Global drivers and vulnerabilities of coral reef fish functions

Preserving the integrity of ecosystems – their functioning – is a critical challenge of the 21st century. The world's coral reefs are highly diverse, productive, and valuable to millions of people[1], and protecting their ecosystem functioning under the escalating pressures of fishing[2,3] and climate change[5] underpins most of today's coral reef conservation efforts. However, a quantitative assessment of multiple, process-based functions on reefs is lacking. Here, we show that five key functions mediated by fishes — nitrogen and phosphorus cycling, biomass production, herbivory, and piscivory— exhibit critical trade-offs that are driven by diverging community structures and species dominance. Contrary to the concept of maximizing functioning as a single currency, no reef fish community worldwide can simultaneously sustain high values across all functions. Furthermore, functions are locally dominated by few species, but worldwide, 70% of the 1110 species in our dataset contribute disproportionally to functioning in at least one local community. This leads to disparate vulnerabilities of functions to anthropogenic stressors. Our findings reveal a profound challenge in coral reef conservation: there is no clear objective for protecting coral reef functioning and there are no global keystone species to target. Instead, quantitative assessments of multiple functions are necessary to design informed conservation strategies.

The flow of elements through biological communities fuels all life on Earth[6]. Preserving these fluxes, often termed ecosystem functions, is critical for the integrity of ecosystems[6]. For millennia, resources have been managed with an economic mindset to maximize desirable functions such as the production of plant or animal biomass[7]. Sustaining multiple functions likely requires both high species richness and a variety of species assemblages across a landscape[8]. However, in agro-ecological settings, efforts to maximize one function negatively impact another (e.g. timber production vs. erosion control), shedding light on the existence of trade-offs between functions[8,9]. An understanding of such trade-offs is required to make informed management decisions[10], but simultaneously quantifying multiple ecosystem functions is challenging. Therefore, trade-offs between functions, their drivers, and functional vulnerability are poorly understood in many ecosystems[11].

Coral reefs are among the most diverse and productive ecosystems on Earth and provide essential ecosystem services to humanity[1]. As coral reefs thrive in nutrient-poor waters, efficient cycling of elements is key to their high productivity[12]. Yet, the integrity of coral reefs is threatened by a plethora of anthropogenic stressors, such as exploitation and climate change[5]. Over the past decades, severe declines in coral reef habitat quality and fish biomass as well as shifts in community structure have brought coral reef functioning and services to the forefront of scientific discourse[11,13,14]. However, our capacity to quantitatively evaluate, monitor, and compare reef functioning primarily relies on static proxies of functions, such as live coral cover, standing stock biomass of reef fishes, or functional richness based on qualitative species traits[15–17]. Conversely, we know comparatively little about elemental fluxes and their drivers (but see[18]). This constitutes a severe limitation to effective conservation management of coral reefs[11].

Here, we integrate biogeochemistry and community ecology to advance our understanding of the elemental fluxes that underpin reef fish functioning. Using empirically-collected species-specific data on basic organismal processes and Bayesian phylogenetic models to extrapolate to species that lack data, we parameterize individual-level bioenergetic models to estimate five key ecosystem functions: nitrogen (N) excretion, phosphorus (P) excretion, biomass production, herbivory (daily consumption of primary producers, expressed as mass of carbon), and piscivory (daily consumption of fishes, expressed as mass of carbon). We apply these bioenergetic models to all individuals across 9,118 reef fish communities in 585 sites worldwide (Extended Data Table 1) to: (1) quantify community-level reef fish functions and their trade-offs, (2) extract the community- and species-level effects on these functions, and (3) gauge the vulnerability of reef fish functioning in the Anthropocene.

The five key ecosystem functions performed by fishes across the world's reefs exhibit high variability (Fig. 1). Biomass is the most commonly employed indicator of coral reef functioning[11,16,19], and as can be expected from the cumulative nature of community-level ecosystem functioning, we indeed observed a strong relationship between fish standing stock biomass and all five functions (Extended Data Fig. 1a-e, Extended Data Fig. 2). However, our analyses demonstrate that reef fish functions vary remarkably after accounting for biomass, as functions of communities with similar biomass may differ with two orders of magnitude and a two-fold difference in biomass can yield similar levels for

functions (i.e. biomass production, Extended Data Fig. 1a-f). Thus, using biomass as a sole proxy for functioning can mask differences in critical community-level functions. Further, we demonstrate strong trade-offs among the five functions, independent of biomass (Fig.1, Extended Data Fig. 1g). For example, high herbivory rates or nitrogen excretion negatively correlate with rates of phosphorus excretion. As a consequence, for a given value of standing stock biomass, no reef can yield above average values across all five functions. In other words, while many reefs may stand out as hotspots for one function, none can maximize functioning as a holistic, multifaceted concept (Fig. 1).

Community structure and species-specific traits clearly impact rates of functioning. First, using community-level ecological predictors known to affect elemental fluxes[20] (body size, trophic level, species richness, biomass, temperature, and age structure; Fig. 2), we show that correlations between functions are mediated by contrasting aspects of community structure (Fig. 2; Extended Data Table 2; Extended Data Fig. 2). For example, phosphorus excretion is higher in communities with many large-bodied, mature fishes that occupy high trophic levels, thus creating a positive relationship between phosphorus excretion and piscivory (Extended Fig. 1g; See 21). In contrast, biomass production is highest in communities dominated by small and/or immature fishes at low trophic levels, creating a trade-off between biomass production and phosphorus excretion. Metabolic theory predicts that small-bodied individuals have higher mass-specific metabolic rates, leading to elevated consumption rates and disproportional contributions to functions that rely on rapid energetic turnover[12,22,23]. Conversely, fishes in early life stages or with a nutrient-poor diet are often limited by phosphorus[20], resulting in low contributions of these individuals to phosphorus excretion. Thus, due to variations in organismal physiology and life-history traits[18,20], fish community structure can significantly impact ecosystem-wide functioning[24].

Secondly, alongside features of community structure, ecosystem functioning may also be influenced by specific high-performing taxa. Certain species or entire families tend to contribute more or less to a particular function, relative to their contribution to standing biomass or due to specific characteristics (Fig. 3a; Extended Data Fig. 3,4). Further, abundant species can impact rates of functioning at the community level[25,26]. Therefore, we assessed the role of each species by quantifying their relative contributions and the degree to which they disproportionally induce (i.e. dominate) different functions in each community. We

93 show that functions consistently hinge a few dominant species (Fig. 3b). Specifically, more
94 than 50% of a given function, on average, is upheld by only 12% of the species present
95 within a local community. However, the identity of these species varies dramatically (Fig.
96 3c), despite evidence for some species assuming key functional roles across widely-
97 dispersed locations[27]. While few high-performing taxa dominate functioning in each
98 location, there are almost no species that are important across their entire range, and most
99 species are locally important in at least one community. Indeed, 70% of all species
100 contributed disproportionally to a specific function in at least one reef fish community.
101 Despite high species richness on coral reefs, researchers often report the existence of
102 functionally-dominant "key species"[28]. Our results reveal that while functional dominance is
103 indeed prevalent, the identity of local, dominant species vary strongly across different
104 locations, suggesting that maintaining high levels of species richness across coral reef
105 regions is essential to sustain global ecosystem functioning[26].

106 The critical importance of both reef fish community structure and species-specific
107 contributions shines new light on the vulnerability of coral reef functioning in our changing
108 world. Anthropogenic stressors have caused severe changes in reef fish biomass and
109 community structure[5,13,15], and our findings suggest that these changes will have strong
110 effects on ecosystem functioning. For example, intensive fishing and associated reductions
111 in biomass of large fishes truncates the size, age, and trophic structure of fish communities[2].
112 When accounting for the effect of biomass, these effects can enhance nitrogen excretion
113 and production[23], while negatively impacting phosphorus excretion herbivory, and piscivory
114 (Fig. 2). On the other hand, declines in coral cover related to climate change are often
115 associated with a shift toward herbivores, which may help avoid algal domination[3,29].
116 However, herbivores have a minor contribution to phosphorus excretion[20,21], so a shift to
117 herbivore dominance in fish communities and the subsequent decline of community-level
118 phosphorus excretion may change the balance of nutrient cycling on reefs. Higher N:P ratios
119 may favor algal growth[30], and promote algal symbiont dominance within the coral
120 holobiont[31]. Shifts in elemental ratios may be particularly problematic when considered in
121 concert with climate change since phosphorus starvation in corals reduces their
122 temperature threshold for bleaching[32]. Thus, considering multiple functions paints a more

123 nuanced, holistic picture of how human-induced shifts in reef fish community structure
124 impact coral reef ecosystems.

125 Similarly, the species-specific vulnerability of functionally-dominant species heavily affects
126 functioning. By combining species-level vulnerability scores to fishing and climate change
127 induced coral loss[33] and the contributions of each species to each function, we demonstrate
128 that the loss of individuals most vulnerable to fishing will have greatest impacts on piscivory,
129 followed by phosphorus excretion (Fig. 4). Conversely, the loss of individuals due to climate
130 change and consequent coral mortality may disproportionally reduce phosphorus excretion,
131 nitrogen excretion, and biomass production. Fishing and the loss of live coral impact species
132 important for phosphorus excretion. Surprisingly, although fishing pressure can negatively
133 impact large herbivores such as parrotfishes[34], herbivory is the least vulnerable function.
134 This may be due to the high variability in ecosystem roles within the comparatively large
135 pool of herbivorous fish species. While small herbivores are abundant and not particularly
136 vulnerable to fishing, larger herbivorous species are frequently targeted and prone to
137 functional extinction in areas with high fishing pressure[35]. While herbivores of all body sizes
138 and functional groups are combined in our assessment, their realized contributions to
139 herbivory are strongly complementary, and it is important to recognize the specific
140 functional roles performed by small subsets of herbivorous species (e.g. bioerosion[34,36]).
141 Thus, our results reflect the overall vulnerability of functions; yet, this does not consider the
142 distinct roles played by small groups of species nested within a broad function such as
143 herbivory.

144 Conserving biomass, diversity, and ecosystem functioning are important objectives of most
145 current conservation initiatives[16]. Even though safeguarding fish biomass enhances coral
146 reef functioning, the strong trade-offs between key ecosystem functions reveals a critical
147 challenge for coral reef conservation, where actions to enhance one function may
148 negatively impact another. For example, the establishment of marine protected areas,
149 which are one of the primary conservation tactics for coral reefs[37], may protect herbivorous
150 species and thus provide benefits for herbivory. However, marine protected areas do not
151 protect reefs from the pervasive effects of climate change[37], and community shifts towards
152 domination of herbivores may result in the decline of phosphorus excretion. Thus,
153 measuring conservation success with biomass or solely one function (e.g. herbivory) can

mask the collapse of other essential functions. It is necessary to gauge the state of reef ecosystems based on multiple, complementary, process-based functions[11], as well as making informed decisions on local needs and stressors. Finally, while there is a general consensus on the role of diversity in enhancing functioning[8,38], we highlight the overarching importance of community structure and the identity of dominant species at the local scale. While, maintaining the diversity of fishes is critical for coral reef functioning, at local scales, species richness only has a minor impact on individual functions. Rather, dissimilarity between local communities may be the most critical factor in maintaining functioning across the seascape since no species consistently provides high contributions for all functions or across its range[8,39,40].

Overall, through coral reef fishes, we demonstrate that the striking variability in processes that govern the cycling of elements presents a critical, unrecognized challenge for coral reef ecosystem functioning. Management strategies that call for the enhancement of coral reef functioning via an economic mindset, where higher functioning is better, are not feasible. Instead, conserving coral reef ecosystem functioning will require a more nuanced approach that considers processes that vary beyond the effect of standing stock biomass and are subject to variable, local trade-offs, drivers, and anthropogenic threats.

Figures

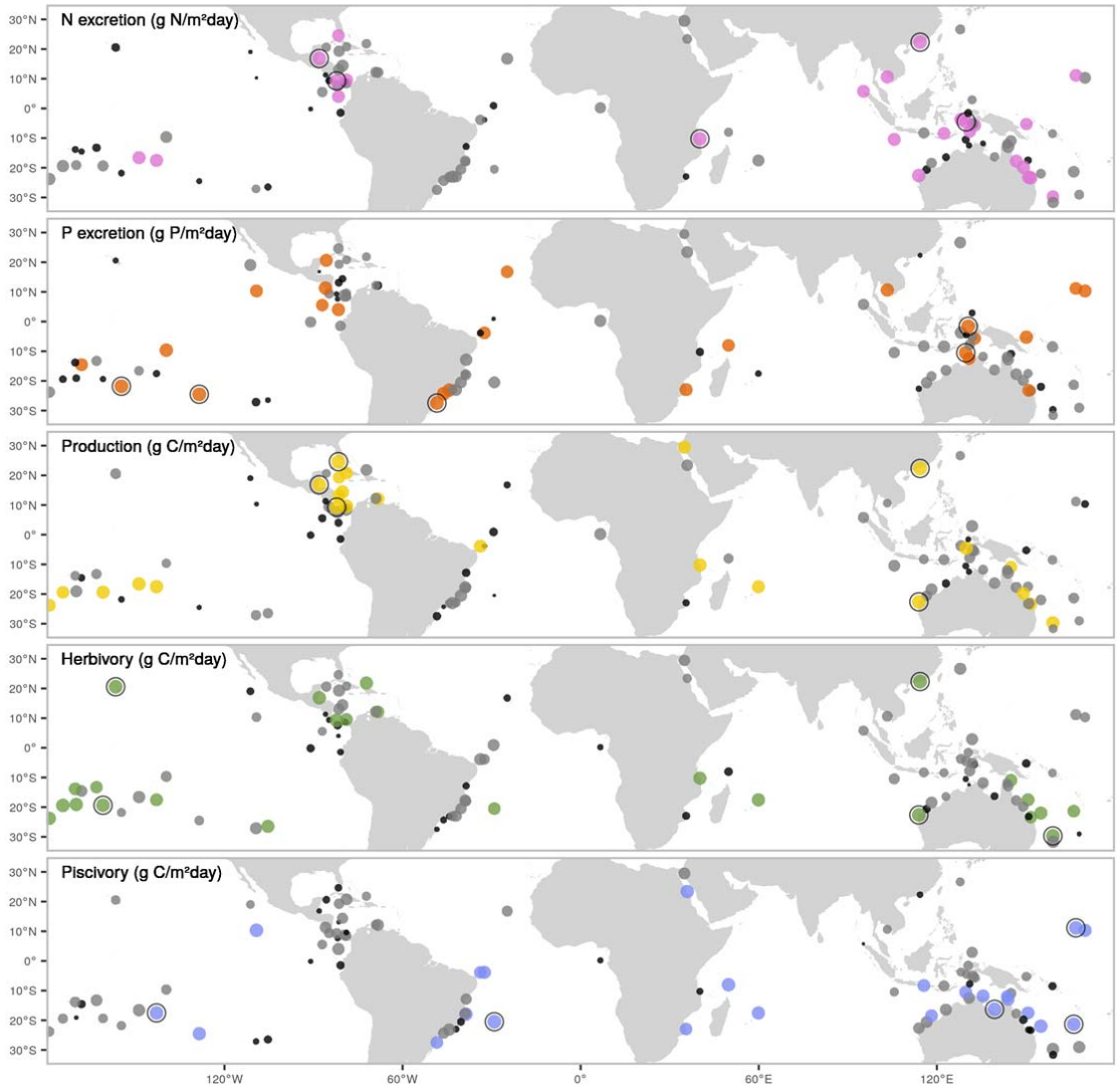


Figure 1: Spatial variation in five key, biomass-corrected ecosystem functions. Dots indicate locations of field surveys, with dot sizes representing the ranked values of biomass-corrected function, and color scales showing categorical assignments (black = lower 25%, grey = 25-75%, color = >75%). Black circular outlines highlight the five locations with the highest values of each biomass-corrected function.

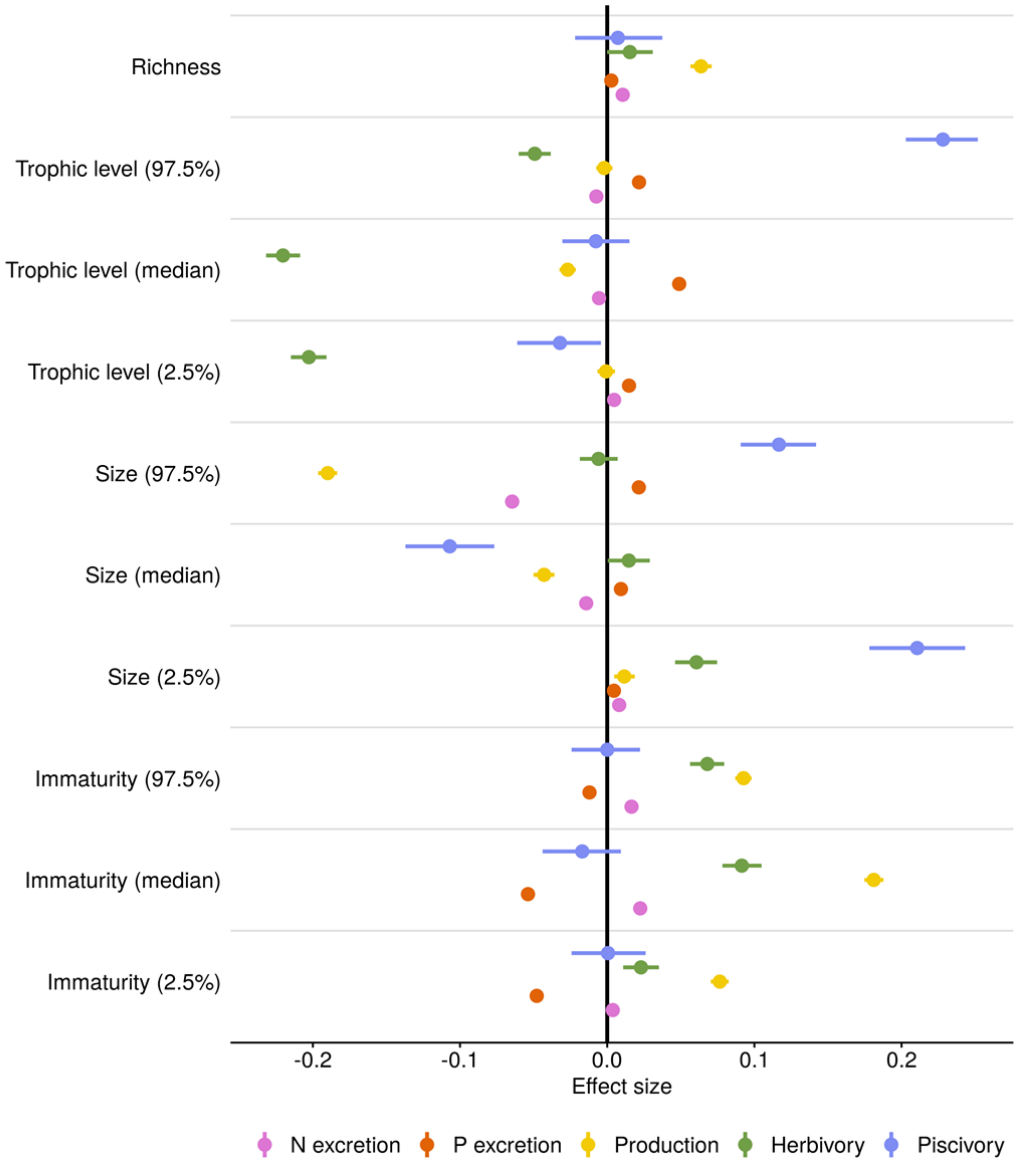


Figure 2: Fixed effect values from Bayesian linear regressions that examine the effects of community variables on five functions. Community variables include species richness, and descriptors of the trophic level, size, and immaturity of fishes. To represent both the median and the spread of trophic level, size, and immaturity across individuals inside a community, we included lower and upper 95% quantile values of these three traits as community variables. All data were log-transformed and standardized to compare across functions and variables (see Extended Data Table 2 for parameter values on non-standardized data). Dots represent the average effect size estimate, and horizontal lines indicate the 95% credible interval.

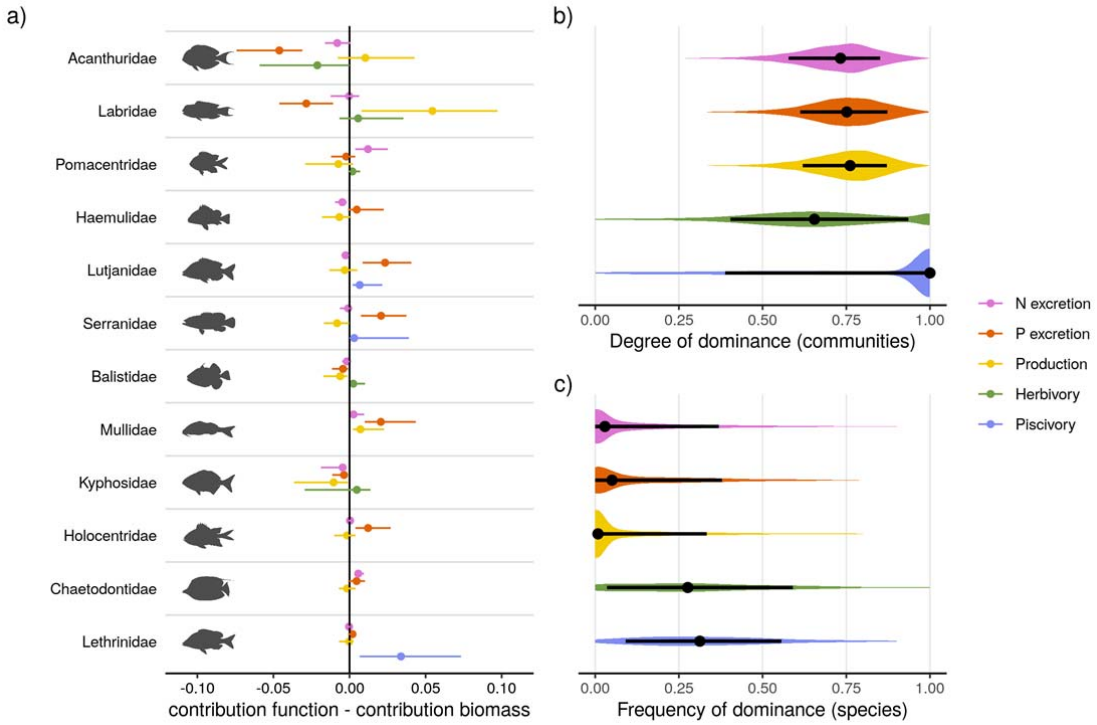
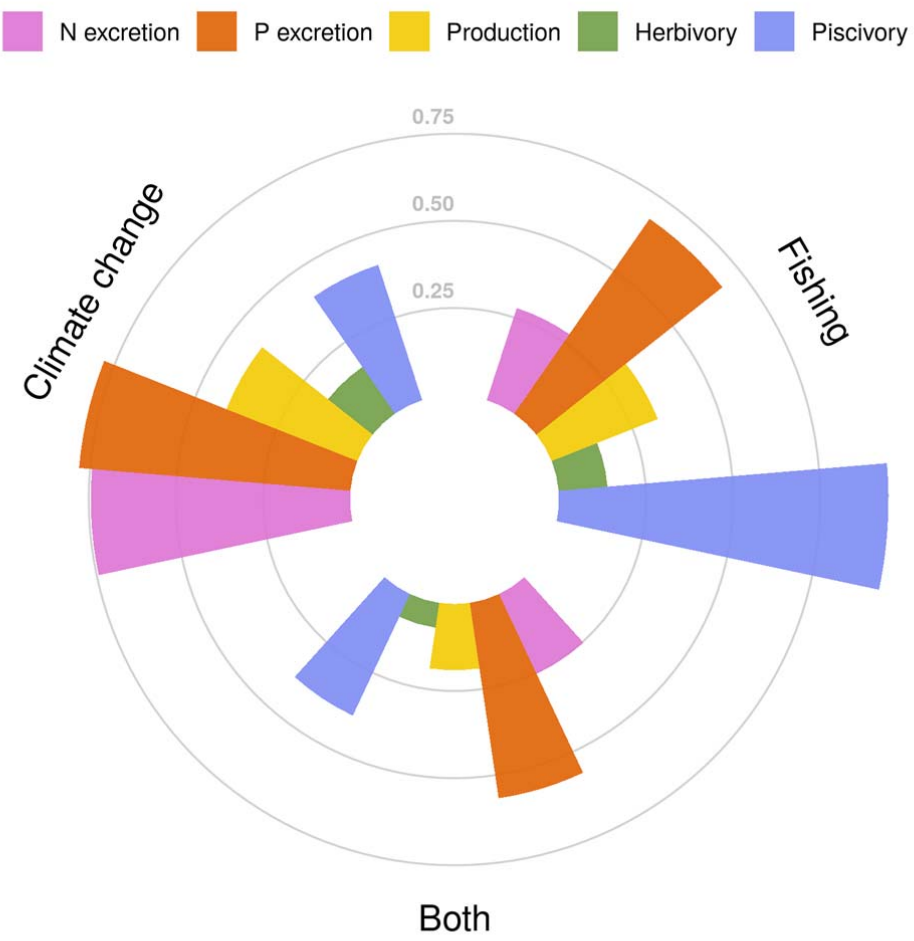


Figure 3: a) The median family-level contributions to each function relative to their contribution to standing stock biomass. The twelve included families are ordered by their median contribution to biomass. b) The distribution of the degree of dominance of communities for each function. Degrees of dominance range between zero (all species contribute equally) and one (a single species is the sole contributor to a given function). c) Species-specific frequencies of dominance in each function across all communities, ranging from zero (species are never dominant) to one (dominant wherever present). A species is categorized as dominant in a community if its contribution to a function is higher than a scenario in which all species are equal (i.e. one divided by the number of species that contribute to the function). Shaded areas show the distribution of the values. Dots represent the median value, and lines indicate the interquartile range.



202

203 Figure 4: Vulnerability of five critical functions to fishing, climate change-induced coral loss,
204 and both stressors combined. Vulnerability is presented as the proportion of communities
205 (filled bars) in which functional vulnerability is higher than vulnerability based on fish
206 biomass (i.e. not accounting for species contributions to each function).

207

References

1. Teh, L. C. L. A. S., Louise S. L. AND Teh. A global estimate of the number of coral reef fishers. *PLOS ONE* **8**, 1–10 (2013).
2. Graham, N. A. *et al.* Human Disruption of Coral Reef Trophic Structure. *Current Biology* **27**, 231–236 (2017).
3. Ceccarelli, D. M., Emslie, M. J. & Richards, Z. T. Post-Disturbance Stability of Fish Assemblages Measured at Coarse Taxonomic Resolution Masks Change at Finer Scales. *PLOS ONE* **11**, e0156232 (2016).
4. Stuart-Smith, R. D., Brown, C. J., Ceccarelli, D. M. & Edgar, G. J. Ecosystem restructuring along the Great Barrier Reef following mass coral bleaching. *Nature* **560**, 92–96 (2018).
5. Hughes, T. P. *et al.* Global warming and recurrent mass bleaching of corals. *Nature* **543**, 373–377 (2017).
6. Welti, N. *et al.* Bridging food webs, ecosystem metabolism, and biogeochemistry using ecological stoichiometry theory. *Frontiers in Microbiology* **8**, 1298 (2017).
7. Weisser, W. W. *et al.* Biodiversity effects on ecosystem functioning in a 15-year grassland experiment: Patterns, mechanisms, and open questions. *Basic and Applied Ecology* **23**, 1–73 (2017).
8. Zavaleta, E. S., Pasari, J. R., Hulvey, K. B. & Tilman, G. D. Sustaining multiple ecosystem functions in grassland communities requires higher biodiversity. *Proceedings of the National Academy of Sciences* **107**, 1443–1446 (2010).
9. Tallis, H., Kareiva, P., Marvier, M. & Chang, A. An ecosystem services framework to support both practical conservation and economic development. vol. 105 9457–9464 (2008).
10. Rodríguez, J. P. *et al.* Trade-offs across space, time, and ecosystem services. *Ecology and Society* **11**, (2006).
11. 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).
12. Brandl, S. J. *et al.* Demographic dynamics of the smallest marine vertebrates fuel coral reef ecosystem functioning. *Science* **364**, 1189–1192 (2019).
13. Williams, G. J. *et al.* Coral reef ecology in the Anthropocene. *Functional Ecology* **33**, 1014–1022 (2019).
14. Bellwood, D. R., Streit, R. P., Brandl, S. J. & Tebbett, S. B. The meaning of the term 'function' in ecology: A coral reef perspective. vol. 33 948–961 (2019).
15. Mora, C. *et al.* Global Human Footprint on the Linkage between Biodiversity and Ecosystem Functioning in Reef Fishes. *PLoS Biology* **9**, e1000606 (2011).
16. Cinner, J. E. *et al.* Meeting fisheries, ecosystem function, and biodiversity goals in a human-dominated world. *Science* **368**, 307–311 (2020).
17. 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).
18. Allgeier, J. E., Valdivia, A., Cox, C. & Layman, C. A. Fishing down nutrients on coral reefs. *Nature Communications* **7**, (2016).
19. MacNeil, M. A. *et al.* Recovery potential of the world's coral reef fishes. *Nature* **520**, 341–344 (2015).

20. 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).
21. Allgeier, J. E., Layman, C. A., Mumby, P. J. & Rosemond, A. D. Consistent nutrient storage and supply mediated by diverse fish communities in coral reef ecosystems. *Global Change Biology* **20**, 2459–2472 (2014).
22. Barneche, D. R. & Allen, A. P. The energetics of fish growth and how it constrains food-web trophic structure. *Ecology Letters* **21**, 836–844 (2018).
23. Morais, R. A., Connolly, S. R. & Bellwood, D. R. Human exploitation shapes productivity–biomass relationships on coral reefs. *Global Change Biology* **26**, 1295–1305 (2020).
24. Schramski, J. R., Dell, A. I., Grady, J. M., Sibly, R. M. & Brown, J. H. Metabolic theory predicts whole-ecosystem properties. *Proceedings of the National Academy of Sciences* **112**, 2617–2622 (2015).
25. Ruttenberg, B. I., Adam, T. C., Duran, A. & Burkepile, D. E. Identity of coral reef herbivores drives variation in ecological processes over multiple spatial scales. *Ecological Applications* **29**, e01893 (2019).
26. Topor, Z. M., Rasher, D. B., Duffy, J. E. & Brandl, S. J. Marine protected areas enhance coral reef functioning by promoting fish biodiversity. *Conservation Letters* **12**, e12638 (2019).
27. Bellwood, D. R., Hoey, A. S. & Choat, J. H. Limited functional redundancy in high diversity systems: Resilience and ecosystem function on coral reefs. *Ecology Letters* **6**, 281–285 (2003).
28. Bellwood, D. R., Hughes, T. P. & Hoey, A. S. Sleeping Functional Group Drives Coral-Reef Recovery. *Current Biology* **16**, 2434–2439 (2006).
29. Graham, N. A. *et al.* Dynamic fragility of oceanic coral reef ecosystems. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 8425–8429 (2006).
30. Burkepile, D. E. *et al.* Nutrient supply from fishes facilitates macroalgae and suppresses corals in a Caribbean coral reef ecosystem. *Scientific Reports* **3**, 1493 (2013).
31. Allgeier, J. E. *et al.* Rewiring coral: Anthropogenic nutrients shift diverse coral–symbiont nutrient and carbon interactions toward symbiotic algal dominance. *Global Change Biology* **26**, 5588–5601 (2020).
32. Ezzat, L., Maguer, J. F., Grover, R. & Ferrier-Pagès, C. Limited phosphorus availability is the Achilles heel of tropical reef corals in a warming ocean. *Scientific Reports* **6**, 31768 (2016).
33. Graham, N. A. *et al.* Extinction vulnerability of coral reef fishes. *Ecology Letters* **14**, 341–348 (2011).
34. Bellwood, D. R., Hoey, A. S. & Hughes, T. P. Human activity selectively impacts the ecosystem roles of parrotfishes on coral reefs. *Proceedings of the Royal Society B: Biological Sciences* **279**, 1621–1629 (2012).
35. Bellwood, D. R., Hoey, A. S. & Hughes, T. P. Human activity selectively impacts the ecosystem roles of parrotfishes on coral reefs. *Proceedings of the Royal Society B: Biological Sciences* **279**, 1621–1629 (2012).
36. Green, A. & D.R., B. *Monitoring functional groups of herbivorous reef fishes as indicators of coral reef resilience – A practical guide for coral reef managers in the Asia Pacific region.* 70 (IUCN working group on Climate Change; Coral Reefs. IUCN, Gland, Switzerland., 2009).
37. Graham, N. A. *et al.* Changing role of coral reef marine reserves in a warming climate. *Nature Communications* **11**, 1–8 (2020).

299 38. Hooper, D. U. *et al.* Effects of biodiversity on ecosystem functioning: A consensus of
300 current knowledge. *Ecological Monographs* **75**, 3–35 (2005).
301 39. Duffy, J. E. Why biodiversity is important to the functioning of real-world ecosystems.
302 *Frontiers in Ecology and the Environment* **7**, 437–444 (2009).
303 40. Lefcheck, J. S. *et al.* Tropical fish diversity enhances coral reef functioning across multiple
304 scales. *Science Advances* **5**, (2019).

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 transects[1]. 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/m^2), P excretion rate (gP/day/m^2), production of body mass through growth (gC/day/m^2), herbivory, i.e. ingestion rate of macrophytes (gC/day/m^2), and piscivory, i.e. ingestion rate of fishes (g/day/m^2). 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 limitation[2]. 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 guild[3]. 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 framework[3]. Further, as a comparison, we quantified herbivory and piscivory rates using two alternative trophic guild classifications based on expert opinion[3,4] (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 (y_j):

$$y_j \sim N(\mu_j, \sigma_j),$$

$$\mu_j = \beta_0 + \beta_1 x_{log(biomass),j} + \beta_2 x_{sst,j}$$

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 functions[2]. 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:

$$immaturity_i = \kappa(l_\infty - l_i),$$

where κ is the species-specific growth rate parameter and l_∞ is the species-specific asymptotic adult length, and l_i 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 fishbase[5]. 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).

$$y_j \sim N(\mu_j, \sigma_j),$$

$$\mu_j = \beta_0 + \beta_1 x_{log(biomass),j} + \beta_2 x_{sst,j} + \beta_3 x_{richness,j} + \beta_4 x_{size_m,j} + \beta_5 x_{size_{2.5},j} + \beta_6 x_{size_{97.5},j} + \beta_7 x_{troph_m,j} + \beta_8 x_{troph_{2.5},j} + \beta_9 x_{troph_{97.5},j} + \beta_{10} x_{imm_m,j} + \beta_{11} x_{imm_{2.5},j} + \beta_{12} x_{imm_{97.5},j}$$

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 priors[6].

5. Species dominance and contributions to functions

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

$$contribution_{f,i,j} = \frac{\sum F_{f,i,j}}{\sum F_{f,j}},$$

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:

$$DD = \frac{A - A_{min}}{A_{max} - A_{min}},$$

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

$$A_{min} = \frac{R^2 - 1}{2R},$$

$$A_{max} = R - 1,$$

$$A = \sum_{i=2}^{R} \frac{C_i + C_{i-1}}{2},$$

where C_i 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 pressure[7]. 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 size[7]. 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.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.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.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).

Global drivers and vulnerabilities of coral reef fish functions: Supplementary methods

To apply the bioenergetic model that estimates fluxes of carbon (C), nitrogen (N), and phosphorus (P), a number of parameters are required[1]. Here, we describe how these parameters were quantified for all 1110 species in our database, with a combination of literature, empirical measures, and Bayesian models. All analyzes were carried out in R v.3.6.3 and Bayesian modes were run using Stan[2] and the R package brms[3].

1. Growth parameters

1.1 Data compilation

We first compiled maximum lengths for all species with Fishbase[4] and used these lengths for the l_∞ . For κ, we used a standardized coefficient that describes the potential growth trajectory of an individual if l_∞ were to be equal to its maximum length[5]. $t0$ was kept constant at 0 for all species.

We extracted the data for k_{max} from Morais et al. (2018)[5] and filtered out only the species of our species list. As the Lenth-Frequency method consistently overestimates kmax, we omitted the k_{max} estimates coming from this method. In total, this selection process resulted in 439 observations of kmax for different species and temperatures.

Further, we collected additional otolith data, including measurements of fishes from five Polynesian islands. We collected data across four archipelagos, including six distinct islands: Mo'orea and Manuae (Society Islands), Hao and Mataiva (Tuamotus), Mangareva (Gambiers), and Nuku Hiva (Marquesas) between 2014 and 2018. All fishes were collected in the lagoon and/or outer slope, depending on the accessibility of the respective habitats.

For each species, otoliths were cut transversely, using a diamond disc saw (Presi Mecatome T210) to obtain a section of 500 µm. Sections were then fixed on a glass side with thermoplastic glue (Crystalbond TM). Small otoliths were directly embedded in the thermoplastic glue and polished until obtaining a transversal section. Otoliths were sanded with abrasive discs of decreasing grain size (2,400 and 1,200 grains cm- 130 2) and polished with a 0.25 µm diamond suspension in order to be closest to the nucleus. All sections were

photographed under a Leica DM750 light microscope with a Leica ICC50 HD microscope camera and LAS software (Leica Microsystems).

A standardized transect across the otoliths (from the nucleus to the edge) was chosen for each species, and distances between annual growth increments were measured using the software ImageJ. This procedure was performed twice by two independent researchers to prevent biases induced by a single observer. When the coefficient of variation between the two observers was greater than 5%, a common reading was reached by averaging the measurements for each section.

We then used the Modified Fry back-calculation model (MF)[6] to estimate fish length at previous ages, modified to also investigate the uncertainty around the obtained length estimates using a Bayesian approach (R package fishgrowbot).

Finally, we fitted the Von Bertalanffy growth models to all species at each location for which there were at least 3 individuals. We fitted the models using Bayesian hierarchical regression models provided by the R package fishgrowbot.

After combining the two data sources, we obtained 496 estimates of k_{max} for 181 species.

1.2 Data analysis and extrapolation

Aside from phylogeny, k_{max} is mostly determined by body size and temperature[5].
We applied a Bayesian hierarchical model to predict the growth rate of fishes as a function of body size, temperature and phylogeny:

$$lnkmax = (\beta_0 + \gamma_{0phy}) + \beta_1 lnsizemax + \beta_2 sst + \epsilon,$$

where $lnkmax$ represents the natural log-transformed kmax value, β_0 is the fixed-effect intercept, γ_{0phy} is the vector of random-effect coefficients that account for the residual intercept variation, based on the relatedness as described by the phylogeny, β_1 is the slope for the natural transformed maximum body size, β_2 is the slope for the average ambient sea surface temperature, ϵ is the residual variation. We used uninformative priors and ran the model for 2000 iterations with a warm-up of 1000 iteration for 4 chains. The model fit confirmed a negative relationship of $lnkmax$ with $lnsize$, and a positive relationship with sea surface temperature. The Bayesian R2 of the model was 0.738 (95%CI: 0.702-0.769). The

phylogenetic heritability (equivalent to Pagel'λ) was estimated as the proportion of total variance, conditioned on the effects, attributable to the phylogeny(i.e. $\lambda = \frac{sd(\gamma_{0phy})^2}{sd(\gamma_{0phy})^2+\epsilon^2}$). This calculation resulted in a phylogenetic signal of 0.74 (95% CI: 0.70 - 0.77).

We extrapolated $kmax$ for all species across the full temperature range in which those species occur in the database, with temperature rounded to the °C, which results in 4712 unique temperature and species combinations.

There is currently no streamlined method to make predictions for new species from a phylogenetic regression model. We circumvented the issue by extracting draws of the phylogenetic effect, γ_{0phy} for each species included in the model. We subsequently predicted these phylogenetic effects for missing species with the help of the function phyEstimate in the picante package for R[7]. This function uses phylogenetic ancestral state estimation to infer trait values for new species on a phylogenetic tree by rerooting the tree to the parent edge for the node to be predicted[8]. We repeated this for all 100 trees and 1000 draws. Per draw, we averaged the extrapolated values per species for the hundred trees. Then, by combining the predicted phylogenetic effects with the global intercept and slopes for body size and temperatures for each draw, we predicted κ for each species. We only use one chain in order to keep computational time reasonable. Finally, we summarised all κ predictions per sst per species by taking the mean and standard deviation across the 1000 draws.

2 Body stoichiometry

2.1 Data collection

1633 individuals of 108 species and 25 families were collected between 2015 and 2017 in Mo'orea, the Caribbean, and Palmyra. Their gut contents were removed, and the whole body was freeze-dried and ground to powder with a Precellys homogenizer. Q_k (%) were then measured in the lab using standard methods. Ground samples were analysed for %C and %N content using a CHN Carlo-Erba elemental analyzer (NA1500) for %P using dry oxidation-acid hydrolysis extraction followed by a colorimetric analysis[9]. Elemental content was calculated based on dry mass.

2.2 Data analysis and extrapolation

The CNP% content of organisms is known to be highly conserved within families[10]. We therefore use phylogeny to extrapolate these values. We fitted C, N and P contents (%) through a hierarchical phylogenetic multivariate normal model with phylogenetic effects and random effects per species.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} \sim MVNormal\left(\begin{bmatrix} mu1 \\ mu2 \\ mu3 \end{bmatrix}, S\right),$$

$$mu_{n\times k} = \beta_{0k} + \gamma_{0phy\times k} + \gamma_{0sp\times k},$$

where Y_1, Y_2 and Y_3 are the % content of C, N, and P respectively, $mu_{n\times k}$ represents the average % content of element k (C, N, and P) per species, β_{0k} is the fixed-effect intercept for each element k, $\gamma_{0phy\times k}$ is the matrix of random-effect coefficients that account for the intercept variation, based on the relatedness as described by the phylogeny per element k, $\gamma_{0sp\times k}$ is the matrix of random-effect coefficients that account for the residual species-level intercept variation per element k.

We used uninformative priors and ran the model for 2000 iterations with a warm-up of 1000 iteration for 4 chains. The Bayesian R2 of the model was 0.39 (95%CI: 0.36-0.42), 0.50 (95%CI: 0.48-0.53), and 0.43 (95%CI: 0.40-0.46) for C, N and P respectively. The phylogenetic heritability was 0.41 (95%CI: 0.28-0.55), 0.58 (95%CI: 0.4-0.66), and 0.57 (95%CI: 0.46-0.69) for C, N, and P respectively.

As before, we used 1000 fitted draws for each species, and 100 phylogenetic trees to extrapolate to all species with unknown body stoichiometry. Specifically, we used the phylopars function from the Rphylopars package[11]. This function uses ancectral state reconstruction and brownian motion, and takes the correlation between C, N and P into account.

3 Diet

3.1 Data collection

We collected 571 adult individuals of 51 species between 2018 and 2019 in Mo'orea and Tetiaroa, and Mangareva, three Polynesian islands. We extracted the stomach content and stored it in a 2ml tube. After freezing the samples, we dry-froze all samples for at least 24 hours, and ground to powder. Then, samples were sent to the lab for CNP content analysis using similar methods as for the fish body stoichiometry.

3.2 Data analysis and extrapolation

We used trophic guilds defined by 12. We fitted a multivariate Bayesian regression model to summarize CNP% content data per trophic guild with random effects at the species level. This model had a median Bayesian R2 of 0.62, 0.62, and 0.48 for C, N and P respectively. Next, we extracted 1000 draws of the predicted the CNP% per trophic guild. Parravicini et al. 2020[12] provides the probability of reef fish species to be assigned to each of the eight defined trophic guilds(i.e. sessile invertivores; herbivores, microvores, and detrivores; corallivores; piscivores; microinvertivores; macroinvertivores; crustacivores; planktivores). By combining these probabilities with the predicted diet contents per trophic guild, we finally estimated the diet CNP% for each species in our database. We then took the average and standard deviation across all 1000 draws. While we recognize the bias of using diet CNP% estimates of a dataset in one region, we argue that variability between food categories e.g. animal material and primary producers is likely to be higher than regional differences within trophic categorizations. Further, as the used trophic guild classification includes probabilities to belong to each group, variation is included when the trophic categorization is not well known. For example, if a species has a 50% probability to be a herbivore and a 50% probability to be a sessile invertivore this uncertainty will be reflected the estimation of the diet CNP%.

4 Metabolic parameters

4.1 Data collection

In the period between 2018 and 2019, we collected 1393 individuals of 61 species and 18 families with a minimum of 3 replicates per species. Individuals were collected using handnets and clove oil by scuba divers.

4.2 Metabolic rate

To quantify standard metabolic rate (SMR) and maximum metabolic rate (MMR), we conducted intermittent-closed respirometry experiments at 28°C[13,14]. After an acclimatization and fasting period of 48 h in aquaria, the fish were individually transferred to a water-filled tub at 28°C and manually chased by the experimenter until exhausted[15,16]. Then, they were placed in respirometry chambers submersed in an ambient and temperature-controlled tank, where they were left for ~23 h. The intermittent respirometry cycles started immediately after a fish was placed in its respirometry chamber. The cycles consisted of a measurement (sealed) period followed by a flush period during which the respirometry chambers were flushed with fully aerated water from the ambient tank. Because fish were exhausted right before entering the respirometry chambers, it is possible to measure the approximate MMR. Depending on fish size, 8 respirometry chambers ranging in volume (including tubes and pumps) from 0.4 to 4.4 L were run in parallel, and measurement and flush periods lasted between 3 to 15 min and 3 to 5 min, respectively. SMR was calculated as the average of the 10 % lowest values measured during the entire period, after the removal of outliers[17]. MMR was calculated from the slope of the first measurement period.

4.3 Data analysis and extrapolation

To retrieve the parameters f_0 (Metabolic normalisation constant independent of body mass; $gCg^{-\alpha}d^{-1}$) and α (mass-scaling exponent), and θ (factorial activity scope), we fitted a Bayesian mixed effect model predicting the log10-transformed metabolic rate with the log10-transformed biomass including random effects of family, species, and mr type (SMR or MMR) on both the intercept and the species. We ran the model for 4000 iterations, with a warm-up of 2000 iterations. Further, we used an informative prior for the slope

($\alpha \sim normal(0.8,0.5)$. The model had a Bayesian R2 of 0.973 (95%CI: 0.972-0.974). We then extracted the family-level α by summing the slope of the model with the effects of the family on the slope of the SMR. We did this for 1000 iterations and then took the mean and standard deviation. In a similar way we extracted the family-level intercept for SMR, and then quantified mean and standard deviation of f_0 after the back-transformation of 1000 iterations of the intercept. Finally, θ was quantified as followed, based on the assumption that fishes rest 12h a day and they on average spend the remaining 12 hours at a metabolic rate that is the average of their SMR and MMR:

$$\theta = \frac{3SMR + MMR}{4SMR},$$

where 1000 iterations of the back-transformed family-level intercepts were used for SMR and MMR. We then summarized these predictions by taking the mean and standard deviation. We used the family-level estimates for these three parameters for all species in our database. For families that were not represented in our respirometry dataset, we used an average across all families.

5. Additional parameters

We retrieved the parameters lw_a, lw_b, h, and r from fishbase[4]. For the mass-specific turnover rates for N and P(F_{0Nz}; F_{0Pz}), we used the estimates provided in Schiettekatte et al. (2020)[1].

References

1. 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).
2. Carpenter, B. *et al.* Stan : A Probabilistic Programming Language. *Journal of Statistical Software* **76**, 1–31 (2017).
3. Bürkner, P.-C. brms : An R Package for Bayesian Multilevel Models using Stan. *Journal of Statistical Software* **80**, 1–28 (2017).
4. Froese, R. & Pauly, D. FishBase. *World Wide Web electronic publication.* (2018).
5. Morais, R. A. & Bellwood, D. R. Global drivers of reef fish growth. *Fish and Fisheries* **19**, 874–889 (2018).
6. Vigliola, L., Harmelin-Vivien, M. & Meekan, M. G. Comparison of techniques of back-calculation of growth and settlement marks from the otoliths of three species of Diplodus from the Mediterranean Sea. *Canadian Journal of Fisheries and Aquatic Sciences* **57**, 1291–1299 (2000).
7. Kembel, S. W. *et al.* Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**, 1463–4 (2010).
8. Kembel, S. W., Wu, M., Eisen, J. A. & Green, J. L. Incorporating 16S Gene Copy Number Information Improves Estimates of Microbial Diversity and Abundance. *PLoS Computational Biology* **8**, e1002743 (2012).
9. Allen, S. E., Grimshaw, H. M., Parkinson, J. A. & Quarmby, C. *Chemical analysis of ecological materials,* 565 (Blackwell Scientific Publications, 1974).
10. Allgeier, J., Wenger, S. & Layman, C. Taxonomic identity best explains variation in body nutrient stoichiometry in a diverse marine animal community. *Scientific reports* **10**, (2020).
11. Bruggeman, J., Heringa, J. & Brandt, B. W. PhyloPars: Estimation of missing parameter values using phylogeny. *Nucleic Acids Research* **37**, W179–W184 (2009).
12. 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.X
13. Steffensen, J. F. Some errors in respirometry of aquatic breathers: How to avoid and correct for them. *Fish Physiology and Biochemistry* **6**, 49–59 (1989).
14. Clark, T. D., Sandblom, E. & Jutfelt, F. Aerobic scope measurements of fishes in an era of climate change: Respirometry, relevance and recommendations. vol. 216 2771–2782 (2013).
15. Norin, T. & Malte, H. Repeatability of standard metabolic rate, active metabolic rate and aerobic scope in young brown trout during a period of moderate food availability. *Journal of Experimental Biology* **214**, 1668–1675 (2011).
16. Clark, T. D. *et al.* Physiological benefits of being small in a changing world: Responses of coho salmon (Oncorhynchus kisutch) to an acute thermal challenge and a simulated capture event. *PLoS ONE* **7**, 1–8 (2012).
17. Chabot, D., Steffensen, J. F. & Farrell, A. P. The determination of standard metabolic rate in fishes. *Journal of Fish Biology* **88**, 81–121 (2016).

Global drivers and vulnerabilities of coral reef fish functions: Extended data figures and tables

Tables

Extended Data table 2: Overview of localities of UVC transects, used in this study, including number of sites and number of transects

bioregion	locality	n_sites	n_transects
c_indopacific	aceh	4	50
c_indopacific	ambon	1	10
c_indopacific	bali	1	18
c_indopacific	cambodia	1	5
c_indopacific	christmas_island	2	17
c_indopacific	dampier_archipelago	3	44
c_indopacific	darwin_(nt)	1	18
c_indopacific	flores	3	13
c_indopacific	hong_kong_island	1	12
c_indopacific	kai_ketjil	2	8
c_indopacific	kimberley	1	11
c_indopacific	mornington_island	3	16
c_indopacific	ningaloo_marine_park	8	250
c_indopacific	northern_territory_(other)	11	76
c_indopacific	north_west_shelf	26	284
c_indopacific	offshore_shoals	3	11
c_indopacific	okinawa	1	8
c_indopacific	palau	1	25
c_indopacific	papua_new_guinea	1	6
c_indopacific	pualu_kaimeer	1	4
c_indopacific	pulau_jamdena	2	8
c_indopacific	pulau_naira	1	12
c_indopacific	raja_ampat	15	259
c_indopacific	solomon	47	322
c_pacific	ailuk_atoll	3	14

bioregion	locality	n_sites	n_transects
c_pacific	austral_islands	1	12
c_pacific	capricorn_group	6	64
c_pacific	central_coral_sea	22	233
c_pacific	central_gbr	20	145
c_pacific	cook_islands	3	14
c_pacific	cook_islands_sp	1	24
c_pacific	elizabeth_and_middleton_reefs	2	66
c_pacific	fiji	3	316
c_pacific	french_polynesia	18	226
c_pacific	hawaii	12	521
c_pacific	lord_howe_island	2	530
c_pacific	marquesas_islands	2	6
c_pacific	minerva_reefs	2	17
c_pacific	new_caledonia	12	884
c_pacific	niue	1	8
c_pacific	norfolk_island	2	38
c_pacific	northern_coral_sea	10	157
c_pacific	northern_gbr	9	97
c_pacific	pitcairn	4	254
c_pacific	queensland_(other)	10	66
c_pacific	rapa_nui	4	65
c_pacific	rongalap_atoll	2	9
c_pacific	rose_atoll	1	16
c_pacific	salaz_y_gomez	1	63
c_pacific	samoa	8	358
c_pacific	society_islands	5	21
c_pacific	southern_coral_sea	7	109
c_pacific	southern_gbr	1	25
c_pacific	tonga	9	293
c_pacific	whitsundays	1	8

bioregion	locality	n_sites	n_transects
e_atlantic	cverde	1	97
e_atlantic	stome	5	38
e_pacific	clipperton	1	80
e_pacific	cocos	1	178
e_pacific	coiba	15	188
e_pacific	costa_rica	5	48
e_pacific	galapagos	13	139
e_pacific	las_perlas	6	47
e_pacific	machalilla	3	19
e_pacific	malpelo	1	70
e_pacific	nicaragua_tep	5	58
e_pacific	panama_pacific	1	6
e_pacific	revillagigedo	3	116
w_atlantic	abrolhos	1	91
w_atlantic	arraial	1	347
w_atlantic	belize	2	37
w_atlantic	bocas_del_toro	3	30
w_atlantic	bonaire	3	14
w_atlantic	cuba	1	3
w_atlantic	curacao	4	117
w_atlantic	florida_keys	4	33
w_atlantic	grand_cayman	1	3
w_atlantic	guarapari	2	114
w_atlantic	ilha_gde	5	25
w_atlantic	l_santos	1	57
w_atlantic	mexico_caribbean	2	31
w_atlantic	neb	3	22
w_atlantic	noronha	1	61
w_atlantic	rio_de_janeiro	1	2
w_atlantic	rocas	1	51

bioregion	locality	n_sites	n_transects
w_atlantic	salvador_bts	2	49
w_atlantic	san_blas	1	13
w_atlantic	santa_catarina	6	253
w_atlantic	seaflower_marine_reserve	3	47
w_atlantic	southwestern_caribbean	2	6
w_atlantic	stpauls_rocks	1	27
w_atlantic	trindade	2	238
w_atlantic	turks_and_caicos_islands	1	4
w_indian	eilat	1	5
w_indian	mozambique	7	30
w_indian	red_sea	1	5
w_indian	seychelles	6	165
w_indian	tanzania	1	8

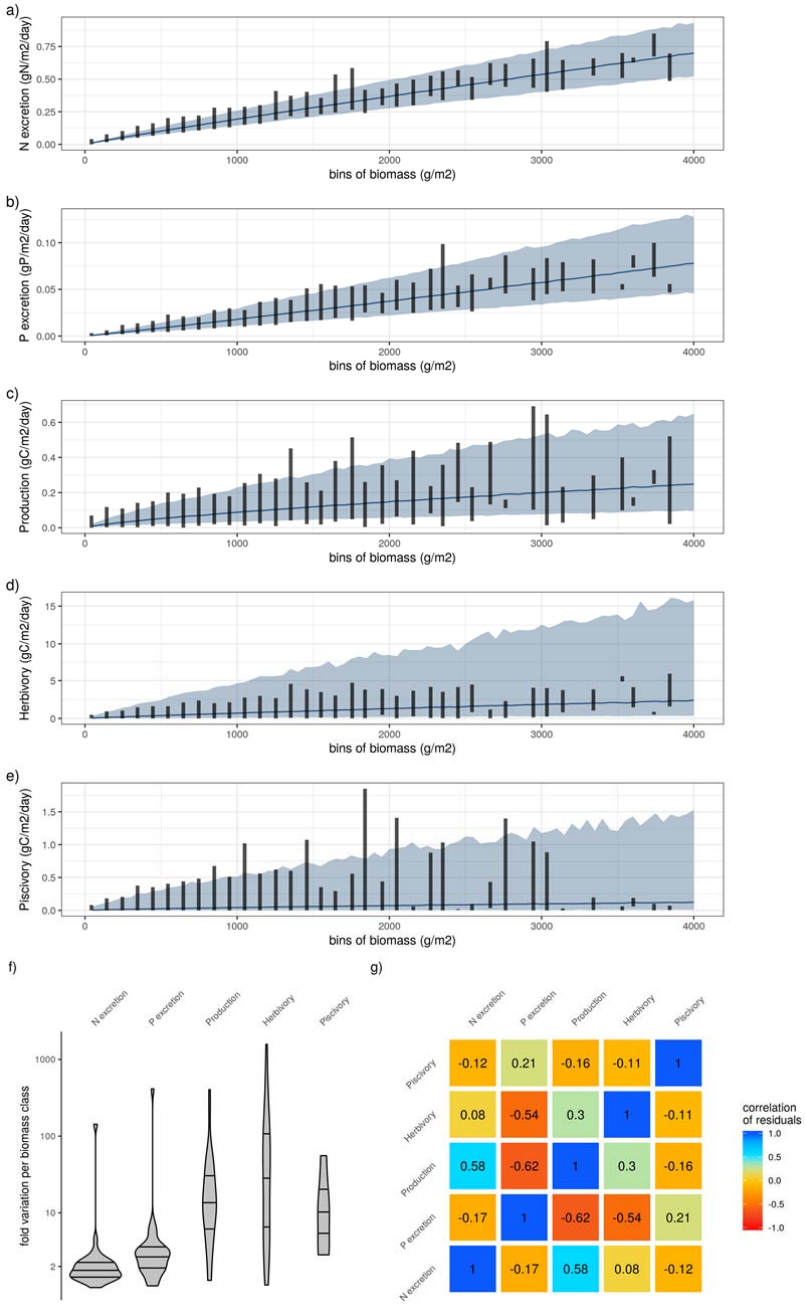
Extended Data table 2: Overview of parameters of the regressions relating the five functions to the community structure variables

response	term	estimate	std.error	lower	upper
log(N excretion)	intercept	-8.7502	0.0264	-8.7927	-8.7061
	sst	0.0290	0.0007	0.0279	0.0301
	log(biomass)	0.9664	0.0013	0.9643	0.9686
	richness	0.0010	0.0001	0.0008	0.0012
	size (mean)	-0.0040	0.0004	-0.0046	-0.0034
	trophic level (mean)	-0.0145	0.0028	-0.0190	-0.0098
	immaturity (mean)	0.0185	0.0010	0.0169	0.0201
	size (97.5)	-0.0087	0.0002	-0.0089	-0.0084
	trophic level (97.5%)	-0.0385	0.0052	-0.0471	-0.0300
	immaturity (2.5%)	0.0055	0.0016	0.0027	0.0081
	immaturity (97.5%)	0.0099	0.0006	0.0089	0.0110
	trophic level (2.5%)	0.0199	0.0048	0.0121	0.0278
	size (2.5%)	0.0034	0.0006	0.0025	0.0044
log(P excretion)	intercept	-12.7600	0.0454	-12.8354	-12.6861
	sst	0.0265	0.0011	0.0246	0.0284
	log(biomass)	1.0130	0.0023	1.0092	1.0167
	richness	0.0003	0.0002	-0.0001	0.0007
	size (mean)	0.0031	0.0006	0.0020	0.0041
	trophic level (mean)	0.1468	0.0046	0.1393	0.1543
	immaturity (mean)	-0.0510	0.0016	-0.0536	-0.0482
	size (97.5)	0.0033	0.0003	0.0029	0.0038
	trophic level (97.5%)	0.1285	0.0093	0.1129	0.1437
	immaturity (2.5%)	-0.0810	0.0028	-0.0857	-0.0765
	immaturity (97.5%)	-0.0083	0.0011	-0.0100	-0.0065
	trophic level (2.5%)	0.0716	0.0081	0.0583	0.0858
	size (2.5%)	0.0022	0.0010	0.0006	0.0037
log(Production)	intercept	-9.2949	0.0646	-9.4005	-9.1877
	sst	0.0371	0.0016	0.0344	0.0398
	log(biomass)	0.8809	0.0033	0.8755	0.8865
	richness	0.0053	0.0003	0.0047	0.0058

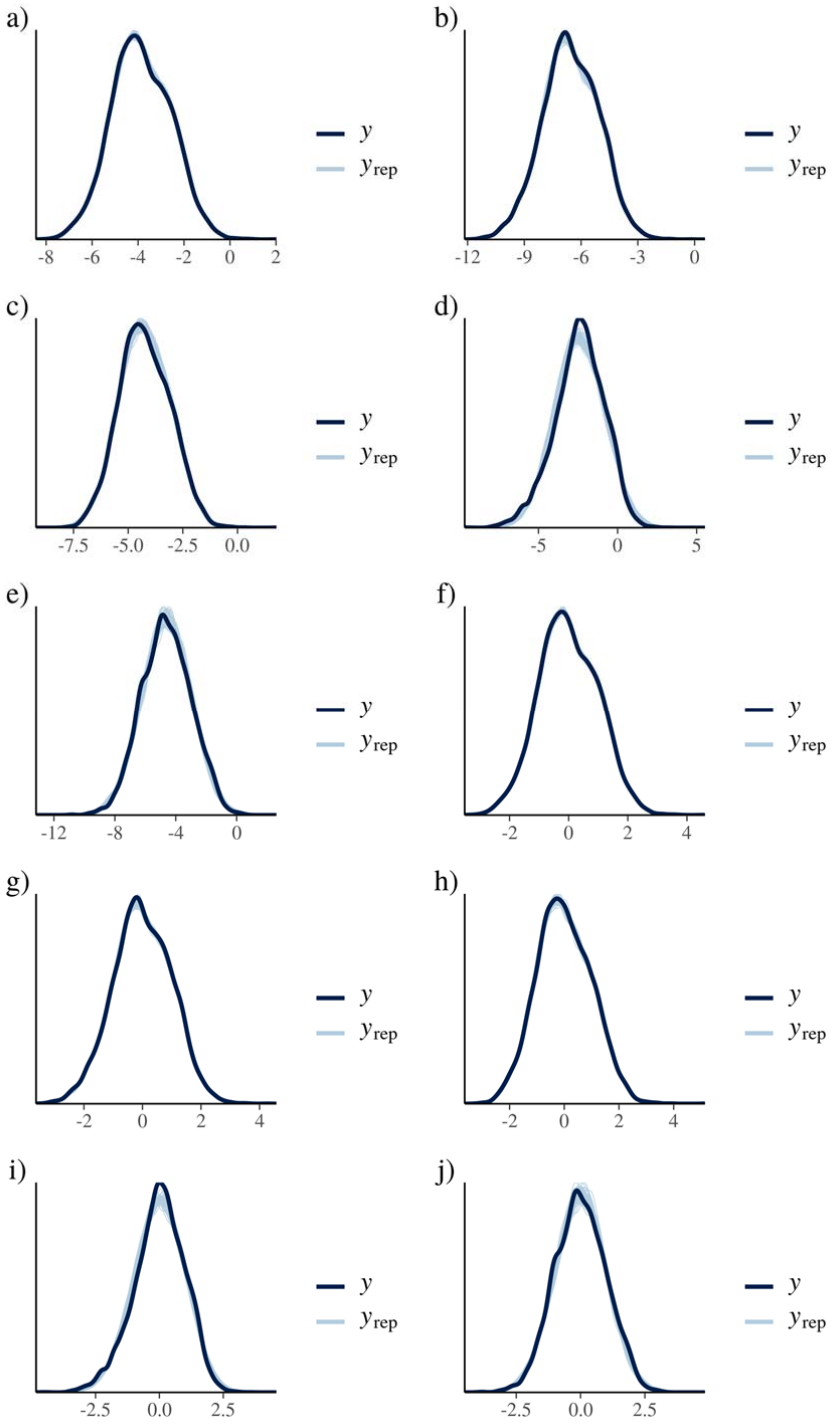
response	term	estimate	std.error	lower	upper
	size (mean)	-0.0109	0.0009	-0.0124	-0.0094
	trophic level (mean)	-0.0632	0.0064	-0.0737	-0.0528
	immaturity (mean)	0.1344	0.0024	0.1304	0.1385
	size (97.5)	-0.0230	0.0004	-0.0236	-0.0223
	trophic level (97.5%)	-0.0100	0.0131	-0.0320	0.0116
	immaturity (2.5%)	0.1014	0.0041	0.0946	0.1083
	immaturity (97.5%)	0.0501	0.0015	0.0476	0.0526
	trophic level (2.5%)	-0.0031	0.0116	-0.0214	0.0159
	size (2.5%)	0.0044	0.0013	0.0022	0.0066
log(Herbivory)	intercept	-4.3397	0.1756	-4.6184	-4.0594
	sst	0.0895	0.0045	0.0821	0.0969
	log(biomass)	0.9258	0.0091	0.9110	0.9407
	richness	0.0017	0.0009	0.0003	0.0031
	size (mean)	0.0050	0.0025	0.0009	0.0092
	trophic level (mean)	-0.7042	0.0174	-0.7335	-0.6762
	immaturity (mean)	0.0922	0.0067	0.0813	0.1032
	size (97.5)	-0.0009	0.0011	-0.0027	0.0009
	trophic level (97.5%)	-0.3129	0.0377	-0.3757	-0.2511
	immaturity (2.5%)	0.0416	0.0115	0.0223	0.0605
	immaturity (97.5%)	0.0502	0.0043	0.0433	0.0574
	trophic level (2.5%)	-1.1401	0.0342	-1.1950	-1.0836
	size (2.5%)	0.0321	0.0038	0.0258	0.0384
log(Piscivory)	intercept	-13.0583	0.4678	-13.8237	-12.3086
	sst	-0.0807	0.0104	-0.0980	-0.0635
	log(biomass)	0.7567	0.0192	0.7253	0.7885
	richness	0.0007	0.0016	-0.0019	0.0035
	size (mean)	-0.0334	0.0051	-0.0416	-0.0250
	trophic level (mean)	-0.0240	0.0376	-0.0867	0.0377
	immaturity (mean)	-0.0178	0.0149	-0.0420	0.0062
	size (97.5)	0.0194	0.0022	0.0158	0.0231
	trophic level (97.5%)	1.5393	0.0839	1.4012	1.6814

response	term	estimate	std.error	lower	upper
	immaturity (2.5%)	-0.0003	0.0260	-0.0434	0.0428
	immaturity (97.5%)	-0.0001	0.0093	-0.0155	0.0150
	trophic level (2.5%)	-0.1561	0.0700	-0.2705	-0.0398
	size (2.5%)	0.1072	0.0083	0.0938	0.1210

Figures

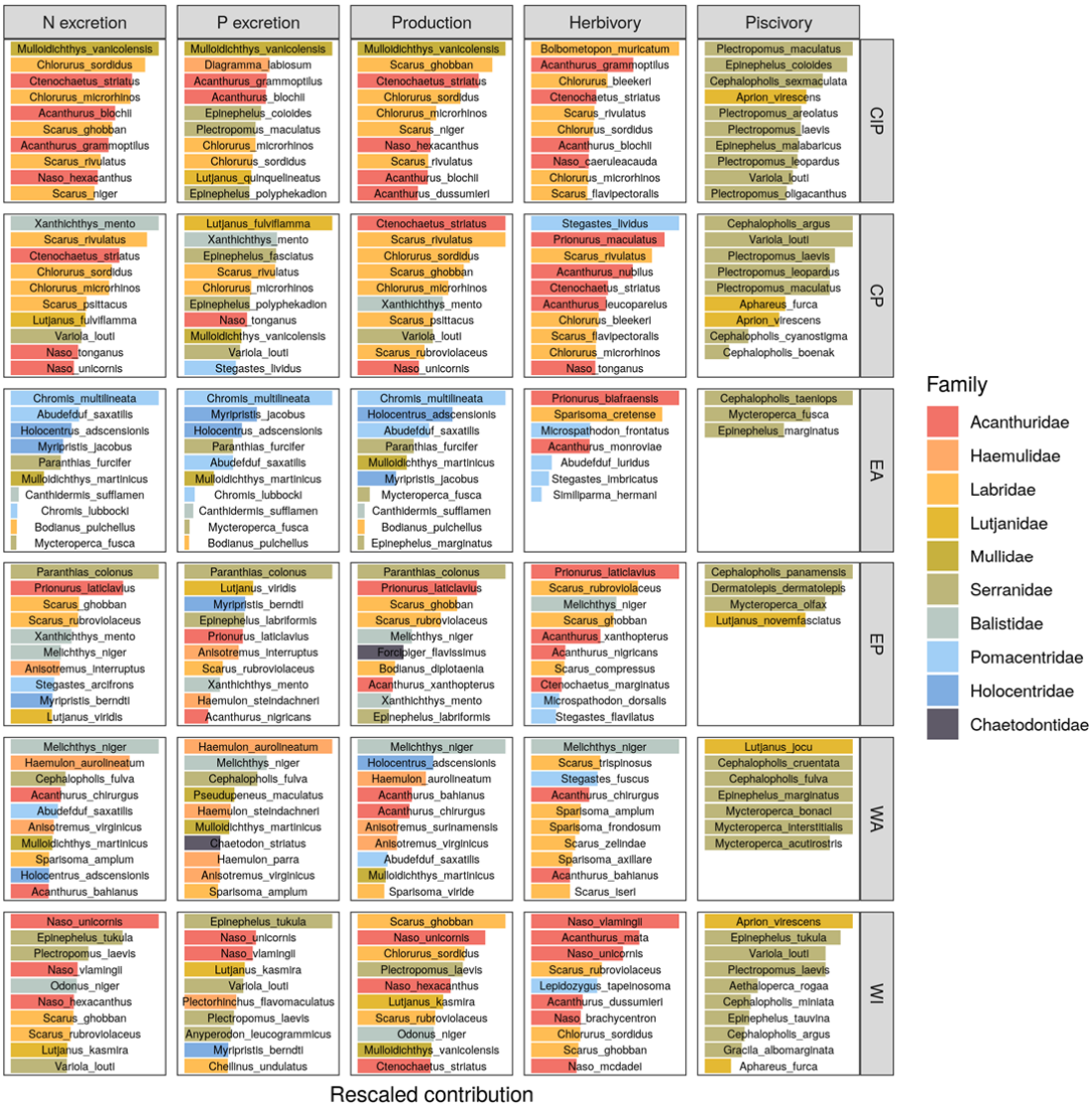


Extended Data Figure 1: a-e) Relationship between biomass and the five functions. Lines and shaded areas show the average and 95% credible interval of the predicted functions respectively, for a constant sea surface temperature of 26°C (the average across all sites). Vertical lines show the range of the estimated functions across fish communities per biomass class of 100g/m2. f) Fold variation of each function per biomass class of 100g/m2 across fish communities. g) Correlation matrix of the residuals of the five functions after regression with biomass and sea surface temperature. Standard deviations of correlation coefficients did not exceed 0.01.

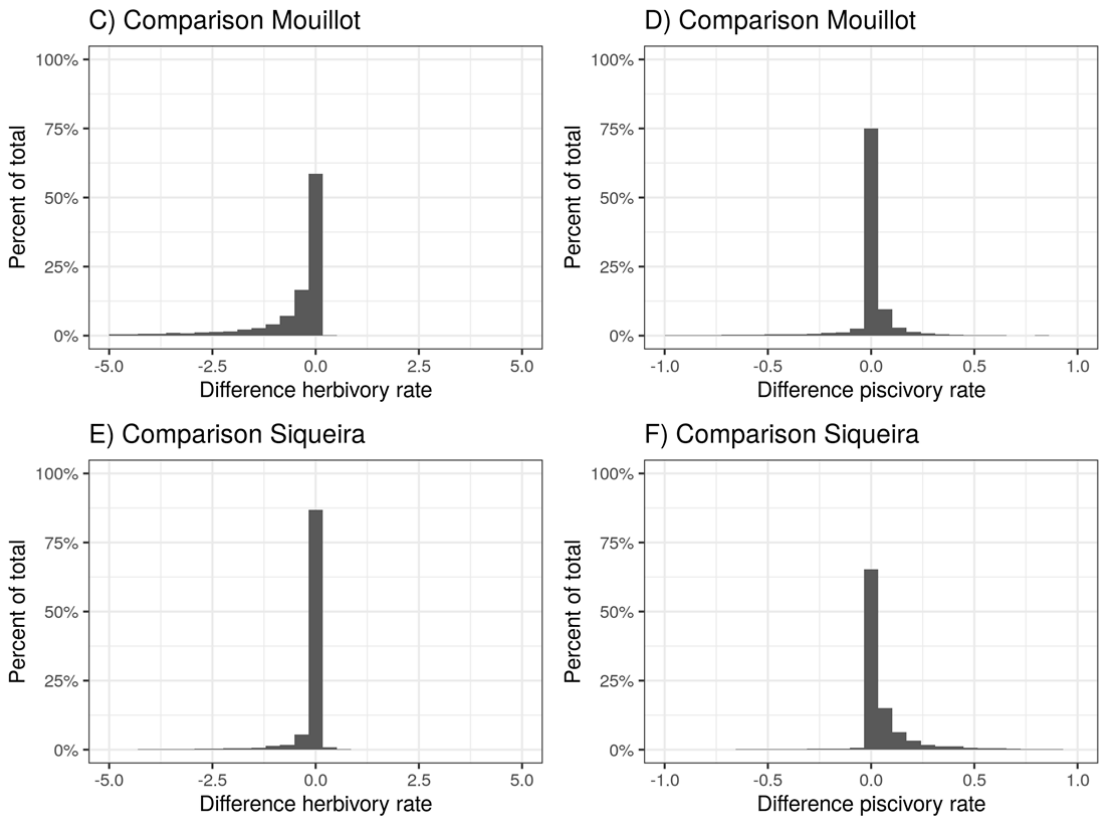


Extended Data Figure 2: Posterior predictive checks of the five models relating functions with biomass and sea surface temperature only (a) N excretion, b) P excretion, c) Production, d) Herbivory, e) Piscivory), and the five models relating functions with community variables (f) N excretion, g) P excretion, h) Production, i) Herbivory, j) Piscivory)

Extended Data Figure 3: Average relative contribution of fish families to all five functions per biogeographical ocean basin. CIP = Central-Indo-Pacific, CP = Central Pacific, EA = Eastern Atlantic, WA = Western Atlantic, WI = Western Indian



Extended Data Figure 4: Average relative contribution of the top ten most contributing species to all five functions per biogeographical ocean basin. CIP = Central-Indo-Pacific, CP = Central Pacific, EA = Eastern Atlantic, WA = Western Atlantic, WI = Western Indian



Extended Data Figure 5: Comparison herbivory and piscivory rates when using alternative diet classifications from Mouillot et al. (2014) and Siqueira et al. 2020