# 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 required1. 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 Stan2 and the R package brms3.

## 1. Growth parameters

### 1.1 Data compilation

We first compiled maximum lengths for all species with Fishbase4 and used these lengths for the . For , we used a standardized coefficient that describes the potential growth trajectory of an individual if were to be equal to its maximum length5. was kept constant at 0 for all species.

We extracted the data for 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 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 for 181 species.

### 1.2 Data analysis and extrapolation

Aside from phylogeny, is mostly determined by body size and temperature5.

We applied a Bayesian hierarchical model to predict the growth rate of fishes as a function of body size, temperature and phylogeny:

where represents the natural log-transformed kmax value, is the fixed-effect intercept, is the vector of random-effect coefficients that account for the residual intercept variation, based on the relatedness as described by the phylogeny, is the slope for the natural transformed maximum body size, is the slope for the average ambient sea surface temperature, 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 with , 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’) was estimated as the proportion of total variance, conditioned on the effects, attributable to the phylogeny(i.e. ). This calculation resulted in a phylogenetic signal of 0.74 (95% CI: 0.70 - 0.77).

We extrapolated 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, 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 R7. 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 predicted8. 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 for each species. We only use one chain in order to keep computational time reasonable. Finally, we summarised all 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. (%) 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 analysis9. 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 families10. 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.

where , and are the % content of , , and respectively, represents the average % content of element (, , and ) per species, is the fixed-effect intercept for each element , 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, 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 package11. 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. 202012 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°C13,14. 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 exhausted15,16. 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 outliers17. MMR was calculated from the slope of the first measurement period.

### 4.3 Data analysis and extrapolation

To retrieve the parameters (Metabolic normalisation constant independent of body mass; ) and (mass-scaling exponent), and (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 (. The model had a Bayesian R2 of 0.973 (95%CI: 0.972-0.974). We then extracted the family-level 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 after the back-transformation of 1000 iterations of the intercept. Finally, 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:

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 , , , and from fishbase4. For the mass-specific turnover rates for N and P(; ), 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 <i>Diplodus</i> 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](https://doi.org/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).