Community-wide scan identifies fish species associated with coral reef services across the Indo-Pacific

Authors: Eva Maire[1,2,*], Sébastien Villéger[1], Nicholas A. J. Graham[3], Andrew S. Hoey[2], Joshua Cinner[2], Sebastian C. A. Ferse[4], Catherine Aliaume[1], David J. Booth[5], David A. Feary[6], Michel Kulbicki[7], Stuart A. Sandin[8], Laurent Vigliola[9] & David Mouillot[1,2]

Author affiliations:

[1] MARBEC, UMR IRD-CNRS-UM-IFREMER 9190, Université Montpellier, 34095 Montpellier Cedex, France.

[2] Australian Research Council Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, QLD 4811 Australia.

[3] Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK

[4] Leibniz Centre for Tropical Marine Ecology (ZMT), Fahrenheitstrasse 6, D-28359 Bremen, Germany

[5] School of Life Sciences, University of Technology, Sydney, New South Wales 2007, Australia

[6] Ecology & Evolution Group, School of Life Sciences, University Park, University of Nottingham, Nottingham NG7 2RD, UK

[7] UMR Entropie, Labex Corail, –IRD, Université de Perpignan, 66000 Perpignan, France

[8] Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093, USA

[9] Institut de Recherche pour le Développement, UMR IRD-UR-CNRS ENTROPIE, Laboratoire d'Excellence LABEX CORAIL, BP A5, 98848 Nouméa Cedex, New Caledonia

***Corresponding author:** Eva Maire

Contact: eva.maire@umontpellier.fr / +33 4 67 14 40 93

Address: Laboratoire Marine Biodiversity Exploitation and Conservation, Université Montpellier, Place Eugène Bataillon, 34095 Montpellier, France.

Author Contributions. D.M. conceived the study with support from E.M., S.V. and N.A.J.G.; E.M. developed and ran the analyses; E.M. wrote the first draft of the manuscript and all the authors made substantial contributions.

Data accessibility statement. If the manuscript is accepted, we confirm that the data supporting the results will be archived in an appropriate public repository such as Dryad.

Competing interests. The authors declare no competing interests.

Funding. Data collection was supported, amongst others, by NOAA’s Coral Reef Conservation Program and the US National Science Foundation for the Moorea Coral Reef LTER site (OCE 1637396, 1236905, 1026851, 0417412).

Article type: Research article

Keywords: ecosystem functioning, biodiversity, fish community, key species, ecosystem services

Number of words in the Abstract: 157

Number of words in main body of the paper: 4357

Number of figures: 5

Number of references: 61

ABSTRACT

Determining whether all species or only a subset are necessary to maintain ecosystem functioning and services is still an unresolved issue. Identifying such key species remains challenging, especially in the tropics where many species co-occur. Here, we develop a new community-wide scan (CWS) approach, analogous to the genome-wide scan, to identify fish species that contribute disproportionately to fish biomass and live coral cover across the Indo-Pacific. We found that only a limited set of species (51 out of ~400 or 13%), belonging to various functional groups and evolutionary lineages, are strongly and positively associated to fish biomass and live coral cover. Many of these species have not previously been identified as functionally important and thus, may be involved in unknown, yet important, biological mechanisms that help sustain healthy and productive coral reefs. CWS studies can uncover the species that contribute to ecosystem functioning and services and help guide experiments to decipher underlying ecological mechanisms.

MAIN TEXT

INTRODUCTION

Within the context of global changes and biodiversity loss, effective ecosystem management relies on a better understanding of the causal pathways between ecological communities and the myriad of services they sustain (Cardinale *et al.* 2012; Cheung *et al.* 2016; Ricketts *et al.* 2016; Ratcliffe *et al.* 2017). Experiments that manipulate community compositions have unambiguously demonstrated the positive effect of species diversity on ecosystem functioning over short and long timescales (Cardinale *et al.* 2006; Isbell *et al.* 2011; Isbell *et al.* 2015; Isbell *et al.* 2018). Recent studies have also convincingly shown that natural species-rich communities are more productive and can deliver higher rates of ecosystem services than impoverished communities (Grace *et al.* 2016; Duffy *et al.* 2017). Beyond the mere number of species, the diversity of species traits and evolutionary histories have been related to enhanced ecosystem functioning in both controlled experiments and natural communities (Cadotte *et al.* 2009; Flynn *et al.* 2011; Mora *et al.* 2014; Gross *et al.* 2017). In parallel, another line of evidence suggests that particular species are key to ecosystem functioning as they contribute disproportionally to certain processes when present (Bellwood *et al.* 2012; Reich 2012; Bozec *et al.* 2016; Tobner *et al.* 2016; Meyer *et al.* 2018). However, identifying key species remains highly challenging in diverse ecosystems, such as tropical reefs or rainforests, where many species co-occur and individual species can have multiple or unique contributions to ecosystem functions and services (Bozec *et al.* 2016; Pigot *et al.* 2016).

To tackle this challenge, ecologists can now use the increasing availability of extensive and standardized databases that have compiled environmental, social, biological and ecosystem service information across space and time (Cinner *et al.* 2016; Duffy *et al.* 2017). This emergence of large socio-ecological databases parallels what happened 20 years ago in genetics with advances in genome sequencing, generating millions of genetic variants for individual loci. To identify genetic variants among this myriad of sequences that are more frequent in people with a particular disease or traits of biomedical significance, genome-wide scan or genome-wide association studies (GWAS) were developed (Bush & Moore 2012). Such an approach is powerful to relate a given biological feature or trait to its underlying genetics, based on the simple idea that if a genetic variant increases the frequency of a given trait it should be more frequent in individuals with this trait than expected by chance (Visscher *et al.* 2017). Although this approach does not attribute causality, it can uncover previously unsuspected, yet important, potential biological mechanisms and pathways (McCarthy & Hirschhorn 2008). Although similar approaches have not been used in ecology, they hold much promise in empirical community ecology where only a few, among dozens or even hundreds of species (counterparts of genetic variants) can disproportionally drive ecosystem functioning and the delivery of services (counterparts of diseases, traits or phenotypes) (Bellwood *et al.* 2006; Straub & Snyder 2006; Bozec *et al.* 2016; Meyer *et al.* 2018). This approach could also reveal the unknown level of ecological pleiotropy in communities, i.e. the propensity that a single species can be key to many ecological functions and services (Hooper *et al.* 2005; Gascon *et al.* 2015). This term was initially coined by Strauss & Irwin (2004) [48] by analogy to genetic pleiotropy when one gene can influence two or more seemingly unrelated phenotypic traits. Under ecological pleiotropy a few species may underpin many different ecosystem functions or services and would deserve particular conservation actions.

Identifying functionally important or key species is particularly challenging in biodiverse ecosystems, due largely to the complexity of interactions among species and with their environment including human disturbances. For example, despite the large body of research on coral reefs, the identification of fish species that disproportionally drive ecosystem functioning is still in its infancy (Hoey & Bellwood 2009; Bellwood *et al.* 2012). The functional importance of most coral reef fishes is still poorly understood, and no study has scanned entire fish communities to detect potential links with ecosystem functioning and services at large scale. Here, we develop a new community-wide scan (CWS) approach, analogous to the GWAS approach, to identify key fish species that contribute to the delivery of services on coral reef ecosystems. Here ‘key’ has a different meaning than ‘keystone’ which corresponds to a “species whose effect is large, and disproportionately large relative to its abundance" (Paine 1966; Power *et al.* 1996). We define key species as those consistently and significantly associated, so above a certain statistical threshold, to a certain level of ecosystem functioning or services.

More precisely, we propose a statistical framework and use empirical data from 1,824 Indo-Pacific coral reefs hosting ~400 fish species to determine species whose presence disproportionately influences fish biomass and live coral cover. We then place those key species in a functional trait space (Villeger *et al.* 2008), and on a reef fish phylogeny to show the extent of species traits and evolutionary lineages that are necessary to sustain these two services on coral reefs. Identifying key species can provide new research priorities to elucidate ecological processes by which such candidate species positively affect coral reefs and to motivate a diversification of management options to maintain fish communities and their associated services in the face of a highly uncertain future.

MATERIAL AND METHOD

General framework

The Community-Wide Scan (CWS) framework to identify species that are associated with higher levels of ecosystem services involves three steps (Figure 1): i) collecting environmental, socio-economic, species abundance and/or biomass, and indicators of ecosystem services data across many sites; ii) modeling ecosystem services as functions of this large set of predictor variables (socio-economic, environmental conditions, and species richness). The relevance of this comprehensive initial, or reference, model (M_0) is validated according to its R^2 and its Akaike Information Criterion (AIC_{M0}); iii) testing the effect of each species separately on each ecosystem service beyond the effect of previous variables including species richness. For this, the presence of a given candidate species (binary variable) is added as an explanatory variable to M_0. The resulting model M_1 is evaluated according to its AIC (AIC_{M1K}). A species is declared as a potential key contributor to the ecosystem service if ΔAIC (AIC_{M0}-AIC_{M1k}) > 4 and if its partial effect is positive (Figure 1).

Coral reef data

Coral reef services. The proxies for coral reef services we considered are fish biomass and live coral cover which support, among many others, food security, shoreline protection and recreational value (Burke *et al.* 2011; Kittinger *et al.* 2012; Harris *et al.* 2018). Coral cover and fish biomass are already monitored at the global scale using visual censuses as well as underwater video surveys (e.g. Reef Life Survey, Catlin Sea Survey) and are highly sensitive to local human activities (e.g. fishing, habitat destruction, pollution) and global climate change (Hughes *et al.* 2018) and thus, can be considered as key variables for the health and productivity of coral reefs (Bozec *et al.* 2016; Cinner *et al.* 2016; Hughes *et al.*

2017). We used data from 1,824 coral reefs in 26 nations, states, or territories located across the Indo-Pacific which include fish biomass and live coral cover estimates.

Fish Biomass. Reef fish biomass estimates were based on instantaneous visual counts from 4,694 surveys collected from 1,824 reefs (Fig. S1). All surveys used two census methods (belt-transects or distance sampling) and were conducted between 2004 and 2013. On average 2.4 transects (sd: 1.32; 1-10) were performed on each reef. Within each survey area, reef associated fishes were identified to species level, abundance counted, and total length (TL) estimated.

To make estimates of biomass from these transect-level data comparable among studies, we:

i) Considered only Indo-Pacific reefs and retained families that were consistently included in surveys and were above a minimum size cut-off. Thus, we retained counts of non-cryptic reef fish species >10cm in total length, that are reef-associated (30 families, 748 species) (Table S1). We did not include sharks as they were often excluded from visual surveys. We calculated total biomass of fishes on each reef using published species-level length-weight relationship parameters or those available on FishBase (Froese & Pauly 2014). When length-weight relationship parameters were not available for a species, we used the parameters for a closely related species or genus.

ii) Directly accounted for depth and habitat as covariates in the model (see “environmental drivers” section in Supplementary Material).

iii) Accounted for any potential bias among census method by including census method and sampling area as covariates in the model.

Biomass values were calculated at the reef scale and demonstrates a wide distribution (mean: 1,055 kg.ha^{-1}; 2-25,910).

Coral cover. Percent cover of live coral was based on 1,715 point-intercept transects or quadrats collected from 741 reefs (Fig. S2). All surveys were conducted between 2008 and 2013. On average 2.3 transects (sd: 1.03; 1-4) were performed on each reef. To make estimates of coral cover from these transect-level data comparable among studies, we included depth, habitat and census method as covariates in the model (see “drivers” section in Supplementary Material). Coral cover values were calculated at the reef scale and demonstrates a wide distribution (mean: 27%; 3-94).

Initial models and species candidates. For each of the 1,824 reefs located in the Indo-Pacific we collected and used 12 relevant social and environmental variables (listed below), together with the occurrence, abundance and size of 748 reef fish species (Cinner *et al.* 2016). To provide the initial model (M_0) and reference Akaike Information Criterion (AIC_{M0}) we modeled fish biomass and live coral cover using linear mixed models (LMM) with the complete set of socio-economic and environmental conditions plus species richness as predictor variables. For each of the 748-fish species present in this dataset, we calculated its occurrence on each reef. Then, we computed the number of reefs where a fish species occurs and excluded those present on less than 1% (threshold commonly used for rare species) of the reefs so 18 and 7 reefs for fish biomass and coral cover, respectively. Thus, 381 fish species grouped into 116 genera and 30 families were considered as potential candidate species (binary variable as presence/absence in the model M_1).

Identifying potential key contributors to ecosystem services. Each of the 381 species were tested as candidates for improving prediction of reef fish biomass and live coral cover given the socioeconomic and environmental conditions. More precisely, we tested presence of each candidate species as an additional explanatory binary variable to M_0 to compute model M_1 and its associated AIC (AIC_{M1K}). Finally, a species was identified as a key contributor to the ecosystem service if, when included, $\Delta AIC > 4$ and if its partial effect was positive (positive coefficient in the model). The binary variable describing the presence/absence of species was computed according to its occurrence (i.e. presence of at least 1 individual) but could be also determined using any relative abundance threshold (Fig. 1 and supplemental materials).

Environmental and socio-economic variables. The variables included in the models were environmental: 1) oceanic productivity, 2) habitat type, 3) depth and socio-economic: 4) Management 5) local human population growth rate, 6) gravity of local population, 7) gravity of markets, 8) levels of human development (Human Development Index), 9) human population size, 10) levels of tourism, 11) degree of voice and accountability of citizens, and 12) reef fish landings (tons)/km^2 of reef (definitions and detail on data collection for these variables are provided in the Supplementary Material).

Statistical analyses. We first computed two linear mixed models (LMM), which predicted fish biomass and live coral cover respectively, while accounting for the different scales at which the data were collected as random effects (reef location, site, and nation/state, see Supplementary Material), 11 key environmental and socio-economic variables expected to influence reef conditions (Cinner *et al.* 2016) and fish species richness as fixed effects (Supplementary Material). To evaluate the fit of the two linear mixed models, we checked the relationship between observed and predicted values. Model validation and quality control procedures are described in the Supplementary Material.

In order to quantify the potential net benefit of each identified key species, we extracted the net effect of each key species for biomass and live coral cover using a partial plot from linear mixed models while the other variables were held constant.

We next investigated whether reefs with several key species show high levels of fish biomass and live coral cover. To control for the effects of species richness we compared modeled estimates of fish biomass and live coral between reefs while increasing the number of key species. We defined the number of key species for every reef and chose the richest quartile as a threshold (i.e. 4 and 6 key species for biomass and live coral cover respectively). We next created 3 categories of reefs: those with no key species, those with at least one key species but below the richness threshold (4 and 6 for biomass and live coral cover respectively) and those with more key species than the threshold.

Functional space and entities. The 381-fish candidates species were functionally described using six traits: (1) size, coded using 3 ordered categories: 10-30cm, 30.1-50cm, >50cm ; (2) mobility, coded using 3 ordered categories: sedentary, mobile within a reef and mobile between reefs; (3) period of activity, coded using 3 ordered categories: diurnal, both diurnal and nocturnal, and nocturnal ; (4) schooling, coded using 5 ordered categories: solitary, pairing, or living in small (3-20 individuals), medium (20-50 individuals) or large groups (>50 groups) ; (5)

vertical position in the water column, coded using 3 ordered categories: benthic, bentho-pelagic and pelagic; (6) diet, coded using 7 trophic categories: herbivorous-detritivorous, macro-algal browser, invertivorous targeting sessile invertebrates, invertivorous targeting mobile invertebrates, planktivorous, piscivorous, and omnivorous, i.e. fish for which both plant and animal material are important in their diet. Values for these six traits were taken from the global trait database on tropical reef fishes from Mouillot *et al.* (2014) (Mouillot *et al.* 2014). Since all traits were categorical, species with identical traits were grouped into functional entities. Most functional entities comprise species from different genera (Mouillot *et al.* 2014).

We assessed the functional richness (FRic), e.g. the functional space occupied by the key fish species for biomass and coral cover respectively, using the convex hull volume index proposed by Cornwell *et al.* (2006) (Cornwell *et al.* 2006). This volume corresponds to the amount of functional space filled by key species, where axes are defined by species traits.

Fish phylogeny. We used a time-calibrated phylogeny of Acanthomorph fishes (Near *et al.* 2013) which covers the 33 major reef fish families. (i.e. Labridae, Chaetodontidae, Gobiidae, Blenniidae, Apogonidae, Epinephelinae, Serranidae, Pomacentridae, Acanthuridae, Balistidae, Lutjanidae, Callionymidae, Carangidae, Cirrhitidae, Haemulidae, Hemiramphidae, Holocentridae, Lethrinidae, Monacanthidae, Mullidae, Nemipteridae, Ostraciidae, Pempheridae, Pinguipedidae, Platycephalidae, Pomacanthidae, Pseudochromidae, Scorpaenidae, Siganidae, Sparidae, Synanceiidae, Tetraodontidae, Tripterygiidae). Some fish genera (e.g. *Elagatis* and *Parupeneus* for example) recorded on reefs were missing in this phylogeny.

RESULTS

Predictability of fish biomass and coral cover

The two initial (M0) models explained 79% and 61% of the variance in fish biomass and live coral cover, respectively (Fig. S3 and Supplementary Material). The residuals of the two models were normally distributed (Fig S3). In total, 8 and 6 variables had the highest importance (Akaike weight = 1) to predict fish biomass and live coral cover respectively (see Tables S2 & S3). Fish species richness, oceanic productivity, population size, tourism and census method were the main predictors of both fish biomass and coral cover. Depth, management, and sampling area were also important predictors of fish biomass while habitat type was important to predict coral cover (see Tables S2 & S3).

Key species associated to reef fish biomass

Among the 381 fish species considered as candidates, only 26 species (7%) were significantly related to fish biomass beyond the initial set of variables (ΔAIC > 4 and positive effect) considering their presence (at least 1 individual) (Table S4). Those 26 key species covered a wide breadth of phylogenetic lineages (Fig. 2), representing 16 out of 116 genera and 8 out of 31 families (i.e. Acanthuridae, Carangidae, Labridae, Lethrinidae, Lutjanidae, Mullidae, Scombridae, Serranidae).

When considering functional traits, we found that those 26 key species represented 24 different functional entities (Table S6) demonstrating a very low functional redundancy with 1.1 species per functional entity (range 1-2). In addition, key species had contrasting functional traits with all body sizes (from 10cm to > 50cm) and all diets (7 trophic categories) represented (Table S6). Together these 26 key species filled 20% of the whole functional space defined by the 240 functional entities corresponding to the 381 candidate species (FRic = 0.20; Fig. 3).

Key species associated to live coral cover

We found that 28 reef fish species out of 381 (7%), were significantly and positively related to coral cover (ΔAIC > 4) considering their presence (at least 1 individual) (Table S5). Those 28 key species also encompassed a wide breadth of phylogenetic lineages (Fig. 2), representing 15 out of 116 genera and 8 out of 31 families (i.e. Acanthuridae, Chaetodontidae, Cirrhitidae, Haemulidae, Labridae, Lutjanidae, Monacanthidae, Serranidae).

When considering functional traits, we found that the 28 key fish species were distributed among 17 different functional entities (Table S6). Key fish species for coral cover showed some degree of functional redundancy with, on average, 1.6 key species per functional entity (range 1-6). This higher functional redundancy translated into a more restricted functional space filled by these key species (only 5% with FRic = 0.05; Fig. 3). Species of all sizes (from 10cm to > 50cm) and almost all diets (6 diet categories out of 7) were significantly associated to live coral cover. However, large mobile predators and large herbivorous fishes were not considered as key for live coral cover (Table S6).

Low overlap between species key to fish biomass and coral cover

The two sets of key fish species associated to total fish biomass and live coral cover (26 and 28 key species, respectively) represented all together less than 10% of the 381 fish species tested as candidates. Only three species (*Acanthurus albipectoralis, Lutjanus bohar, Lutjanus gibbus)*, four genera (*Acanthurus, Chlorurus, Lutjanus, and Scarus*) and four families (Acanthuridae, Labridae, Lutjanidae and Serranidae) were common to both sets and thus, significantly associated to the two proxies of coral reef services (Fig. 2, Table S6).

Only six functional entities were common and significantly associated to both biomass and live coral cover (Fig. 3), namely small and medium herbivores, small planktivores, medium and large fishes targeting mobile invertebrates and meso-predators (Table S6).

The net benefit of key species for fish biomass and live coral cover

When present, each key species belonged to a community with a median level of fish biomass higher (560 kg/ha, range 439-773 kg/ha) than the median biomass observed when absent (370 kg/ha, range 337-385). Similarly, live coral cover was estimated at a median value of 50% (range 36-82) when each key species was present against 34% (range 26-38) when absent (Fig. 4). For clarity, we only presented the net effect of the four most significant key species (lowest AIC compared to AIC(M_0), see Tables S4 & S5) associated to biomass and live coral cover (Fig. 4). It is important to note that these four most significant key species (lowest AIC) did not necessarily provide the highest level of biomass and coral cover. Highest levels of biomass were obtained when either *Elagatis bipinnulata, Naso tuberosus, Bolbometopon muricatum* or *Plectropomus laevis* were present in the model, while highest levels of coral cover were reached when either *Plectorhinchus picus, Acanthurus leucosternon, Chaetodon guttatissimus* or *Chaetodon meyeri* were present (Fig S4).

It is not only individual key species, but also the accumulation of key species that promoted high levels of ecosystem services. For instance, reefs with more than four key species reached a median level of biomass of 1,150 kg/ha (range 362-3715), i.e. three times the median biomass observed (370 kg/ha, range 86-1380) in reefs with an intermediate number of key species (from 1 to 3 key species) and more than seven times higher than the median level of fish biomass reached in reefs having no key species (156 kg/ha, range 12-812). Although less pronounced, reefs with at least 6 key fish species showed a median live coral cover of 40% (range 20-68) while reefs with no key species had a median level of 31% (range 18-54) live coral cover (Fig. 5).

DISCUSSION

Sustaining healthy and productive coral reefs

In the present study, many different fish species (~400 species candidates) were scanned and only 26 and 28 species were identified as positively related to fish biomass and live coral cover respectively, with only three species being common to both. In total, these 51 species (i.e. ~13% of the species pool tested), represent 35 distinct functional entities (out of 240 so 15%) and are widespread in the functional space. While large-bodied species may be expected to disproportionately contribute to fish biomass, our results indicate that only 25% (7 out of 26, see Table S6) of key species for fish biomass were large-bodied

(>50cm), which is directly comparable to the percentage of large-bodied species among the initial candidate species (20% or 49 functional entities out of 240). We also show that 35% of key species for fish biomass are smaller than 30cm (9 out of 26, see Table S6). The positive association with fish biomass is thus independent of body size.

It comes as no surprise that some key fish species identified in this study have already aroused considerable interest in coral reef ecology. Herbivorous fish support coral reef resilience by controlling algal growth, influencing competitive interactions between corals and macroalgae, and preventing coral-algal phase shifts (Bellwood *et al.* 2004; Mumby *et al.* 2006; Hughes *et al.* 2007; Rasher *et al.* 2013; Graham *et al.* 2015; Bozec *et al.* 2016), and therefore may contribute to the maintenance of high coral cover and fish biomass. In particular, scarinine parrotfishes (i.e., *Bolbometopon, Chlorurus, Hipposcarus* and *Scarus,* see Tables S4,S5&S6) play critical roles as grazers and bioeroders of the reef substratum (Bellwood *et al.* 2003; Mumby 2006), and their abundances have strong positive effects on cover of corals and hence accretion rates of the reef (Cramer *et al.* 2017). Further, grazing and detritivorous acanthurids (i.e., *Acanthurus* and *Ctenochaetus*, see Tables S4,S5&S6) intensely graze epilithic algal turfs (Marshell & Mumby 2012; Rasher *et al.* 2013), while benthic-feeding unicornfishes (i.e., *Naso,* see Tables S4&S6) play a significant role in macroalgal removal (Fox & Bellwood 2008; Hoey & Bellwood 2009).

By contrast, some key species identified in the present study have not previously been identified as playing significant roles. While predation is a key process shaping prey behavior and populations (Rasher *et al.* 2017), structuring ecological communities (Boaden & Kingsford 2015), and promoting nutrient capacity and primary production (Allgeier *et al.* 2016; Allgeier *et al.* 2017), no individual predator species have been expressly identified as beneficial for total fish biomass and coral cover. Here, we show that the predatory species of *Aprion, Caranx, Cephalopholis, Elagatis, Gymnosarda, Lethrinus, Lujtanus, Oxycheilinus,* and *Plectropomus* (Fig 4, Tables S4&S5, Fig S4) may play a critical role for fish biomass and live coral cover, although the exact pathways through which they act remain to be elucidated.

Low ecological pleiotropy on coral reefs

The finding that a limited number of functionally and evolutionary different species are positively related to high levels of fish biomass and coral cover (Fig. 2,3&S4) supports the idea that sustaining ecosystem services may require a large breadth of particular attributes beyond the number of species (Soliveres *et al.* 2016; Gross *et al.* 2017). The limited overlap between the two sets of species significantly associated to two proxies of key services (3 species, 4 genera and 6 functional entities) suggests a low level of ecological pleiotropy (Strauss & Irwin 2004), i.e. that a single species, genus or functional entity cannot be key to many independent ecosystem functions and services. Extended to the community level, we show that ecological pleiotropy, the exact opposite of functional redundancy, is not the norm on coral reefs. This finding explains why the multi-functionality of ecosystems relies more strongly on biodiversity than do single functions (Isbell *et al.* 2011; Lefcheck *et al.* 2015; Soliveres *et al.* 2016; Gross *et al.* 2017), since some species play unique and thus irreplaceable roles in ecosystems (Bellwood *et al.* 2006; Petchey *et al.* 2008). We suggest that this ecological pleiotropy reconciles two opposing views in Biodiversity and Ecosystem Functioning **(**BEF) research since many complementary species groups and lineages, and hence a large amount of biodiversity, are necessary to sustain ecosystem multi-functionality and associated services. Once combined, those key species appear to provide high benefits in terms of fish biomass and live coral cover (Fig 5). Maintaining habitat heterogeneity and associated processes is thus a major component of management and conservation. Our results call for more species-focused management strategies such as the banning of fishing species considered as key for the ecosystem (Bozec *et al.* 2016). Alternatively, sustaining multi-functionality also requires a broader portfolio approach which may reduce local extinction risk by securing the biodiversity level in an increasingly uncertain future (Webster *et al.* 2017).

CWS as a flexible framework to link biodiversity to ecosystem functioning and services

The community-wide scan (CWS) approach can be adapted for a wide range of ecosystems, combinations of taxa or interactions and services. Here we only tested the presence of key species, while it would be possible to look for key species groups (pairs or more), key evolutionary lineages or even key biotic interactions. For instance, interactions among local neighbors promote overyielding in species mixtures of tree communities (Fichtner *et al.* 2018). Since those interactions are potentially multiple in species-rich communities they cannot be experimentally tested, but they can emerge from empirical data using the WCS approach. In the same vein, positive effects of some species may only be revealed beyond some thresholds. For instance, herbivorous fish are both important for standing biomass and ecosystem resilience, and it has been shown that maintaining high fish abundance may increase behavioral coupling and thus, promote the consumption of algae by fish (Gil & Hein 2017) while maintaining high parrotfish biomass and implementing a minimum harvest size may be critical in sustaining the health and the productivity of coral reefs (Bozec *et al.* 2016). The way candidates are tested can be modulated while respecting independence between predicted and explanatory variables. For example, presence data can be determined by any abundance threshold such as a minimum number of individuals, cover rate, biomass or level of interactions.

On coral reefs, defining species presence based on intraspecific biomass distribution (using upper percentiles or deciles) can promote small-bodied species but can also disadvantage species that are not commonly encountered or have skewed biomass distributions since species need to occur sufficiently to be tested (see supplemental materials and Table S7-10). Rather, defining species presence as a proportion of the intracommunity biomass can be applied independently of the species biomass distribution, but large species may be more likely to reach the biomass threshold. However, we found consistent results between these two procedures since the majority of species detected as key species using the intracommunity approach are also significant using the intraspecific approach reinforcing the robustness of our findings.

The CWS framework thus offers a new and flexible way to analyze empirical data relating biodiversity to ecosystem functioning and services. In genetics, the greatest challenge is no longer the identification of the association signals themselves, but the discovery of molecular mechanisms through which genes influence disease risk and phenotypic expression. In ecology, CWS studies can only be considered as initial forays into a better understanding of the complex relationships between particular species, species groups or interactions and ecosystem functioning and associated services. A large number of false-positives, species being detected as key while they are not, may be revealed and no causality is determined in this approach; the main merit is to identify unsuspected and statistically significant positive associations. The logical progression would be to conduct experiments focusing on potential key species or interactions with the ultimate aim of highlighting the underlying ecological or biological processes and pathways that potentially sustain healthy and productive ecosystems.

CONCLUSION

The community-wide scan (CWS) approach has the potential to reveal unsuspected contributions to ecosystem functioning and its associated services, especially in complex and biodiverse ecosystems where the detection of such contributions remains challenging. Using coral reefs as our model system, we have identified a phylogenetically and functionally wide set of fish species strongly associated with high levels of fish biomass and live coral cover. While these findings cannot be interpreted as causative, they highlight a limited pool of species (58 fish species), many of which appear to be involved in unsuspected, yet important, biological mechanisms and pathways that could sustain healthy and productive reefs upon which human welfare depends. The CWS approach holds much promise in empirical BEF studies where only a few species, functional or phylogenetic groups, can disproportionally drive ecosystem functioning and the delivery of services. Ultimately, the key species identified form tractable conservation targets, for example the implementation of bans on fishing species considered as key for the ecosystem (Bozec *et al.* 2016). Given the growing interest in the assessment and consequences of the ongoing extinction crisis on ecosystem functioning, such a framework is extremely timely and widely applicable. Our framework offers a new and flexible way to analyze the ongoing massive empirical data relating biodiversity to ecosystem functioning and services with the potential to reconcile two opposing views: species identity vs. diversity.

Acknowledgments

We thank M. Beger, E. Brokovich, A. J. Brooks, P. Chabanet, A. Green, C. Gough, C. Mora, M. H. Tupper and I. Williams for compiling and providing essential data for this study. We are grateful to M. Troussellier for his great support and A-S. Tribot for providing many fish silhouettes used in the figures.

References

1.

Allgeier, J.E., Burkepile, D.E. & Layman, C.A. (2017). Animal pee in the sea: consumer-mediated nutrient dynamics in the world's changing oceans. *Global Change Biology*, 23, 2166-2178.

2.

Allgeier, J.E., Valdivia, A., Cox, C. & Layman, C.A. (2016). Fishing down nutrients on coral reefs. *Nat Commun*, 7, 12461.

3.

Bellwood, D.R., Hoey, A.S. & Choat, J.H. (2003). Limited functional redundancy in high diversity systems: resilience and ecosystem function on coral reefs. *Ecology Letters*, 6, 281-285.

4.

Bellwood, D.R., Hoey, A.S. & Hughes, T.P. (2012). Human activity selectively impacts the ecosystem roles of parrotfishes on coral reefs. *Proceedings of the Royal Society B-Biological Sciences*, 279, 1621-1629.

5.

Bellwood, D.R., Hughes, T.P., Folke, C. & Nystrom, M. (2004). Confronting the coral reef crisis. *Nature*, 429, 827-833.

6.

Bellwood, D.R., Hughes, T.P. & Hoey, A.S. (2006). Sleeping functional group drives coral-reef recovery. *Current Biology*, 16, 2434-2439.

7.

Boaden, A.E. & Kingsford, M.J. (2015). Predators drive community structure in coral reef fish assemblages. *Ecosphere*, 6, 1-33.

8.

Bozec, Y.-M., O'Farrell, S., Bruggemann, J.H., Luckhurst, B.E. & Mumby, P.J. (2016). Tradeoffs between fisheries harvest and the resilience of coral reefs. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 4536-4541.

9.

Burke, L., Reytar, K., Spalding, M. & Perry, A. (2011). *Reefs at risk revisited*.

10.

Bush, W.S. & Moore, J.H. (2012). Chapter 11: Genome-Wide Association Studies. *PLOS Computational Biology*, 8, e1002822.

11.

Cadotte, M.W., Cavender-Bares, J., Tilman, D. & Oakley, T.H. (2009). Using Phylogenetic, Functional and Trait Diversity to Understand Patterns of Plant Community Productivity. *Plos One*, 4.

12.

Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P. *et al.* (2012). Biodiversity loss and its impact on humanity. *Nature*, 486, 59-67.

13.

Cardinale, B.J., Srivastava, D.S., Duffy, J.E., Wright, J.P., Downing, A.L., Sankaran, M. *et al.* (2006). Effects of biodiversity on the functioning of trophic groups and ecosystems. *Nature*, 443, 989-992.

14.

Cheung, W.W.L., Reygondeau, G. & Frölicher, T.L. (2016). Large benefits to marine fisheries of meeting the 1.5°C global warming target. *Science*, 354, 1591.

15.

Cinner, J.E., Huchery, C., MacNeil, M.A., Graham, N.A.J., McClanahan, T.R., Maina, J. *et al.* (2016). Bright spots among the world's coral reefs. *Nature*, 535, 416-419.

16.

Cornwell, W.K., Schwilk, D.W. & Ackerly, D.D. (2006). A trait-based test for habitat filtering: Convex hull volume. *Ecology*, 87, 1465-1471.

17.

Cramer, K.L., O'Dea, A., Clark, T.R., Zhao, J.-x. & Norris, R.D. (2017). Prehistorical and historical declines in Caribbean coral reef accretion rates driven by loss of parrotfish. *Nat Commun*, 8, 14160.

18.

Duffy, J.E., Godwin, C.M. & Cardinale, B.J. (2017). Biodiversity effects in the wild are common and as strong as key drivers of productivity. *Nature*, 549, 261.

19.

Fichtner, A., Härdtle, W., Bruelheide, H., Kunz, M., Li, Y. & von Oheimb, G. (2018). Neighbourhood interactions drive overyielding in mixed-species tree communities. *Nat Commun*, 9, 1144.

20.

Flynn, D.F.B., Mirotchnick, N., Jain, M., Palmer, M.I. & Naeem, S. (2011). Functional and phylogenetic diversity as predictors of biodiversity–ecosystem-function relationships. *Ecology*, 92, 1573-1581.

21.

Fox, R.J. & Bellwood, D.R. (2008). Remote video bioassays reveal the potential feeding impact of the rabbitfish Siganus canaliculatus (f: Siganidae) on an inner-shelf reef of the Great Barrier Reef. *Coral Reefs*, 27, 605-615.

22.

Froese, R. & Pauly, D. (2014). FishBase. version (2/2011). *World Wide Web electronic publication. Available at: http://www.fishbase. org (accessed 30 May 2014).*

23.

Gascon, C., Brooks, Thomas M., Contreras-MacBeath, T., Heard, N., Konstant, W., Lamoreux, J. *et al.* (2015). The Importance and Benefits of Species. *Current Biology*, 25, R431-R438.

24.

Gil, M.A. & Hein, A.M. (2017). Social interactions among grazing reef fish drive material flux in a coral reef ecosystem. *Proceedings of the National Academy of Sciences*, 114, 4703.

25.

Grace, J.B., Anderson, T.M., Seabloom, E.W., Borer, E.T., Adler, P.B., Harpole, W.S. *et al.* (2016). Integrative modelling reveals mechanisms linking productivity and plant species richness. *Nature*, 529, 390-+.

26.

Graham, N.A.J., Jennings, S., MacNeil, M.A., Mouillot, D. & Wilson, S.K. (2015). Predicting climate-driven regime shifts versus rebound potential in coral reefs. *Nature*, 518, 94-97.

27.

Gross, N., Bagousse-Pinguet, Y.L., Liancourt, P., Berdugo, M., Gotelli, N.J. & Maestre, F.T. (2017). Functional trait diversity maximizes ecosystem multifunctionality. *Nature Ecology & Evolution*, 1.

28.

Harris, D.L., Rovere, A., Casella, E., Power, H., Canavesio, R., Collin, A. *et al.* (2018). Coral reef structural complexity provides important coastal protection from waves under rising sea levels. *Science Advances*, 4.

29.

Hoey, A.S. & Bellwood, D.R. (2009). Limited Functional Redundancy in a High Diversity System: Single Species Dominates Key Ecological Process on Coral Reefs. *Ecosystems*, 12, 1316-1328.

30.

Hooper, D.U., Chapin, F.S., Ewel, J.J., Hector, A., Inchausti, P., Lavorel, S. *et al.* (2005). Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. *Ecological Monographs*, 75, 3-35.

31.

Hughes, T.P., Anderson, K.D., Connolly, S.R., Heron, S.F., Kerry, J.T., Lough, J.M. *et al.* (2018). Spatial and temporal patterns of mass bleaching of corals in the Anthropocene. *Science*, 359, 80.

32.

Hughes, T.P., Kerry, J.T., Alvarez-Noriega, M., Alvarez-Romero, J.G., Anderson, K.D., Baird, A.H. *et al.* (2017). Global warming and recurrent mass bleaching of corals. *Nature*, 543, 373-377.

33.

Hughes, T.P., Rodrigues, M.J., Bellwood, D.R., Ceccarelli, D., Hoegh-Guldberg, O., McCook, L. *et al.* (2007). Phase shifts, herbivory, and the resilience of coral reefs to climate change. *Current Biology*, 17, 360-365.

34.

Isbell, F., Calcagno, V., Hector, A., Connolly, J., Harpole, W.S., Reich, P.B. *et al.* (2011). High plant diversity is needed to maintain ecosystem services. *Nature*, 477, 199-U196.

35.

Isbell, F., Cowles, J., Dee Laura, E., Loreau, M., Reich Peter, B., Gonzalez, A. *et al.* (2018). Quantifying effects of biodiversity on ecosystem functioning across times and places. *Ecology Letters*, 0.

36.

Isbell, F., Craven, D., Connolly, J., Loreau, M., Schmid, B., Beierkuhnlein, C. *et al.* (2015). Biodiversity increases the resistance of ecosystem productivity to climate extremes. *Nature*, 526, 574-577.

37.

Kittinger, J., Finkbeiner, E., Glazier, E. & Crowder, L. (2012). Human dimensions of coral reef social-ecological systems. *Ecology and Society*, 17.

38.

Lefcheck, J.S., Byrnes, J.E.K., Isbell, F., Gamfeldt, L., Griffin, J.N., Eisenhauer, N. *et al.* (2015). Biodiversity enhances ecosystem multifunctionality across trophic levels and habitats. *Nat Commun*, 6, 6936.

39.

Marshell, A. & Mumby, P.J. (2012). Revisiting the functional roles of the surgeonfish Acanthurus nigrofuscus and Ctenochaetus striatus. *Coral Reefs*, 31, 1093-1101.

40.

McCarthy, M.I. & Hirschhorn, J.N. (2008). Genome-wide association studies: potential next steps on a genetic journey.

41.

Meyer, S.T., Ptacnik, R., Hillebrand, H., Bessler, H., Buchmann, N., Ebeling, A. *et al.* (2018). Biodiversity–multifunctionality relationships depend on identity and number of measured functions. *Nature Ecology & Evolution*, 2, 44-49.

42.

Mora, C., Danovaro, R. & Loreau, M. (2014). Alternative hypotheses to explain why biodiversity-ecosystem functioning relationships are concave-up in some natural ecosystems but concave-down in manipulative experiments. *Scientific Reports*, 4, 5427.

43.

Mouillot, D., Villeger, S., Parravicini, V., Kulbicki, M., Ernesto Arias-Gonzalez, J., Bender, M. *et al.* (2014). 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, 13757-13762.

44.

Mumby, P.J. (2006). The Impact Of Exploiting Grazers (Scaridae) On The Dynamics Of Caribbean Coral Reefs. *Ecological Applications*, 16, 747-769.

45.

Mumby, P.J., Dahlgren, C.P., Harborne, A.R., Kappel, C.V., Micheli, F., Brumbaugh, D.R. *et al.* (2006). Fishing, trophic cascades, and the process of grazing on coral reefs. *Science*, 311, 98-101.

46.

Near, T.J., Dornburg, A., Eytan, R.I., Keck, B.P., Smith, W.L., Kuhn, K.L. *et al.* (2013). Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 12738-12743.

47.

Paine, R.T. (1966). Food Web Complexity and Species Diversity. *Am Nat*, 100, 65-+.

48.

Petchey, O.L., Eklof, A., Borrvall, C. & Ebenman, B. (2008). Trophically unique species are vulnerable to cascading extinction. *Am Nat*, 171, 568-579.

49.

Pigot, A.L., Bregman, T., Sheard, C., Daly, B., Etienne, R.S. & Tobias, J.A. (2016). Quantifying species contributions to ecosystem processes: a global assessment of functional trait and phylogenetic metrics across avian seed-dispersal networks. *Proceedings of the Royal Society B: Biological Sciences*, 283.

50.

Power, M.E., Tilman, D., Estes, J.A., Menge, B.A., Bond, W.J., Mills, L.S. *et al.* (1996). Challenges in the quest for keystones. *Bioscience*, 46, 609-620.

51.

Rasher, D.B., Hoey, A.S. & Hay, M.E. (2013). Consumer diversity interacts with prey defenses to drive ecosystem function. *Ecology*, 94, 1347-1358.

52.

Rasher, D.B., Hoey, A.S. & Hay, M.E. (2017). Cascading predator effects in a Fijian coral reef ecosystem. *Scientific Reports*, 7, 15684.

53.

Ratcliffe, S., Wirth, C., Jucker, T., van der Plas, F., Scherer-Lorenzen, M., Verheyen, K. *et al.* (2017). Biodiversity and ecosystem functioning relations in European forests depend on environmental context. *Ecol Lett*, 20, 1414-1426.

54.

Reich, P.B. (2012). Key canopy traits drive forest productivity. *Proceedings of the Royal Society B: Biological Sciences*, 279, 2128.

55.

Ricketts, T.H., Watson, K.B., Koh, I., Ellis, A.M., Nicholson, C.C., Posner, S. *et al.* (2016). Disaggregating the evidence linking biodiversity and ecosystem services. *Nat Commun*, 7.

56.

Soliveres, S., van der Plas, F., Manning, P., Prati, D., Gossner, M.M., Renner, S.C. *et al.* (2016). Biodiversity at multiple trophic levels is needed for ecosystem multifunctionality. *Nature*, 536, 456-459.

57.

Straub, C.S. & Snyder, W.E. (2006). Species identity dominates the relationship between predator biodiversity and herbivore suppression. *Ecology*, 87, 277-282.

58.

Strauss, S.Y. & Irwin, R.E. (2004). Ecological and Evolutionary Consequences of Multispecies Plant-Animal Interactions. *Annual Review of Ecology, Evolution, and Systematics*, 35, 435-466.

59.

Tobner, C.M., Paquette, A., Gravel, D., Reich, P.B., Williams, L.J. & Messier, C. (2016). Functional identity is the main driver of diversity effects in young tree communities. *Ecology Letters*, 19, 638-647.

60.

Villeger, S., Mason, N.W.H. & Mouillot, D. (2008). New multidimensional functional diversity indices for a multifaceted framework in functional ecology. *Ecology*, 89, 2290-2301.

61.

Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A. *et al.* (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation.

62.

Webster, M.S., Colton, M.A., Darling, E.S., Armstrong, J., Pinsky, M.L., Knowlton, N. *et al.* (2017). Who Should Pick the Winners of Climate Change? *Trends in Ecology & Evolution*, 32, 167-173.

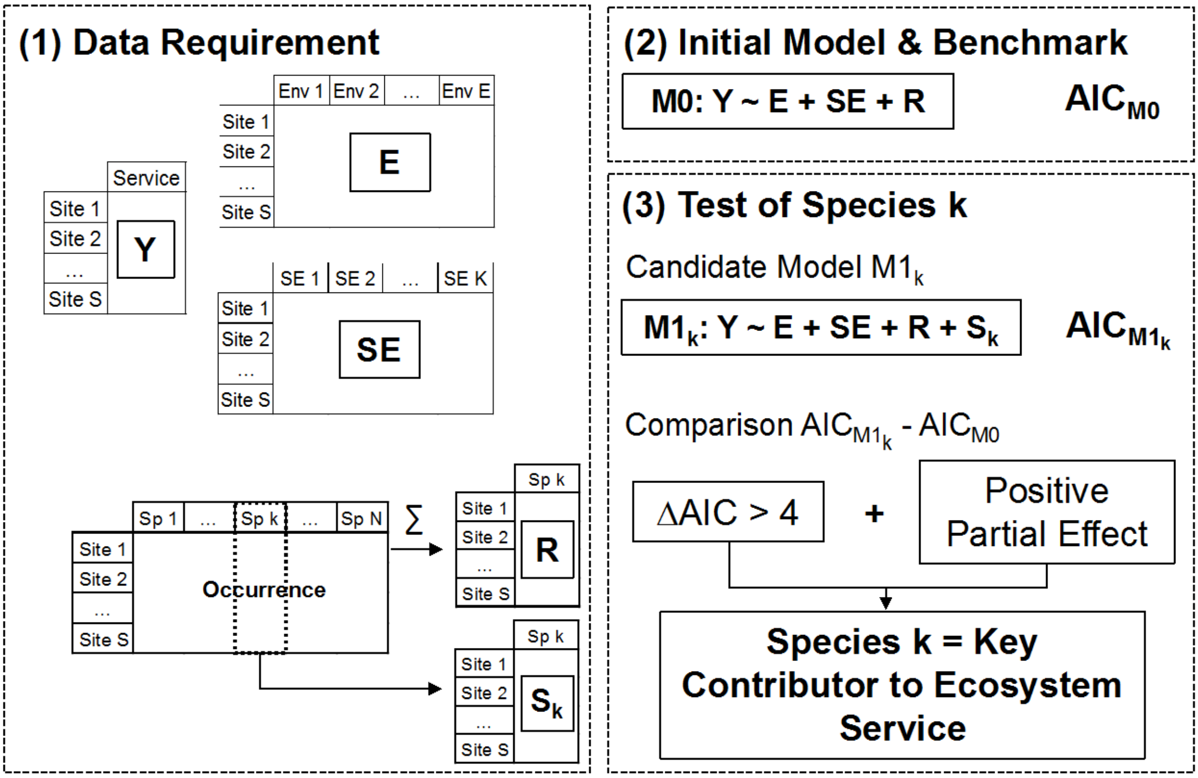

Figure 1: Statistical framework to assess the significant contribution of species to ecosystem services beyond the effects of environmental and socio-economic conditions and species richness.

Step 1: Collecting datasets: for a (large) set of sites, variables describing a given ecosystem service (Y), Environmental (E) and Socio-Economic conditions (SE), and the occurrence of species. Species richness (R) is computed for each site from the Sites-Species matrix as well as the vector (S_k) with presence-absence of each species in sites.

Step 2: The goal is to model a given ecosystem service (Y) according to Environmental (E) and Socio-Economic conditions (SE) and species richness (R); to check its relevance according to its explanatory power and to save its Akaike Information Criterion (AIC_{M0}) as a reference for the next step.

Step 3: The goal is to identify species key for the studied ecosystem service (Y) adding each candidate species (presence-absence, S_k) as an additional explanatory variable to M0 to compute model M1 and its associated AIC (AIC_{M1k}). Finally, a species is declared as a key contributor to the ecosystem service if ΔAIC (AIC_{M0}-AIC_{M1k}) > 4 and if its partial effect is positive (positive coefficient in the model).

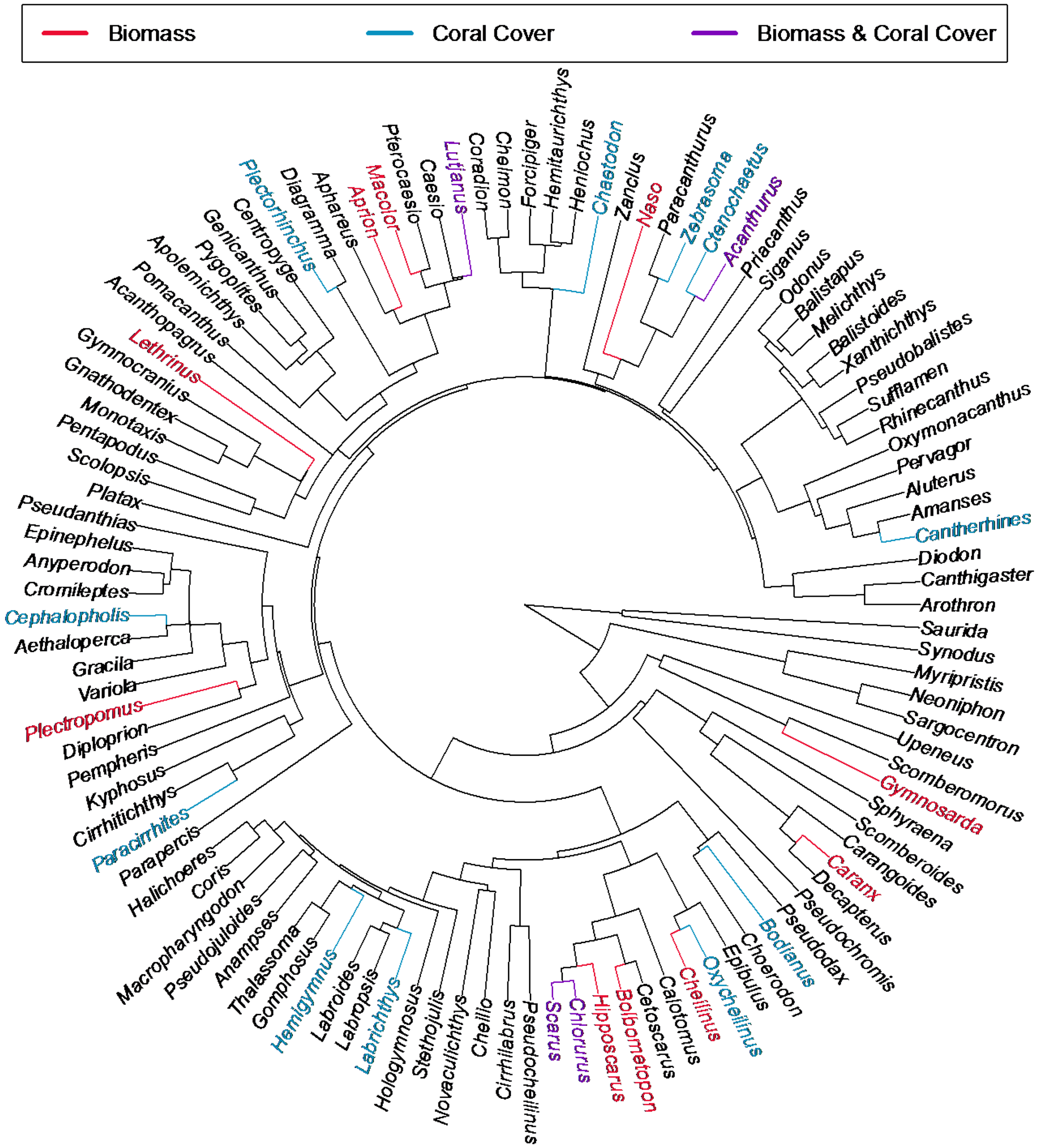

Figure 2. Positions of key species for biomass (red), live coral cover (blue) or both of them (purple) represented as their corresponding fish genera in the Tree of Life of Coral Reef Fishes, adapted from Near et al. (2013). The 26 key species for biomass represent 16 genera while the 28 key species for coral cover represent 15 genera with 4 common genera. *Elagatis* and *Parupeneus* genera are missing.

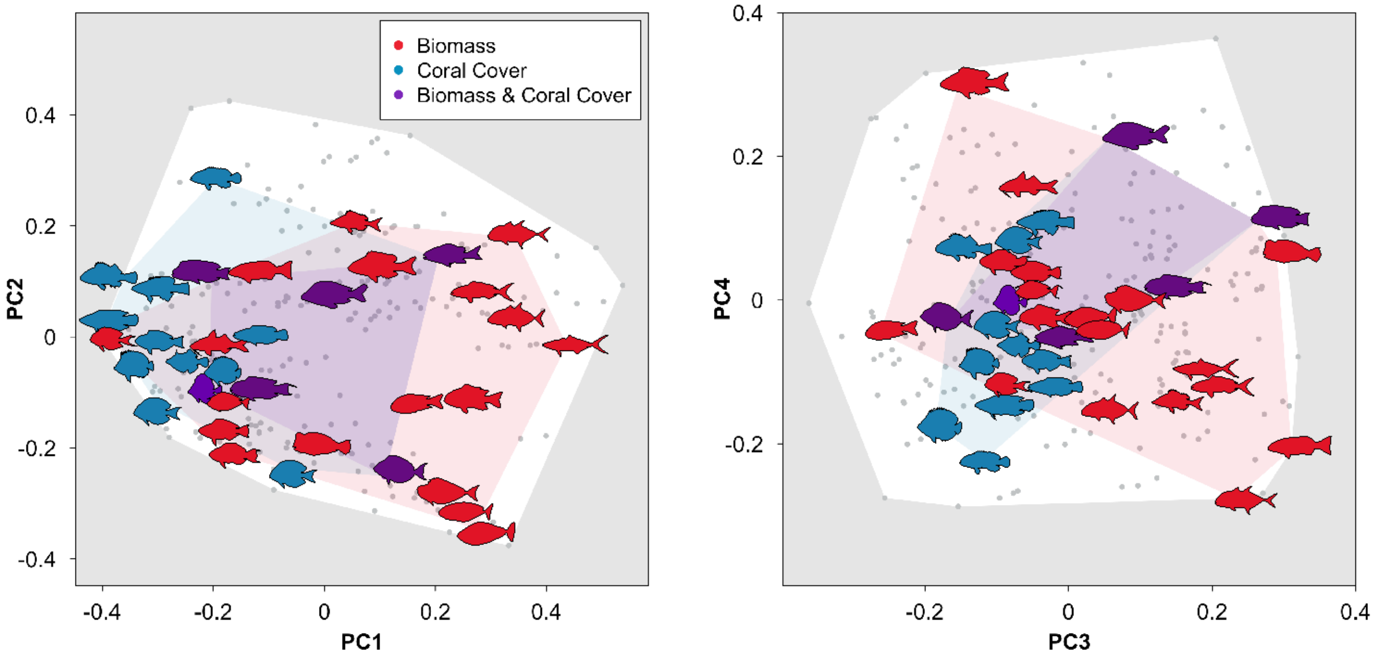

Figure 3. Functional attributes of key fish species. In total, 51 fish species which correspond to 35 out of 240 functional entities (15%) have been identified as strongly related to high biomass (18 red shapes), high live coral cover (11 blue shapes) or both of them (6 purple shapes). The positions of each of those 35 functional entities in the 4-dimensional functional space is defined according to species trait values. Fish shapes were chosen to illustrate the main genus of the species comprised in each functional entity. Other functional entities are represented with grey dots. Colored areas represent the functional volume filled by the functional entities that have been identified as strongly related to high biomass (red, FRic = 0.2), high live coral cover (blue, FRic = 0.05), both of them (purple, FRic = 0.01) or all functional entities (e.g. all species, white) present in the dataset. See Table S4 for details of species traits.

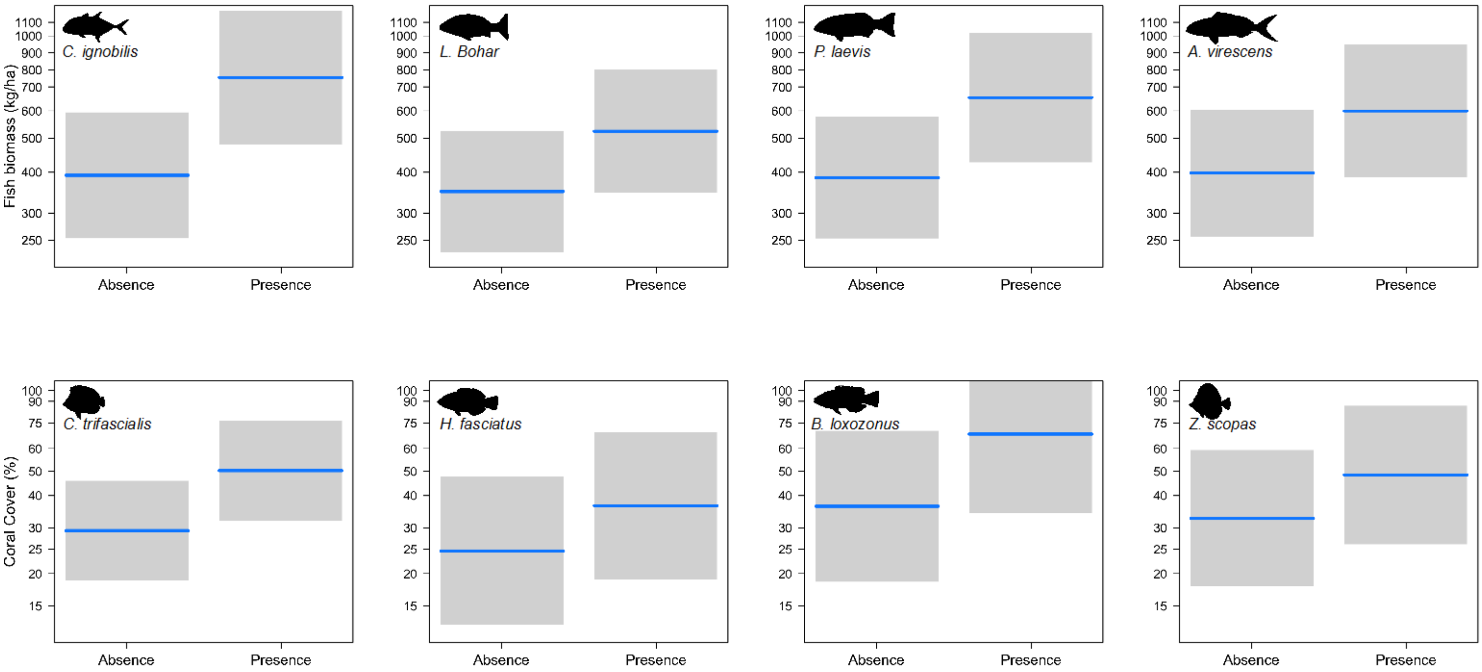

Figure 4. Net effect of the 4 most significant key fish species (lowest AIC) for fish biomass and live coral cover (among the 26 and 28 key species respectively) using a partial plot from the LMMs while the other variables are held constant. When present, each key species ensures median level of biomass and live coral cover significantly ($p<0.05$) higher than level observed when absent.

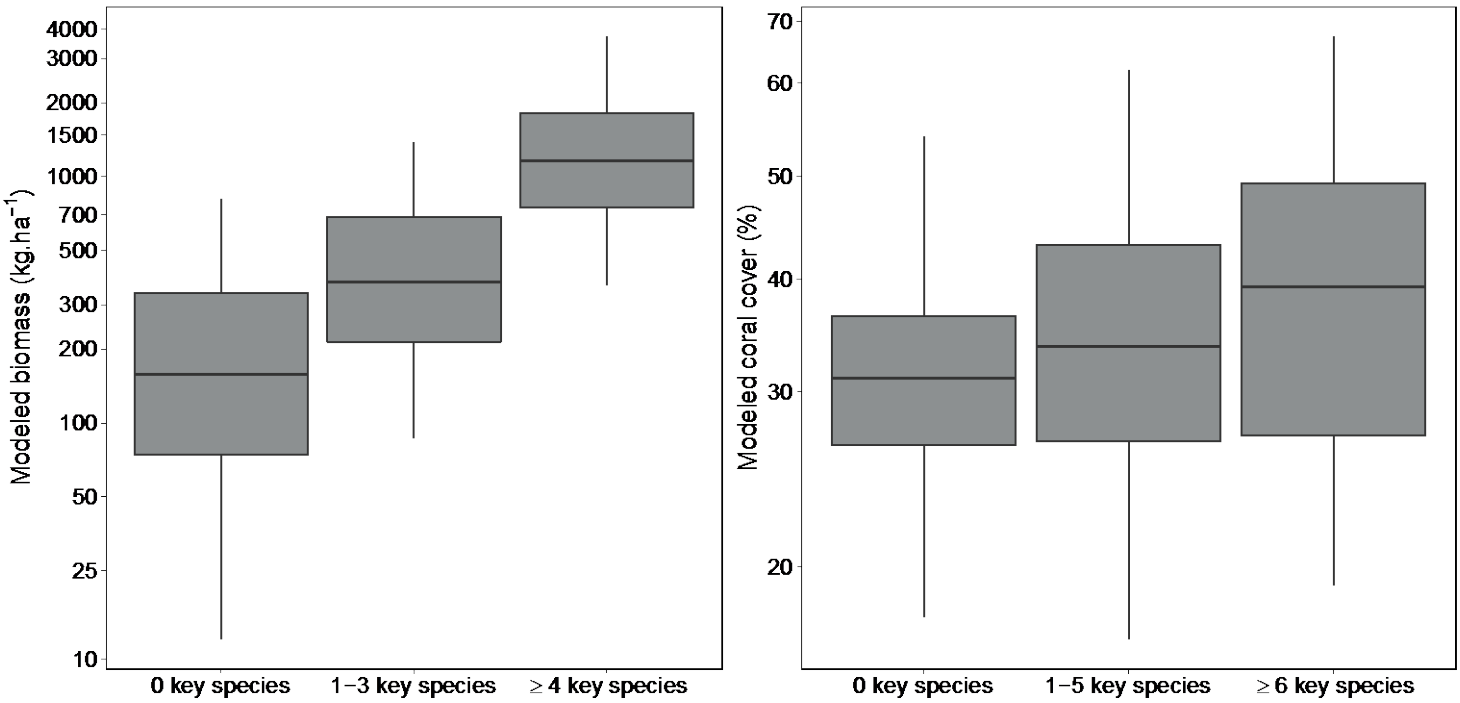

Figure 5. The accumulation of key species on coral reefs is positively related to fish biomass and live coral cover. To control for positive effect of species richness on ecosystem functioning (biomass and live coral cover), we compared modeled estimates of fish biomass and live coral between reefs while increasing the number of key species. 3 categories of reefs were considered based on the 3rd-quartile of the number of key species as threshold (≥4 and ≥6 key species for biomass and coral cover, respectively): i) reefs with no key species, ii) reefs with at least one key species but below the threshold and iii) reefs above the threshold. Reef with the highest number of key species reached higher level of biomass (1,150 kg.ha^{-1}) and coral cover (40%) than their counterparts having no key species (156 kg.ha^{-1}, 31% respectively). Distributions are represented using 95 percent confidence intervals.