**Methods**

1. **Underwater visual census database**

We used a published global database of reef fish abundances and sizes collected along belt transects1. This database encompasses 9118 transects across 585 sites (within 98 localities) in the Central Indo-Pacific, Central Pacific, Eastern Pacific, Western Indian, Eastern Atlantic, and Western Atlantic. Sites are defined as small islands or stretches of continuous reefs in larger coastlines and localities encompass sites that belong to the same biogeographic sub-provinces1. The database only includes transects 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 discarded the species inside families for which we did not have body stoichiometry data, individuals that were smaller than 7cm to minimize the bias related to the identification of small individuals, and rare species for which less than 20 individuals were ever recorded across all transects. The dataset then included 1110 species belonging 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). Sea surface temperature (SST) for each site was obtained from daily time‐series data from the National Oceanicand Atmospheric Administration of the USA (NOAA) covering a 5‐year period (°C; 0.25° resolution)2 (available from <https://www.esrl.noaa.gov/psd/data/gridded/data.noaa.oisst.v2.highres.html>). Further, for each transect, we calculated the species richness and estimated the total standing stock biomass of fishes by using Bayesian length-weight relationships available from Fishbase3. All data wrangling and analyses were performed in the software program R (version 4.0.2; R core team 2020).

2. **Quantification of functions**

For each transect, we estimated five key process-based functions mediated by fishes: nitrogen excretion rate (gN m-2 day -1), phosphorus excretion rate (gP m-2 day -1 ), production of biomass through growth (gCm-2 day -1), herbivory, i.e. ingestion rate of macrophytes (gCm-2 day -1), and piscivory, i.e. ingestion rate of fishes (gm-2 day -1)4. These five functions were estimated for each transect using individual-based bioenergetic models predicting fluxes of carbon (C), nitrogen (N), and phosphorus (P) (e.g. daily C intake rates, N and P excretion rates, growth rates)5. This bioenergetic model framework integrates elements of metabolic theory, stoichiometry, and flexible elemental limitation5. 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 then ran a unique bioenergetic model for each combination of species identity, body size, and sea surface temperature (n = 30668) to get the contribution of each individual to each function in each transect. Finally, we summarized functions on the community level by summing up all individual contributions inside a transect and dividing the sum by the surface area. Each function is thus expressed as dry mass (of C, N, or P) 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 guild as defined by6. 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 framework6. Further, as a comparison, we quantified herbivory and piscivory rates using two alternative trophic guild classifications based on expert opinion6,7 (Fig. S5). Both the herbivory and piscivory rates are congruent with the expert opinion trophic guild classifications. Finally, we estimated multifunction, i.e. one measure that combines all five functions by taking the geometric average of the five functions ( nomalized to range between zero and 100). We used the geometric mean because functions are dependent on each other and may vary considerably.

3. **Community structure variables**

We quantified a set of variables that characterize the fish community structure. These variables describe the size, age, and trophic distribution of the community, as these may all affect functions5. 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 fishbase8.

4. **Multivariate regression models**

We fitted three multivariate Bayesian models with all five functions to (1) predict functions on the locality level, (2) investigate the effects of bioimass and sst, and the correlations among functions independent of biomass and sst, and (3) estimate the effects of the community structure on each function. For each model, functions were log-transformed to ensure the normal distribution of residuals and to allow the allometric relationship with biomass, hypothesized by metabolic theory9.

In the underwater visual transect database, 291 transects (3%) did not contained herbivores and 4467 transects (49%) did not contain piscivores yielding false zeros for herbivory and piscivory, respectively. We considered these zeros as missing values because the absence of the observation of a herbivore or piscivore does not prove the actual absence of herbivores and piscivores. In fact, it is highly unlikely that a coral reef fish community contains no herbivores and piscivores at all. To avoid removing all transects with missing values for herbivory or piscivory (n = 4620) from our database when running multivariate analyses, we imputed the missing values by adding these observations as parameters in the multivariate models.

First, we performed a multivariate intercept-only regression model with the five log-transformed functions to estimate the functions per locality. The model structure includes intercepts and random effects for localities and sites:

where i is the index of the transect, yExN,i is the N excretion rate of transect i, yExP,i is the P excretion rate, yProd,i is the biomass production rate, yHerb,i is the herbivory rate, yExN,i is the piscivory rate, represents the residual error of each function (ExN, ExP, Prod, Herb, and Pisc), R is the correlation matrix of the residuals. Locality- and site-level effects are also structured including covariation among functions. There are thus three correlation matrices in total, or in other words the model will estimate the correlation between functions (independent of biomass and SST) on three levels: locality-level, site-level, and transect-level. We used non-centered parametrisation for site and location effects and all standard deviations had the following prior: . We used a lkj\_corr prior for each of the three correlation matrices ().

Second, we ran a mixed-effect model to simultaneously investigate the effects of biomass and sst on all functions and the correlations among functions (independent of biomass and sst). The standing stock biomass of communities is inevitably positively related to all functions because of the additive nature of the quantification and metabolic theory.9 Further, because of the known relationship between temperature and parameters related to growth and respiration (see suppl. methods), functions are expected to be affected by temperature. We thus fitted a multivariate Bayesian mixed-effect model using transect-level log-transformed functions that included random effects for sites and localities:

,

,

where are the fixed effects of the log-transformed biomass, and are the fixed effects of sst. Locality- and site-level effects are now thus structured including covariation among functions, independent of biomass and sst. Similarly, the residual variation of functions incorporates the correlations between functions, without the effect of biomass and sst. We used similar priors as described above, and we used weakly-informative normal priors for the model slopes (, ).

Finally, to investigate the effect of the community structure while still accounting for the effects of standing biomass and sst, we fitted a mixed effect multivariate model similar to the model specified above, but adding all community structure variables:

where richness is the species richness, is the total length, is the trophic level, is the immaturity, and , , and respectively represent the 50%, 2.5%, and 97.5% quantiles across the fish community. For these models, we used weakly informative priors for the fixed effect parameters () and the same priors as described above for other parameters.

All Bayesian models were fitted using the R package *brms*10 which uses Stan, a C++ package for performing full Bayesian inference11. The posterior distributions of model parameters were estimated using Markov chain Monte Carlo (MCMC) methods by using four chains of 2,000 samples, including 1,000 samples as a warm‐up, meaning that a total of 4,000 samples were used to estimate posterior distributions. The convergence and fit of the models were verified by examining the Rhat, parameter trace plots, and posterior prediction plots (Fig S2).

5. **Species dominance and contributions to functions**

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

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

Then, we quantified the degree of species dominance per function for each site. 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, we quantified the frequency of dominance per species, i.e. the number of sites in which a species is dominant for a given function divided by the total number of sites in which that species is observed. A species is considered dominant for a certain function in a given site 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.

**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. Reynolds, R. W. *et al.* Daily High-Resolution-Blended Analyses for Sea Surface Temperature. J. Climate, 20, 5473-5496. (2007).

3. Froese, R., Thorson, J. T. & Reyes, R. B. A Bayesian approach for estimating length-weight relationships in fishes. *Journal of Applied Ichthyology* **30**, 78–85 (2014).

4. Brandl, S. J. *et al.* Coral reef ecosystem functioning: Eight core processes and the role of biodiversity. *Frontiers in Ecology and the Environment* **17**, 445–454 (2019).

5. Schiettekatte, N. M. D. *et al.* Nutrient limitation, bioenergetics and stoichiometry: A new model to predict elemental fluxes mediated by fishes. *Functional Ecology* **34**, 1857–1869 (2020).

6. 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

7. 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).

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

9. Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M. & West, G. B. Toward a metabolic theory of ecology. *Ecology* **85**, 1771–1789 (2004).

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

11. Carpenter, B. *et al.* Stan : A Probabilistic Programming Language. *Journal of Statistical Software* **76**, 1–31 (2017).