Proceedings of the Royal Society B: Biological Sciences
April 2022, Volume 289 Issue 1973 Pages 20220162 (10p.)
https://doi.org/10.1098/rspb.2022.0162
https://archimer.ifremer.fr/doc/00766/87808/

Cross-ocean patterns and processes in fish biodiversity on coral reefs through the lens of eDNA metabarcoding

Mathon Laetitia [1, 2, *], Marques Virginie [1, 3], Mouillot David [3, 4], Albouy Camille [5], Andrello Marco [3, 17], Baletaud Florian [2, 3, 6], Borrero-Pérez Giomar H. [7], Dejean Tony [8], Edgar Graham J. [9], Grondin Jonathan [8], Guerin Pierre-Edouard [1], Hocdé Régis [3], Juhel Jean-Baptiste [3], Kadarusman [10], Maire Eva [3, 11], Mariani Gael [3], McLean Matthew [12], Polanco F. Andrea [7], Pouyaud Laurent [13], Stuart-Smith Rick D. [9], Sugeha Hagi Yulia [14], Valentini Alice [8], Vigliola Laurent [2], Vimono Indra B. [14], Pellissier Loïc [15, 16], Manel Stéphanie [1]

[1] CEFE, Univ. Montpellier, CNRS, EPHE-PSL University, IRD, Montpellier, France
[2] ENTROPIE, Institut de Recherche pour le Développement (IRD), Univ. Réunion, UNC, CNRS, Q1 IFREMER, Nouméa, New Caledonia, France
[3] MARBEC, Univ Montpellier, CNRS, IFREMER, IRD, Montpellier, France
[4] Institut Universitaire de France, France
[5] DECOD (Ecosystem Dynamics and Sustainability), IFREMER, INRAE, Institut Agro - Agrocampus Ouest, Nantes, France
[6] SOPRONER, groupe GINGER, 98000 Noumea, New Caledonia, France
[7] Programa de Biodiversidad y Ecosistemas Marinos, Museo de Historia Natural Marina de Colombia (MHNMC), Instituto de Investigaciones Marinas y Costeras- INVEMAR, Santa Marta, Colombia
[8] SPYGEN, Le Bourget-du-Lac, France
[9] Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania, Australia
[10] Politeknik Kelautan dan Perikanan Sorong, KKD BP Sumberdaya Genetik, Konservasi dan Domestikasi, Papua Barat, Indonesia
[11] Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK
[12] Department of Biology, Dalhousie University, Halifax NSB3H4R2, Canada
[13] ISEM, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France
[14] Research Center for Oceanography, National Research and Innovation Agency, Jl. Pasir Putih 1, Ancol Timur, Jakarta Utara 14430, Indonesia
[15] Landscape Ecology, Institute of Terrestrial Ecosystems, Department of Environmental Systems Science, ETH Zürich, Zürich, Switzerland
[16] Unit of Land Change Science, Swiss Federal Research Institute WSL, Birmensdorf, Switzerland
[17] Institute for the study of Anthropic Impacts and Sustainability in the marine environment, National Research Council (CNR-IAS), Rome, Italy

* Corresponding author : Laetitia Mathon, email address : laetitia.mathon@gmail.com

Abstract :

Increasing speed and magnitude of global change threaten the world's biodiversity and particularly coral reef fishes. A better understanding of large-scale patterns and processes on coral reefs is essential to prevent fish biodiversity decline but it requires new monitoring approaches. Here, we use environmental DNA metabarcoding to reconstruct well-known patterns of fish biodiversity on coral reefs and uncover hidden patterns on these highly diverse and threatened ecosystems. We analysed 226 environmental

DNA (eDNA) seawater samples from 100 stations in five tropical regions (Caribbean, Central and Southwest Pacific, Coral Triangle and Western Indian Ocean) and compared those to 2047 underwater visual censuses from the Reef Life Survey in 1224 stations. Environmental DNA reveals a higher (16%) fish biodiversity, with 2650 taxa, and 25% more families than underwater visual surveys. By identifying more pelagic, reef-associated and crypto-benthic species, eDNA offers a fresh view on assembly rules across spatial scales. Nevertheless, the reef life survey identified more species than eDNA in 47 shared families, which can be due to incomplete sequence assignment, possibly combined with incomplete detection in the environment, for some species. Combining eDNA metabarcoding and extensive visual census offers novel insights on the spatial organization of the richest marine ecosystems.

Keywords : eDNA metabarcoding, coral reef fish, biogeographic patterns, visual census

1. Introduction

Coral reefs host the highest fish diversity on earth despite covering less than 0.1% of the ocean's surface (1,2). They are also severely threatened (3), with near future outlooks predominantly pessimistic (4). Data syntheses over decades of surveys estimate the total number of coral reef fishes from 2,400 to 8,000 species (5,6), distributed among approximately 100 families (7). Typically, coral reef biodiversity displays clear spatial patterns, including longitudinal and latitudinal gradients outwards the Indo-Australian Archipelago (8,9), also known as the 'Coral Triangle', hosting the world's highest level of marine biodiversity (10). The exceptional biodiversity in the Coral Triangle has recently been suggested to strongly relate to higher diversity among fish families that feed on plankton (11). Other trophic groups are also very important on coral reefs but are often undetected because they are transient or hidden (12,13). Intriguingly, the proportions of fish species among families are shown to be strongly conserved across the Indo-Pacific (8). The spatial patterns of coral reef fishes are also marked by strong variations in taxonomic composition (species turnover or β diversity), often due to isolation (14). Many species on coral reefs are geographically localized, but can sometimes be locally abundant, while others are widespread (15).

Coral reef fishes have evolved in a physically complex environment and present a wide range of forms and functions (16). Small cryptic species, hereafter called crypto-benthic, that live inside the reef structure, can be very difficult to sample or survey using non-destructive methods (17), yet represent half of the fish diversity on coral reefs (13). Even though fishes are among the best-studied taxa inhabiting coral reefs (18), our knowledge of their biodiversity is only partial (19), the taxonomy is complex, uncertain for many species (5), and countless species remain undescribed.

Environmental DNA (eDNA) metabarcoding, a method retrieving and analyzing DNA naturally released by organisms in their environment (20), provides an opportunity to not only better understand classical biodiversity patterns, but also uncover novel ones hidden by our incomplete taxonomic and biogeographic coverage (21). Environmental DNA is particularly powerful in aquatic ecosystems (22) and is now well established for marine microorganisms (23,24). By contrast, its potential to provide an integrated biodiversity assessment of macroorganisms, including vertebrates of all trophic levels (from crypto-benthic to large pelagic fish species), is only shown at local (25) and regional (26–30) scales but not yet at spatial scales including more than one biogeographic region or multiple ocean basins.

Here, we investigate how a cross-ocean basin snapshot of eDNA sampling could describe the distribution of fish biodiversity on coral reefs, reveal unknown patterns, and challenge well-established assembly rules. From 226 eDNA seawater samples (2,712 PCR replicates) collected in 100 stations at 26 sites covering five tropical regions (Southeast Polynesia, Tropical Northwestern Atlantic, Tropical Southwestern Pacific, Western Indian Ocean and Western Coral Triangle) across the Indian, Pacific and Atlantic Oceans (figure S1-S2), we produced a final dataset of 189,350,273 mitochondrial 12S rRNA gene sequence reads (see Methods), clustered into 2,023 molecular operational taxonomic units (MOTUs), and assigned to Actinopterygii (bony fishes) and Chondrichthyes (cartilaginous fishes) taxa (tables S1- S2). We then compared fish biodiversity patterns obtained from eDNA to those observed from 2,047 standardized visual surveys of reef fishes in 1224 stations at 219 sites within 24 tropical regions (31).

2. Results

(a) Global estimates of fish biodiversity on coral reefs

We estimated total fish diversity on coral reefs using the asymptote of a multi-model accumulation curve for both eDNA MOTUs (32) and visual census species (Methods). The asymptote estimated from 100 eDNA stations distributed in five regions sampled over a 28-month period reaches 2,650 MOTUs (figure 1*a*). This detectable fish MOTUs diversity, including also MOTUs unassigned at the species-level, is 16% higher than the estimate from visual census data, which reaches an asymptote at 2,268 fish species from 2,047 tropical transects surveyed during 13 years (figure 1*b*). The asymptotic estimation of family richness obtained with eDNA reaches 147 families, 25% more than the asymptotic number of families estimated with visual census data (118 families, figure 1*c-d*). Among the 71 families shared between both datasets, 24 have a higher number of MOTUs from eDNA survey than species from visual survey while 47 have more species from visual survey than MOTUs from eDNA survey (figure 1*e*). Families with more taxa identified using eDNA include those often associated with reef-adjacent habitats such as mangroves or soft sediments like Mugilidae *(*e.g. *Mugil rubrioculus),* Elopidae and Gerreidae (33, *e.g. Gerres oyena),* and crypto-benthic species that live hidden in crevices (e.g. Gobiidae) or nocturnal fish species (34, *e.g.* Congridae). Families with more taxa with visual census include Acanthuridae, Chaetodontidae, Blenniidae, Labridae, Pomacentridae and Scaridae. Fifty-five families are detected only with eDNA, including Myctophidae, Engraulidae, Atherinidae and Exocoetidae, while 24 families are detected only by visual census, including Caesionidae, Chaenopsidae, Labrisomidae and Microdesmidae. Environmental DNA estimates a diversity of crypto-benthic species 13% higher than with visual census, and, among many others, includes species such as the elegant firefish (*Nemateleotris decora),* which lives on the outer reef slope between 25 to 70 m (figure 2*a*). Yet, the difference in fish diversity assessment between the two methods is the strongest for pelagic and wide-ranging species, for which eDNA reveals more than 7 times higher richness than with

visual census. These species mainly belong to Scombridae (e.g. *Katsuwonus pelamis*), Clupeidae, Carcharhinidae (e.g. *Carcharhinus leucas*, *Sphyrna lewini*) and Belonidae (figure 2*b*).

MOTU richness per fish family retrieved with eDNA is strongly correlated with fish species richness within families recorded in visual census data (Pearson correlation = 0.84, $p < 0.001$, $n = 71$, figure 1*e*). Highly diverse families seen on coral reefs are also well represented in eDNA samples, with Gobiidae, Labridae and Pomacentridae containing more than 100 MOTUs each, together representing about 20% of MOTUs (figure 1*f*, figures S3- S4). The slope of the log-log relationship between MOTUs richness per family and species richness per family is equal to 0.8 showing that the relationship is not proportional but saturating. The richest fish families contain more MOTUs detected with eDNA than species detected with visual surveys.

(b) Biogeography of eDNA sequences

The spatial distribution of MOTUs follows clear biogeographic patterns, with a peak in the coral triangle and lower values of MOTU richness toward Southeast Polynesia (figure S5). The richest region (West Papua, Indonesia, Western Coral Triangle) contains ~50% of the global pool of fish MOTUs while the poorest region (Fakarava, French Polynesia, Southeast Polynesia) contains only 9% of the global pool (figures S6-S7 and table S2). Distance-based Redundancy Analysis (dbRDA) was performed on fish family proportions at each site (i.e. number of MOTUs or species assigned to each family in each site, see Methods) for eDNA and visual surveys with the region and the site MOTU/species richness as explanatory variables, including their interaction (figure 3, table S3). For eDNA, the dbRDA explains up to 42% of variation in family proportions between pairs of sites with region and MOTU/species richness both having significant effects ($F = 4.1$ and 5.7, respectively, $p < 0.001$), but no significant interaction ($F = 1.99$, $p>0.05$). The partial dbRDA

on eDNA showed a significant effect of region while controlling for MOTU richness ($F = 2.79$, $p < 0.001$). The first axis explains 17.2% of variation in family proportions and separates the Western Coral Triangle from other regions (figure 3*a-b*). The first axis shows a higher proportion of Lutjanidae but lower proportions of Labridae and Gobiidae in sites of the Western Coral Triangle. It also confirms the longitudinal diversity gradient from the Coral Triangle. The second axis explains 11.2% of variation and discriminates the Tropical Northwestern Atlantic from the Western Indian Ocean, due to a higher proportion of Clupeidae and Carangidae in the Atlantic Ocean and a higher proportion of Acanthuridae in the Indian Ocean. The dbRDA performed on visual census data explained greater variation ($R^2 = 0.5$, $p < 0.001$) and the region also had a significant, albeit weaker than for MOTUs, effect on fish family proportions ($F = 17.7$, $p < 0.01$), while species richness and interaction between the two variables also had significant effects ($F =$ 6.28 and 2, $p < 0.01$ respectively). The first axis explains 41.6% of variance in family proportions and separates the Tropical Northwestern Atlantic from the other regions with a higher proportion of Gobiidae and Serranidae. The second axis explains 5.7% of variance in family proportions and separates the Southeast Polynesia from Indo-Pacific regions, and is mostly driven by the higher proportion of Pomacentridae in the Indo-Pacific (figure 3*c-d*).

(c) Global patterns of fish turnover and rarity

Our eDNA survey shows that a majority of MOTUs are geographically restricted, with 85% of the MOTUs detected in only one region (figure 4*a*), and 35% in only one site (figure S8). Geographic restriction is one aspect of species rarity but is shown to play a primary role in determining extinction risk while local abundance and habitat specialization have secondary roles (35). We hierarchically partitioned the global MOTU diversity (γ_{global}) into additive diversity components

(i.e. dissimilarity) due to difference between regions ($\beta_{inter-region}$), mean difference between sites within regions ($\overline{\beta}_{inter-site}$), mean difference between stations within sites ($\overline{\beta}_{inter-station}$) and mean station diversity ($\overline{\alpha}_{station}$) (36). As a consequence of the geographic restriction of most MOTUs to one region, the total fish MOTU (γ) diversity is mainly due to inter-region β-diversity (~74%) followed by inter-site (14.8%) and inter-station (5.9%) β-diversity (figure 4*b*). The same partitioning using different site delineations (10 and 20 km) provides similar results (table S4). Diversity partitioning of crypto-benthic fish MOTUs only or pelagic fish MOTUs only reveals similar patterns (table S5). The partitioning diversity of species detected by visual census also revealed similar patterns but with a stronger effect of $\beta_{inter-region}$ (84%) and lower (3x) $\overline{\beta}_{inter-site}$ and $\overline{\beta}_{inter-station}$ (table S5, figure S9).

Beyond the hierarchical partitioning of diversity, we compared the distribution of fish MOTUs and species visual occurrences independently of the survey method and sampling effort using global species abundance distributions (gSAD) (37). We fitted the fish MOTU and species visual occurrences to three distributions (log-series, Pareto and Pareto with exponential finite adjustment, *i.e*. Pareto Bended, see Methods) and estimated the parameters by maximum likelihood. For the visual census gSAD, the best fit was obtained with the log-series and Pareto distributions (table S6) with a slope of -0.95 (confidence interval at 95% [-0.98;-0.92]) (figure S10). This suggests a distribution of geographically restricted or rare species close to the neutral theory (β close to -1). By contrast, the best fit for fish MOTUs was obtained with the Pareto Bended distribution with a slope β = -0.76 (confidence interval at 95% [-0.85;-0.65]) and then with the log-series distribution, suggesting a lower prevalence of rarity than under the neutral theory, in agreement with previous tests based on species distributions on coral reefs (38).

3. Discussion

Environmental DNA allows the detection and identification of more taxa than traditional techniques (26,39), but further offers novel insights on the spatial organization of the richest marine ecosystem at large scale. Over a timespan of 2.3 years, in major tropical ocean basins, eDNA metabarcoding reveals a higher proportion of crypto-benthic, pelagic and soft-sediment-associated fishes on coral reefs than detected in the most extensive visual census over 13 years. We found a high local MOTU turnover, but we were not able to conclude if it is due to an insufficient sampling at the station level, or if it suggests that differences in fish species composition may exist between adjacent reefs that are not detected by visual surveys (40), so that fish biodiversity is more patchy than previously thought on coral reefs.

We were also able to retrieve well-known patterns of fish diversity on coral reefs such as the biogeographic boundaries between the Atlantic and Pacific oceans, the longitudinal diversity gradient from the center of the Coral Triangle, with Southeast Polynesia being the least diverse region and Western Coral Triangle the richest, and that Gobiidae, Labridae, Pomacentridae and Apogonidae are the most diverse fish families on coral reefs (8). We found a lower proportion of rare MOTUs than expected under the neutral theory with eDNA, which is in agreement with the findings of a previous study from coral reefs in the Indo-Pacific (38), while visual census data suggests higher rarity close to that predicted from the neutral theory. More surprising, our study calls into question the pattern of fish family stability composition across the Indo-Pacific that was revealed more than 20 years ago (8), and the recent finding that planktivore families drive fish biodiversity patterns on coral reefs (11). We found significant effects of species richness and region on family composition, which appears less stable than previously thought.

Environmental DNA identified many pelagic, deepwater and crypto-benthic species not seen by divers. Among the pelagic species identified with eDNA, many belong to the Scombridae and Carcharhinidae families, which likely avoid divers or are not permanent residents on coral reefs so can be missed in visual surveys (41). Some crypto-benthic or reef-associated species, hidden in the reef, can also be missed by divers so were also more represented in eDNA than in visual surveys. Crypto-benthic species also have a crucial role for coral reef functioning, by promoting biomass production and fueling the reef trophodynamics (42), but their diversity has been underestimated so far (13). Transient, pelagic and deep-water species may be very important for reef functioning, through pelagic larval stages or nocturnal migration up the reef slope (12,43,44), but their presence and role need further investigation. In contrast, visual census also detected many families not detected, or not identified, by eDNA, such as Acanthuridae, Blenniidae, Caesionidae, Chaenopsidae, Chaetodontidae, Labrisomidae, Labridae or Microdesmidae. This limited identification by eDNA can be due to the very low representation of these families in 12S reference databases (between 0 and 12%), or to the low resolution of the teleo marker for species of these families, so several species can share the same sequence and be grouped under the same MOTU. Environmental DNA may also be inappropriate to detect these species in the environment.

The finding of a strong regional effect on both species composition (figure 3) and species differentiation (figure 4) at a large scale is in agreement with visual surveys and previous knowledge (45), while the suggestion of a strong turnover at the local scale may be an unexpected result for coral reef fishes. This predominant role of large-scale bioregional differentiation explains the exceptional fish diversity on coral reefs, probably associated with long-term geological isolation (2). Overall, the Tropical Northwestern Atlantic region has a very distinct MOTU

composition compared to the four other regions (figure 3) with only 1.2% of MOTUs being shared between the Tropical Northwestern Atlantic and any other region, while 20% of MOTUs are shared between at least two Indo-Pacific regions (figure 4*a*). The isolation of the Tropical Northwestern Atlantic region can be explained by the hard vicariant barrier of the Isthmus of Panama (14,46), and a limited suitable area for coral reefs during the past quaternary glaciation. By contrast, the Indo-Pacific maintained extensive coral reef refuges that have served as centers of survival during ice-age periods (9).

The greater local compositional dissimilarity of reef fishes among adjacent stations with eDNA than with visual census may correspond to local environmental or habitat differences, to stochastic or random processes (47), or may be due to an insufficient sampling at the station level (Supplementary Analyses Fig. 6). A higher number of replicates per station would be necessary to characterize exhaustively the diversity at the station level and more confidently conclude on the local turnover hypothesis.

While our results confirm the potential of eDNA to monitor biodiversity in marine ecosystems, some limitations should be addressed in the future to fully exploit this potential. Completing public reference databases would improve the accuracy of taxonomic assignment, which is essential for a better estimation of biodiversity patterns. At such a large spatial scale, reference databases are far from exhaustive with only up to 13% of fish species sequenced on our marker (52), preventing assignment to the species level for 81% of our eDNA sequences. Using multiple markers is an alternative to the database limitation (53,54), but it is much more expensive. For these reasons, we used MOTUs curated by a combination of a clustering algorithm and conservative abundance-based post-clustering filters. While un-curated MOTUs are prone to overestimate real diversity (55) and a given MOTU can represent several species within one cluster or several MOTUs

belonging to one species, MOTUs with conservative curation have been shown to reflect the true level of fish diversity across scales in streams (56,57). Additionally, some species share the same barcode sequence due to insufficient genetic differentiation on such a small mitochondrial marker (54). This lack of taxonomic resolution combined with a conservative curated MOTUs pipeline can underestimate MOTUs richness. Moreover, some crypto-benthic or rare fish families are still underrepresented in public databases, and their diversity is potentially underestimated with eDNA (i.e.Blenniidae, Gobiescocidae, Chaenopsidae, Aploactinidae).

Differences in sampling method and in sample size might influence the detected biodiversity with eDNA. The lower volume of water sampled in the Western Coral Triangle region (2L per sample, so 4L per station using point-sampling instead of 2-km transect with 30L elsewhere), could underestimate fish biodiversity. However, previous studies show that MOTU accumulation curves based on this dataset were close to the total fish diversity reported in this region (32). Furthermore, β-diversity between samples within stations in each region indicates that dissimilarity between samples is not greater in the Western Coral Triangle than in other regions (figure S11). To account for differences in sample size and obtain a balanced design, we performed sensitivity analyses by rarefying our complete dataset to i) 4 stations for all sites and ii) 4 sites per region after removing the lowest sampled region (Southeast Polynesia) (Supplementary Analyses Fig. 1-4). We obtained similar patterns even after subsampling stations or sites. However, our site-based and station-based accumulation curves do not reach plateaus suggesting that our sampling effort was not sufficient to exhaustively estimate fish biodiversity for each site (Supplementary Analyses Fig. 5) and station (Supplementary Analyses Fig. 6). Twenty-five replicates (so, 12 stations in case of field duplicates) could accurately estimate biodiversity regionally due to high local turnover (58). A higher number of eDNA samples would be necessary here to reach MOTU accumulation per site and station.

The transport and degradation of eDNA can also impact species detection. As some evidence suggests that eDNA from pelagic fishes degrades slower than from inshore species (59), we cannot exclude that eDNA from pelagic and deep-water families (*e.g.* Myctophidae) might disperse sufficiently with sea currents such that species living close to reef habitats are detected. Environmental DNA transport could also explain the detection of some freshwater fish families (*i.e.* Centrarchidae, Osphronemidae or Channidae) in a few samples located near an estuary or in an enclosed bay with freshwater inputs.

Better understanding and anticipating the effects of multiple threats to the marine environment depends on the temporal and spatial extent of our monitoring capacity in the vast ocean. Environmental DNA is a powerful tool to investigate biodiversity patterns at large scale and monitor biodiversity, but still benefits from the combination with complementary approaches as visual methods for an exhaustive biodiversity survey across space and time to keep pace with ongoing changes.

4. Methods

(a) Environmental DNA collection and sample processing

Environmental DNA seawater samples were collected between 2017 and 2019, following a hierarchical pattern. A total of 226 eDNA samples (filters) were collected in 100 stations (gathering of replicates at the same location) located in 26 sites (groups of stations separated by at least 35 km) distributed across five tropical regions (figure S1-S2). Three different sampling methods were used comprising a 2km-long sampling transect of 30L (surface or bottom depth) or point samples of 2L (table S7 and Methods S1), and between 12 and 64 samples were collected by

region. Filtration was performed with Polyethersulfone (PES) filters, 0.2 µm pore size. For each sampling campaign, a strict contamination control protocol was followed in both field and laboratory stages (39). Negative field controls were performed in multiple sites, and revealed no contamination from the boat or samplers.

(b) eDNA extraction, amplification and sequencing

DNA extraction was performed in a dedicated DNA laboratory (SPYGEN, www.spygen.com) equipped with positive air pressure, UV treatment and frequent air renewal. Decontamination procedures were conducted before and after all manipulations. Detailed protocols of DNA extraction, amplification and sequencing can be found in Method S2 and in (32,39). A teleost-specific 12S mitochondrial rRNA primer pair (teleo, forward primer - ACACCGCCCGTCACTCT, reverse primer – CTTCCGGTACACTTACCATG (39)) was used for the amplification of metabarcode sequences. As we analysed our data using MOTUs as a proxy for species to overcome genetic database limitations, we chose to amplify only one marker. Teleo marker has been shown to be the most appropriate for fish, owing to its high interspecific variability, and its short size allowing us to detect rare and degraded DNA reliably (39,54,60,61). Twelve DNA amplifications PCR per sample were performed.

(c) Bioinformatic analysis

Following sequencing, reads were processed using clustering and post-clustering cleaning to remove errors and estimate the number of species using Molecular Operational Taxonomic Units (MOTUs) (56). First, reads were assembled using *VSEARCH* (62), then demultiplexed and trimmed using *CUTADAPT* (63) and clustering was performed using *SWARM* v.2 (64) with a

minimum distance of 1 mismatch between clusters. Taxonomic assignment of MOTUs was carried out using the Lower Common Ancestor (LCA) algorithm *ecotag* implemented in the OBITOOLS toolkit (65) and the European Nucleotide Archive (ENA) as a reference database (release 143, March 2020). Details on the bioinformatics analysis can be found in Methods S3. Taxonomic assignments obtained from the LCA algorithm at the species level were accepted if the percentage of similarity with the reference sequence was 100%, at the genus level if the similarity was between 90 and 99%, and at the family level if the similarity was > 85% following previous studies (32,66). If these criteria were not met, the MOTU was left unassigned. Only 21% of assigned MOTUs are assigned to the family level with a similarity between 85 and 90% (Table S8).

(d) Visual census data

The visual census survey data used here is a subset (2047 transects, in 219 sites, figure S1) of the complete visual census data (3027 transects) provided by the RLS (31), and comprises all species observed on standardized 50 m surveys at sites in tropical biogeographic realms between 2006 and 2017 (Methods S4) (67). We selected only the most recent survey for each station and only transects with more than five percent of coral cover. Two different sampling protocols were adapted to detect both reef and crypto-benthic fishes.

(e) Statistical analysis

More details on the statistical analysis are available in Methods S5.

Accumulation curves were calculated for species per 500 m^2 transect, MOTUs per eDNA sample, and families per transect and sample. We used the functions “specaccum” and “fitspecaccum” from the R package “vegan” which calculates the expected species accumulation curve using a

sample-based rarefaction method and fit a nonlinear accumulation model. In order to assess the impact of the irregular sampling on the estimates measured with accumulation curves, we subset randomly half of the transects in the 3 most sampled regions in Australia, and calculated again the accumulation curves for species and families (figure S12). The results were unchanged.

Linear regression models were fitted between the number of MOTUs per family in the eDNA dataset and the number of species per family in the visual census dataset, after log(x+1) transformation (figure 1*e*).

Accumulation curves were also calculated by sub-setting MOTUs belonging to crypto-benthic orders, or to pelagic families, for both datasets (figure 2). The asymptote was calculated as described above.

We performed distance-based Redundancy Analysis (dbRDA) on family proportions, with *region* and *site richness* as explanatory variables, using the function *capscale* from the *vegan* package. We subset the Visual Census to select only the 68 sites that fell into the 5 regions in common with the eDNA dataset. Total dbRDA provided the effects of each of the variables and their interaction. We then calculated partial dbRDA to measure the effect of the Region while correcting for the effect of site richness (figure 3, table S3).

We applied an additive partitioning framework (68) to separate the total MOTUs diversity at the global scale (ɣ global) into contributions at smaller scales from regions to local richness : $\gamma_{global} = \beta_{inter\text{-}region} + \text{mean } \beta_{inter\text{-}site} + \text{mean } \beta_{inter\text{-}station} + \text{mean } \bar{\alpha}_{station}$. In this additive framework, the three levels of biodiversity (69) (i.e. α, β and γ) are expressed with the same unit and consequently the contribution of α and β diversity to total diversity (ɣ) can be directly compared (70).

We analyzed the distribution of fish MOTU and species occurrences using global species abundance distribution (gSAD) which plots, on a log-log scale, the number of species as a function of the number of observations (37).

Funding

The sampling in the Caribbean, Indian Ocean and Polynesia and the sequencing were funded by Monaco Explorations. Fieldwork in Indonesia and laboratory activities were supported by the Lengguru 2017 Project (www.lengguru.org), conducted by the French National Research Institute for Sustainable Development (IRD), National Research and Innovation Agency (BRIN) with the Research Center for Oceanography (RCO), the Politeknik Kelautan dan Perikanan Sorong), the University of Papua (UNIPA) with the help of the Institut Français in Indonesia (IFI), funding from Monaco Explorations, and corporate sponsorship from the Total Foundation and TIPCO company. Fieldwork and laboratory activities in New-Caledonia were supported by the projects ANR SEAMOUNTS and CIFRE REEF 3.0 conducted by the French National Institute for Sustainable Development (IRD) and GINGER-BURGEAP-SOPRONER company with funding from Monaco Explorations. Fieldwork and laboratory activities in Colombia were supported by Monaco Explorations, ETH Global grant and the project Reefish, conducted in collaboration with the Instituto de Investigaciones Marinas y Costeras – INVEMAR. Monaco Explorations supported also sampling and sequencing in the Caribbean. Fieldwork in the French Scattered islands was supported by the Terres Australes et Antartiques Françaises (TAAF).

Acknowledgements

The authors thank all staff and students involved in fieldwork and acknowledge SPYGEN staff for the technical support in the eDNA laboratory.

References

1. Parravicini V, Kulbicki M, Bellwood DR, Friedlander AM, Arias-Gonzalez JE, Chabanet P, et al. Global patterns and predictors of tropical reef fish species richness. Ecography (Cop). 2013;36(12):1254–62.
2. Cowman PF, Bellwood DR. The historical biogeography of coral reef fishes: Global patterns of origination and dispersal. J Biogeogr. 2013;40(2):209–24.
3. Cinner JE, Zamborain-mason J, Gurney GG, Graham NAJ, Macneil MA, Hoey AS, et al. Meeting fisheries, ecosystem function, and biodiversity goals in a human-dominated world. Science (80-). 2020;368(April):307–11.
4. Hoegh-Guldberg O, Jacob D, Taylor M, Guillén Bolaños T, Bindi M, Brown S, et al. The human imperative of stabilizing global climate change at 1.5°C. Science (80-). 2019;365(6459).
5. Victor BC. How many coral reef fish species are there? Cryptic diversity and the new molecular taxonomy. In: Ecology of fishes on coral reefs Cambridge University Press, Cambridge. 2015. p. 76–88.
6. Siqueira AC, Morais RA, Bellwood DR, Cowman PF. Trophic innovations fuel reef fish diversification. Nat Commun. 2020;11(1):1–11.
7. Bellwood D, Wainwright P. The history and biogeography of fishes on coral reefs. In: Sale P, editor. Coral Reef Fishes: Dynamics and Diversity in a Complex Ecosystem. Academic P. San Diego; 2002.
8. Bellwood DR, Hughes TP. Regional-scale assembly rules and biodiversity of coral reefs. Science (80-). 2001;292(5521):1532–4.
9. Pellissier L, Leprieur F, Parravicini V, Cowman PF, Kulbicki M, Litsios G, et al. Quaternary coral reef refugia preserved fish diversity. Science (80-). 2014;344(6187):1016–20.
10. Veron J, Devantier LM, Turak E, Green AL, Kininmonth S, Stafford-Smith M, et al. Delineating the Coral Triangle. Galaxea, J Coral Reef Stud. 2009;11(2):91–100.
11. Siqueira AC, Morais RA, Bellwood DR, Cowman PF. Planktivores as trophic drivers of global coral reef fish diversity patterns. PNAS. 2021;118(9).
12. Morais RA, Bellwood DR. Pelagic Subsidies Underpin Fish Productivity on a Degraded Coral Reef. Curr Biol. 2019;29(9):1521–7.
13. Brandl SJ, Goatley CHR, Bellwood DR, Tornabene L. The hidden half : ecology and evolution of cryptobenthic fishes on coral reefs. Biol Rev. 2018;93:1846–73.
14. Bender MG, Leprieur F, Mouillot D, Kulbicki M, Parravicini V, Pie MR, et al. Isolation drives taxonomic and functional nestedness in tropical reef fish faunas. Ecography (Cop). 2017;40(3):425–35.
15. Hughes TP, Bellwood DR, Connolly SR, Cornell H V., Karlson RH. Double jeopardy and global extinction risk in corals and reef fishes. Curr Biol. 2014;24(24):2946–51.
16. Mouillot D, Villéger S, Parravicini V, Kulbicki M, Arias-González JE. Functional over-redundancy and high functional vulnerability in global fish faunas on tropical reefs. PNAS. 2014;111(38):13757–62.
17. Alzate A, Zapata FA, Giraldo A. A comparison of visual and collection-based methods for assessing community structure of coral reef fishes in the Tropical Eastern Pacific. Rev Biol Trop. 2014;62(February):359–71.
18. Bellwood D, Renema W, Rosen BB. Biodiversity hotspots, evolution and coral reef biogeography: a review. In: Biotic Evolution and Environmental Change in Southeast Asia Cambridge: Cambridge University Press. 2012. p. 216–45.
19. Mora C. Ecology of fishes on coral reefs. Ecology of Fishes on Coral Reefs. 2015. 1–374 p.
20. Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH. Environmental DNA. Mol Ecol. 2012;21(8):1789–93.
21. Boulanger E, Loiseau N, Valentini A, Arnal V, Boissery P, Dejean T, et al. Environmental DNA metabarcoding reveals and unpacks a biodiversity conservation paradox in Mediterranean marine reserves, Dryad, Dataset. Proc R Soc B. 2021;288(20210112).
22. Harrison JB, Sunday JM, Rogers SM. Predicting the fate of eDNA in the environment and implications for studying biodiversity. Proc R Soc B Biol Sci. 2019;286(1915):1–9.
23. Cordier T, Alonso-Saez L, Apothéloz-Perret-Gentil L, Aylagas E, Bohan DA, Bouchez A, et al. Ecosystem monitoring powered by environmental genomics: A review of current strategies with an implementation roadmap. Mol Biol Evol. 2020;may:1–22.
24. De Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, et al. Eukaryotic plankton diversity in the sunlit ocean. Science (80-). 2015;348(6237):1–11.
25. Valdivia-Carrillo T, Rocha-Olivares A, Reyes-Bonilla H, Domínguez-Contreras JF, Munguia-Vega A. Integrating eDNA metabarcoding and simultaneous underwater visual surveys to describe complex fish communities in a marine biodiversity hotspot. Mol Ecol Resour. 2021;(March):1558–74.

26. West K, Travers MJ, Stat M, Harvey ES, Richards ZT, DiBattista JD, Newman SJ, Harry A, Skepper CL, Heydenrych M, Bunce M. Large-scale eDNA metabarcoding survey reveals marine biogeographic break and transitions over tropical north-western Australia. Diversity and Distributions. 2021 Oct;27(10):1942-57.
27. Kume M, Lavergne E, Ahn H, Terashima Y, Kadowaki K, Ye F, et al. Factors structuring estuarine and coastal fish communities across Japan using environmental DNA metabarcoding. Ecol Indic [Internet]. 2021;121(November 2020):107216. Available from: https://doi.org/10.1016/j.ecolind.2020.107216
28. Fraija-Fernández N, Bouquieaux MC, Rey A, Mendibil I, Cotano U, Irigoien X, et al. Marine water environmental DNA metabarcoding provides a comprehensive fish diversity assessment and reveals spatial patterns in a large oceanic area. Ecol Evol. 2020;10(14):7560–84.
29. Aglieri G, Baillie C, Mariani S, Cattano C, Calò A, Turco G, et al. Environmental DNA effectively captures functional diversity of coastal fish communities. Mol Ecol. 2020;(August):1–13.
30. DiBattista JD, Berumen ML, Priest MA, De Brauwer M, Coker DJ, Sinclair-Taylor TH, et al. Environmental DNA reveals a multi-taxa biogeographic break across the Arabian Sea and Sea of Oman. Environ DNA. 2021;(December 2020):1–16.
31. Edgar GJ, Stuart-Smith RD. Systematic global assessment of reef fish communities by the Reef Life Survey program. Sci Data. 2014;1:1–8.
32. Juhel JB, Utama RS, Marques V, Vimono IB, Sugeha HY, Kadarusman, et al. Accumulation curves of environmental DNA sequences predict coastal fish diversity in the coral triangle. Proceedings Biol Sci. 2020;287(1930):1–10.
33. Castellanos-Galindo GA, Krumme U, Rubio EA, Saint-Paul U. Spatial variability of mangrove fish assemblage composition in the tropical eastern Pacific Ocean. Rev Fish Biol Fish. 2013;23(1):69–86.
34. Willis TJ, Anderson MJ. Structure of cryptic reef fish assemblages: Relationships with habitat characteristics and predator density. Mar Ecol Prog Ser. 2003;257:209–21.
35. Harnik PG, Simpson C, Payne JL. Long-term differences in extinction risk among the seven forms of rarity. Proc R Soc B Biol Sci. 2012;279(1749):4969–76.
36. Crist TO, Veech JA. Additive partitioning of rarefaction curves and species-area relationships: Unifying α-, β- and γ-diversity with sample size and habitat area. Ecol Lett. 2006;9(8):923–32.
37. Enquist BJ, Feng X, Boyle B, Maitner B, Newman EA, Jørgensen PM, et al. The commonness of rarity: Global and future distribution of rarity across land plants. Sci Adv. 2019;5(11):1–14.
38. Dornelas M, Connolly SR, Hughes TP. Coral reef diversity refutes the neutral theory of biodiversity. Nature. 2006;440(7080):80–2.
39. Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, et al. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Mol Ecol. 2016;25(4):929–42.
40. West K, Travers MJ, Stat M, Harvey ES, Richards ZT, Dibattista JD, et al. Large-scale eDNA metabarcoding survey reveals marine biogeographic break and transitions over tropical north- western Australia. Divers Distrib. 2021;(00):1–16.
41. Boussarie G, Kiszka JJ, Mouillot D, Bonnin L, Manel S, Kulbicki M, et al. Environmental DNA illuminates the dark diversity of sharks. Sci Adv. 2018;4(5):1–8.
42. Brandl SJ, Tornabene L, Goatley CHR, Casey JM, Morais RA, Côté IM, et al. Demographic dynamics of the smallest marine vertebrates fuel coral-reef ecosystem functioning. Science (80-). 2019;(May):799–802.
43. Kimmerling N, Zuqert O, Amitai G, Gurevich T, Armoza-Zvuloni R, Kolesnikov I, et al. Quantitative species-level ecology of reef fish larvae via metabarcoding. Nat Ecol Evol. 2018;2(2):306–16.
44. Beckley LE, Holliday D, Sutton AL, Weller E, Olivar MP, Thompson PA. Structuring of larval fish assemblages along a coastal-oceanic gradient in the macro-tidal, tropical Eastern Indian Ocean. Deep Res Part II. 2019;161(March 2018):105–19.
45. McLean M, Stuart-Smith RD, Villéger S, Auber A, Edgar GJ, MacNeil MA, et al. Trait similarity in reef fish faunas across the world's oceans. Proc Natl Acad Sci. 2021;118(12):e2012318118.
46. Gaboriau T, Leprieur F, Mouillot D, Hubert N. Influence of the geography of speciation on current patterns of coral reef fish biodiversity across the Indo-Pacific. Ecography (Cop). 2018;41(8):1295–306.
47. Ahmadia GN, Tornabene L, Smith DJ, Pezold FL. The relative importance of regional, local, and evolutionary factors structuring cryptobenthic coral-reef assemblages. Coral Reefs. 2018;37(1):279–93.
48. Coker DJ, DiBattista JD, Sinclair-Taylor TH, Berumen ML. Spatial patterns of cryptobenthic coral-reef fishes in the Red Sea. Coral Reefs. 2018;37(1):193–9.
49. Goatley CHR, González-Cabello A, Bellwood DR. Reef-scale partitioning of cryptobenthic fish assemblages across the Great Barrier Reef, Australia. Mar Ecol Prog Ser. 2016;544(February):271–80.
50. Wang S, Loreau M. Biodiversity and ecosystem stability across scales in metacommunities. Ecol Lett. 2016;19:510–8.

51. Benkwitt C, Wilson S, Graham NAJ. Biodiversity increases ecosystem functions despite multiple stressors on coral reefs. Nat Ecol Evol. 2020;4:916–26.
52. Marques V, Milhau T, Albouy C, Dejean T, Manel S, Mouillot D, et al. GAPeDNA : Assessing and mapping global species gaps in genetic databases for eDNA metabarcoding. 2020;(April):1–13.
53. Ruppert KM, Kline RJ, Rahman MS. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. Glob Ecol Conserv. 2019;17:e00547.
54. Polanco A, Richards FE, Flück B, Valentini A, Altermatt F, Jean- SB, et al. Comparing the performance of 12S mitochondrial primers for fish environmental DNA across ecosystems. Environ DNA. 2021;(June):1–15.
55. Brandt MI, Trouche B, Quintric L, Günther B, Wincker P, Poulain J, et al. Bioinformatic pipelines combining denoising and clustering tools allow for more comprehensive prokaryotic and eukaryotic metabarcoding. Mol Ecol Resour. 2021;21(6):1904–21.
56. Marques V, Guérin PÉ, Rocle M, Valentini A, Manel S, Mouillot D, et al. Blind assessment of vertebrate taxonomic diversity across spatial scales by clustering environmental DNA metabarcoding sequences. Ecography (Cop). 2020;43:1–12.
57. Sales NG, Wangensteen OS, Carvalho DC, Deiner K, Praebel I, McDevitt A, et al. Space-time dynamics in monitoring neotropical fish communities using eDNA metabarcoding. bioRxiv. 2020;1–38.
58. Stauffer S, Jucker M, Keggin T, Marques V, Andrello M, Bessudo S, et al. How many replicates to accurately estimate fish biodiversity using environmental DNA on coral reefs ? Authors : bioRxiv. 2021;
59. Collins RA, Wangensteen OS, O'Gorman EJ, Mariani S, Sims DW, Genner MJ. Persistence of environmental DNA in marine systems. Commun Biol. 2018;1(185):1–12.
60. Collins RA, Bakker J, Wangensteen OS, Soto AZ, Corrigan L, Sims DW, et al. Non-specific amplification compromises environmental DNA metabarcoding with COI. Methods Ecol Evol. 2019;10(11):1985–2001.
61. Zhang S, Zhao J, Yao M. A comprehensive and comparative evaluation of primers for metabarcoding eDNA from fish. Methods Ecol Evol. 2020;2020(January):1609–25.
62. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 2016;4:1–22.
63. Martin. M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 1994;17(1):10–2.
64. Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M. Swarm v2: highly-scalable and high-resolution amplicon clustering. PeerJ. 2015;3:1–12.
65. Boyer F, Mercier C, Bonin A, Le Bras Y, Taberlet P, Coissac E. obitools: A unix-inspired software package for DNA metabarcoding. Mol Ecol Resour. 2016;16(1):176–82.
66. Polanco Fernández A, Marques V, Fopp F, Juhel J, Borrero-Pérez GH, Cheutin M, et al. Comparing environmental DNA metabarcoding and underwater visual census to monitor tropical reef fishes. Environ DNA. 2020 Oct 2;1–15.
67. Spalding MD, Fox HE, Allen GR, Davidson N, Ferdaña ZA, Finlayson M, et al. Marine Ecoregions of the World: A Bioregionalization of Coastal and Shelf Areas. Bioscience. 2007;57(7):573–83.
68. Escalas A, Troussellier M, Yuan T, Bouvier T, Bouvier C, Mouchet MA, et al. Functional diversity and redundancy across fish gut, sediment and water bacterial communities. Environ Microbiol. 2017;19(8):3268–82.
69. Whittaker RH. Evolution and measurement of species diversity. Taxon. 1972;21:213–51.
70. Veech JA, Summerville KS, Crist TO, Gering JC. The additive partitioning of species diversity: Recent revival of an old idea. Oikos. 2002;99(1):3–9.

Figures

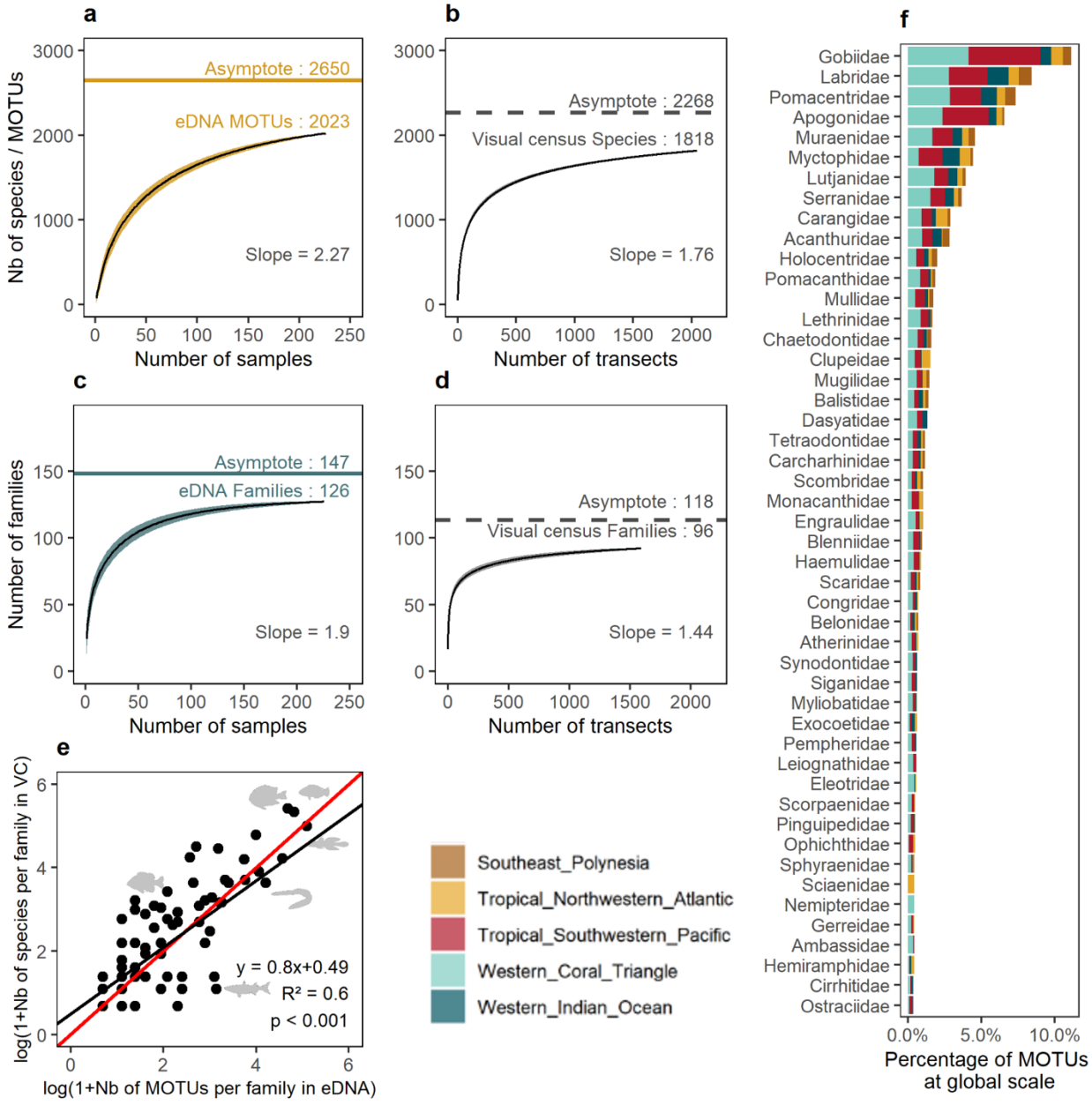
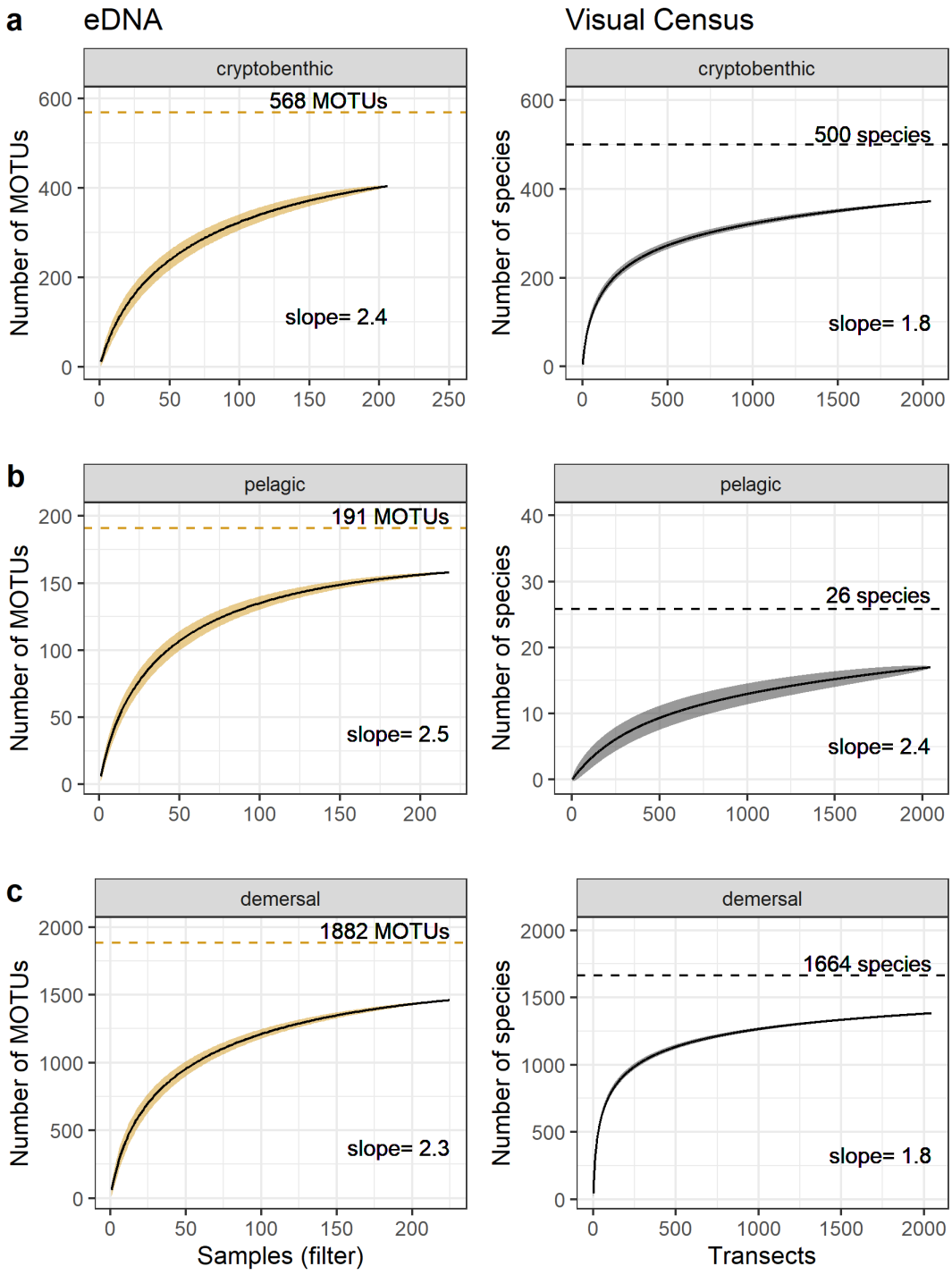


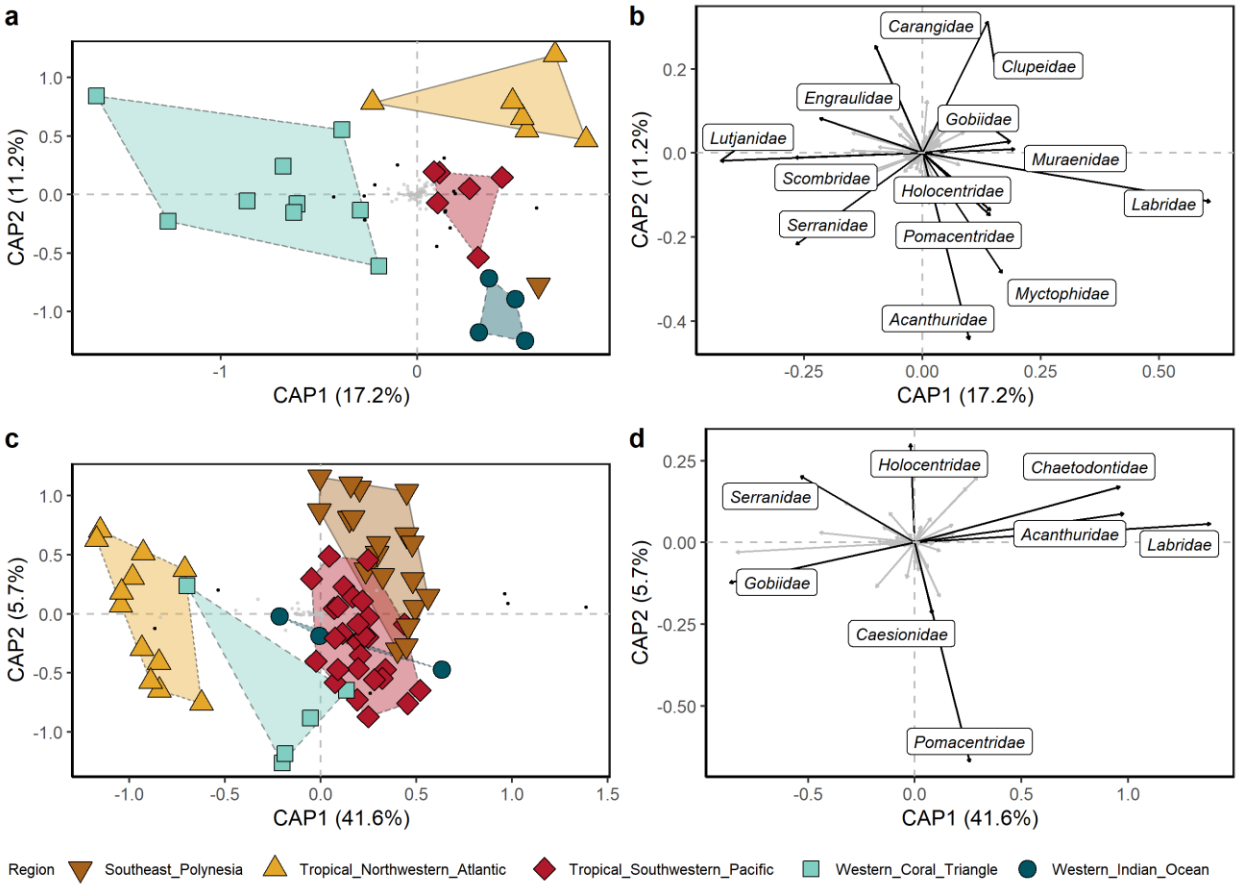
Figure 1. Estimates of overall fish richness from environmental DNA (eDNA) and visual census. (*a*) accumulation curve of molecular operational taxonomic units from eDNA (eDNA MOTUs), (*b*) accumulation curve of species from the visual census database, (*c*) accumulation curve of eDNA families, (*d*) accumulation curve of visual census families. For (*a-d*), Species accumulation model is fitted according to Lomolino method (see methods). (*e*) linear regression (black line) between the number of species per family in visual census data and the number of MOTUs per family in eDNA (log(x+1) transformation) over *n* = 77 families. Each point is a family. Red line is x=y. (*f*) percentage of MOTUs assigned to each family at global scale, and proportion in each region.



579

580 **Figure 2. Estimates of overall fish richness from eDNA and visual census across habitat**
581 **categories**. (*a*) accumulation curve of crypto-benthic eDNA MOTUs (left) and visual census
582 species (right), (*b*) accumulation curve of pelagic MOTUs (left) and visual census species (right),
583 c, accumulation curve of demersal MOTUs (left) and visual census species (right). Accumulation
584 model is fitted with a nonlinear Lomolino model (see Methods).

585



586

587 **Figure 3. Partial Distance-based Redundancy analysis of MOTU proportions of each family**
588 **in each site.** (*a*) dbRDA on eDNA dataset, with 133 families in 26 sites (R^2=0.21, F=3.11,
589 p=0.001), (*b*) families with scores > 95% of scores distribution on each axis for eDNA, (*c*) dbRDA
590 on a subset of Visual Census dataset to select only the sites in the same regions as in the eDNA
591 dataset, with 76 families in 68 sites (R^2=0.5, F=15.8, p=0.001), (*d*) families with scores > 95% of
592 scores distribution on each axis for Visual Census. Axis labels indicate the percentage of variance
593 explained by the 2 first dbRDA dimensions (CAP1 and CAP2).

594

Supplementary information for manuscript

Cross-ocean patterns and processes in fish biodiversity on coral reefs through the lens of eDNA metabarcoding

Laetitia Mathon[1,2,8]*†, Virginie Marques[1,3]†, David Mouillot[3,4], Camille Albouy[5], Marco Andrello[3,17], Florian Baletaud[2,3,6], Giomar H. Borrero-Pérez[7], Tony Dejean[8], Graham J. Edgar[9], Jonathan Grondin[8], Pierre-Edouard Guerin[1], Régis Hocdé[3], Jean-Baptiste Juhel[3], Kadarusman[10], Eva Maire[3,11], Gael Mariani[3], Matthew McLean[12], Andrea Polanco F.[7], Laurent Pouyaud[13], Rick D. Stuart-Smith[9], Hagi Yulia Sugeha[14], Alice Valentini[8], Laurent Vigliola[2], Indra B Vimono[14], Loïc Pellissier[15,16]‡, Stéphanie Manel[1]*‡

In order to verify that our unbalanced eDNA sampling did not bias our results and patterns, we ran the analyses after performing two types of rarefaction:

- We randomly sampled 4 stations in all the other sites (random sampling repeated 50 times), as there is only 4 stations in the site in Southeast Polynesia, the lowest sampled region
- We removed the lowest sampled region, Southeast Polynesia,from our dataset and we rarefied the dataset according to the second least sampled regions (Western Indian, 4 sites) by randomly sampling 4 sites in all other regions (random sampling repeated 50 times)

Original complete dataset richness : 2023 MOTUs and 127 families

Rarefaction - 4 stations per site

Number of MOTUs: 1928 +/- 30 (SD)
Number of families: 124 +/- 2 (SD)

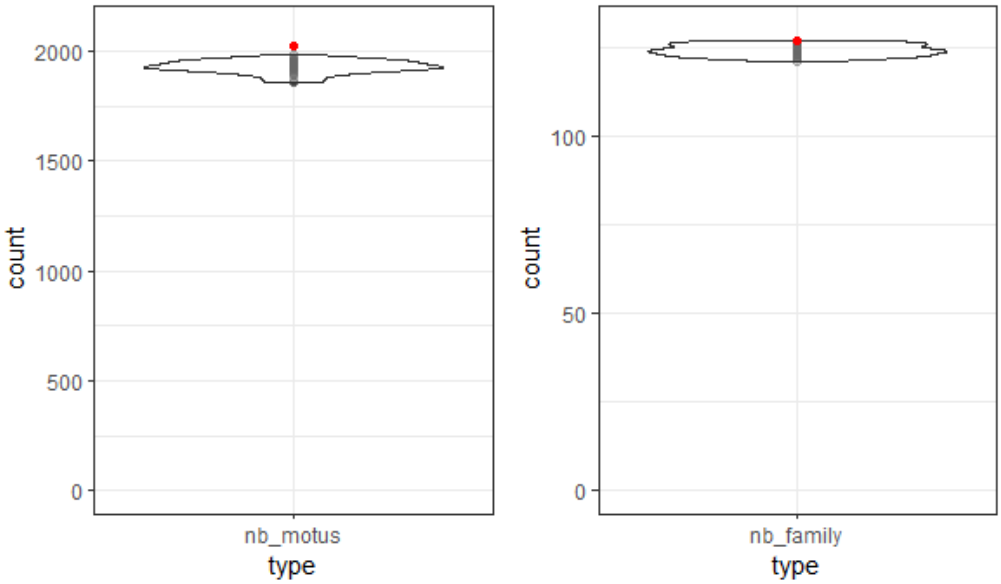


Fig. 1 *Richness of MOTUs (left) and families (right) in the 50 datasets rarefied by sites. Red dots are the richness in the original dataset.*

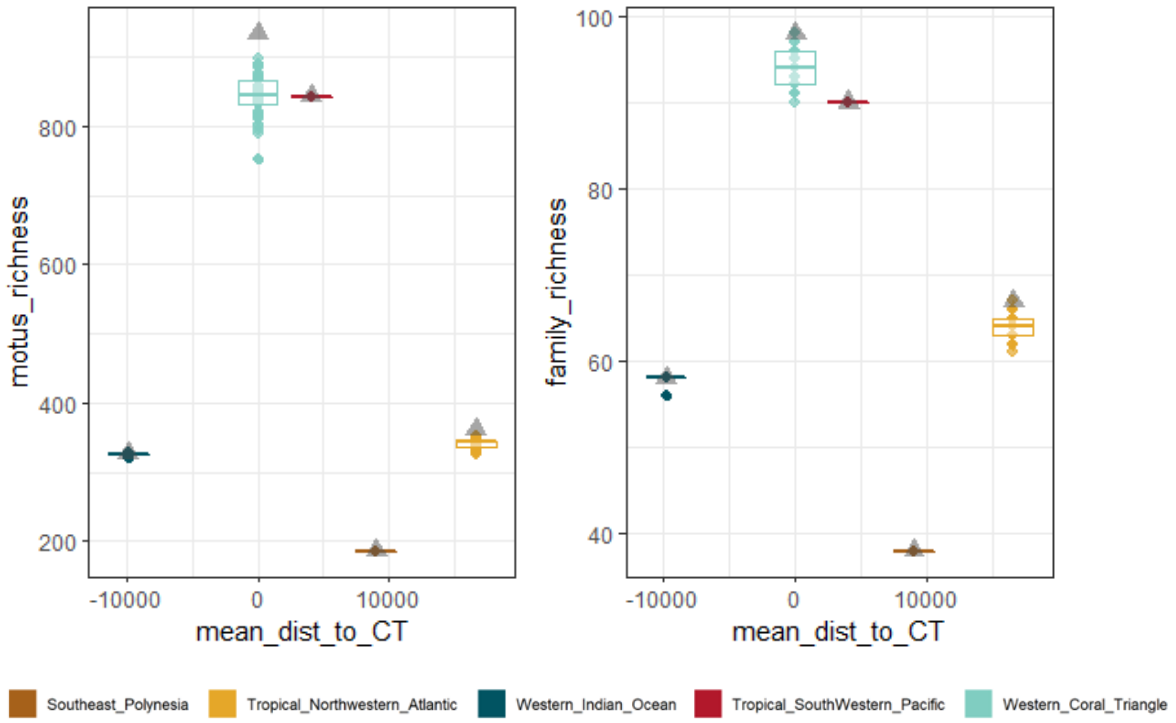


Fig. 2 *Regional MOTUs richness (left) and family richness (right). Boxplots represent the distribution across the 50 rarefied datasets, sampled randomly. Gray triangles represent regional richness in the original dataset.*

Rarefaction - 4 sites per region

Number of MOTUs: 1616 +/- 114 (SD)
Number of families: 116 +/- 5 (SD)

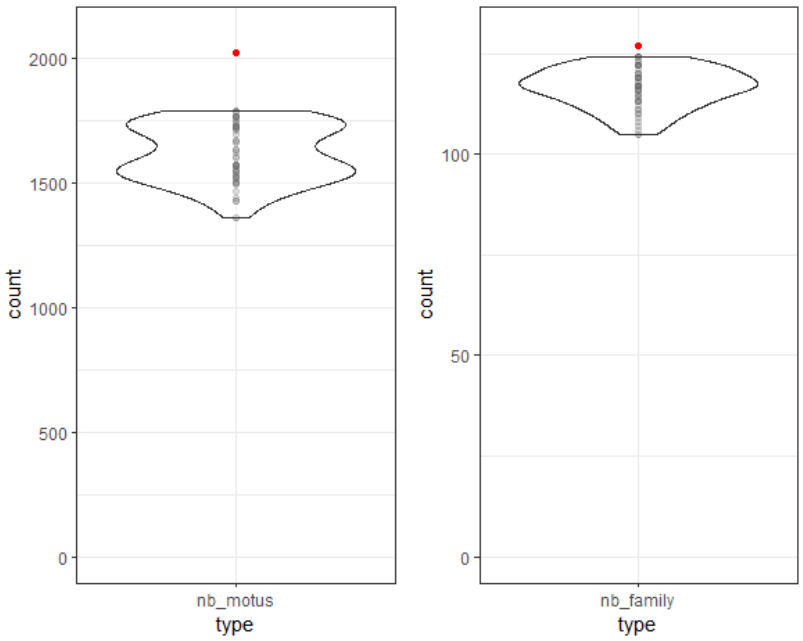


Fig. 3 *Richness of MOTUs (left) and families (right) in the 50 datasets rarefied by regions. Red dots are the richness in the original dataset.*

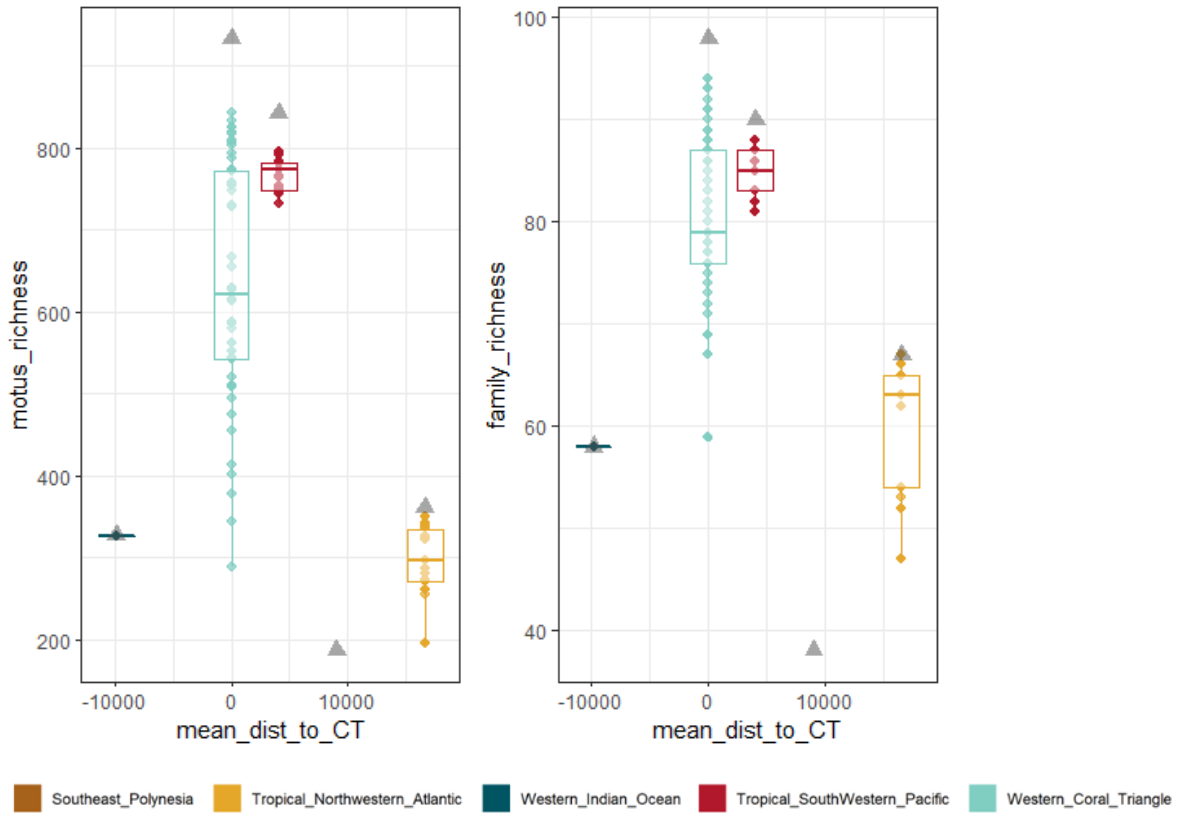


Fig. 4 *Regional MOTUs richness (left) and family richness (right). Boxplots represent the distribution across the 50 rarefied datasets, sampled randomly. Gray triangles represent regional richness in the original dataset.*

Table 1. Diversity partitioning across scales calculated for the two types of rarefactions.

component	Full dataset	Rarefaction 4 stations per site (%)	Rarefaction 4 sites per region(%)
$\underline{\alpha}_{station}$	5.7 %	6 ±0.08	7.16 ±0.39
$\underline{\beta}_{inter-station}$	5.9 %	5.77 ±0.06	7.69 ±0.45
$\underline{\beta}_{inter-site}$	14.8 %	14.84 ±0.07	16.41 ±0.51
$\beta_{inter-region}$	73.7 %	73.67 ±0.09	68.73 ±0.51

To investigate the impact of the sampling effort at each site on diversity estimations, we built MOTUs accumulation curves at the site level.

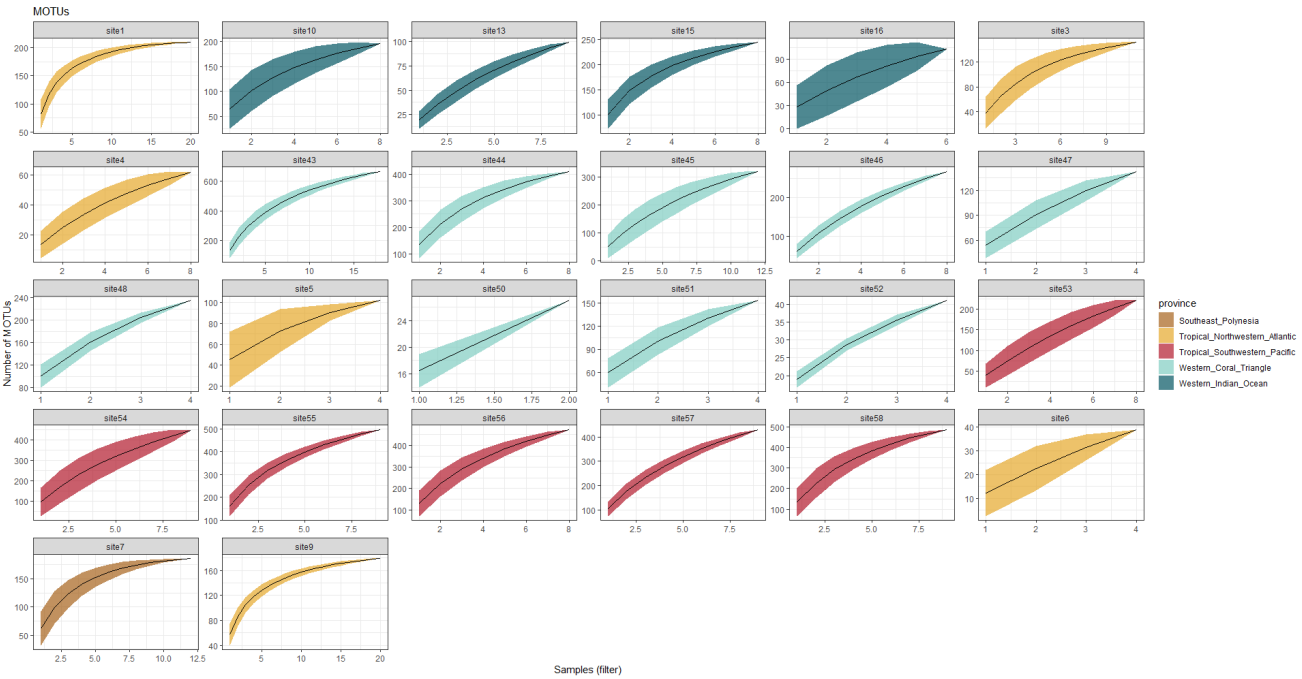


Fig. 5 Accumulation curve of molecular operational taxonomic units from eDNA at the site level (colors indicate the region of each site). Species accumulation model is fitted according to Lomolino method.

A)

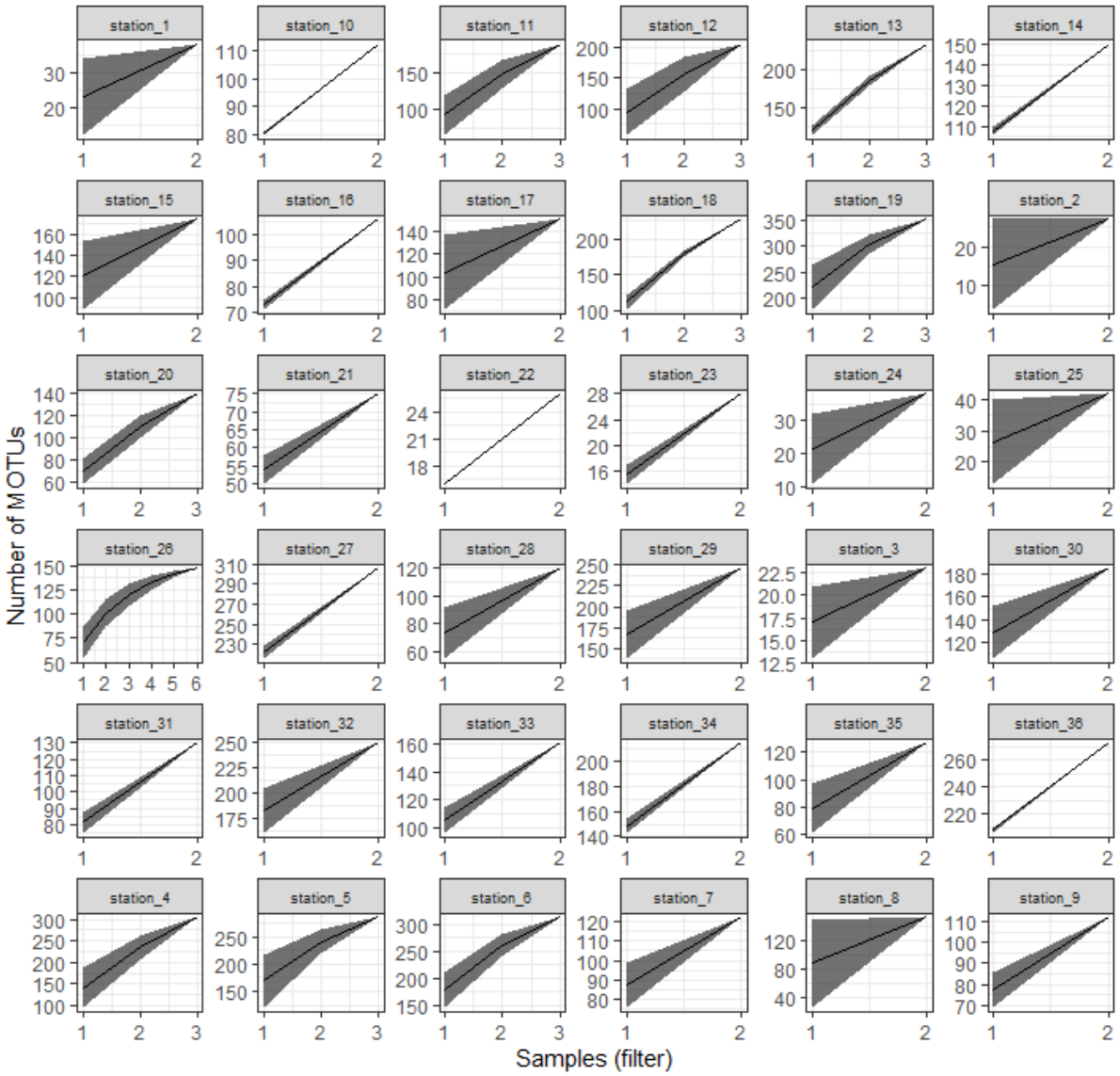
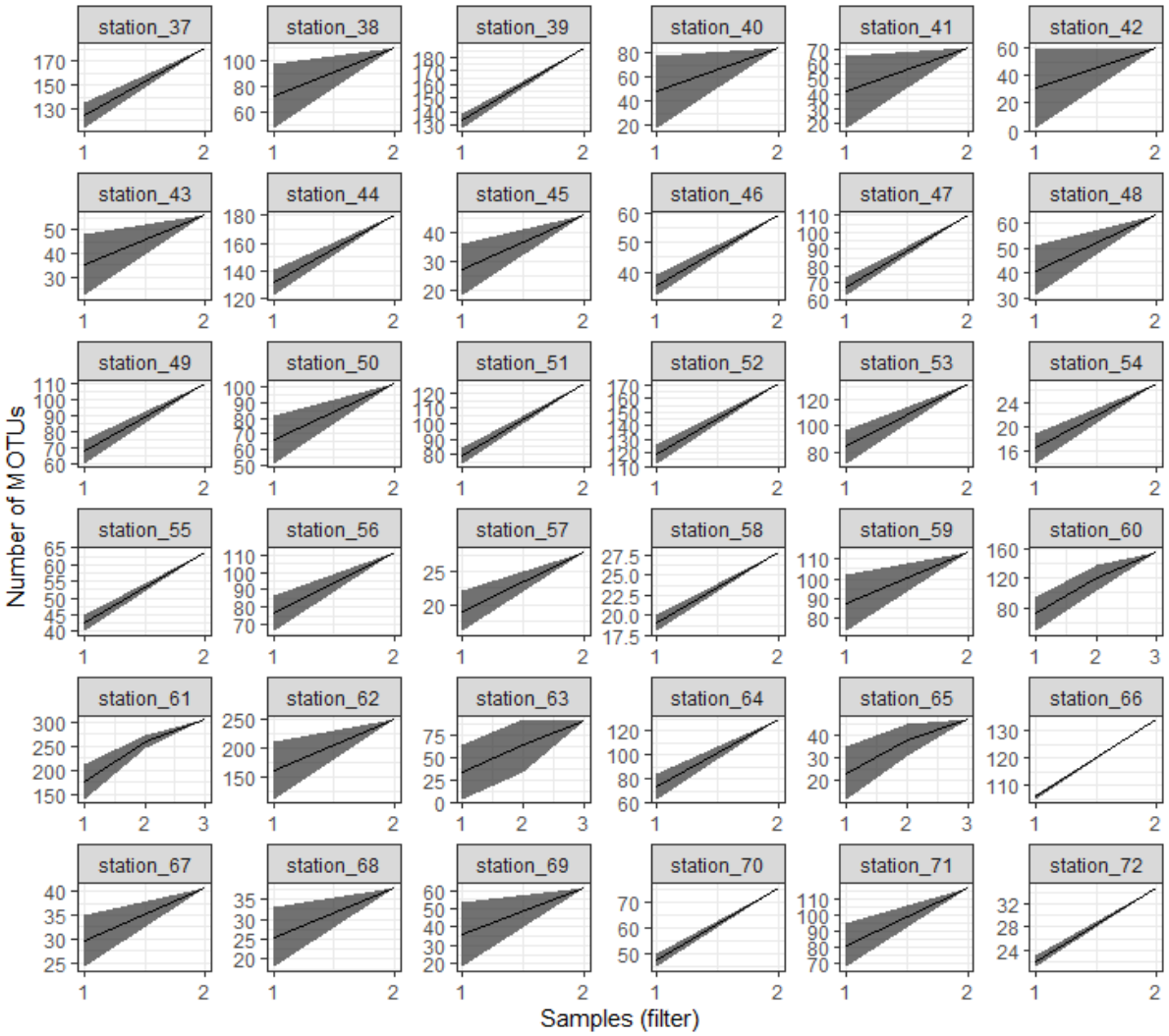
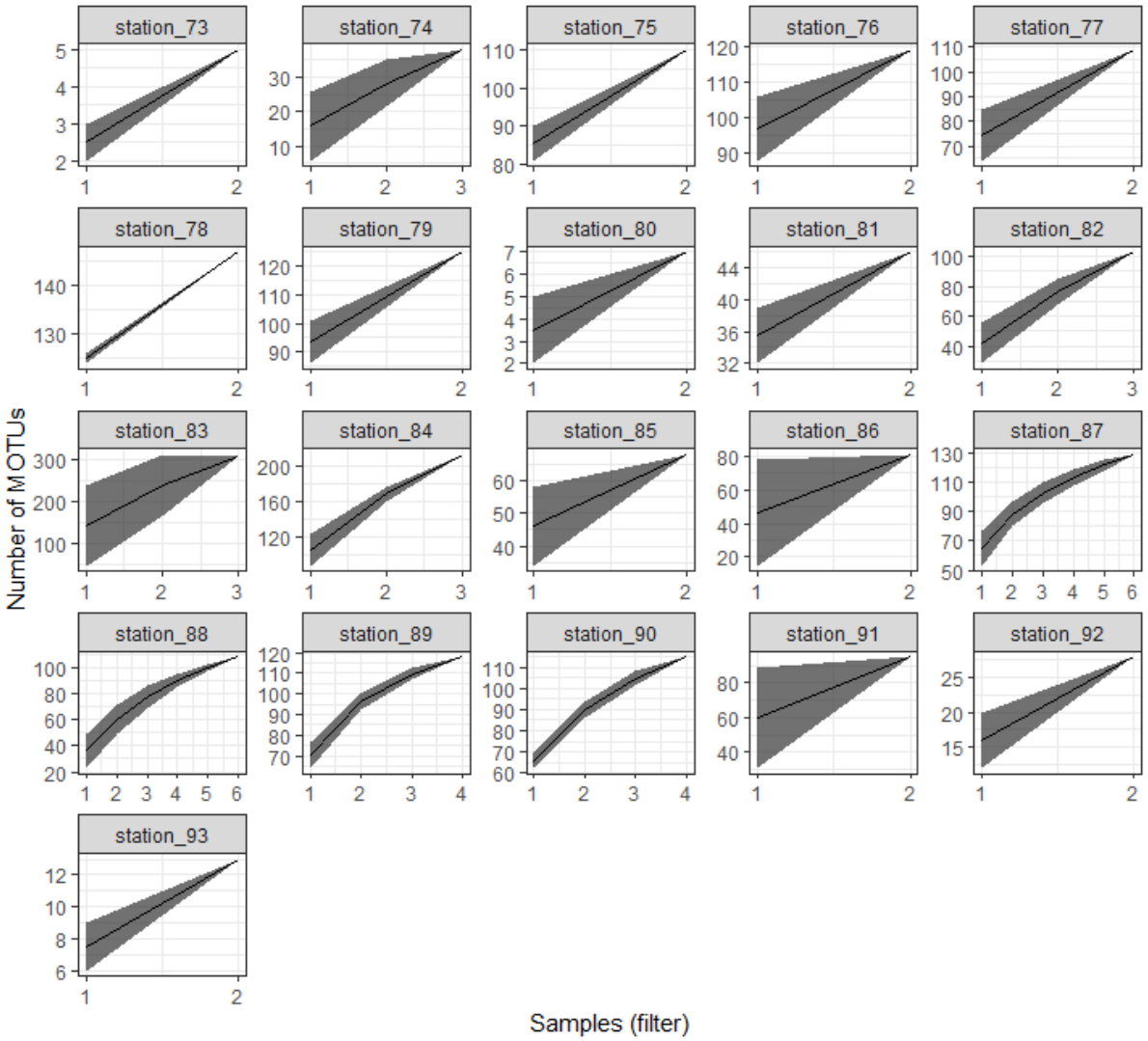


Fig. 6 Accumulation curve of molecular operational taxonomic units from eDNA at the station level. Species accumulation model is fitted according to Lomolino method. A) stations 1 to 36, B) stations 37 to 72, C) stations 73 to 93

B)


station_37
station_38
station_39
station_40
station_41
station_42
station_43
station_44
station_45
station_46
station_47
station_48
station_49
station_50
station_51
station_52
station_53
station_54
station_55
station_56
station_57
station_58
station_59
station_60
station_61
station_62
station_63
station_64
station_65
station_66
station_67
station_68
station_69
station_70
station_71
station_72
Number of MOTUs
Samples (filter)

C)


station_73
station_74
station_75
station_76
station_77
station_78
station_79
station_80
station_81
station_82
station_83
station_84
station_85
station_86
station_87
station_88
station_89
station_90
station_91
station_92
station_93
Number of MOTUs
Samples (filter)

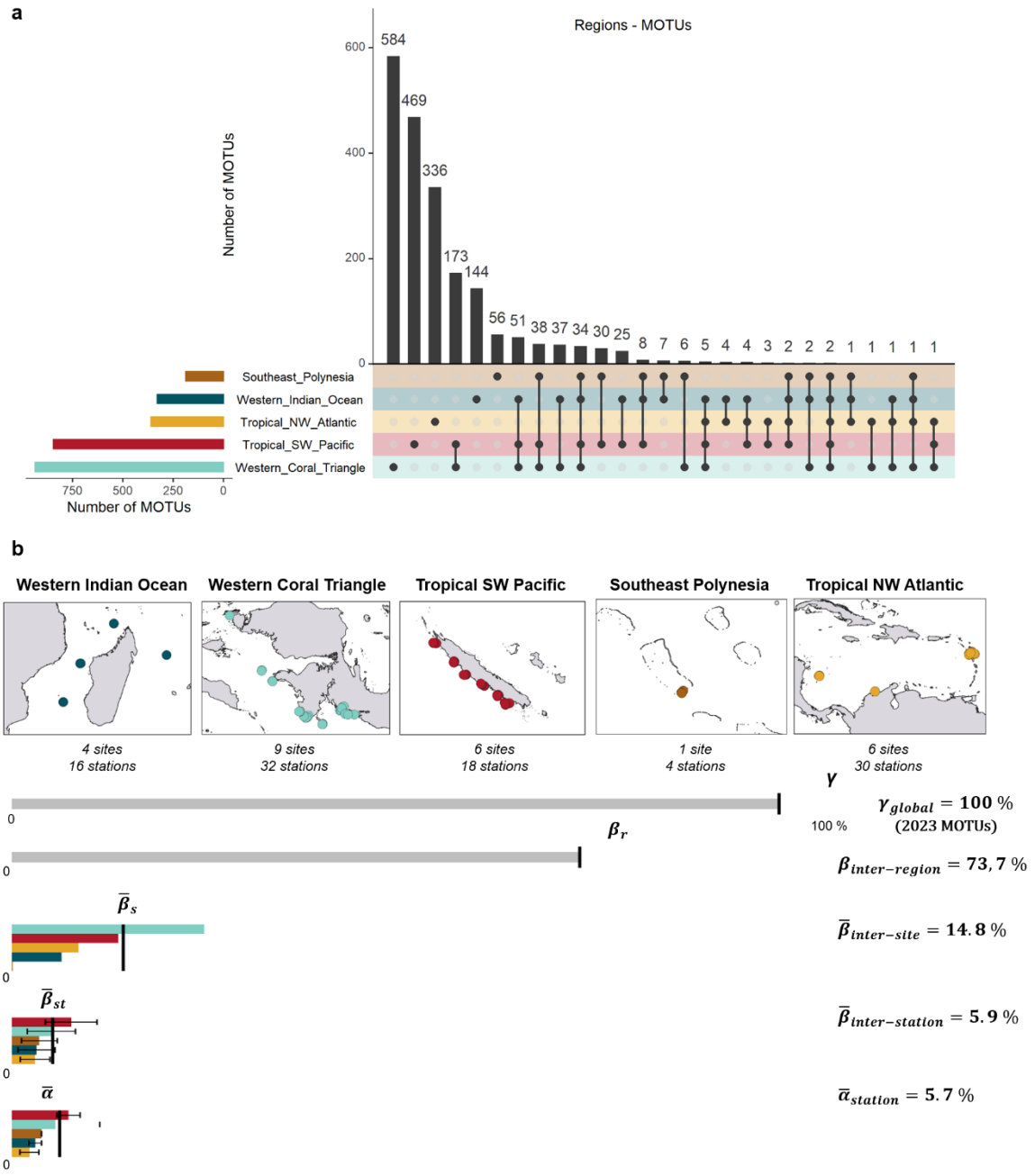


Figure 4. Hierarchical partitioning of MOTU occurrences across spatial scales. (*a*) Number of MOTUs found in only one region, or shared between 2, 3, 4 or all 5 regions. Histograms indicate the number of MOTUs present in all the regions identified by the dots in the lower part. (*b*) Global fish diversity (γ_{global}) is partitioned into $\beta_{inter\text{-}region}$ + mean $\beta_{inter\text{-}site}$ + mean $\beta_{inter\text{-}station}$ + mean $\bar{\alpha}_{station}$. Mean values at global scales are indicated with the black vertical segments. For $\beta_{inter\text{-}site}$, $\beta_{inter\text{-}station}$ and $\bar{\alpha}_{station}$, mean values are given for each region (colored bars) with the standard errors. $\beta_{inter\text{-}region}$ contributes the highest to gamma global (73.7%).

Supplementary information for manuscript

Cross-ocean patterns and processes in fish biodiversity on coral reefs through the lens of eDNA metabarcoding

Laetitia Mathon[1,2,8]*†, Virginie Marques[1,3]†, David Mouillot[3,4], Camille Albouy[5], Marco Andrello[3,17], Florian Baletaud[2,3,6], Giomar H. Borrero-Pérez[7], Tony Dejean[8], Graham J. Edgar[9], Jonathan Grondin[8], Pierre-Edouard Guerin[1], Régis Hocdé[3], Jean-Baptiste Juhel[3], Kadarusman[10], Eva Maire[3,11], Gael Mariani[3], Matthew McLean[12], Andrea Polanco F.[7], Laurent Pouyaud[13], Rick D. Stuart-Smith[9], Hagi Yulia Sugeha[14], Alice Valentini[8], Laurent Vigliola[2], Indra B Vimono[14], Loïc Pellissier[15,16]‡, Stéphanie Manel[1]*‡

Method S1. Environmental DNA collection and sample processing

Environmental DNA (eDNA) samples of seawater were collected in five marine regions, encompassing 26 sites (defined as groups of stations separated by at least 35 km), 100 stations and 226 samples (figure S1-S2, (1)). Three different sampling methods were used: collection of 2 L of water in DNA-free sterile plastic bags on the surface water from a dinghy as well as close circuit rebreather diving (depths between 10 – 40 m) as close as possible to the habitat (1); 2-km long filtration transect with two replicates (one on each side of a boat at each station for 30 min), for a total of 30 of water just under the surface; 2 km-long filtration of water along a transect, approximately 5 m above the substrate, using a long pipe, from the boat. Details on which sampling method was used in each region are provided in table S7. For each sample collected with the first sampling protocol, 2 L of seawater were filtered with sterile Sterivex filter capsules (Merck© Millipore; pore size 0.22µm) using disposable sterile syringes. Immediately after, the filter units were filled with CL1 Conservation buffer (SPYGEN, le Bourget du Lac, France) and stored in 50 mL screw-cap tubes at room temperature. The eDNA filtration device for the other two sampling protocols was composed of an Athena® peristaltic pump (Proactive Environmental Products LLC, Bradenton, Florida, USA; nominal flow of 1.0 $L.min^{-1}$), a VigiDNA® 0.2µM cross flow filtration capsule with a polyethersulfone membrane (SPYGEN, le Bourget du Lac, France) and disposable sterile tubing for each filtration capsule. At the end of each filtration, the water inside the capsules were emptied, and the capsules were filled with 80 mL of CL1 Conservation buffer (SPYGEN, le Bourget du Lac, France) and stored at room temperature. For each sampling campaign, a strict contamination control protocol was followed in both field and laboratory stages (2,3), and each water sample processing included the use of disposable gloves and single-use filtration equipment. Negative field controls were performed in multiple sites across all sampling locations, and revealed no

contamination from the boat or samplers. A large number of extraction and amplification negative controls were performed for each sample (see next section).

Method S2. eDNA extraction, amplification and sequencing

DNA extraction was performed in a dedicated DNA laboratory (SPYGEN, www.spygen.com) equipped with positive air pressure, UV treatment and frequent air renewal. Decontamination procedures were conducted before and after all manipulations. Each filtration capsule was agitated for 15 min on a S50 Shaker (Cat Ingenieurbüro™) at 800 rpm. For sterivex filters, the buffer was retrieved using a 3 mL BD Disposable Syringe with Luer-Lok™ tips, emptied into a 50 mL tube containing 33 mL of ethanol and 1.5 mL of 3M sodium acetate and, finally, stored for at least one night at -20°C. The tubes were centrifuged at 15,000 × g for 15 min at 6°C, and the supernatants were discarded. After this step, 720 µL of ATL buffer from the DNeasy Blood & Tissue Extraction Kit (Qiagen) was added to each tube. Each tube was then vortexed, and the supernatant was transferred to a 2-mL tube containing 20 µL of Proteinase K. The tubes were finally incubated at 56°C for two hours. Subsequently, DNA extraction was performed using NucleoSpin® Soil (MACHEREY-NAGEL GmbH & Co., Düren Germany) starting from step 6 and following the manufacturer's instructions. The elution was performed by adding 100 µL of SE buffer twice. For VigiDNA 0.2 µM filters, each capsule, containing the CL1 buffer, was agitated for 15 min on an S50 shaker (cat Ingenieurbüro™) at 800 rpm and then the buffer was emptied into two 50-mL tube before being centrifuged for 15 min at 15,000×g. The supernatant was removed with a sterile pipette, leaving 15 mL of liquid at the bottom of each tube. Subsequently, 33 mL of ethanol and 1.5 mL of 3M sodium acetate were added to each 50-mL tube and stored for at least one night at -20°C. The DNA extraction was performed as described above except that the two 50 mL tubes per filtration capsule were extracted separately then the two DNA samples were pooled before the amplification step. A teleost-specific 12S mitochondrial rRNA primer pair (teleo, forward primer - ACACCGCCCGTCACTCT, reverse primer – CTTCCGGTACACTTACCATG (2)) was used for the amplification of metabarcode sequences. As we analysed our data using MOTUs as a proxy for species to overcome genetic

database limitations, we chose to amplify only one marker. Twelve DNA amplifications PCR per sample were performed in a final volume of 25 µL, using 3 µL of DNA extract as the template. The amplification mixture contained 1 U of AmpliTaq Gold DNA Polymerase (Applied Biosystems, Foster City, CA), 10 mM Tris-HCl, 50 mM KCl, 2.5 mM MgCl2, 0.2 mM each dNTP, 0.2 µM of each primers, 4 µM human blocking primer for the "teleo" primers (2) and 0.2 µg/µL bovine serum albumin (BSA, Roche Diagnostic, Basel, Switzerland). The PCR mixture was denatured at 95°C for 10 min, followed by 50 cycles of 30 s at 95°C, 30 s at 55°C, 1 min at 72 °C and a final elongation step at 72°C for 7 min. The teleo primers were 5'-labeled with an eight-nucleotide tag unique to each PCR replicate with at least three differences between any pair of tags, allowing the assignment of each sequence to the corresponding sample during sequence analysis. The tags for the forward and reverse primers were identical for each PCR replicate. Negative extraction controls and negative PCR controls (ultrapure water) were amplified (with 12 replicates as well) and sequenced in parallel to the samples to monitor possible contaminations. After amplification, samples were titrated using capillary electrophoresis (QIAxcel; Qiagen GmbH, Hilden, Germany) and purified using a MinElute PCR purification kit (Qiagen GmbH, Hilden, Germany). The purified PCR products were pooled in equal volumes, to achieve a theoretical sequencing depth of 1,000,000 reads per sample. Library preparation and sequencing were performed at Fasteris (Geneva, Switzerland). A total of 18 libraries were prepared using MetaFast protocol. A paired-end sequencing (2x125 bp) was carried out using an Illumina HiSeq 2500 sequencer with the HiSeq Rapid Flow Cell v2 using the HiSeq Rapid SBS Kit v2 (Illumina, San Diego, CA, USA) or a MiSeq (2x125 bp, Illumina, San Diego, CA, USA) using the MiSeq Flow Cell Kit v3 (Illumina, San Diego, CA, USA) or a NextSeq sequencer (2x125 bp, Illumina, San Diego, CA, USA) with the NextSeq Mid kit following the manufacturer's instructions. This generated an average of 1,335,896 sequence reads (paired-end Illumina) per sample.

Methods S3. Bioinformatic analysis

Following sequencing, reads were processed using clustering and post-clustering cleaning to remove errors and estimate the number of species using Molecular Operational Taxonomic Units (MOTUs) (4). First, reads were assembled using *vsearch* (5), then demultiplexed and trimmed using *cutadapt* (6) and clustering was performed using *Swarm* v.2 (7) with a minimum distance of 1 mismatch between clusters. Taxonomic assignment of MOTUs was carried out using the Lower Common Ancestor (LCA) algorithm *ecotag* implemented in the Obitools toolkit (8) and the European Nucleotide Archive (ENA (9)) as a reference database (release 143, March 2020). It assigns a taxonomy to sequences even when the sequence match is not perfect, based on NCBI taxonomic tree of species to consider the current knowledge on molecular diversity per branch and assign a taxonomy at the lowest possible rank. If the sequence matches several identifications with equal percentages of similarity, *ecotag* assigns to the upper taxonomic level common between all possible matches. We then applied quality filters to be conservative in our estimates. We discarded all observations with less than 10 reads, and present in only one PCR per site to avoid spurious MOTUs originating from a PCR error. Then, errors generated by index-hopping (10) were filtered using a threshold empirically determined per sequencing batch using experimental blanks (combinations of tags not present in the libraries) (11), and tag-jump (12) was corrected using a threshold of 0.001 of occurrence for a given MOTU within a library. Taxonomic assignments at the species level were accepted if the percentage of similarity with the reference sequence was 100%, at the genus level if the similarity was between 90 and 99%, and at the family level if the similarity was > 85%. If these criteria were not met, the MOTU was left unassigned. The post-LCA algorithm correction threshold of 85% similarity for family assignment was chosen to include a maximum of correct family assignment while minimizing the risk of adding wrong family assignments in the family

detections, and only 21% of assigned MOTUs are assigned to the family level with a similarity between 85 and 90% (Table S8).

Methods S4. Visual Census data

The visual census survey data used here is a subset (2047 transects, figure S1)) of the complete visual census data (3027 transects) provided by the Reef Life Survey (13), and comprises all species observed on standardized 50 m surveys at sites in tropical biogeographic realms (14). We selected only the most recent survey for each transect and only transects with more than five percent of coral cover. The visual census method involves divers surveying duplicate 5-m-wide blocks along each 50 m transect in which all fish species sighted are recorded, and then in duplicate 1-m-wide blocks in which the divers closely search the substrate (including in crevices) for smaller crypto-benthic fishes (13). The full list of fish species for each survey from both methods was used for this study. Full details of the methods are provided in an online methods manual at www.reeflifesurvey.com.

Methods S5. Statistical analysis

Accumulation curves were calculated for species per 500 m^2 transect, MOTUs per eDNA sample, and families per transect and sample. We used the function “specaccum” from the R package “vegan” v.2.5-6, with the "exact" method, which calculates the expected species accumulation curve using a sample-based rarefaction method. We then used the function “fitspecaccum” to fit five nonlinear species accumulation models (Lomolino, Michaelis-Menten, Gompertz, Asymp and Logis). The best model was selected based on AIC, and its asymptote recorded. Sampling effort varied between regions in the Visual Census dataset, with Australia having twice as many transects as other regions. In order to assess the impact of this irregular sampling on the estimates measured with accumulation curves, we randomly subset half of the transects in the 3 most sampled regions in Australia, and calculated again the accumulation curves for species and families (figure S12). The results were unchanged.

Pearson’s correlation coefficient was calculated between the number of MOTUs per family in the eDNA dataset and the number of species per family in the visual census dataset. Linear regression models were fitted between the number of MOTUs per family in the eDNA dataset and the number of species per family in the visual census dataset, after log(x+1) transformation (figure 1*e*).

Accumulation curves were also calculated by sub-setting MOTUs belonging to crypto-benthic orders, or to pelagic families, for both datasets (figure 2). The asymptote was calculated as described above.

MOTU proportions of each fish family (i.e. family proportions) were calculated as the number of MOTUs assigned to each family in each site for eDNA and species assigned to each family in each site for the Visual Census. We performed distance-based Redundancy Analysis (dbRDA) on these family proportions, with *region* and *site richness* as explanatory variables,

using the function *capscale* from the *vegan* package. We subset the Visual Census to select only the 68 sites that fell into the 5 regions in common with the eDNA dataset. Total dbRDA provided the effects of each of the variables and their interaction. We then calculated partial dbRDA to measure the effect of the Region while correcting for the effect of site richness (figure 3, table S3).

As eDNA is rapidly degraded in tropical inshore waters (15,16), and based on caged experiments in marine ecosystems (17), we assume the eDNA signal comes from individuals present in close proximity to the filtering station. Thus, the detection of species not typically considered as coral reef fishes may reveal their use of reef habitats from time to time (18).

We applied an additive partitioning framework (19,20) to separate the total MOTUs diversity at the global scale (ɣ global) into contributions at smaller scales from regions to local richness. More precisely, global MOTUs diversity was expressed as the sum of inter-region difference, the mean inter-site difference, the mean inter-station difference and mean station MOTUs diversity with: $\gamma_{global} = \beta_{inter\text{-}region} + \text{mean } \beta_{inter\text{-}site} + \text{mean } \beta_{inter\text{-}station} + \text{mean } \bar{\alpha}_{station}$. In this additive framework, the three levels of biodiversity (21) (i.e. α, β and γ) are expressed with the same unit and consequently the contribution of α and β diversity to total diversity (ɣ) can be directly compared (22,23). The diversity partitioning in figure 4 has been calculated with sites defined as groups of stations distant from 35km. In order to assess the influence of the spatial scale in site definition on the diversity partition, we repeated the diversity analysis with sites defined as groups of stations distant from 10 or 20 km (table S4). The results were similar.

We analyzed the distribution of fish MOTU and species occurrences using global species abundance distribution (gSAD) which plots, on a log-log scale, the number of species as a function of the number of observations (24). This representation has the advantage of being comparable between datasets sampled with different methods and allowing the testing of

several species assembly rules and models at large scale. For example, the unified neutral theory of biogeography (UNTB) (25) would produce a gSAD following a log series model or Pareto model with a slope $\beta = -1$ while niche-based processes would provide β values indicating more or less rare species than under the UNTB if β values are respectively higher or lower than -1. Testing whether the gSAD is best fit by a log series or a Pareto distribution (where β is allowed to vary) provides a test of neutral dynamics. Additionally, a third model, coined the Pareto with exponential finite adjustment, adds an exponential "bending" parameter to the Pareto model allowing the right tail to drop down because of finite sample size. Thus, fitting the Pareto or the Pareto with exponential finite adjustment provides a test of neutral or niche dynamics with a β value $\neq -1$ rejecting the neutral theory while a β value <-1 indicates more rare species than under neutrality and >-1 fewer. We summed all fish MOTUs and species observations across all samples obtained with eDNA and visual census data to build gSAD that were fitted with a log series, Pareto and Pareto with exponential finite adjustment (Pareto bended) distribution using maximum likelihood estimation.

References

1. Hocdé R, Vimono I, Suruwaki A, Tuti Y, Utama R, Mohammad A, et al. Mission report : LENGGURU 2017 Expedition ' Biodiversity assessment in reef twilight zone and cloud forests ', R / V AIRAHA 2 , 1st October 2017 – 30th November 2017 , Kaimana Regency , West Papua , Indonesia [Research Report]. Institut de Recherche pour le Développement (IRD), France; Pusat Penelitian Oseanografi, Lembaga Ilmu Pengetahuan (LIPI-P2O), Indonesia; Politeknik Kelautan Dan Perikanan Sorong (Politeknik-KP-Sorong), Papua Barat, Indonesia. 2020.
2. Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, et al. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Mol Ecol. 2016;25(4):929–42.
3. Goldberg CS, Turner CR, Deiner K, Klymus KE, Thomsen PF, Murphy MA, et al. Critical considerations for the application of environmental DNA methods to detect aquatic species. Methods Ecol Evol. 2016;7(11):1299–307.
4. Marques V, Guérin PÉ, Rocle M, Valentini A, Manel S, Mouillot D, et al. Blind assessment of vertebrate taxonomic diversity across spatial scales by clustering environmental DNA metabarcoding sequences. Ecography (Cop). 2020;43:1–12.
5. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 2016;4:1–22.
6. Martin. M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 1994;17(1):10–2.
7. Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M. Swarm v2: highly-scalable and high-resolution amplicon clustering. PeerJ. 2015;3:1–12.
8. Boyer F, Mercier C, Bonin A, Le Bras Y, Taberlet P, Coissac E. obitools: A unix-inspired software package for DNA metabarcoding. Mol Ecol Resour. 2016;16(1):176–82.
9. Leinonen R, Akhtar R, Birney E, Bower L, Cerdeno-Tárraga A, Cheng Y, et al. The European nucleotide archive. Nucleic Acids Res. 2011;39(SUPPL. 1):44–7.
10. MacConaill LE, Burns RT, Nag A, Coleman HA, Slevin MK, Giorda K, et al. Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. BMC Genomics. 2018;19(1):1–10.
11. Taberlet P, Bonin A, Coissac E, Zinger L. Environmental DNA: For biodiversity research and monitoring. 2018.
12. Schnell IB, Bohmann K, Gilbert MTP. Tag jumps illuminated - reducing sequence-to-sample misidentifications in metabarcoding studies. Mol Ecol Resour. 2015;15(6):1289–303.
13. Edgar GJ, Stuart-Smith RD. Systematic global assessment of reef fish communities by the Reef Life Survey program. Sci Data. 2014;1:1–8.
14. Spalding MD, Fox HE, Allen GR, Davidson N, Ferdaña ZA, Finlayson M, et al. Marine Ecoregions of the World: A Bioregionalization of Coastal and Shelf Areas. Bioscience. 2007;57(7):573–83.
15. Harrison JB, Sunday JM, Rogers SM. Predicting the fate of eDNA in the environment and implications for studying biodiversity. Proc R Soc B Biol Sci. 2019;286(1915):1–9.
16. Collins RA, Wangensteen OS, O'Gorman EJ, Mariani S, Sims DW, Genner MJ. Persistence of environmental DNA in marine systems. Commun Biol. 2018;1(185):1–12.
17. Murakami H, Yoon S, Kasai A, Minamoto T, Yamamoto S, Sakata MK, et al.

Dispersion and degradation of environmental DNA from caged fish in a marine environment. Fish Sci. 2019;85(2):327–37.
18. Sambrook K, Hoey AS, Andréfouët S, Cumming GS, Duce S, Bonin MC. Beyond the reef: The widespread use of non-reef habitats by coral reef fishes. Fish Fish. 2019;20(5):903–20.
19. Belmaker J, Ziv Y, Shashar N, Connolly SR. Regional variation in the hierarchical partitioning of diversity in coral-dwelling fishes. Ecology. 2008;89(10):2829–40.
20. Escalas A, Troussellier M, Yuan T, Bouvier T, Bouvier C, Mouchet MA, et al. Functional diversity and redundancy across fish gut, sediment and water bacterial communities. Environ Microbiol. 2017;19(8):3268–82.
21. Whittaker RH. Evolution and measurement of species diversity. Taxon. 1972;21:213–51.
22. Lande R. Statistics and Partitioning of Species Diversity , and Similarity among Multiple Communities. Oikos. 1996;76(1):5–13.
23. Veech JA, Summerville KS, Crist TO, Gering JC. The additive partitioning of species diversity: Recent revival of an old idea. Oikos. 2002;99(1):3–9.
24. Enquist BJ, Feng X, Boyle B, Maitner B, Newman EA, Jørgensen PM, et al. The commonness of rarity: Global and future distribution of rarity across land plants. Sci Adv. 2019;5(11):1–14.
25. Hubbell SP. The unified neutral theory of biodiversity and biogeography. Princeton University Press., editor. Vol. 32. 2001.

Supplementary information for manuscript

Cross-ocean patterns and processes in fish biodiversity on coral reefs through the lens of eDNA metabarcoding

Laetitia Mathon[1,2,8]*†, Virginie Marques[1,3]†, David Mouillot[3,4], Camille Albouy[5], Marco Andrello[3,17], Florian Baletaud[2,3,6], Giomar H. Borrero-Pérez[7], Tony Dejean[8], Graham J. Edgar[9], Jonathan Grondin[8], Pierre-Edouard Guerin[1], Régis Hocdé[3], Jean-Baptiste Juhel[3], Kadarusman[10], Eva Maire[3,11], Gael Mariani[3], Matthew McLean[12], Andrea Polanco F.[7], Laurent Pouyaud[13], Rick D. Stuart-Smith[9], Hagi Yulia Sugeha[14], Alice Valentini[8], Laurent Vigliola[2], Indra B Vimono[14], Loïc Pellissier[15,16]‡, Stéphanie Manel[1]*‡

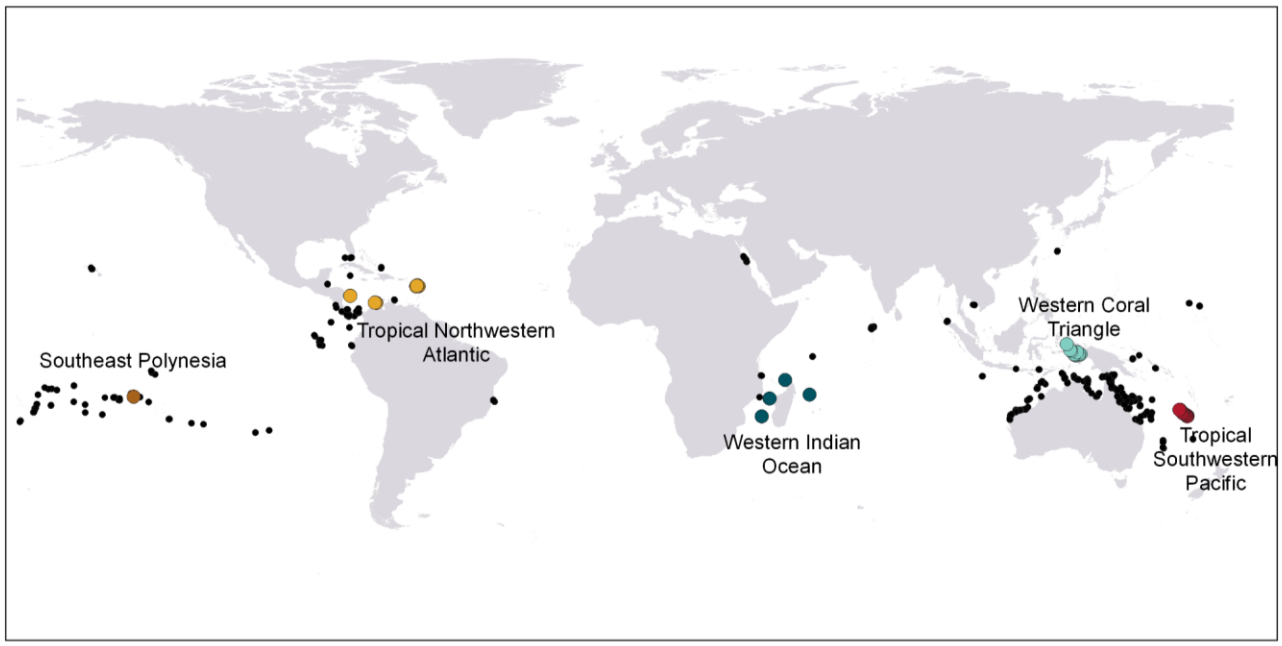


Figure S1. Global map of the sampling locations. The 26 eDNA sampling sites (including 100 stations) are represented by colored dots (colors represent regions). The 219 UVC sampling sites (including 2,047 transects) are represented by the black dots.

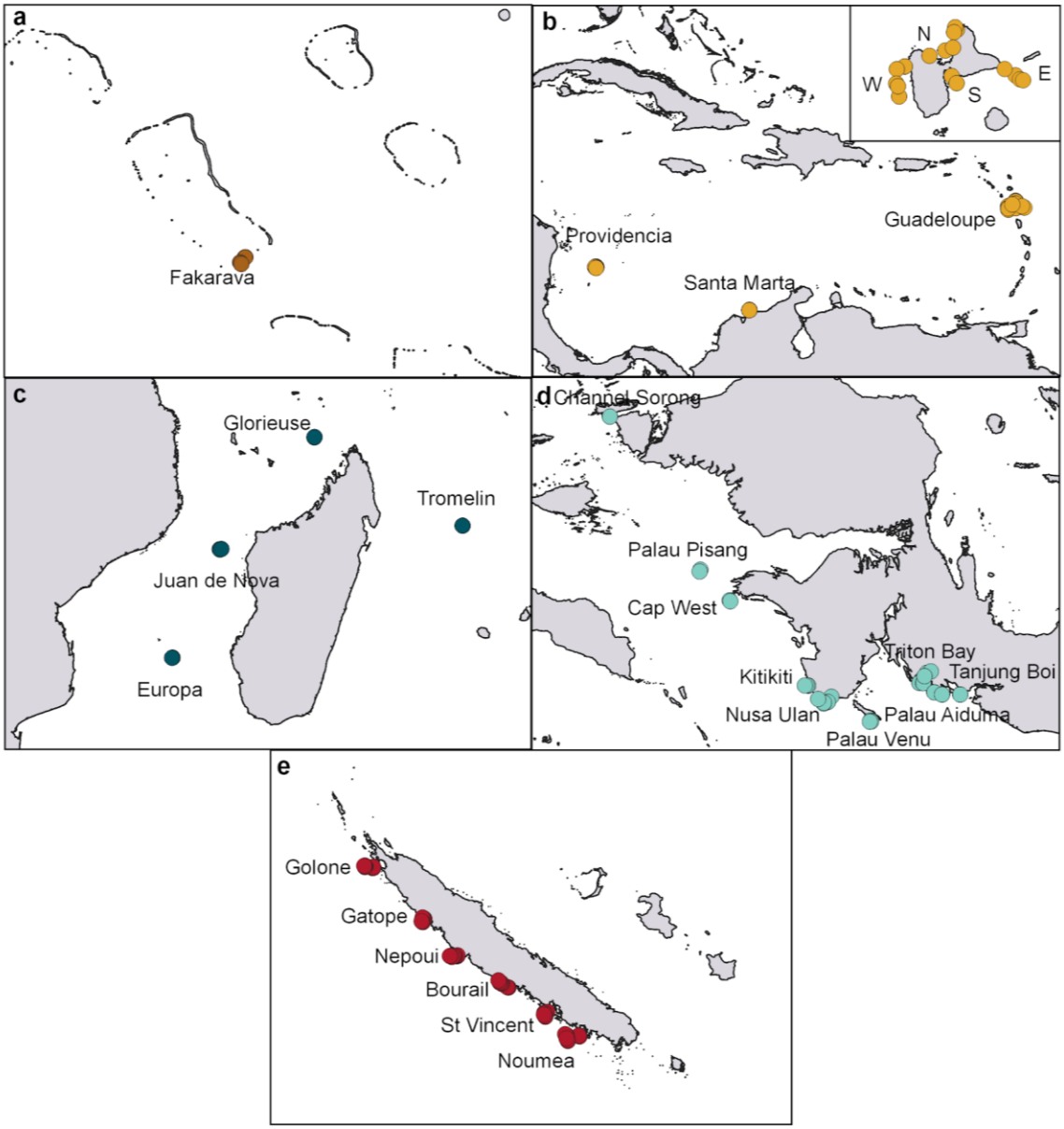


Figure S2. Sampling locations. Map of our sampling in the 5 regions including 26 sites: (*a*) 1 site in Southeast Polynesia, (*b*) 4 sites in Western Indian Ocean, (*c*) 6 sites in the Tropical Northwestern Atlantic, (*d*) 9 sites in Western Coral Triangle and (*e*) 6 sites in Tropical Southwestern Pacific. The 100 eDNA stations are overlapping at each site.

Table S1. Number of reads, MOTUs or assignment to species in the global dataset after each bioinformatic treatment: 1) without any treatment, 2) after removing sequences with less than 10 reads per sample, and sequences being identified as chimeras, 3) after removing MOTUs found in PCR blanks, 4) after removing MOTUs that do not belong to fish taxa, 5) after removing reads outside the size limits of 30-100bp, 5) after removing MOTUs found in only one PCR in the total dataset, 6) after cleaning with LULU (ie = total MOTUs richness in our study), and 7) number of species detected. As only 16% of 12S rDNA reference barcodes from reef-associated fish species are currently available, only 382 of the 2,023 MOTUs (19%) could be assigned to particular species. Of the remaining MOTUs, 1446 (71%) could be assigned to a particular family, representing 126 families in total.

Step	Reads	MOTUs	Species
Before	238,322,711	77,065	474
Tenreads	238,120,827	5,595	449
Blanks	238,101,674	5,212	449
Fishonly	199,261,204	3,900	442
Readlength	199,258,587	3,891	442
PCR_all	189,436,754	2,375	382
LULU	189,350,273	2,023	382
LULU_family	157,425,418	1,446	382

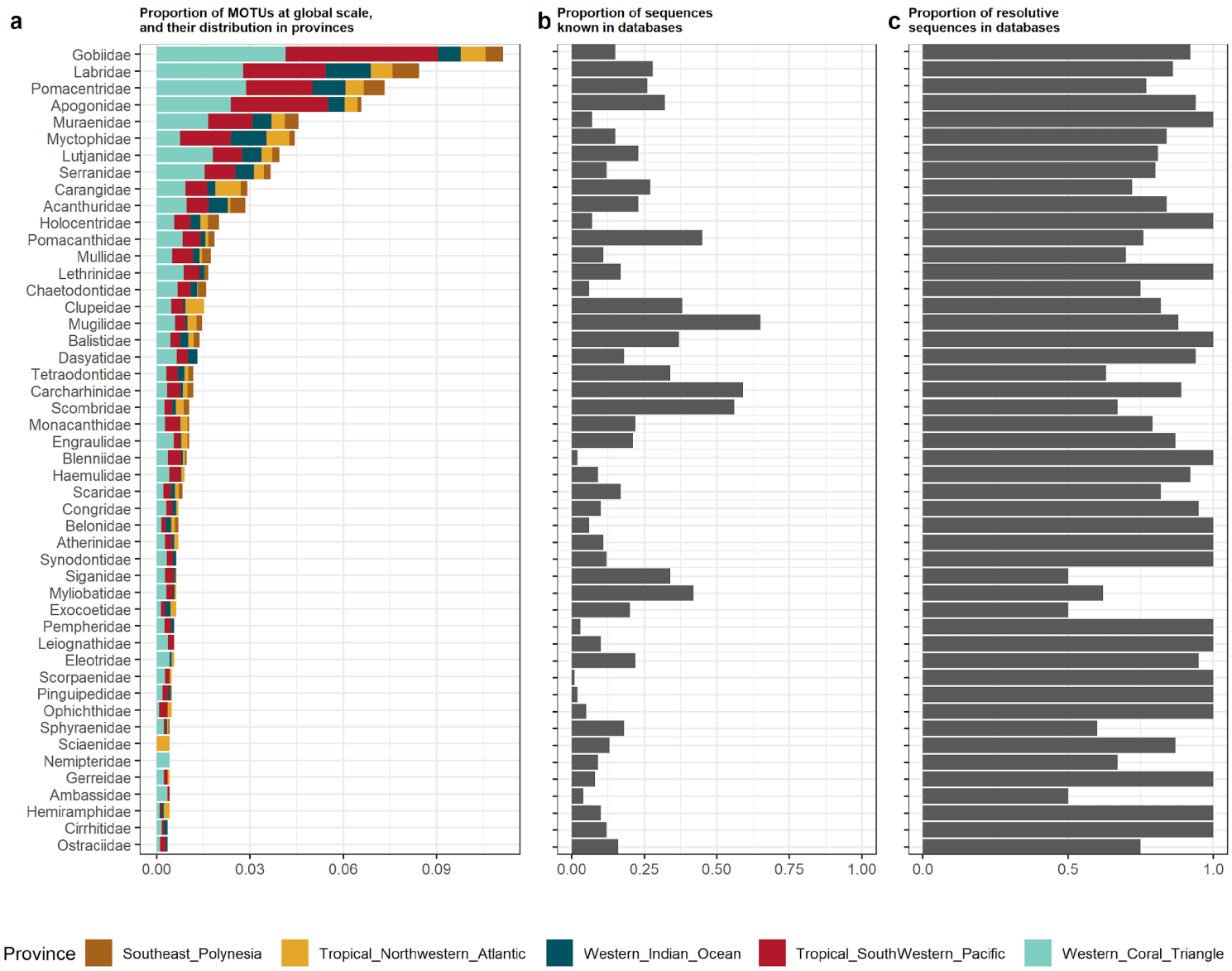


Figure S3. Characteristics of the 48 families identified in our study, with more than 5 MOTUs. (*a*) Proportion of MOTUs assigned to each family at global scale, and proportion in each region (*b*) proportion of 12S sequences in databases for each family. (*c*) Proportion of resolutive sequences in the reference database of our barcode for each family (= distinguishable species). All families with less than 100% of resolution (45% of all families), might result in an underestimated MOTU richness, due to a perfect genetic match of the 12S rRNA teleo marker between some species within these families.

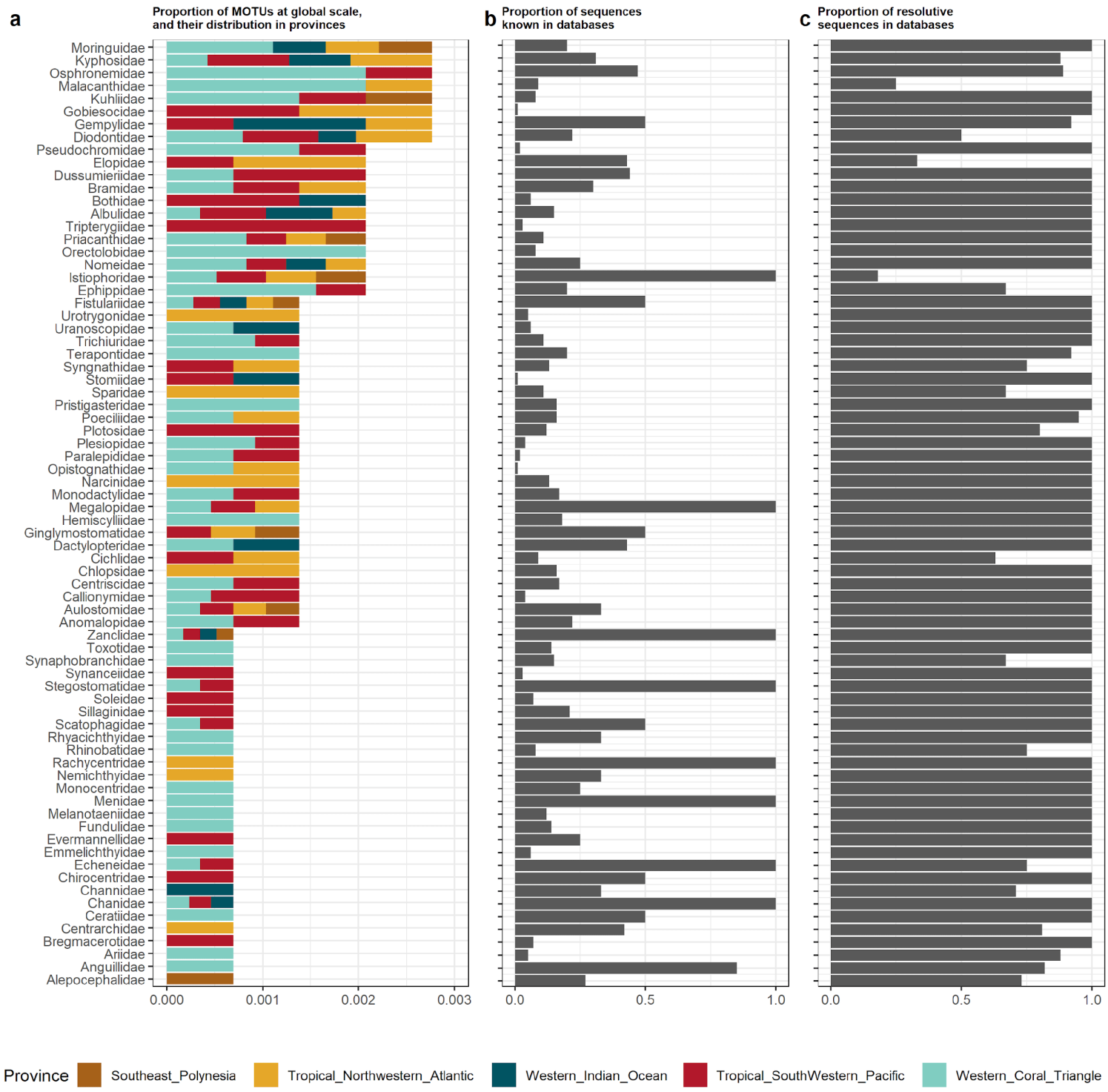


Figure S4. Characteristics of the 78 families identified in our study, with less than 5 MOTUs. (*a*) Proportion of MOTUs assigned to each family at global scale, and proportion in each region. (*b*) proportion of 12S sequences in databases for each family. (*c*) Proportion of resolutive sequences in the reference database of our barcode for each family (= distinguishable species). All families with less than 100% of resolution (45% of all families), might result in an underestimated MOTU richness, due to a perfect genetic match of the 12S rRNA teleo marker between some species within these families.

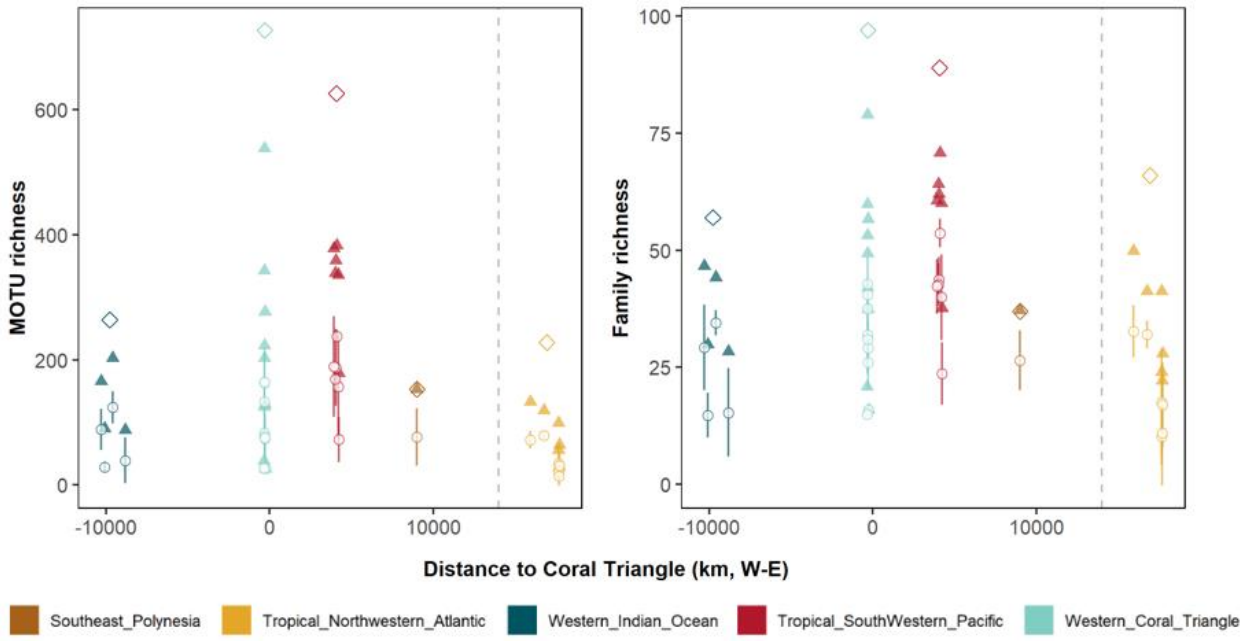


Figure S5. MOTUs and Family richness according to the distance to the coral triangle. Mean MOTUs (left) and mean Family (right) richness per station in each site ± standard deviation (empty circles and vertical bars), total site richness (filled triangles) and total region richness (empty diamonds) as a function of the distance from the center of the coral triangle (in km); the vertical dashed line represents the delimitation between the Indo-Pacific and the Atlantic Ocean basins. Kruskal-Wallis test showed significant differences in site MOTU richness between regions (Dunn post-hoc test showed Western Coral Triangle and Tropical Southwestern Pacific richest than the three other regions). (Kruskal Wallis test among sites: $p < 0.001$, n=26, Dunn test of pairwise comparisons: $p < 0.001$)

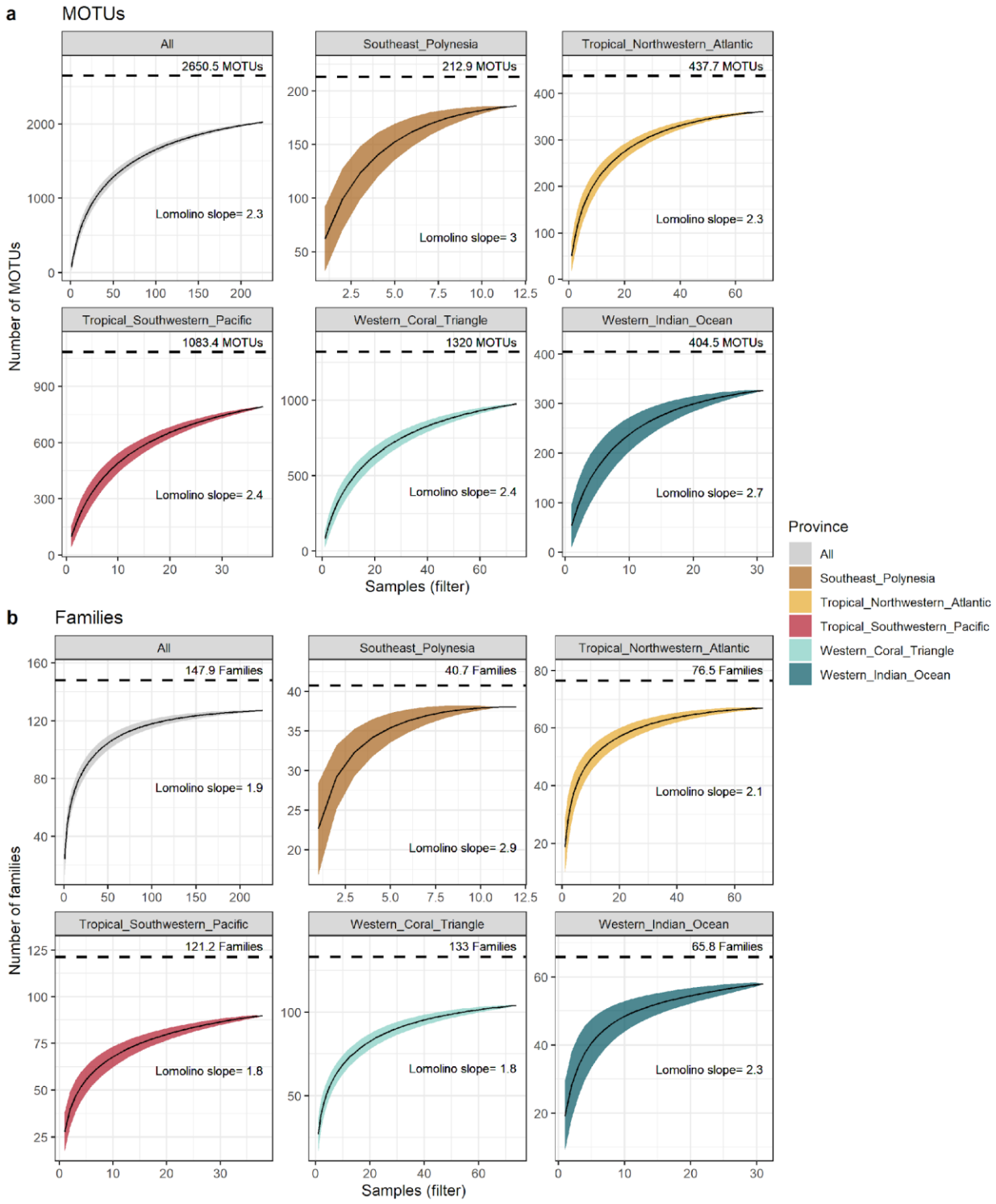


Figure S6. Accumulation curves per region. (*a*) of MOTUs and (*b*) of families in each region, according to the number of samples. Accumulation model is fitted with a nonlinear lomolino model (see methods).

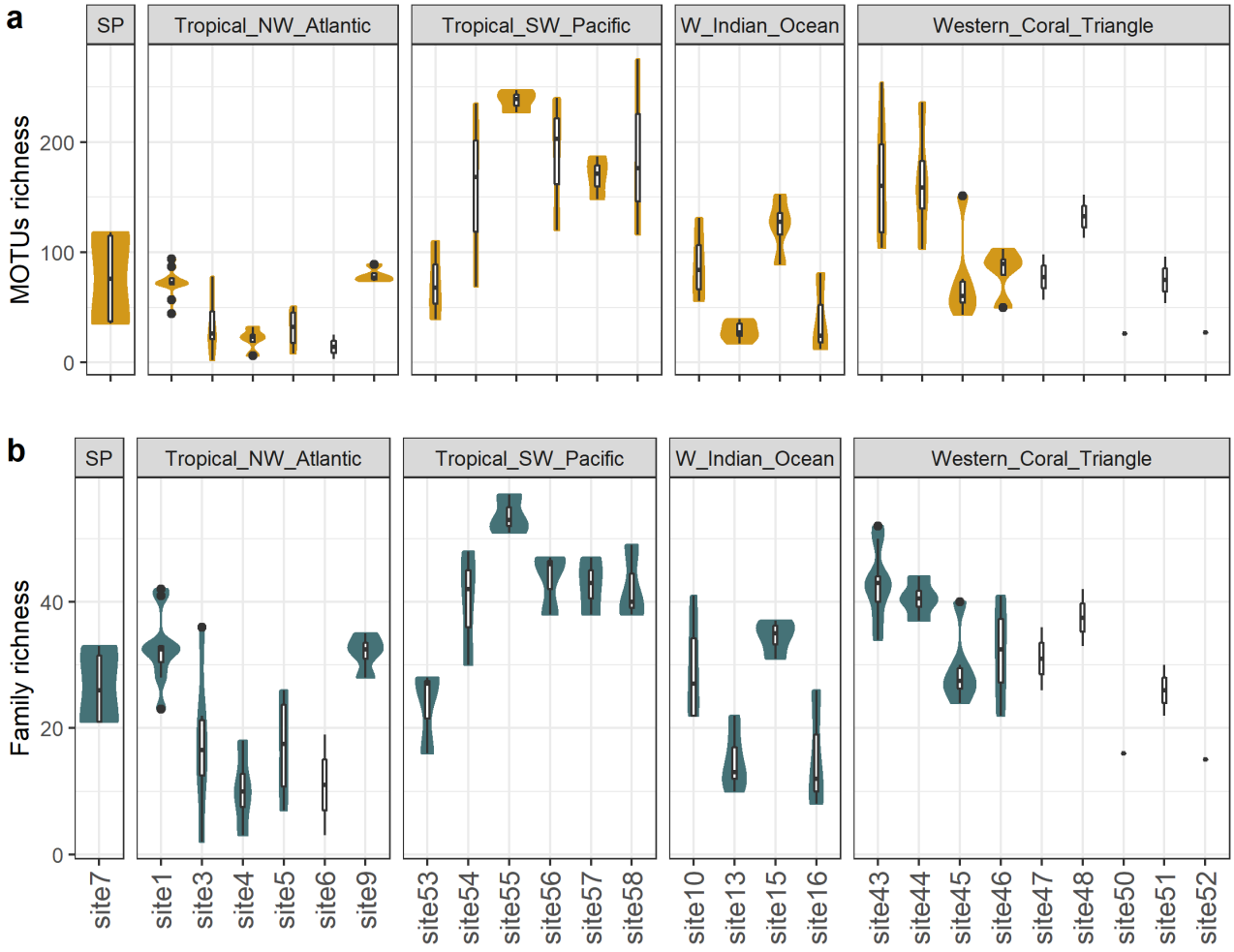


Figure S7. Richness per site in each region. (*a*) MOTU richness, (*b*) family richness. Boxplots represent median and quartiles of richness per station. Violin plots represent the density of probabilities of richness values among stations.

Table S2. Number of reads, MOTUs or assignment to species in each region after each bioinformatic treatment: 1) without any treatment, 2) after removing sequences with less than 10 reads per sample, and sequences being identified as chimeras, 3) after removing MOTUs found in PCR blanks, 4) after removing MOTUs that do not belong to fish taxa, 5) after removing reads outside the size limits of 30-100bp, 5) after removing MOTUs found in only one PCR in the total dataset, 6) after cleaning with LULU, and 7) number of species detected

Region	Step	Reads	MOTUs	Species
Southeast_Polynesia	before	7,450,199	2,308	86
	tenreads	7,445,477	370	75
	PCR_blanks_chimeras	7,445,443	367	75
	fishonly	7,132,341	306	74
	readlength	7,132,341	306	74
	PCR_all	6,806,780	195	61
	LULU	6,801,947	186	61
	LULU_family	6,174,191	153	61
Tropical_Northwestern_Atlantic	before	27,106,054	7,952	102
	tenreads	27,078,315	827	79
	PCR_blanks_chimeras	27,073,034	785	79
	fishonly	24,495,949	634	76
	readlength	24,495,509	633	76
	PCR_all	22,881,429	402	65
	LULU	22,866,586	361	65
	LULU_family	17,871,554	228	65
Tropical_Southwestern_Pacific	before	36,064,726	10,614	218
	tenreads	36,039,813	1,370	214
	PCR_blanks_chimeras	36,039,240	1,352	214
	fishonly	32,637,229	1,181	211
	readlength	32,637,157	1,179	211
	PCR_all	31,372,335	873	189
	LULU	31,368,868	843	188
	LULU_family	23,448,388	626	188
Western_Coral_Triangle	before	149,448,618	51,069	293

	tenreads	149,318,822	3,314	279
	PCR_blanks_chimeras	149,306,439	3,022	279
	fishonly	119,045,200	2,097	273
	readlength	119,043,095	2,091	273
	PCR_all	113,385,473	1,210	240
	LULU	113,323,213	1,035	237
	LULU_family	96,458,866	787	237
Western_Indian_Ocean	before	18,253,114	7,011	106
	tenreads	18,238,400	702	104
	PCR_blanks_chimeras	18,237,518	670	104
	fishonly	15,950,485	543	101
	readlength	15,950,485	543	101
	PCR_all	14,990,737	349	86
	LULU	14,989,659	327	86
	LULU_family	13,472,419	264	86

Table S3. Summary of ANOVA on Distance-based Redundancy Analysis models.

		eDNA			Visual Census		
		Df	SS	F	Df	SS	F
Total dbRDA	Region	4	1.02	4.1***	4	1.83	17.7***
	Richness	1	0.35	5.77***	1	0.16	6.28**
	Region*Richness	3	0.25	1.99	1	0.21	2**
	Residuals	17	1.05		58	1.49	
Partial dbRDA with Regions	Region	4	0.74	2.79***	4	1.74	15.7**
	Residuals	20	1.32		62	1.71	
Partial dbRDA with Richness	Richness	1	0.35	5.38***	1	0.16	5.9**
	Residuals	20	1.32		62	1.71	

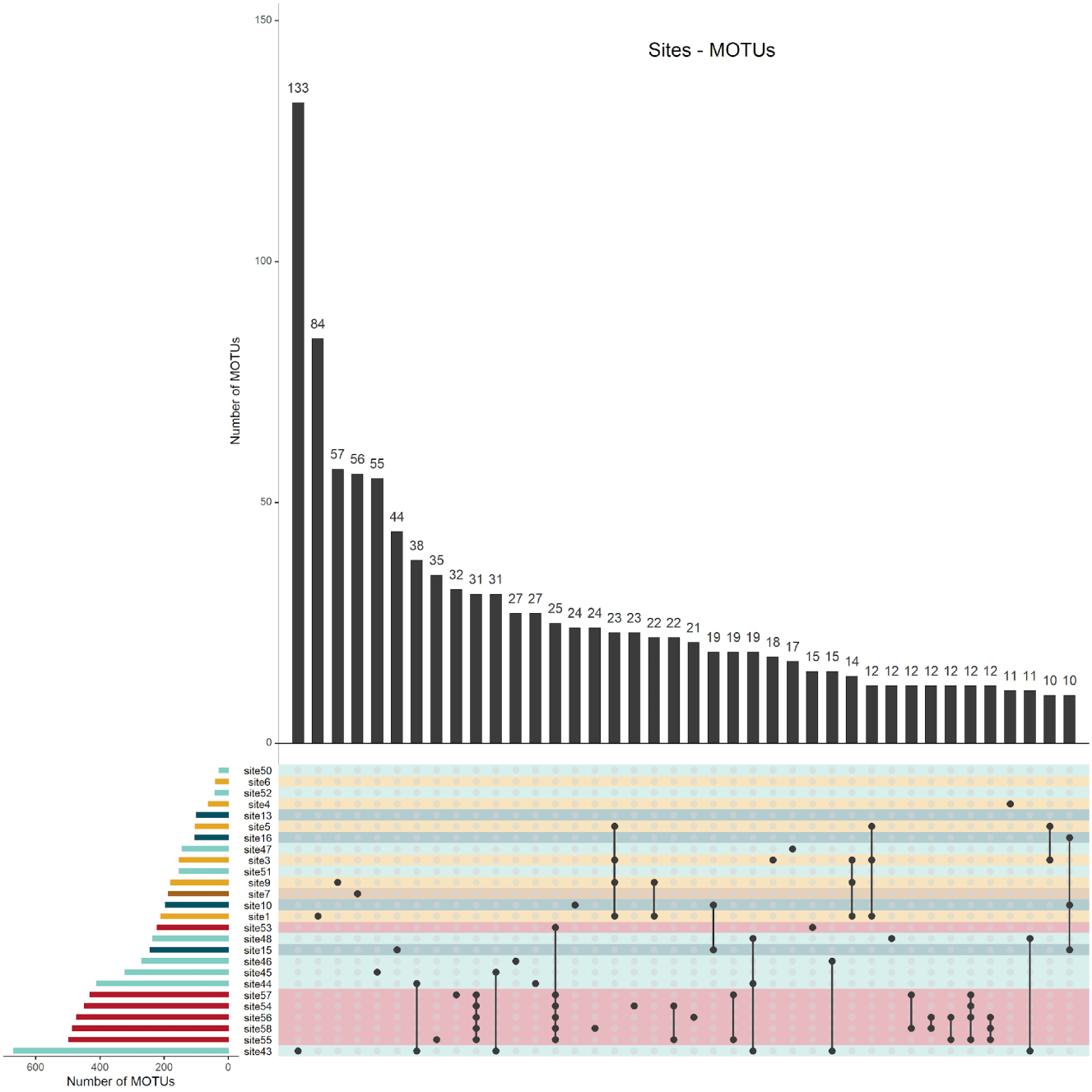


Figure S8. Distribution of MOTUs across sites. Upset plot representing the number of MOTUs found in only one site, or shared between 2 to all 25 sites. Histograms in the upper part and numbers on top indicate the number of MOTUs present in all the sites identified by the dots in the lower part. The black lines in the lower part link the sites where the MOTUs are present, for visual simplicity. Colors show regions of each site. Horizontal histograms in the lower part indicate the MOTU richness of each site.

Table S4. eDNA MOTUs and RLS species diversity partitioning across scales, with 2 different sites definition : groups of stations distinct from 10km and groups of stations distinct from 20km

Site definition	Beta scale	eDNA MOTUs	RLS species
Site = Groups of station distinct from 10km	$\beta_{inter-region}$	73.9%	84.5%
	$\underline{\beta}_{inter-site}$	16.6%	10.2%
	$\underline{\beta}_{inter-station}$	4%	2.2%
	$\underline{\alpha}_{station}$	5.5%	3.1%
Site = Groups of station distinct from 20km	$\beta_{inter-region}$	73.9%	84.6%
	$\underline{\beta}_{inter-site}$	15.6%	9.5%
	$\underline{\beta}_{inter-station}$	5.5%	2.9%
	$\underline{\alpha}_{station}$	5%	3%

Table S5. Partitioning of different subsets across spatial scales. Partitioning for all MOTUs and Visual Census species, and for MOTUs assigned to crypto-benthic families, pelagic families and to species level.

	γ_{global}	$\beta_{inter-region}$	$\underline{\beta}_{inter-site}$	$\underline{\beta}_{inter-station}$	$\underline{\alpha}_{station}$
All MOTUs	2023	73.7%	14.8%	5.9%	5.7%
Crypto-benthic MOTUs	404	76.7%	14.3%	4.8%	4.2%
Pelagic MOTUs	158	73%	15%	6%	6%
Demersal MOTUs	1461	73%	14.9%	6.1%	6%
eDNA species	396	67.4%	15.6%	8.2%	8.8%
Visual census Species	1818	84.6%	8.9%	3.7%	2.8%

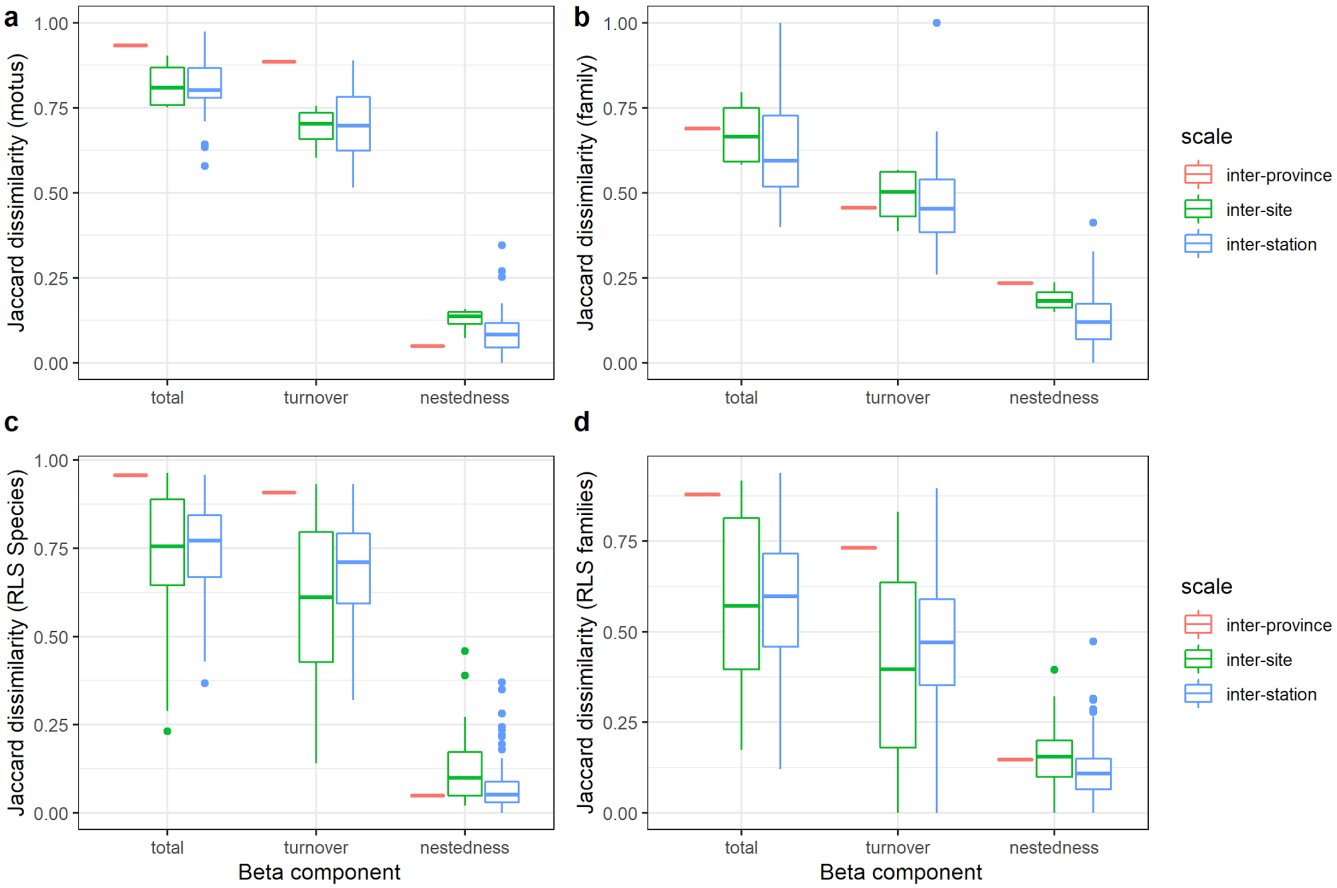


Figure S9. Beta diversity decomposition in turnover and nestedness. (*a*) for eDNA MOTUs, (*b*) for eDNA families, (*c*) for visual census species and (*d*) for visual census families. Beta diversity is measured across spatial scales: between regions, between sites within regions, and between stations/transects within sites. Boxplots represent the median, the 1st and 3rd quartiles, and 1st and 9th deciles.

Table S6. Fit of the three global abundance distribution models on fish species observed in visual census and fish MOTUs detected with eDNA. For each model, parameter values (standard deviation) are provided (intercept, slope, bending) along with the degree of freedom (df) and the Akaike's information criterion (AIC).

		Visual census (Species)				eDNA (MOTUs)			
Model	**df**	**AIC**	**Intercept**	**Slope**	**Bending**	**AIC**	**Intercept**	**Slope**	**Bending**
Log series	3	986	295 (0.01)	-1	0.001 (2.10^{-4})	246	792 (1.10^{-4})	-1	0.06 (0.003)
Pareto	3	991	267 (11)	-0.95 (0.01)	0	286	2213 (9.10^{-5})	-1.79 (0.03)	0
Pareto Bended	4	1038	173 (0.007)	-0.85 (0.01)	0.005 (4.10^{-4})	242	623 (2.10^{-4})	-0.76 (0.05)	0.08 (0.008)

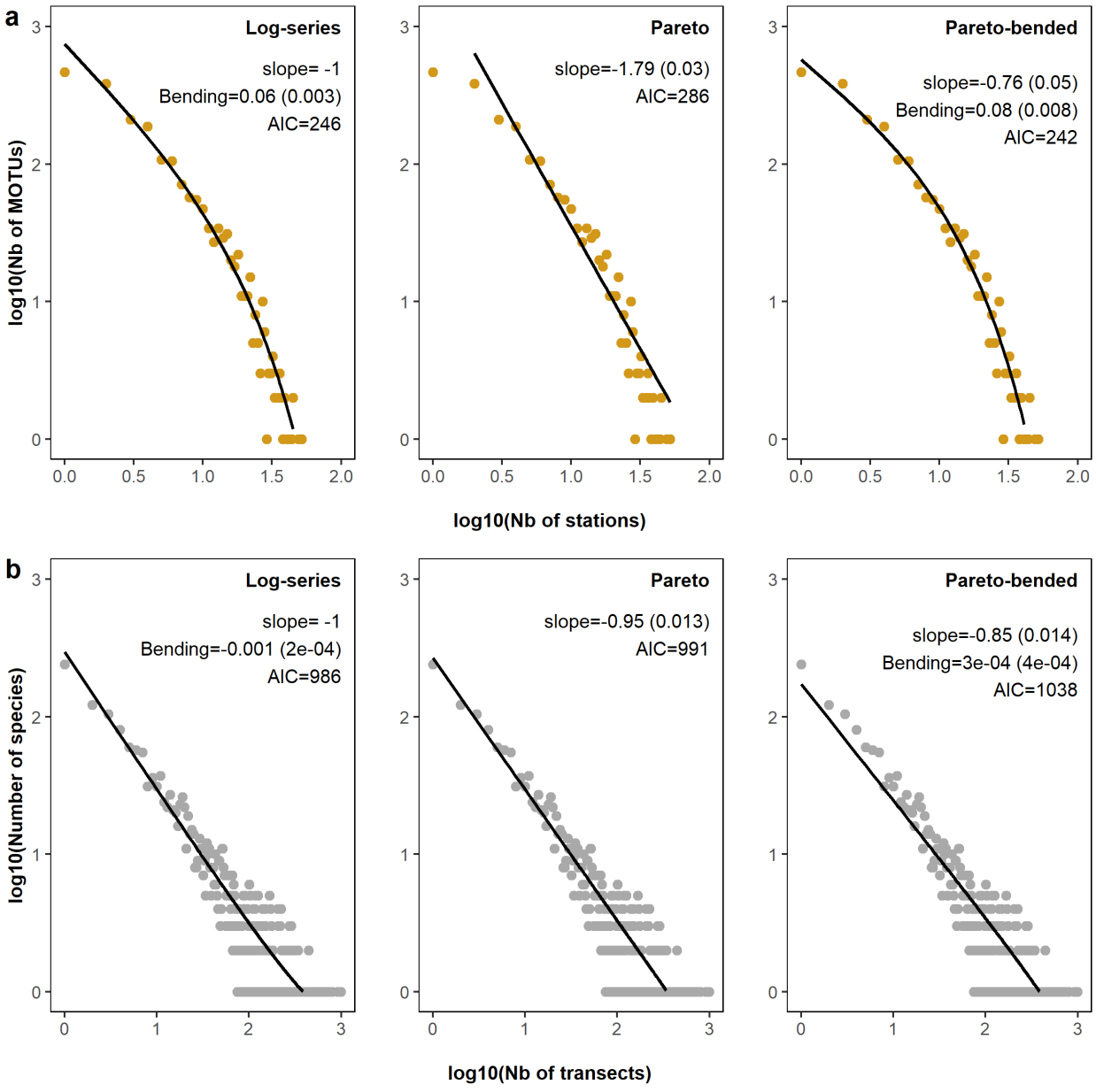


Figure S10. Distribution of the total number of global observations per fish species. (a) Distribution of MOTU occurrences across stations, log-transformed (yellow points). (b) Distribution of visual census occurrences across transects (black points), log-transformed. For both distributions, three abundance distribution models were fitted: Log-series (left), Pareto (middle) and Pareto-bended (with exponential finite adjustment) (right). Slope, confidence interval of the slope (CI) and AIC of the models are given.

Table S7. Environmental DNA sampling information in each of the five regions. Dates of sampling, number of sites per region, number of stations per region, number of filters (samples) per region, the sampling method used, the volume filtered per sample, and the total volume filtered in the region.

Region	Date	Nb sites	Nb stations	Nb filters	Method	Volume per sample	Total Volume filtered
Western Coral Triangle	17/10/17 to 23/11/17	9	32	64	DNA-free plastic bags and Sterivex filters	2L	128L
Tropical Northwestern Atlantic	26/02/18 to 03/03/18; 29/06/18 to 15/07/18; 23/10/18 to 26/10/18	6	30	67	Surface filtration along transect and VigiDNA 0.2 filters	30L	2010L
Western Indian Ocean	8/04/19 to 28/04/19	4	16	31	Surface filtration along transect and VigiDNA 0.2 filters	30L	930L
Southeast Polynesia	19/06/18 to 23/06/18	1	4	12	Surface filtration along transect & DNA-free plastic bags and VigiDNA 0.2 filters	30L	330L
Tropical Southwestern Pacific	08/10/19 to 10/12/19	6	18	52	Bottom filtration along transects and VigiDNA 0.2 filters	32L	1664L

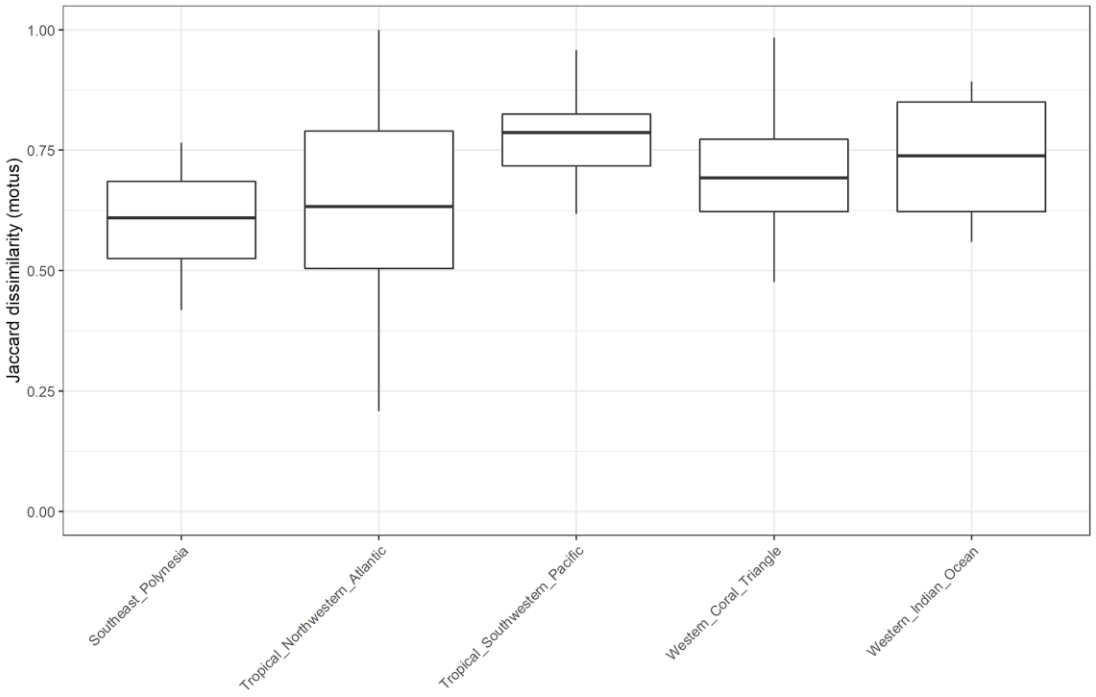


Figure S11. Beta diversity calculated between replicates of each station. The boxplots represent the median, 1st and 3rd quartiles, and 1st and 9th deciles of beta in each region.

Percentage_identity	Number_of_MOTUs	Percentage_of_MOTUs
85-87%	104	7.2
87-89%	140	9.7
89-91%	169	11.7
91-93%	109	7.5
93-95%	96	6.6
95-97%	227	15.7
97-99%	148	10.2
>99%	453	31.3

Table S8. Number and percentage of MOTUs assigned to their taxa, per class of percentage of similarity with the reference sequence.

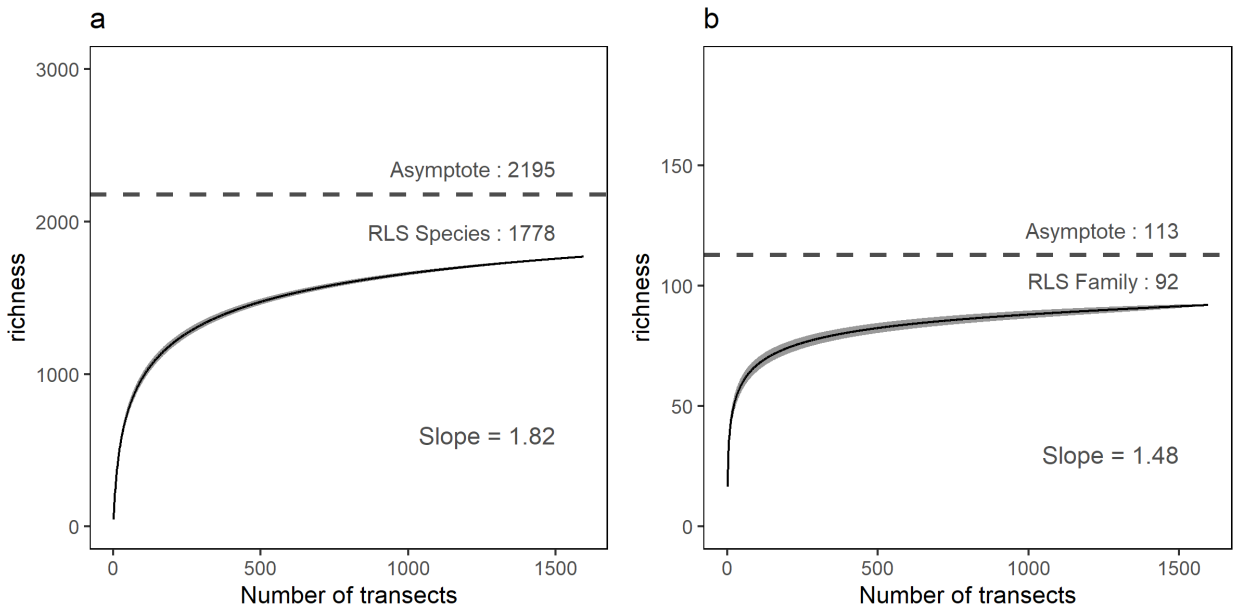


Figure S12. Accumulation curve for (*a*) species and (*b*) families in RLS transects, after a random subset of 169 transects in the 3 regions the most sampled. 169 is the number of transects sampled in the 4th most sampled region.