



RESEARCH PAPER

Global Ecology
and BiogeographyA Journal of
Macroecology

WILEY

Global patterns and predictors of trophic position, body size and jaw size in fishes

R. Keller Kopf^{1,2} | Jian D. L. Yen³ | Dale G. Nimmo² | Sébastien Brosse⁴ | Sébastien Villéger⁵

¹Research Institute for the Environment and Livelihoods, Charles Darwin University, Darwin, Northern Territory, Australia

²Institute for Land Water & Society, Charles Sturt University, Albury, New South Wales, Australia

³School of BioSciences, The University of Melbourne, Parkville, Victoria, Australia

⁴Laboratoire Évolution & Diversité Biologique (EDB UMR5174), Université Paul Sabatier - Toulouse 3, CNRS, IRD, UPS, Toulouse Cedex, France

⁵MARBEQ, University Montpellier, CNRS, IFREMER, IRD, Montpellier, France

Correspondence

R. Keller Kopf, Research Institute for the Environment and Livelihoods, Charles Darwin University, Darwin, Northern Territory 0909, Australia.
Email: keller.kopf@cdu.edu.au

Funding information

Australian Research Council

Editor: Derek Tittensor

Abstract

Aim: The aim of this study was to test whether maximum body mass and jaw length are reliable predictors of trophic position (*TP*) in fishes, and to compare linear and non-linear machine-learning (ML) models incorporating biogeography, habitat and other morphological traits.

Location: Global.

Time period: Modern.

Major taxa studied: Fishes.

Methods: We compiled a global database of *TP* (2.0–4.5), maximum body mass, jaw length, order, ecoregion, habitat and other morphological traits of freshwater, estuarine and diadromous fishes ($n = 1,991$). We used Bayesian linear mixed effects and ML, with r^2 analogues and 10-fold cross-validation, to explain and predict *TP*.

Results: Random forest models outperformed Bayesian models in all comparisons. Jaw length was the most influential predictor of *TP*, but was weakly associated with body mass except in five orders of largely piscivorous fishes. Trophic position did not scale positively with body mass in global ecoregions, riverine fishes, or in 29/30 orders, but scaled positively in lacustrine fishes and Perciformes. Significant negative *TP*–body mass scaling was observed in Characiformes. Best models explained 55% of the global variation in *TP*, but over-estimated the position of herbivores-detritivores, and under-estimated the position of top predators.

Main conclusions: Our study provides support for jaw length as an important mechanism constraining *TP* in one of the world's largest groups of vertebrates. Jaw length and body mass were weakly correlated, and therefore body size was not a strong predictor of *TP*. The diversification of large-bodied herbivores-detritivores and omnivores in freshwater ecosystems, coupled with small predators in species-rich orders (e.g., Cypriniformes, Characiformes) in temperate and tropical rivers explains why *TP* globally shows a weak relationship with body size. Our model validation results underscore the importance of not assuming that explanatory power extends to predictive capacity in macroecology and machine-learning models.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. Global Ecology and Biogeography published by John Wiley & Sons Ltd

KEYWORDS

allometric trophic network models, allometry, body mass, gape limitation, machine learning, predator–prey, random forest, trophic network theory

1 | INTRODUCTION

Body size is a driver of trophic structure and function because it influences energy demand (Rip & McCann, 2011) and predator–prey interactions (Brose, 2010; Emmerson & Raffaelli, 2004). Food webs are considered to be size-structured, whereby predators often consume smaller-bodied prey items (Brose, 2010; Brose et al., 2006). Important exceptions have been described (e.g., parasites; scavengers; hunting in packs), but size-based trophic constraints provide vital parameters that underpin contemporary macroecology (Sibly et al., 2012), fisheries (Andersen, 2019; Blanchard et al., 2017) and food web (Brose et al., 2017) concepts and models. There is an expectation that organisms feed at a high trophic position when larger owing to assumed positive allometric scaling of the morphological and physiological traits (e.g., speed; strength; jaw size) needed to catch, kill and consume large prey, and inherently that larger prey feed at a higher trophic position than smaller prey. As Elton (1927) first pointed out “spiders do not catch elephants in their webs, nor do water scorpions prey on geese”, yet complex food webs are not always well described by simple body size-based rules (Jonsson et al., 2018).

Predators are often larger than their prey (Brose, 2010; Brose et al., 2006), but empirical evidence of positive trophic position (*TP*)–body size scaling has been inconsistent. Results have ranged from no significant relationships between *TP* and body size (Layman et al., 2005; Ou et al., 2017; Tucker & Rogers, 2014), to strong positive (Cohen et al., 2003; Robinson & Baum, 2015; Romero-Romero et al., 2016), hump-shaped (Segura et al., 2015) and even negative scaling patterns (Burruss et al., 2016). Differences in the slope of the relationship have been reported between migratory and non-migratory species (Bloom et al., 2018), invertebrates and vertebrates (Riede et al., 2011), and between species or size-based groupings (Jennings et al., 2001). Different slopes have been observed among terrestrial, marine and freshwater ecosystems (Potapov et al., 2019), whereby positive *TP*–body size relationships have been strongest in marine communities (e.g., Robinson & Baum, 2015; Romero-Romero et al., 2016). Correlations have been weaker or absent in freshwater and terrestrial environments (Keppeler et al., 2020; Potapov et al., 2019; Riede et al., 2011).

Freshwater fishes are useful subjects to test global *TP*–body size scaling hypotheses because these species range in maximum mass from less than 1 to over 3,200,000 g (Froese & Pauly, 2019), have evolved remarkable functional diversity (Toussaint et al., 2016) and occupy trophic niches from algivores and detritivores to top predators (Winemiller, 1990). With over 14,953 fish species inhabiting freshwater systems world-wide (Tedesco et al., 2017), this group represents about 1/4 of extant vertebrates. Studies of freshwater lakes have provided strong evidence of positive *TP*–size

scaling relationships in fish and other consumers (Cohen et al., 2003; Nakazawa et al., 2010). Nevertheless, highly diverse fish communities in tropical rivers of South America and Southeast Asia (Layman et al., 2005; Ou et al., 2017) have reported both non-significant and positive *TP*–body size scaling relationships, largely dependent on the trophic guild (e.g., carnivore versus non-carnivore; Keppeler et al., 2020) and taxonomic order. A global analysis by Romanuk et al. (2011) concluded that *TP* of fishes scaled positively with body mass, but did not test for potential differences between marine and freshwater ecosystems, and their analysis excluded herbivores and detritivores. Large-bodied herbivores and detritivores that feed at low trophic levels are relatively common in river ecosystems (Ou et al., 2017; Winemiller, 1990), which are fuelled by a complex mix of phytoplankton, vascular plants, terrestrial seeds and detritus. Conversely, pelagic systems and lakes are supported mainly by unicellular algae, so food webs are usually longer and strongly size-structured (Keppeler et al., 2020; Potapov et al., 2019). With the exception of large filter feeders, large-bodied animals that feed at the bottom of the food web may be less common in lake or pelagic ecosystems, compared to rivers, due to the energetic and physical handling limitations associated with large animals feeding on microscopic algae. Given the wide range of variability within and among previous studies, and because herbivores and detritivores are a globally diverse and functionally important group of consumers in freshwater food webs (Ou et al., 2017; Winemiller, 1990), it remains uncertain at a global scale how, or if, *TP* scales positively with body size, or other traits of freshwater fishes.

Adaptations associated with jaw, mouth and skull morphology linked to feeding have been vital to the diversification of freshwater fishes and to the vertebrate transition to land (Sallan & Friedman, 2011; Westneat, 2004). In fishes, reptiles and amphibians that often swallow prey whole, mouth gape size is considered to be a fundamental morphological constraint leading consumers to select smaller prey (Arim et al., 2007; Shine, 1991; Wainwright & Barton, 1995). If larger prey items feed at a higher trophic level than smaller prey items, then the *TP* of gape-limited consumers should increase with jaw length or mouth gape size across species and ontogeny (Arim et al., 2007; Mihalitsis & Bellwood, 2017; Mittlebach & Persson, 1998; Wainwright & Barton, 1995). It follows that if jaw length increases positively with body mass (Wainwright & Barton, 1995) then it is logical to assume that *TP* also scales positively with body mass.

The aim of this study was to test whether maximum body mass and jaw length are reliable predictors of *TP* in freshwater fishes globally, and to compare linear and nonlinear machine-learning (ML) models incorporating biogeography, taxonomic order, habitat and morphological traits. ML methods (Ryo & Rillig, 2017) are increasingly used in ecology due to assumed higher accuracy and better ability to predict outcomes of multiple nonlinear interactions

when compared to other statistical methods such as general linear or mixed-effects models (GLMs; GLMMs). Here, we compared the performance of Bayesian linear mixed effects and ML methods in predicting the *TP* of freshwater fishes.

We tested five hypotheses (Table 1) associated with functional mechanisms of trophic position–size scaling theory. We expected: (1) *TP* would scale positively with relative jaw length as predicted by gape-limitation theory (Arim et al., 2007; Wainwright & Barton, 1995); (2) jaw length would scale positively with body mass (Wainwright & Barton, 1995); and therefore (3) *TP* would scale

positively with body mass (Romanuk et al., 2011). We hypothesized that: (4) fish assemblages occurring in lakes (Cohen et al., 2003; Nakazawa et al., 2010) would have a steeper positive *TP*–*body mass* scaling slope compared to fishes inhabiting rivers or streams where large-bodied detritivores and herbivores are common (Layman et al., 2005; Ou et al., 2017; Potapov et al., 2019); and (5) expected the slope of *TP*–*body mass* scaling associations would be negatively correlated with the relative species richness of herbivores–detritivores among orders since these species do not feed at a higher *TP* when larger-bodied. All morphological traits, other than

TABLE 1 Hypotheses tested (1–5) and associated statistical models (1–5) used to explain and predict trophic position (*TP*) of fishes

Hypotheses	Statistical model	Predictor code	Predictor description	Data source	Hypotheses tested
1) <i>TP</i> –jaw length	1, 2	<i>JHd</i>	Maxillary jaw length relative to head depth (unitless ratio)	1, 2	<i>TP</i> would scale positively with jaw length
2) Jaw length–body mass	3, 4	<i>MBM</i>	Maximum body mass (g) of species (<i>MBM</i>)	1, 2	Jaw length would scale positively with body mass
3) <i>TP</i> –body mass	1, 2	<i>MBM</i>	Maximum body mass (g) of species	1	<i>TP</i> would scale positively with body mass
4) <i>TP</i> –body mass:lake/river	1, 2	<i>MBM:river/lake</i>	Interaction between <i>MBM</i> and occurrence in rivers and streams only, lakes only, or both	1	Fish occurring in lakes would have a steeper positive <i>TP</i> –body mass scaling slope compared to fishes inhabiting rivers or streams
5) <i>TP</i> –body mass slope	5	% <i>herbivore</i> . <i>detritivore</i>	Proportion of herbivores–detritivores in each taxonomic order	1	<i>TP</i> –body mass scaling slopes would be negatively correlated with the proportion of herbivores and detritivores among orders
Mediating factors and additional traits modelled:					
Biogeographic region	1–5	<i>ecoregion</i>	Afrotropic; Australasia; IndoMalay; Nearctic; Neotropic; Oceania; Palearctic	3	–
Taxonomic order	1–5	<i>order</i>	Taxonomic order	1	–
Freshwater restricted	1, 2	<i>fresh/marine</i>	Species occurs in freshwater environments only, or also temporarily inhabits marine or estuarine environments	1	–
Eye diameter	1, 2	<i>EdHd</i>	Vertical diameter of the eye relative to head depth (unitless ratio)	2	–
Position of the mouth	1, 2	<i>MoBd</i>	Vertical distance from the top of the maxillary to the bottom of the body relative to maximum body depth (unitless ratio)	2	–
Caudal fin aspect	1, 2	<i>CFdCPd</i>	Maximum depth of the caudal fin relative to minimum depth of the caudal peduncle (unitless ratio)	2	–

Data source 1 = Froese and Pauly (2019); 2 = Toussaint et al. (2016); 3 = Tedesco et al. (2017).

maximum body size, were quantified as ratios rather than raw measurements in order to minimize the potential confounding effects of intraspecific variation.

2 | METHODS

2.1 | Fish species and trait data

A global list of extant freshwater fishes, comprising 14,953 species and their occurrences in seven biogeographic regions (*ecoregion*), was sourced from Tedesco et al. (2017). Species were matched with a trait database described in Toussaint et al. (2016) and with FishBase (Froese & Pauly, 2019; <http://www.fishbase.org>) resulting in complete data on *TP*, order, habitat, maximum body length or mass, and other morphological traits (Table 1) for 1,991 species. FishBase data were sourced using the R package *rfishbase* version 3.0.3 (Boettiger et al., 2012) and used to assign response variables and factors including: *TP*; taxonomic order (*order*); maximum body mass (*MBM*); occurrence in rivers and streams only, lakes only, or both (*lake/river*) and whether the species occurs in freshwater environments only, or if it also inhabits marine or estuarine environments (*fresh/marine*). Our dataset did not include entirely marine species; only diadromous and estuarine or freshwater-obligate species. All species that were listed as temporarily inhabiting marine systems also occurred in rivers, while some species also occurred in lakes, and many occurred in both.

Maximum body mass data were not available for all species, and therefore we estimated $\log_{10}MBM$ from \log_{10} maximum length using a Bayesian linear mixed model and unpublished body elongation data (sensu Froese et al., 2014). The correlation between actual maximum mass and predicted mass in a 10-fold cross-validation yielded an r^2 of .90 and a slope of 0.98 suggesting that the model was suitable to predict maximum mass. *MBM* was represented as the species' maximum total mass (g) averaged across river basins of occurrence from Tedesco et al. (2017). Many species were recorded beyond their native ranges and occurred in multiple river basins in our global dataset, but to avoid pseudo-replication we used one set of averaged trait data for each species in their native biogeographic region (Tedesco et al., 2017). Intraspecific variation in traits is wide-ranging, but could not be examined here due to insufficient data available for all species at the global scale. Therefore, our analyses focused on species-level relationships.

Morphological traits aside from *MBM* were extracted from a database detailed in Toussaint et al. (2016). Morphological traits provided relative measures of external morphological features derived from side view photographs and, for the analyses here, included: maxillary jaw length (JIHd); eye diameter (EdHd); vertical position of the mouth (MoBd); and caudal fin aspect (CFdCPd) (Supporting Information Figure S1). All morphological traits were represented as unitless ratios of another morphological feature (Supporting Information Figure S1), but all were measured independently of body size. Allometric biases associated with morphological ratios

(Albrecht et al., 1993) likely increased uncertainty in our models, but some of these issues were minimized by restricting measurements to photos of adult stages only.

The potential *TP* associations with morphological features including eye diameter, position of the mouth on the head, and caudal fin aspect were included in statistical models (see below) but the direction of the associations were unclear from the literature, and therefore a priori hypotheses (Table 1) were not set for these traits. The caudal fin aspect measured here has been correlated positively with sustained swimming speed and drag reduction in a range of fishes (Langerhans, 2008), which may benefit predators, while prey often exhibit more robust caudal regions (Langerhans et al., 2004) suited to fast-start escape behaviour. Extremely large eye diameter has evolved in some prey taxa in order to detect predators (Nilsson et al., 2012), yet predators also benefit from enhanced visual acuity. Position of the mouth on the head indicates vertical feeding position in the water column, and fishes with superior mouths tend to have higher trophic positions (Keppeler et al., 2020).

2.2 | Trophic position

TP estimates of freshwater fishes were extracted from FishBase (Froese & Pauly, 2019). A fractional *TP*, between 2.0 (Herbivore) and 4.5 (Apex predator), was estimated for each fish species in FishBase. FishBase calculates *TP* following the equation: $TP_i = 1 + \sum_p DC_{ip} \times Troph_p$, where the sum runs from 1 to the total number of prey items in the diet of fish species *i*. *TP* is the trophic position of fish species *i*, $Troph_p$ is the *TP* of food item *p*, while DC_{ip} is the fraction of food item *p* in the diet of fish species *i*. Primary producers, detritus and bacteria were assumed to have a *TP* of 1 which was added to each consumer. Therefore, a fish eating 50% phytoplankton/plants ($TP = 1$) and 50% herbivorous zooplankton ($TP = 2$) will have a *TP* of $1 + (0.5 \times 1 + 0.5 \times 2) = 2.5$. The model used in FishBase to calculate *TP* represents a mean of the species across previously published studies. However, the mean *TP* may not represent the best estimate trophic position for each species. In particular, the trophic position of large predators has been underestimated in recent stable isotope studies (Hussey et al., 2014). Therefore, we undertook sensitivity analyses using the upper 95% confidence interval (Supporting Information Tables S2, S3 and S4) of *TP* to examine whether results were different to those based on mean *TP*. Although estimates of fixed species-level *TP* from FishBase are coarse, they have been shown to correlate well with estimates based on more accurate stable isotope-based methods (Carscallen et al., 2012; Mancinelli et al., 2013). We assumed that for the scale of analyses conducted here, FishBase estimates of *TP* were suitably consistent and reliable measures of species-level differences in *TP*.

Based on the *TP* estimates, each species was grouped into a trophic guild using cut-offs provided on FishBase (Froese & Pauly, 2019): Detritivores-herbivore ($TP_j = 2-2.19$); Omnivore ($TP_j = 2.2-2.79$); Secondary consumer ($TP_j = 2.8-4$) and Top predator ($TP_j > 4$). The four guilds align with generic descriptions representing the diet

composition of adults: herbivore-detritivore—fish that consume only detritus, plants, phytoplankton or algae; omnivore—trophic generalists that consume a range of phytoplankton, algae and aquatic or terrestrial plants and invertebrates, or occasionally higher level consumers; secondary consumer—fish that feed primarily on zooplankton, insects, macroinvertebrates and other crustaceans; top predator—fish that are primarily piscivores, or feed on other higher level consumers and in some cases decapod crustaceans and macroinvertebrates.

2.3 | Statistical analysis

We used Bayesian linear mixed effects and ML random forest models (full details of the models, prior distributions, and software are provided in the Supporting Information statistical analyses; full model code and datasets on github: <https://github.com/jdyen/size-trophic>) with r^2 analogues and 10-fold cross-validation, to explain and predict TP .

To test hypotheses 1, 3 and 4 (Table 1), we used a Bayesian linear mixed effects model with the full ($n = 1,991$) structure (model 1): $\log_{10}TP \sim (\log_{10}MBM | order) + (\log_{10}MBM | ecoregion) + \log_{10}jaw\ length + river/lake * \log_{10}MBM + fresh/marine * \log_{10}MBM + eye\ diameter + position\ of\ the\ mouth + caudal\ fin\ aspect$. We standardized all continuous predictor variables to a mean of zero and a standard deviation of one. The full model included random intercepts and slopes for *ecoregion* and taxonomic *order*. In particular, the body size of freshwater fishes varies widely among orders and biogeographic regions (Blanchet et al., 2010), and therefore we evaluated Hypothesis 3 TP - MBM (Table 1) independent of biogeographic or order-specific differences in body mass. Body mass and jaw length were \log_{10} transformed for all statistical analyses. The interaction between $MBM * river/lake$ was used to test for potential differences in the TP - MBM slopes between lake and river ecosystems (Hypothesis 4). As an interaction term ($fresh/marine * MBM$), we also evaluated the influence of whether the species occurs in freshwater environments only, or temporarily inhabits marine or estuarine environments (see Bloom et al., 2018; Sanchez-Hernandez & Amundsen 2018). However, it was unclear how these movements or environments may influence the TP - $body\ mass$ slopes across orders, and therefore we did not formulate a hypothesis a priori. We fitted Bayesian mixed effects models with the `stan_lmer` function in the `rstanarm` R package version 2.19.2 (Goodrich et al., 2018) in R 3.6.1 (R Core Team, 2019) and present estimates of overall model fit (see Model validation) and posterior distributions of model coefficients.

We used a conditional inference random forest model (model 2; Breiman, 2001; Hothorn et al., 2006) to relate TP to all predictors included in Bayesian linear mixed effects (model 1; hypotheses 1, 3, 4). We trialled four additional ML methods (Supporting Information Table S1) during model development and present conditional inference random forest models because these had the best estimated predictive performance for models of TP (based on cross-validated r^2 values; Supporting Information Table S1). Random forest model 2 used the full dataset ($n = 1,991$) to build an ensemble of decision

trees to relate the response variable (TP) to the set of predictor variables ($\log_{10}MBM$; *ecoregion*; *river/lake*; $\log_{10}jaw\ length$; *fresh/marine*; *eye diameter*; *position of the mouth*; *caudal fin aspect*). This approach implicitly incorporated complex, nonlinear associations and high-order interactions among predictors but did not allow a priori interactions or hierarchical structures to be set. The ability to incorporate multiple complex and higher-order interactions, without setting model structure (e.g., random slopes and intercepts) and interactions a priori, is considered one of the reasons why ML methods have enhanced predictive power (Ryo & Rillig, 2017) when compared to GLMs or GLMMs, which are influenced by subjective decisions about model structure.

We fitted Bayesian linear (model 1) and random forest (model 2) models to the full dataset ($n = 1,991$) including all trophic guilds, but also to subsets of data based on four trophic guilds [herbivores-detritivores ($n = 265$); omnivores ($n = 350$); secondary consumers ($n = 1,609$); and top predators ($n = 254$)] independently. We fitted models in this way to compare the explanatory and predictive power among different trophic guilds and between random forest and Bayesian linear mixed methods (see Model validation below).

To test Hypothesis 2 (Table 1), we used a separate Bayesian linear mixed effect (model 3) and random forest (model 4) model ($n = 1,991$). The Bayesian linear mixed model structure (model 3): $\log_{10}jaw\ length \sim (\log_{10}MBM | order) + (\log_{10}MBM | ecoregion) + (\log_{10}MBM|fresh/marine) + \log_{10}MBM*river/stream$, allowed us to evaluate whether $\log_{10}jaw\ length$ was associated with $\log_{10}MBM$ independent of ecoregional and order-specific differences in body mass, and independent of whether the fish entered marine environments or not. The interaction evaluated whether jaw length- MBM slopes varied between rivers and lakes since this interaction was identified as statistically important in TP - MBM associations. The random forest (model 4) used an ensemble of decision trees to relate the response variable ($\log_{10}jaw\ length$) to the set of predictor variables from model 3 including: $\log_{10}MBM$; *ecoregion*; *order*; *river/lake*; *fresh/marine*, but otherwise followed the same methodology as described earlier in this section. Fitted estimates of TP and $jaw\ length$ among *ecoregions* and *orders* are shown in Figures S3 and S4 and the distribution of residuals in Figure S4.

To test Hypothesis 5, we used a separate Bayesian linear mixed model (model 5) to relate order-specific TP - MBM coefficients ($n = 30$) derived from the Bayesian mixed effects model 1 to the proportion of herbivores and detritivores in a given order. We used the R^2 regularization prior, which specifies prior information on all parameters through a beta prior distribution on the proportion of variation explained by the linear model. We set the location of the R^2 prior to .5, which sets both parameters of the beta distribution equal to half the number of predictors (.5 in our study). We present estimates of overall model fit and posterior distributions of all model coefficients. We fitted the Bayesian linear mixed model with the `stan_lm` function in the `rstanarm` R package (Goodrich et al., 2018) in R 3.6.1 (R Core Team, 2019).

In all random forest models, we estimated the partial dependence of TP on each continuous predictor, and estimated variable importance

to evaluate the strength of associations. Partial dependence is the marginal relationship between model predictions and a subset of predictor variables, which provides a simplified representation of a fitted random forest model. Variable importance reflects the change in model accuracy when a given variable is shuffled randomly during model fitting. We standardized importance values so that their sum over all variables is equal to one. Therefore, importance values lie between zero and one, with larger values indicating a stronger association between a given predictor variable and the model response (Strobl et al., 2007, 2008). We used the train function in the caret R package version 6.0–84 to tune and validate random forest models and the five additional ML methods (Kuhn, 2008). We used the partial_dependence function in the edarf R package version 1.1.1 to estimate partial variable effects (Jones & Linder, 2017), and used the varImp function in the party R package version 1.3–3 to estimate relative variable importance. We fitted random forest models with the cforest function in the party R package (Hothorn et al., 2006).

2.4 | Model validation and sensitivity

To compare the performance of Bayesian linear (model 1) and random forest (model 2) models we used in-sample model fit to estimate explanatory capacity and used 10-fold cross-validation to estimate predictive capacity (Roberts et al., 2017). In both cases, we measured model fit with r^2 values, based on Pearson's r . In 10-fold cross-validation, the dataset was broken into 10 equal-sized folds, and a model was fitted with each fold held out in turn. Model fit was based on the congruence between observed values and those predicted for each holdout dataset. We note that the cross-validation used to tune model parameters was performed internally within each model (i.e., parameters were tuned with cross-validation on the nine folds used for model fitting).

We ran sensitivity analyses (Supporting Information Table S4) to compare variation in explained and predicted TP of the mixed effects linear model (model 1) and random forest model (model 2) for trophic guilds. The sensitivity analyses compared explanatory and predictive power among models that used mean TP , models that used the upper 95% confidence interval (CI) of TP (Upper95TP) and models that excluded herbivores-detritivores (Removed herbivore-detritivore). We undertook sensitivity analyses using the upper 95% CI (Supporting Information Tables S2, S3 and S4) of TP since it may represent a more reliable measure of TP compared to the mean. We removed herbivores-detritivores since the TP of these species is not expected to increase with body mass. Finally, we evaluated whether the removal of herbivores-detritivores reduced bias in TP estimates (Supporting Information Figure S5).

3 | RESULTS

The TP and associated data available for the 1,991 freshwater fishes represented 30 orders and 7 global ecoregions. Cypriniformes

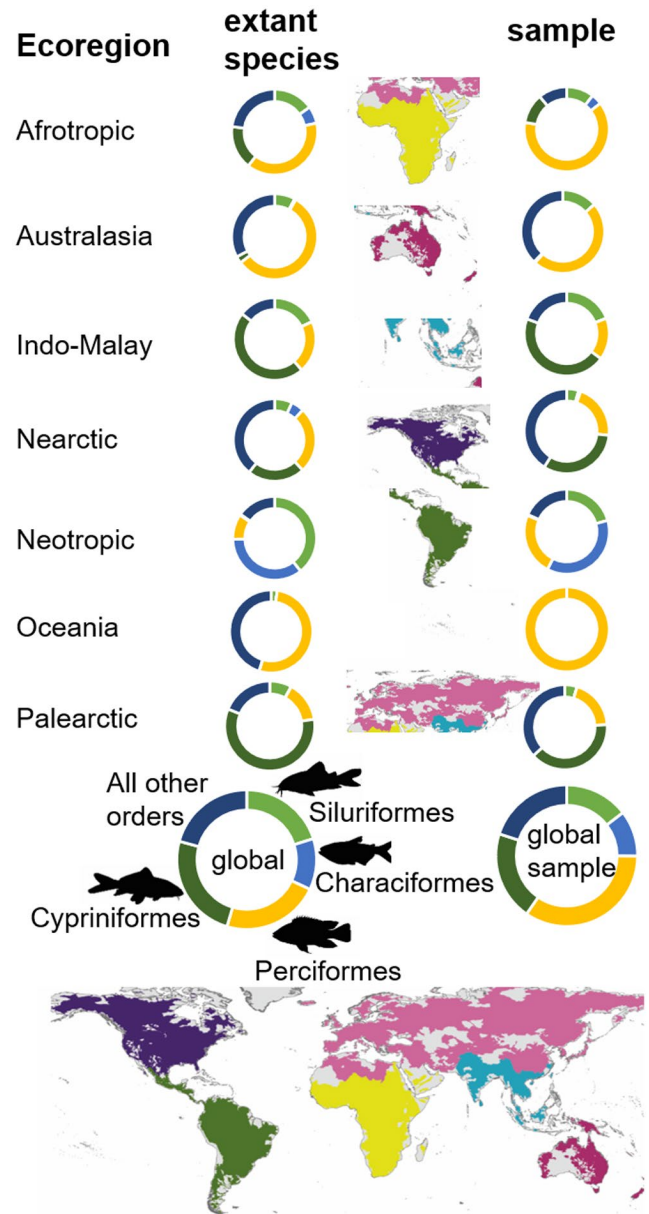


FIGURE 1 Wheel charts displaying the global (global) and ecoregional proportions of extant (left) freshwater fishes within the four most species-rich orders, and their relative proportions in our sample (right) with silhouettes illustrating the four most species-rich orders. No data were available from the regions with no colour [Colour figure can be viewed at wileyonlinelibrary.com]

(minnows and carps), Perciformes (perches), Siluriformes (catfishes) and Characiformes made up 79% (11,850/14,943) of global (Figure 1; global) extant freshwater fishes, and these orders represented 80% (1,593/1,991) of the fishes in our sample (Figure 1; sample). TP ranged from 2.0 to 4.5 in all ecoregions and the modelled median ranged from 2.9 in the Neotropics to 3.1 in the Palearctic (Supporting Information Figures S2 and S3). Most freshwater fishes in our sample were found in rivers only ($n = 885$; 44%) compared to lakes only ($n = 312$; 16%), but many occurred in both or were unknown ($n = 794$; 40%). The majority of species occurred only in

freshwater environments ($n = 1,514$; 76%) compared to those that also entered marine environments ($n = 477$; 24%). The modelled estimates of *jaw length*-*MBM* and *TP*-*MBM* slopes for Oceania and six orders (Amiiformes; Batrachiformes; Elopiformes; Gonorychiformes; Gadiformes; Pleuronectiformes) were not considered to be reliable since these were mostly marine inhabitants, each represented by fewer than five species in our sample.

Jaw length (Hypothesis 1/model 1) had the strongest positive effect size in the Bayesian linear mixed effects model (Figure 2) and was the most influential (variable importance = .46) predictor of *TP* in the random forest equivalent, model 2 (Figures 3a and 4). When compared to the weak positive *TP*-body mass association in the random forest model 2 (Figure 3c), jaw length showed a strong non-linear and positive association (Figure 3a). Top predators with very short jaws were uncommon, as were herbivores-detritivores and omnivores with very long jaws (Figure 3a). The linear mixed model 1 indicated that *TP* increased positively with caudal fin aspect ratio (Figure 2), but this trait had almost no influence (variable importance = .01) in the random forest equivalent (Figure 4).

Jaw length was weakly associated with body mass (Hypothesis 2/model 4) in the random forest model (Figure 3b), but there was a statistically significant positive jaw length-body mass association

among ecoregions in the Bayesian linear model (model 3; Figure 5). There was a significantly positive association (Figure 5) in five orders including: Siluriformes; Osteoglossiformes; Esociformes; Beloniformes; and Salmoniformes. Bayesian linear model 3 ($r^2 = .29$) and random forest model 4 ($r^2 = .34$; Figure 3b) had similarly weak power explaining jaw length-body mass associations. The Neotropical (variable importance = .24) ecoregion was the most influential predictor of jaw length in model 4, followed by Siluriformes (variable importance = .19) and body mass (variable importance = .16).

Bayesian linear model 1 showed that *TP* did not scale linearly, or positively (Hypothesis 3/model 1), with body mass in any of the seven global ecoregions, or in 29/30 orders of freshwater fish (Figure 6). These results were unchanged in sensitivity analyses using the upper 95% CI of *TP* (Supporting Information Tables S2 and S3). The *TP* of Perciformes scaled positively with body mass, Cypriniformes showed no significant relationship (Figure 6), and the *TP* of Characiformes decreased with increasing body mass (Figure 6). Body mass (variable importance = .09) was the fourth most important predictor of *TP* in the random forest model 2 and followed jaw length, Cypriniformes and Characiformes in order of importance (Figure 4).

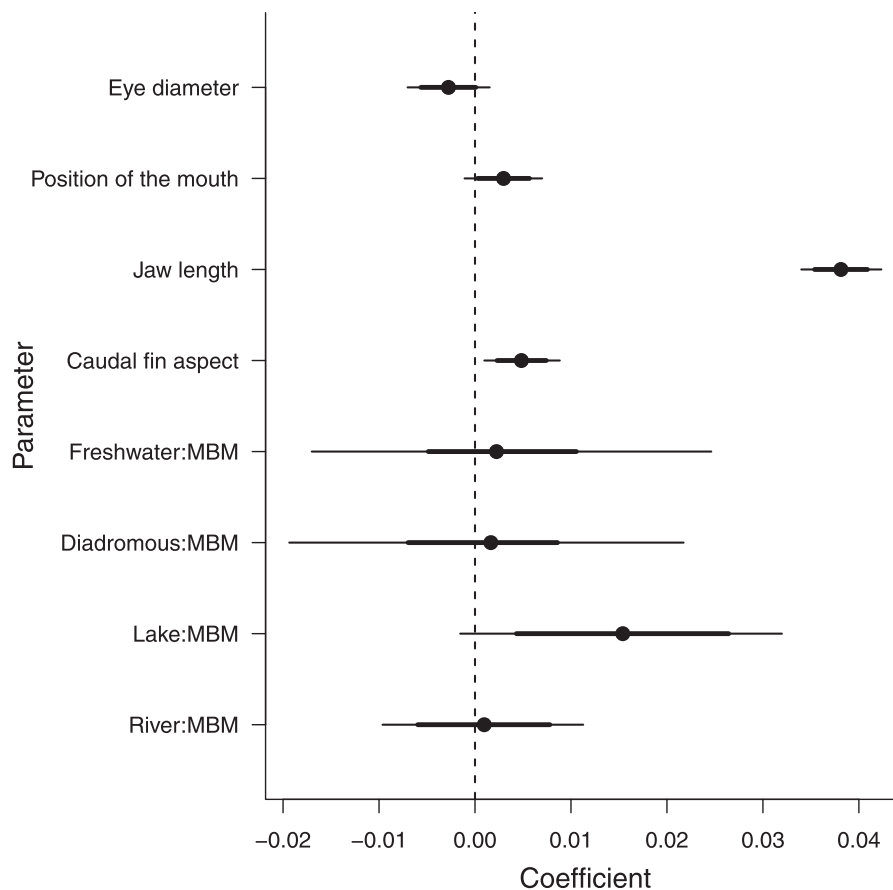


FIGURE 2 Coefficients of fixed effects explaining trophic position (*TP*) of freshwater fishes. Coefficients were from Bayesian linear mixed model 1. Credible intervals not overlapping zero (dashed line) were assumed to be significant and magnitude (positive or negative) of coefficients is associated with effect size. Points are median posterior estimates, thick lines are 80% credible intervals, and thin lines are 95% credible intervals. MBM = maximum body mass

The Bayesian linear mixed model (Hypothesis 4/model 1) showed that fishes inhabiting lakes only (Figure 2; *lake:MBM*) tended to have a more positive *TP*-*body mass* slope compared to riverine species

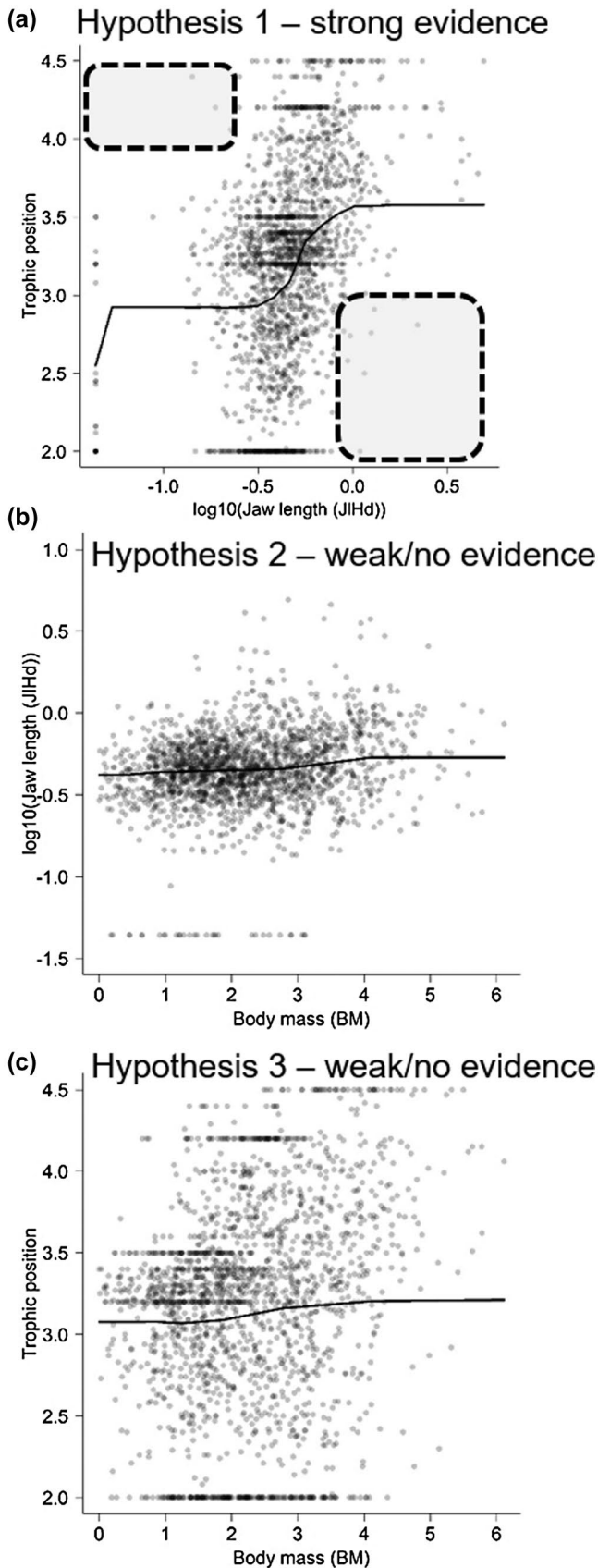


FIGURE 3 Partial dependence of trophic position (*TP*) on \log_{10} jaw length (a), jaw length– \log_{10} maximum body mass (b) and *TP*– \log_{10} maximum body mass (c) from random forest models 2 and 4. Plots display the marginal association between *TP* and a given predictor, marginalizing (averaging) over all other predictor variables included in a model. Black points are observed values of *TP* and predictor variables. Dashed boxes on the jaw length panel (a) qualitatively illustrate the least-occupied trait-spaces of top predators ($TP = 4$ – 4.5) with short jaws, and herbivores-detritivores or omnivores ($TP = 2.0$ – 2.79) with very long jaws

(*river:MBM*), but the effect size was not significant with an alpha of .05. Whether fish inhabited lakes or rivers was a weak overall (combined variable importance = .02) predictor of *TP* in the random forest equivalent model 2 (Figure 4). The slope of the *TP*–*MBM* relationship (Hypothesis 5/model 5) was not significantly (alpha of .05) associated with the proportion of herbivores and detritivores among orders (-0.20 – 0.03 95% CI), but the Bayesian *p*-value .94 suggested a negative association was most likely (Supporting Information Figure S6).

The Bayesian linear mixed effects model 1 explained 31% (based on in-sample model fit) of the global variation in freshwater fish *TP* for all guilds combined (Table 2). In-sample model fit (Table 2) was highest (31%) in the top predator guild and lowest among herbivores-detritivores (20%). Random forest model 2 had 1.9 times better predictive performance (10-fold cross-validation) than the linear model for all trophic guilds combined (Table 2), and had equal or better predictive performance for each guild separately. The random forest model 2 for all guilds explained 55% (r^2) of global variation in *TP*, which was higher than the Bayesian linear mixed effects equivalent (Table 2). All ML methods trialled outperformed the Bayesian linear mixed models, but random forests provided the greatest predictive performance (Supporting Information Table S1).

Despite robust explanatory power and improved predictive performance of nonlinear ML methods, models 1 and 2 both over-estimated the *TP* of herbivores-detritivores and omnivores, and underestimated the *TP* of top predators (Figure 7). Both explanatory and predictive performance were lowest for herbivores and detritivores regardless of the modelling method (Table 2). Bias in low and high *TP*s resulted in poor overall predictive performance (Table 2; 10-fold cross-validation), which was not improved by the exclusion of herbivores-detritivores from analyses (Supporting Information Figure S5). Explanatory and predictive power (Supporting Information Table S4) of random forest models improved slightly, from r^2 of .55 to r^2 of .60 and prediction from .32 to .34, after exclusion of herbivores-detritivores, while Bayesian model results were unchanged (Supporting Information Table S4).

4 | DISCUSSION

Our study suggests that maxillary jaw length is an important constraint on *TP* in fishes globally. Body size determines energy demand (Rip & McCann, 2011) and can influence predator–prey interactions

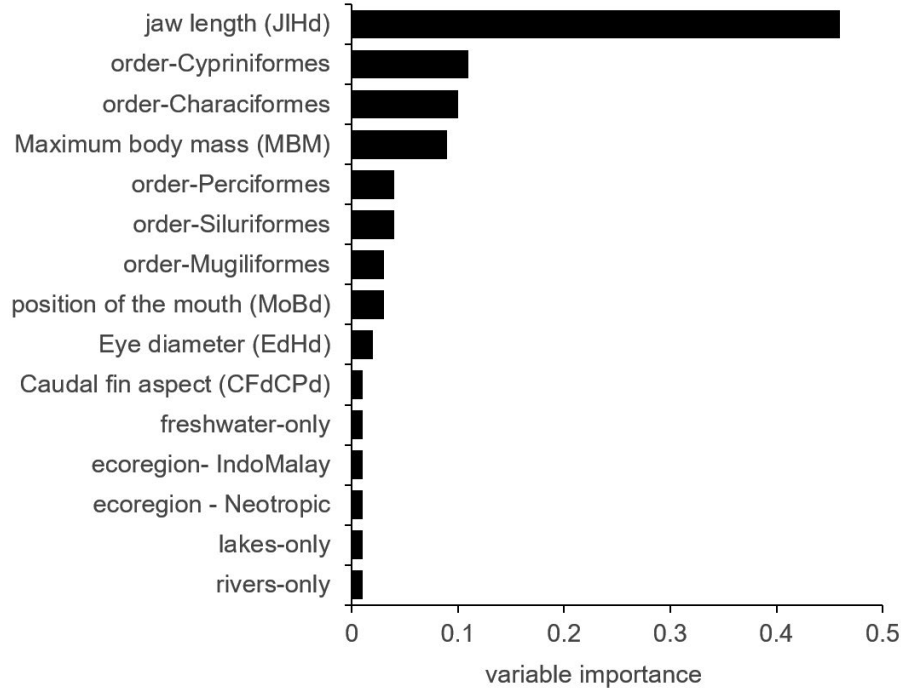


FIGURE 4 Variable importance in random forest model 2 of fish trophic position (*TP*). Variable importance is the proportional decrease in accuracy under permutation of a given variable and values sum to 1. Larger values denote variables that explain more variation in *TP*. Only variables with values $> .01$ are illustrated

(Brose, 2010; Emmerson & Raffaelli, 2004), but in gape-limited consumers it is the jaw and associated mouth and skull adaptations that play a more direct role in the size and type of prey consumed. Despite variation in diet, feeding behaviour, and jaw morphology within and among species (Ferry et al., 2015; Wainwright, 2007), our results underline a steeply positive and nonlinear association between jaw length and *TP* among fishes globally.

In fishes, amphibians and reptiles, which often swallow prey whole, mouth gape can limit prey size-selection (Arim et al., 2007; Shine, 1991; Wainwright & Barton, 1995). Predators generally had longer jaws than fishes feeding at a lower *TP* and, although gape size is not directly related to jaw length, the positive association was consistent with predictions of *TP* increasing with gape size (Arim et al., 2007). Longer jaws can increase gape-distance for prey capture, increase mouth closing speeds, or increase suction feeding velocities in predators (Hulsey & Garcia de Leon, 2005), while shorter jaws maximize bite force needed to feed on hard-bodied prey items (Ferry et al., 2015; Westneat, 2004; Evans et al., 2019). The shape of the nonlinear association showed that top predators rarely had small jaw sizes, while species with extremely long jaws were not commonly omnivorous, herbivorous or detritivorous.

TP was positively associated with jaw length, but jaw length was weakly associated with body mass, and consequently *TP* was generally uncorrelated with body mass among fishes. Given the weak but statistically significant association between jaw length and body mass, it was unsurprising that body mass was a poor predictor of *TP*. Nevertheless, positive jaw length–body size associations were detected among ecoregions and in several orders consisting

primarily of large consumers (e.g., Siluriformes; Osteoglossiformes; Esociformes; Salmoniformes). However, in none of these orders or ecoregions was body size a strong predictor of species-level trophic position. This result suggests that the largest fish in these orders and ecoregions do not hold the highest *TP*s even though they may have large jaws. For example, one of the world's largest freshwater fishes grows to a maximum mass of approximately 350 kg and has a relatively large jaw size—the Mekong giant catfish (Siluriformes)—yet this species is omnivorous and feeds at a low *TP* ($TP = 2.3$).

Freshwater fishes have diversified on all continents to circumvent body size-based trophic constraints. In contrast to widespread assumption and a previous global analysis (Romanuk et al., 2011), our results suggest that body size does not explain or predict *TP* in most orders of freshwater fishes, species inhabiting rivers, or the assemblages in our sample of global ecoregions. Similar to our results showing the best predictive performance in non-herbivore/detritivore trophic guilds, Keppeler et al. (2020) found a weak *TP*–body size association in non-carnivorous fishes and a stronger relationship in carnivorous fishes. The diversification of large-bodied herbivores-detritivores and omnivores in several of the world's most species-rich orders of fish (e.g., Cypriniformes; Characiformes; Siluriformes), may partly explain why *TP* showed a weak relationship with body size. For example, in the most species-rich order of freshwater fish, many large ($> 10,000$ g) Cypriniformes are detritivores and herbivores with low *TP*s, such as European and Asian carps (e.g., *Ctenopharyngodon* sp. and *Hypophthalmichthys* sp., $TP = 2$), whereas most small Cypriniformes are carnivores (German et al., 2009) feeding on zooplankton or invertebrates with *TP*s higher than 3.

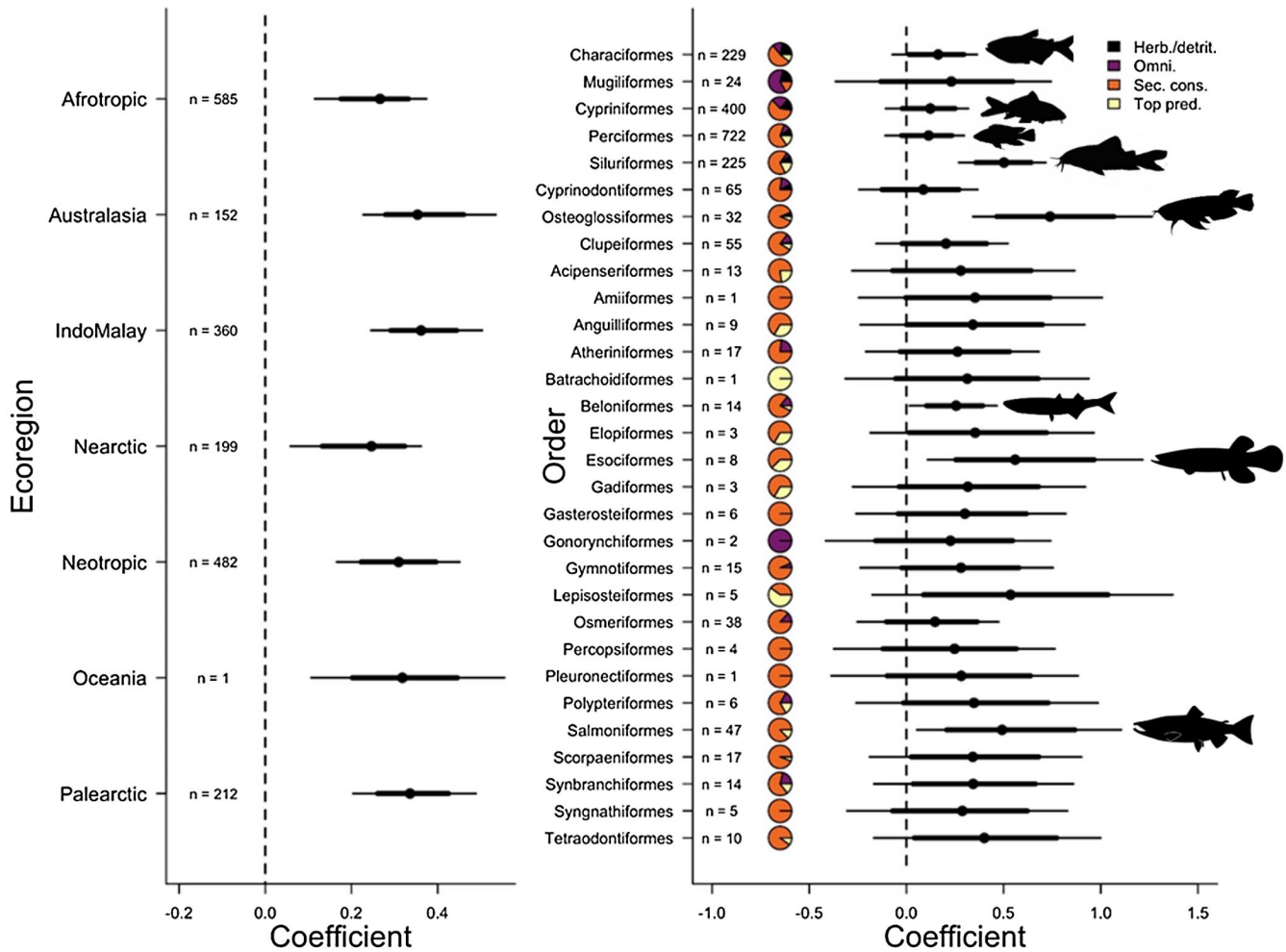


FIGURE 5 Jaw length–body mass associations in freshwater fishes for each global ecoregion (left) and order (right). Coefficient slopes are from the Bayesian linear mixed model 3. Credible intervals not overlapping zero (dashed lines) were assumed to be significant and magnitude (positive or negative) of coefficients is associated with effect size. Points are median posterior estimates, thick lines are 80% credible intervals and thin lines are 95% credible intervals. Pie charts display the proportion of species in each of four trophic guilds (Herbivore–detritivore; Omnivore; Secondary Consumer; Top predator). Silhouettes illustrate the four most dominant orders and those with significant positive associations. [Colour figure can be viewed at wileyonlinelibrary.com]

After we removed herbivores–detritivores from our analyses, the *TP* of Cypriniformes still did not scale positively with body mass, which suggests that other factors are important. A decoupling of jaw length and body mass, and high levels of omnivory in freshwater fishes may explain why *TP* did not scale positively with body mass after excluding herbivores–detritivores. Digestive efficiency of plant material is maximized by increasing gut surface area or length (German et al., 2009), and therefore the large body size of herbivores–detritivores and some omnivores may have evolved to accommodate a large volume intestine required to meet the nutritional and energetic demands of partially plant-based diets. The lack of a positive *TP*–*MBM* relationship for Cypriniformes is consistent with local stable isotope studies on tropical (Ou et al., 2017) and temperate rivers (Burruss et al., 2016) showing negative relationships.

In another species-rich order of freshwater fish, *TP* scaled negatively (Characiformes) with increasing body mass, which is opposite to

the positive association reported by Romanuk et al. (2011). Even after excluding herbivores–detritivores, we could not replicate the result of positive *TP*–body mass scaling for Characiformes. This order is a functionally diverse (Toussaint et al., 2016) group of primarily Neotropical freshwater species including very large fishes that feed at high and low *TP*s, with unique adaptations to eat floodplain plants, nuts and seeds (Correa et al., 2007). Other adaptations of Characiformes include specialized teeth for biting (e.g., piranha *Pygocentrus* sp.; Van der Sleen & Albert, 2017), and fin- and scale-eating strategies (Sazima, 1983)—all of which allow these fishes to feed on relatively large, or high trophic level, prey necessarily without a large body size. Siluriformes in this ecoregion are characterized by a large range of jaw sizes with various adaptations to filter feed, or graze on biofilm and vegetation (Lujan & Armbruster, 2012). Similar to our results, Layman et al (2005) reported no association between *TP* and predator body size in the Cinaruco River of the Neotropical realm.

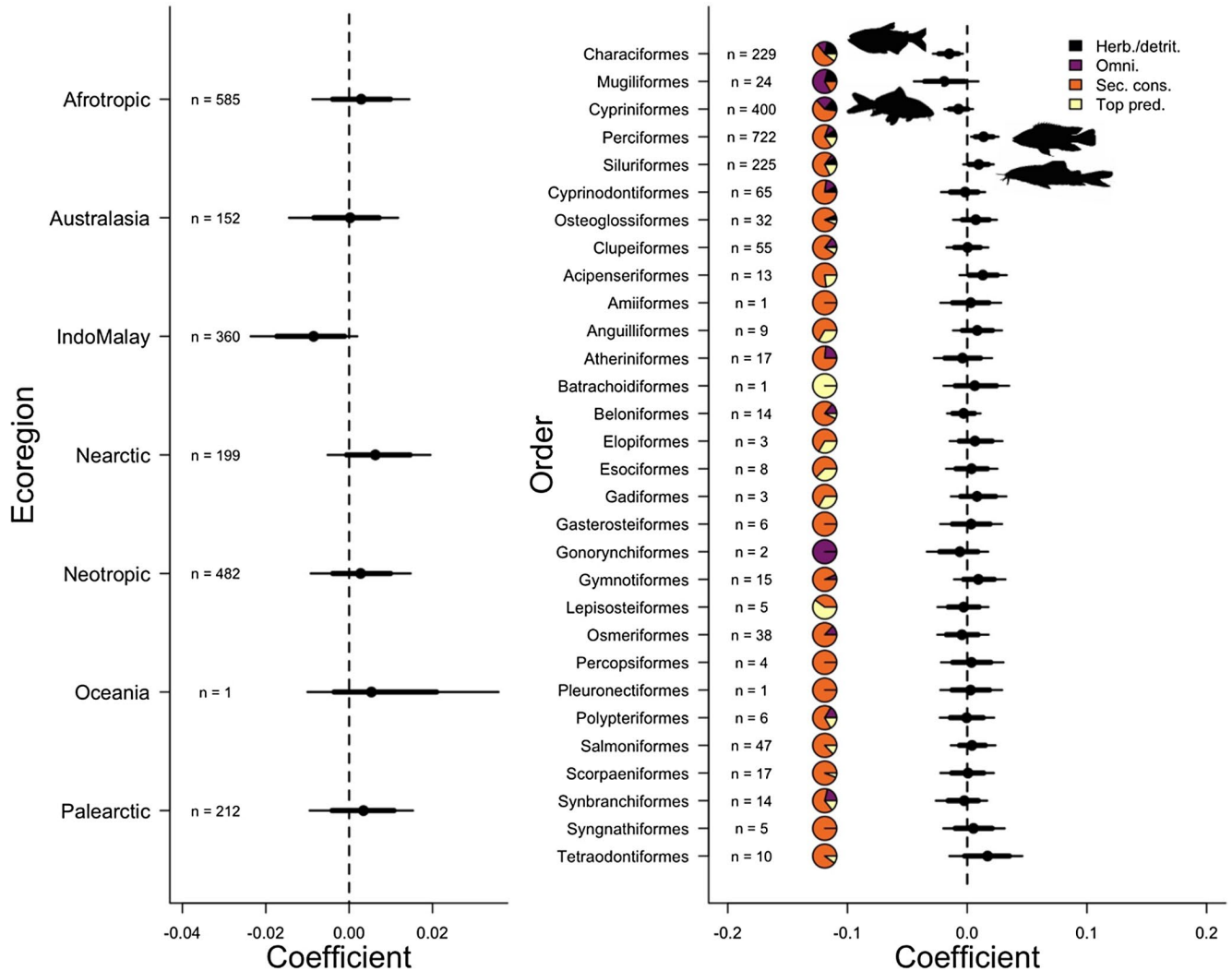


FIGURE 6 Trophic position–body mass scaling slopes in freshwater fishes for each global ecoregion (left) and order (right). Coefficient slopes are from the Bayesian linear mixed model 1. Credible intervals not overlapping zero (dashed lines) were assumed to be significant and magnitude (positive or negative) of coefficients is associated with effect size. Points are median posterior estimates, thick lines are 80% credible intervals and thin lines are 95% credible intervals. Pie charts display the proportion of species in each of four trophic guilds (Herbivore–detritivore; Omnivore; Secondary Consumer; Top predator). Silhouettes illustrate the four most dominant orders and those with significant associations. [Colour figure can be viewed at wileyonlinelibrary.com]

We found evidence of positive allometric scaling of *TP* in Perciformes which is one of the world's most species rich vertebrate orders and most prolific group of fish in marine ecosystems. This result matched several previous studies (Montana & Winemiller, 2013; Ou et al., 2017; Romanuk et al., 2011), and the slope of the *TP*–*MBM* association was steepest in lakes. Perciformes was the dominant order inhabiting lake ecosystems in our sample and this was largely due to radiations of endemic cichlids in the African Rift Valley (Muschick et al., 2012). Since the Afrotropical ecoregion is dominated by perciform fishes, it is reasonable to expect that *TP* may scale positively with body size in this ecoregion, but again we found a weak relationship, and jaw length in this order was not strongly associated with body mass. Perciformes was also dominant in Australasia, but similarly we found little evidence of positive allometric scaling in this ecoregion, possibly due to the influence of fishes in other orders with no discernible *TP*–*MBM* relationship.

Traits other than jaw length and body size examined here were not particularly influential predictors in the random forest models, but caudal fin aspect was significantly and positively associated with *TP* in the Bayesian linear mixed models. Caudal fin aspect is positively correlated with sustained swimming speed and drag reduction (Langerhans, 2008), and therefore this result may indicate that predators are more likely to have faster sustained swimming speeds than fishes feeding at lower *TP*s. However, the strength of the *TP*–caudal fin aspect association in random forest models was very weak. Differences in how predators find and capture prey (e.g., ambush; searching etc.) and other aspects of their ecology (e.g., inhabiting lentic versus lotic environments) and life history (e.g., migratory species versus non-migratory) probably increased variation in *TP*–caudal fin associations. Similar to Bloom et al. (2018) we found little evidence of *TP*–*body mass* scaling in diadromous fishes.

TABLE 2 Proportion of variation in fish trophic position (*TP*) explained and predicted correctly by mixed effects linear (model 1) and random forest model (model 2) for trophic guilds

Trophic guild	Method	Variation explained (r^2) and predicted (in parentheses)
All guilds combined	Mixed effects linear	.31 (.17)
	Random forest	.55 (.32)
Herbivore/detritivore	Mixed effects linear	.20 (.00)
	Random forest	.11 (.00)
Omnivore	Mixed effects linear	.24 (.02)
	Random forest	.49 (.08)
Secondary consumer	Mixed effects linear	.22 (.10)
	Random forest	.48 (.22)
Top predator	Mixed effects linear	.31 (.02)
	Random forest	.48 (.11)

Note: Estimates of variation explained and predicted are based on in-sample model fit r^2 (**bold**) and 10-fold cross-validation (in parentheses), respectively.

Our results contribute to evidence suggesting that size-structured trophic dynamics differ between river and lake ecosystems. The weak association we observed between *TP* and body size of freshwater fishes in rivers is consistent with two previous global food web analyses (Potapov et al., 2019; Riede et al., 2011). Likewise, the more positive allometric scaling of *TP* in lake fishes matched results from lake food web analyses (Cohen et al., 2003; Nakazawa et al., 2010; Potapov et al., 2019; Riede et al., 2011). Brose et al. (2019) reported positive predator–prey mass scaling in lake and stream food webs globally, but noted that data on large species, including fish, were needed to confirm the relationship in streams. Contrasting *TP*–*MBM* scaling relationships between lakes and rivers may be explained partly by size-compartmentalization hypotheses discussed (Potapov et al., 2019). They hypothesized that differences in *TP*–*MBM* slopes among ecosystem types (freshwater, marine, terrestrial) were attributed to differences in the size-structure of primary producers fuelling food webs. For instance, river-floodplain food webs are fuelled by a complex mix of phytoplankton, vascular plants, terrestrial seeds and detritus (Layman et al., 2005; Winemiller, 1990) and generally have not conformed to simple assumptions of positive allometric scaling of *TP*. By contrast, lake ecosystems are more often characterized by longer and linear food chains fuelled by small-sized phytoplankton (Cohen et al., 2003; Vander Zanden & Fetzer, 2007; Vander Zanden et al., 2011) and conform to positive allometric scaling of *TP* (Cohen et al., 2003; Nakazawa et al., 2010; Potapov et al., 2019; Riede et al., 2011). However, many lakes worldwide are

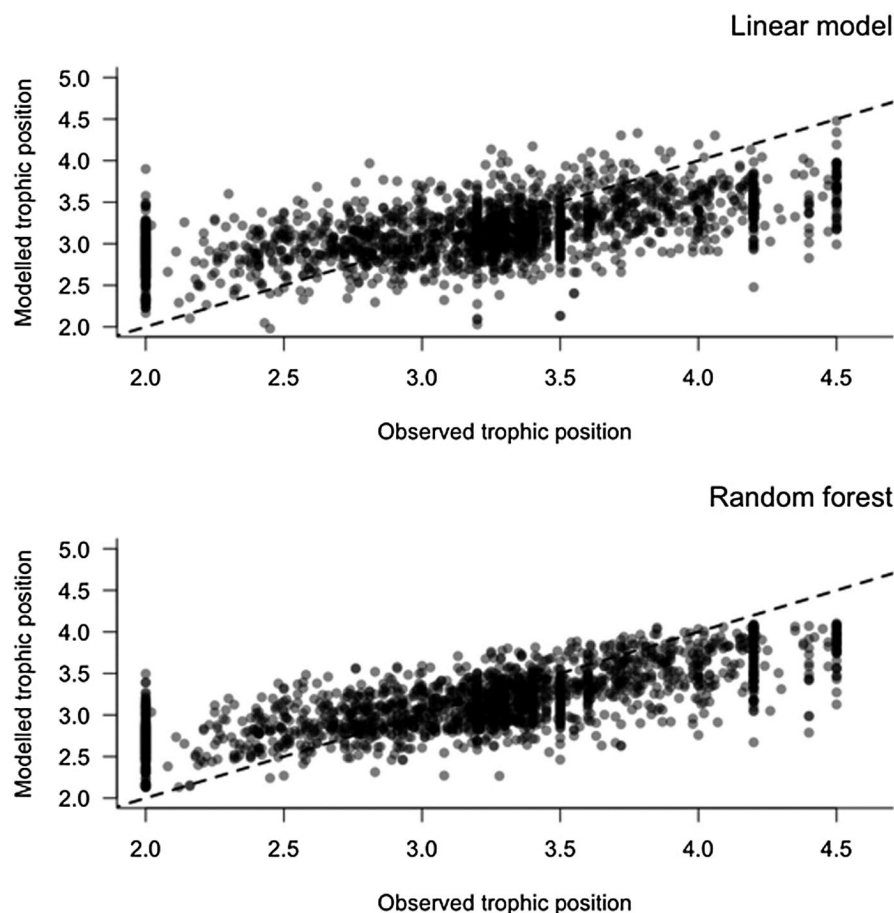


FIGURE 7 Bias in trophic position (*TP*) estimates. Black points are fitted *TP* against observed trophic positions fitted to Bayesian linear mixed model 1 (top panel) and random forest model 2 (bottom panel). The dashed line is the 1:1 line on which points would lie if fitted values were unbiased

shallow and strongly supported by multicellular autotrophs whereby TP -body mass associations may be weak. The contrasting TP - MBM patterns suggest that body size-based assumptions derived from lake or marine ecosystems may not apply to rivers or floodplain ecosystems, and therefore differences among ecosystem types and underlying trophic complexity (Jonsson et al., 2018) should be considered in size-based fisheries, food web and macroecology models (e.g., Andersen, 2019; Blanchard et al., 2017; Brose et al., 2017; Sibly et al., 2012).

With 1.9 times better predictive performance compared to the linear method, our study supports the use of nonlinear ML approaches to improve the predictive capacity of macroecology models. Despite explaining 55% of the global variation in fish TP , our best random forest model over-estimated the position of herbivores-detritivores and omnivores, and under-estimated the position of top predators. A recent review showed that few studies test the predictive capacity of their most supported models (Mac Nally et al., 2018). Our findings reiterate the importance of validating predictive capacity, especially in macroecology and ML models, which often rely on global datasets with large sample sizes. Here, we have highlighted that a series of traits that are clearly associated with TP , supported by our results and mechanistic theory, cannot accurately predict TP per se. Our results caution against the use of traits—including body size—as predictors of TP , at the species level, in freshwater fishes. The addition of jaw length substantially improved performance relative to body size but global predictive capacity remained poor.

Incorporating other functional feeding traits (e.g., jaw shape and mechanics, absolute rather than relative jaw size, gape, digestive tract anatomy), behavioural feeding mode (e.g., Ferry et al., 2015), and intraspecific variation (e.g., Wainwright & Barton, 1995) are likely to yield more accurate predictions of TP . Furthermore, the same fractional TP (e.g., $TP = 3.0$) can be achieved via different diets (e.g., zooplankton, macroinvertebrates), and therefore future studies would benefit from exploring the more direct functional links between consumer morphology and prey morphology and their respective TP s. For example, prey type and size may also be a determinant or constraint on maximum size or gape size in consumers. Therefore, examining consumer-food type traits together in combination with feeding mode (e.g., filter feeding; grazing) and TP may offer useful insight.

ACKNOWLEDGMENTS

Thanks to S. McDonald for assistance with collating datasets. R.K.K. was partially supported by a Charles Sturt University Research Fellowship. J.D.L.Y. was supported by a McKenzie Fellowship from the University of Melbourne. D.G.N. was supported by an Australian Research Council Discovery Early Career Researcher Award. S.B. was supported by Agence Nationale de la Recherche projects: Centre d'Etude de la Biodiversité Amazonienne (CEBA: ANR-10-LABX-25-01) and Toward a unified theory of biotic interactions (TULIP: ANR-10-LABX-0041).

DATA AVAILABILITY STATEMENT

R code and data are available at <https://github.com/jdyen/size-trophic>

ORCID

R. Keller Kopf  <https://orcid.org/0000-0001-5780-0074>

REFERENCES

- Albrecht, G. H., Gelvin, B. R., & Hartman, S. E. (1993). Ratios as a size adjustment in morphometrics. *American Journal of Physical Anthropology*, 91(4), 441–468.
- Andersen, K. H. (2019). *Fish ecology, evolution, and exploitation: A new theoretical synthesis*. Princeton University Press.
- Arim, M., Bozinovic, F., & Marquet, P. A. (2007). On the relationship between trophic position, body mass and temperature: Reformulating the energy limitation hypothesis. *Oikos*, 116, 1524–1530.
- Blanchard, J. L., Heneghan, R. F., Everett, J. D., Trebilco, R., & Richardson, A. J. (2017). From bacteria to whales: Using functional size spectra to model marine ecosystems. *Trends in Ecology and Evolution*, 32, 174–186.
- Blanchet, S., Grenouillet, G., Beauchard, O., Tedesco, P. A., Leprieur, F., Dürr, H. H., Busson, F., Oberdorff, T., & Brosse, S. (2010). Non-native species disrupt the worldwide patterns of freshwater fish body size: Implications for Bergmann's rule. *Ecology Letters*, 13(4), 421–431.
- Bloom, D. D., Burns, M. D., & Schriever, T. A. (2018). Evolution of body size and trophic position in migratory fishes: A phylogenetic comparative analysis of Clupeiformes (anchovies, herring, shad and allies). *Biological Journal of the Linnean Society*, 125, 302–314.
- Boettiger, C., Lang, D. T., & Wainwright, P. (2012). rfishbase: Exploring, manipulating and visualizing FishBase data from R. *Journal of Fish Biology*, 81, 2030–2039.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Brose, U. (2010). Body-mass constraints on foraging behaviour determine population and food-web dynamics. *Functional Ecology*, 24, 28–34.
- Brose, U., Archambault, P., Barnes, A. D., Bersier, L.-F., Boy, T., Canning-Clode, J., Conti, E., Dias, M., Digel, C., Dissanayake, A., Flores, A. A. V., Fussmann, K., Gauzens, B., Gray, C., Häussler, J., Hirt, M. R., Jacob, U., Jochum, M., Kéfi, S., ... Iles, A. C. (2019). Predator traits determine food-web architecture across ecosystems. *Nature Ecology & Evolution*, 3, 919–927.
- Brose, U., Blanchard, J. L., Eklöf, A., Galiana, N., Hartvig, M., R. Hirt, M., Kalinkat, G., Nordström, M. C., O'Gorman, E. J., Rall, B. C., Schneider, F. D., Thébaud, E., & Jacob, U. (2017). Predicting the consequences of species loss using size-structured biodiversity approaches. *Biological Reviews*, 92, 684–697.
- Brose, U., Jonsson, T., Berlow, E. L., Warren, P., Banasek-Richter, C., Bersier, L.-F., Blanchard, J. L., Brey, T., Carpenter, S. R., Blandenier, M.-F., Cushing, L., Dawah, H. A., Dell, T., Edwards, F., Harper-Smith, S., Jacob, U., Ledger, M. E., Martinez, N. D., Memmott, J., ... Cohen, J. E. (2006). Consumer-resource body-size relationships in natural food webs. *Ecology*, 87, 2411–2417.
- Burruss, E. D., Holcomb, J. M., Bonato, K. O., & Armbruster, J. W. (2016). Body size is negatively correlated with trophic position among cyprinids. *Royal Society Open Science*, 3, 150652.
- Carscallen, W. M. A., Vandenberg, K., Lawson, J. M., Martinez, N. D., & Romanuk, T. N. (2012). Estimating trophic position in marine and estuarine food webs. *Ecosphere*, 3, 1–20.
- Cohen, J. E., Jonsson, T., & Carpenter, S. R. (2003). Ecological community description using the food web, species abundance, and body size. *Proceedings of the National Academy of Sciences USA*, 100, 1781–1786.

- Correa, S. B., Winemiller, K. O., Lopez-Fernandez, H., & Galetti, M. (2007). Evolutionary perspectives on seed consumption and dispersal by fishes. *BioScience*, 57, 748–756.
- Elton, C. (1927) *Animal ecology*. Reprint, 2001. University of Chicago Press edition, Sidgwick & Jackson.
- Emmerson, M. C., & Raffaelli, D. (2004). Predator–prey body size, interaction strength and the stability of a real food web. *Journal of Animal Ecology*, 73, 399–409.
- Evans, K. M., Kim, L. Y., Schubert, B. A., & Albert, J. S. (2019). Ecomorphology of neotropical electric fishes: An integrative approach to testing the relationships between form, function, and trophic ecology. *Integrative Organismal Biology*, 1, 1–16.
- Ferry, L. A., Paig-Tran, E. M., & Gibb, A. C. (2015). Suction, ram, and biting: Deviations and limitations to the capture of aquatic prey. *Integrative and Comparative Biology*, 55, 97–109.
- Froese, R., & Pauly, D. (Eds.). (2019). *FishBase*. World Wide Web Electronic Publication. www.fishbase.org, version (08/2019).
- Froese, R., Thorson, J. T., & Reyes, R. B. Jr. (2014). A Bayesian approach for estimating length–weight relationships in fishes. *Journal of Applied Ichthyology*, 30(1), 78–85.
- German, D. P., Nagle, B. C., Villeda, J. M., Ruiz, A. M., Thomson, A. W., Contreras Balderas, S., & Evans, D. H. (2009). Evolution of herbivory in a carnivorous clade of minnows (Teleostei: Cyprinidae): Effects on gut size and digestive physiology. *Physiological and Biochemical Zoology*, 83, 1–18.
- Goodrich, B., Gabry, J., Ali, I., & Brilleman, S. (2018). *rstanarm: Bayesian applied regression modeling via Stan*. R package version 2.17.4. Retrieved from <http://mc-stan.org/>
- Hothorn, T., Buehlmann, P., Dudoit, S., Molinaro, A., & Van Der Laan, M. (2006). Survival ensembles. *Biostatistics*, 7, 355–373.
- Hulse, C., & García de León, F. (2005). Cichlid jaw mechanics: Linking morphology to feeding specialization. *Functional Ecology*, 19, 487–494.
- Hussey, N. E., MacNeil, M. A., McMeans, B. C., Olin, J. A., Dudley, S. F., Cliff, G., Wintner, S. P., Fennessy, S. T., & Fisk, A. T. (2014). Rescaling the trophic structure of marine food webs. *Ecology Letters*, 17(2), 239–250.
- Jennings, S., Pinnegar, J. K., Polunin, N. V., & Boon, T. W. (2001). Weak cross-species relationships between body size and trophic level belie powerful size-based trophic structuring in fish communities. *Journal of Animal Ecology*, 70, 934–944.
- Jones, Z. M., & Linder, F. (2017). *edarf: Exploratory data analysis using random forests*. R package version 1.1.1. Retrieved from <https://CRAN.R-project.org/package=edarf>
- Jonsson, T., Kaartinen, R., Jonsson, M., & Bommarco, R. (2018). Predictive power of food web models based on body size decreases with trophic complexity. *Ecology Letters*, 21, 702–712.
- Keppeler, F. W., Montaña, C. G., & Winemiller, K. O. (2020). The relationship between trophic level and body size in fishes depends on functional traits. *Ecological Monographs*, 90(4), 1–19. <https://doi.org/10.1002/ecm.1415>
- Kuhn, M. (2008). Building predictive models in R using the Caret package. *Journal of Statistical Software*, 28, 1–26.
- Langerhans, R. B. (2008). Predictability of phenotypic differentiation across flow regimes in fishes. *Integrative and Comparative Biology*, 48(6), 750–768.
- Langerhans, R. B., Layman, C. A., Shokrollahi, A. M., & DeWitt, T. J. (2004). Predator-driven phenotypic diversification in *Gambusia affinis*. *Evolution*, 58(10), 2305–2318.
- Layman, C. A., Winemiller, K. O., Arrington, D. A., & Jepsen, D. B. (2005). Body size and trophic position in a diverse tropical food web. *Ecology*, 86, 2530–2535.
- Lujan, N. K., & Armbruster, J. W. (2012). Morphological and functional diversity of the mandible in suckermouth armored catfishes (Siluriformes: Loricariidae). *Journal of Morphology*, 273, 24–39.
- Mac Nally, R., Duncan, R. P., Thomson, J. R., & Yen, J. D. L. (2018). Model selection using information criteria, but is the “best” model any good? *Journal of Applied Ecology*, 55, 1441–1444.
- Mancinelli, G., Vizzini, S., Mazzola, A., Maci, S., & Basset, A. (2013). Cross-validation of $\delta^{15}\text{N}$ and FishBase estimates of fish trophic position in a Mediterranean lagoon: The importance of the isotopic baseline. *Estuarine, Coastal and Shelf Science*, 135, 77–85.
- Mihalitsis, M., & Bellwood, D. R. (2017). A morphological and functional basis for maximum prey size in piscivorous fishes. *PLoS ONE*, 12, e0184679.
- Mittelbach, G. G., & Persson, L. (1998). The ontogeny of piscivory and its ecological consequences. *Canadian Journal of Fisheries and Aquatic Sciences*, 55(6), 1454–1465.
- Montaña, C. G., & Winemiller, K. O. (2013). Evolutionary convergence in Neotropical cichlids and Nearctic centrarchids: Evidence from morphology, diet, and stable isotope analysis. *Biological Journal of the Linnean Society*, 109, 146–164.
- Muschick, M., Indermaur, A., & Salzburger, W. (2012). Convergent evolution within an adaptive radiation of cichlid fishes. *Current Biology*, 22, 2362–2368.
- Nakazawa, T., Sakai, Y., Hsieh, C.-H., Koitabashi, T., Tayasu, I., Yamamura, N., & Okuda, N. (2010). Is the relationship between body size and trophic niche position time-invariant in a predatory fish? First stable isotope evidence. *PLoS ONE*, 5, e9120.
- Nilsson, D. E., Warrant, E. J., Johnsen, S., Hanlon, R., & Shashar, N. (2012). A unique advantage for giant eyes in giant squid. *Current Biology*, 22(8), 683–688.
- Ou, C., Montaña, C. G., & Winemiller, K. O. (2017). Body size–trophic position relationships among fishes of the lower Mekong basin. *Royal Society Open Science*, 4, 160645.
- Potapov, A., Brose, U., Scheu, S., & Tiunov, A. (2019). Trophic position of consumers and size structure of food webs across aquatic and terrestrial ecosystems. *The American Naturalist*, 194(6), 823–839.
- R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Riede, J. O., Brose, U., Ebenman, B., Jacob, U., Thompson, R., Townsend, C. R., & Jonsson, T. (2011). Stepping in Elton’s footprints: A general scaling model for body masses and trophic levels across ecosystems. *Ecology Letters*, 14, 169–178.
- Rip, J., & McCann, K. (2011). Cross-ecosystem differences in stability and the principle of energy flux. *Ecology Letters*, 14, 733–740.
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Aroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40, 913–929.
- Robinson, J. P., & Baum, J. K. (2015). Trophic roles determine coral reef fish community size structure. *Canadian Journal of Fisheries and Aquatic Sciences*, 73, 496–505.
- Romanuk, T. N., Hayward, A., & Hutchings, J. A. (2011). Trophic level scales positively with body size in fishes. *Global Ecology and Biogeography*, 20, 231–240.
- Romero-Romero, S., Molina-Ramírez, A., Höfer, J., & Acuña, J. L. (2016). Body size-based trophic structure of a deep marine ecosystem. *Ecology*, 97, 171–181.
- Ryo, M., & Rillig, M. C. (2017). Statistically reinforced machine learning for nonlinear patterns and variable interactions. *Ecosphere*, 8(11), 1–16.
- Sallan, L. C., & Friedman, M. (2011). Heads or tails: Staged diversification in vertebrate evolutionary radiations. *Proceedings of the Royal Society B: Biological Sciences*, 279, 2025–2032.
- Sánchez-Hernández, J., & Amundsen, P. A. (2018). Ecosystem type shapes trophic position and omnivory in fishes. *Fish and Fisheries*, 19, 1003–1015.

- Sazima, I. (1983). Scale-eating in characoids and other fishes. *Environmental Biology of Fishes*, 9(2), 87–101.
- Segura, A. M., Franco-Trecu, V., Franco-Fraguas, P., & Arim, M. (2015). Gape and energy limitation determine a humped relationship between trophic position and body size. *Canadian Journal of Fisheries and Aquatic Sciences*, 72, 198–205.
- Shine, R. (1991). Why do larger snakes eat larger prey items? *Functional Ecology*, 493–502.
- Sibly, R. M., Brown, J. H., & Kodric-Brown, A. (2012). *Metabolic ecology: A scaling approach*. John Wiley & Sons.
- Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, 9, 307.
- Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8, 25.
- Tedesco, P. A., Beauchard, O., Bigorne, R., Blanchet, S., Buisson, L., Conti, L., Cornu, J.-F., Dias, M. S., Grenouillet, G., Hugueny, B., Jézéquel, C., Leprieur, F., Brosse, S., & Oberdorff, T. (2017). A global database on freshwater fish species occurrence in drainage basins. *Scientific Data*, 4, 170141.
- Toussaint, A., Charpin, N., Brosse, S., & Villéger, S. (2016). Global functional diversity of freshwater fish is concentrated in the Neotropics while functional vulnerability is widespread. *Scientific Reports*, 6, 22125.
- Tucker, M. A., & Rogers, T. L. (2014). Examining predator–prey body size, trophic level and body mass across marine and terrestrial mammals. *Proceedings of the Royal Society B: Biological Sciences*, 281, 20142103.
- Van der Sleen, P., & Albert, J. S. (2017). *Field guide to the fishes of the Amazon, Orinoco, and Guianas*. Princeton Field Guides. Princeton University Press.
- Vander Zanden, M. J., & Fetzer, W. W. (2007). Global patterns of aquatic food chain length. *Oikos*, 116, 1378–1388.
- Vander Zanden, M. J., Vadeboncoeur, Y., & Chandra, S. (2011). Fish reliance on littoral–benthic resources and the distribution of primary production in lakes. *Ecosystems*, 14, 894–903.
- Wainwright, P. C. (2007). Functional versus morphological diversity in macroevolution. *Annual Review of Ecology, Evolution, and Systematics*, 60, 381–401.
- Wainwright, P. C., & Barton, R. A. (1995). Predicting patterns of prey use from morphology of fishes. *Environmental Biology of Fishes*, 44, 97–113.
- Westneat, M. W. (2004). Evolution of levers and linkages in the feeding mechanisms of fishes. *Integrative and Comparative Biology*, 44, 378–389.
- Winemiller, K. O. (1990). Spatial and temporal variation in tropical fish trophic networks. *Ecological Monographs*, 60, 331–367.

BIOSKETCH

The research team includes freshwater, marine and terrestrial ecologists with expertise in applied mathematics, functional diversity and macroecology.

R. Keller Kopf is a lecturer at Charles Darwin University. His research focuses on aquatic macroecology and the conservation of freshwater and marine biodiversity.

Jian D. L. Yen is a researcher at the University of Melbourne. He develops quantitative tools to support biodiversity management, with a particular focus on aquatic ecosystems.

Dale G. Nimmo is an associate professor at Charles Sturt University. His research focuses on ameliorating the impacts of large-scale disturbances on biodiversity. He listens to Townes Van Zandt.

Sébastien Brosse is a professor of Ecology at the University of Toulouse. He is interested in freshwater fish biodiversity patterns and processes across the globe and in the influence of global changes on fish distribution at both macroecological and local scales.

Sébastien Villéger is a researcher at the CNRS in Montpellier. He is interested in the functional ecology of fishes and, in particular, is studying the effect of global change on the functional diversity of coastal fish assemblages.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.

How to cite this article: Kopf RK, Yen JDL, Nimmo DG, Brosse S, Villéger S. Global patterns and predictors of trophic position, body size and jaw size in fishes. *Global Ecol Biogeogr*. 2021;30:414–428. <https://doi.org/10.1111/geb.13227>