributions to function of species. We also calculated community-level vulnerability scores based on biomass contributions as a comparison. Finally, we calculated the proportions of communities that had a higher vulnerability score of functions, compared to the vulnerability score based on biomass alone. In other words, we quantified the proportions of communities that have an increased functional vulnerability.

## References

1. Barneche, D. R. *et al.* Body size, reef area and temperature predict global reef-fish species richness across spatial scales. *Global Ecology and Biogeography* **28**, 315–327 (2019).

2. Schiettekatte, N. M. D. *et al.* Nutrient limitation, bioenergetics, and stoichiometry: a new model to predict elemental fluxes mediated by fishes. *Functional Ecology* (2020) doi:[10.1111/1365-2435.13618](https://doi.org/10.1111/1365-2435.13618).X

3. Parravicini, V., Casey, J. M., Schiettekatte, N. M. D. & Brandl, S. J. Global gut content data synthesis and phylogeny delineate reef fish trophic guilds. *bioRxiv* 0–3 (2020) doi:[10.1101/2020.03.04.977116](https://doi.org/10.1101/2020.03.04.977116).X

4. Mouillot, D. *et al.* Functional over-redundancy and high functional vulnerability in global fish faunas on tropical reefs. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 13757–62 (2014).

5. Froese, R. & Pauly, D. FishBase. *World Wide Web electronic publication.* (2018).

6. Bürkner, P.-C. brms : An R Package for Bayesian Multilevel Models using Stan. *Journal of Statistical Software* **80**, 1–28 (2017).

7. Graham, N. A. *et al.* Extinction vulnerability of coral reef fishes. *Ecology Letters* **14**, 341–348 (2011).

8. Coker, D. J., Wilson, S. K. & Pratchett, M. S. Importance of live coral habitat for reef fishes. vol. 24 89–126 (2014).

9. Cole, A. J., Pratchett, M. S. & Jones, G. P. Diversity and functional importance of coral-feeding fishes on tropical coral reefs. 286–307 (2008) doi:[10.1111/j.1467-2979.2008.00290.x](https://doi.org/10.1111/j.1467-2979.2008.00290.x).X

10. Cheung, W. W., Pitcher, T. J. & Pauly, D. A fuzzy logic expert system to estimate intrinsic extinction vulnerabilities of marine fishes to fishing. *Biological Conservation* **124**, 97–111 (2005).