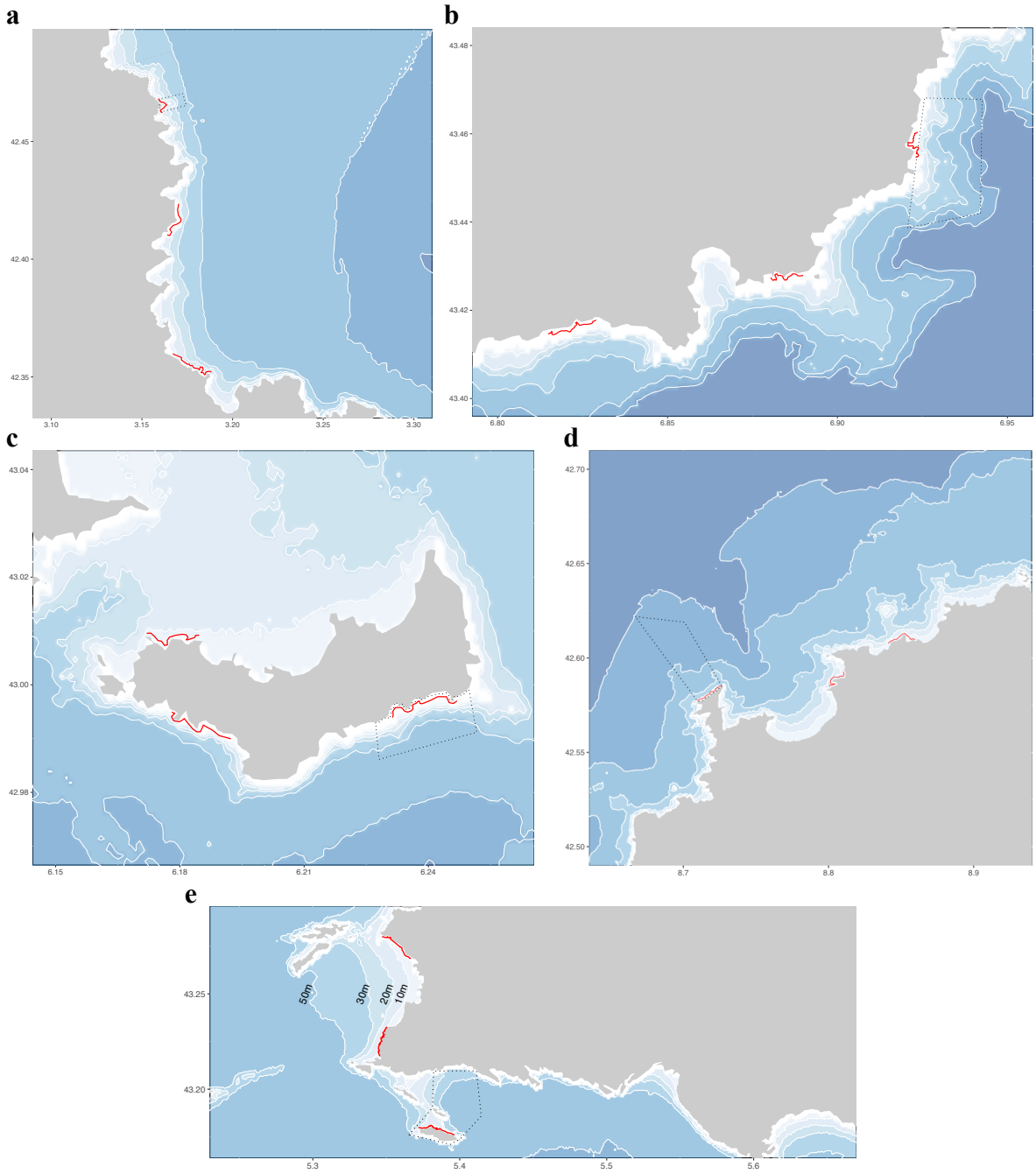
Environmental DNA metabarcoding reveals and unpacks a biodiversity conservation paradox in Mediterranean marine reserves

Emilie Boulanger, Nicolas Loiseau, Alice Valentini, Véronique Arnal, Pierre Boissery, Tony Dejean, Julie Deter, Nacim Guellati, Florian Holon, Jean-Baptiste Juhel, Philippe Lenfant, Stephanie Manel & David Mouillot

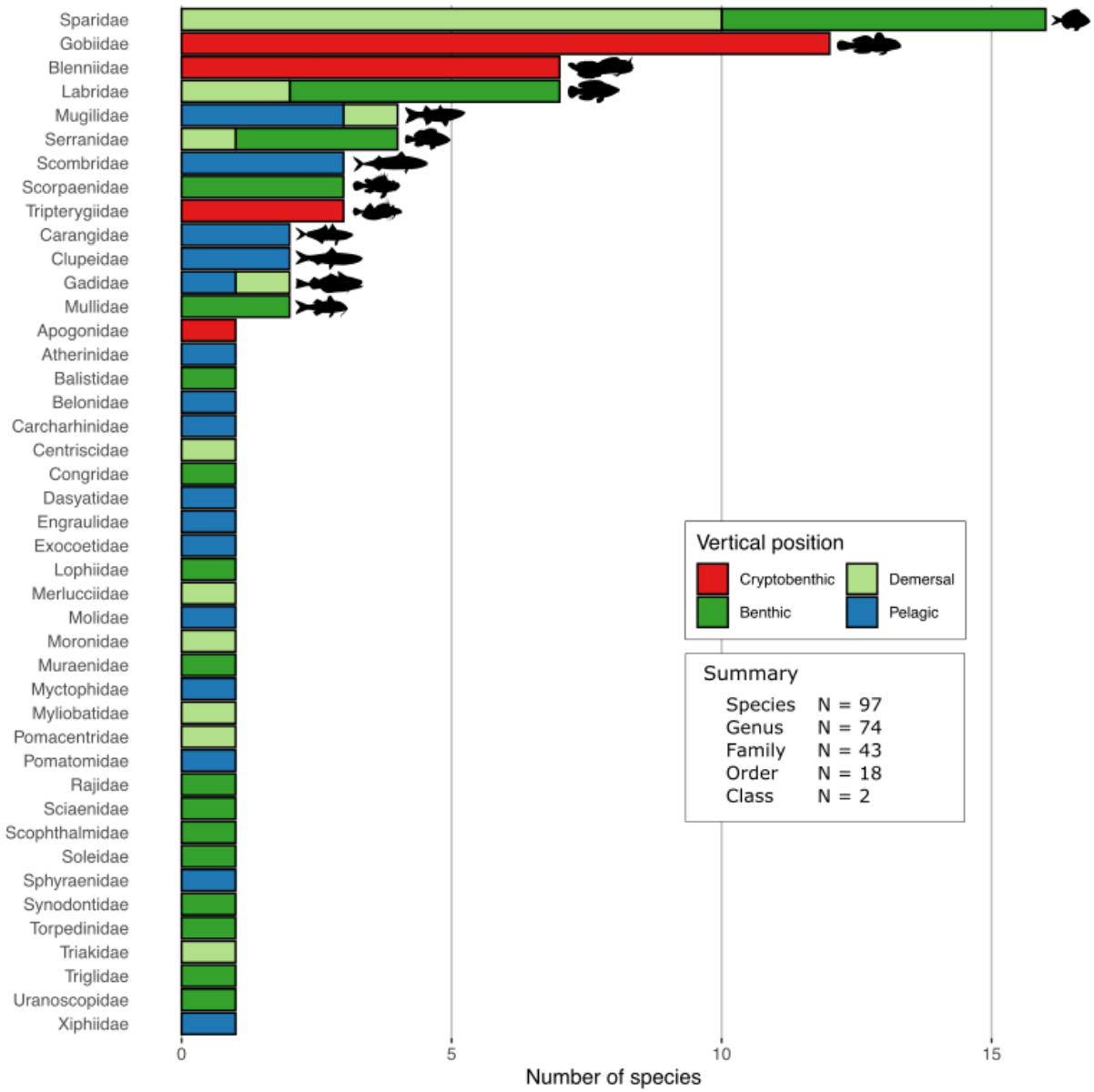
Proceedings of the Royal Society B: Biological Sciences. - DOI 10.1098/rspb.2021.0112

Table of contents

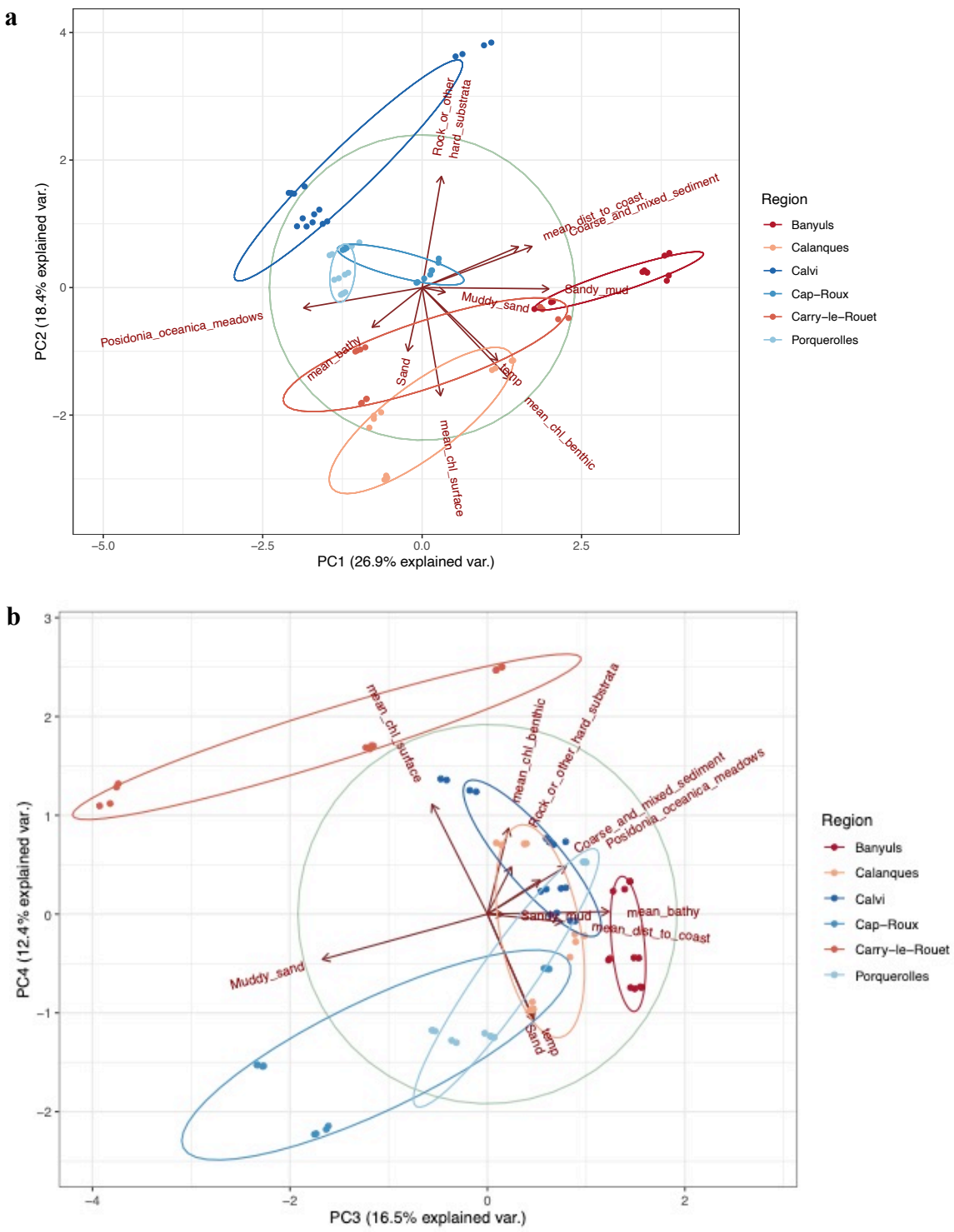
Supplementary Figures



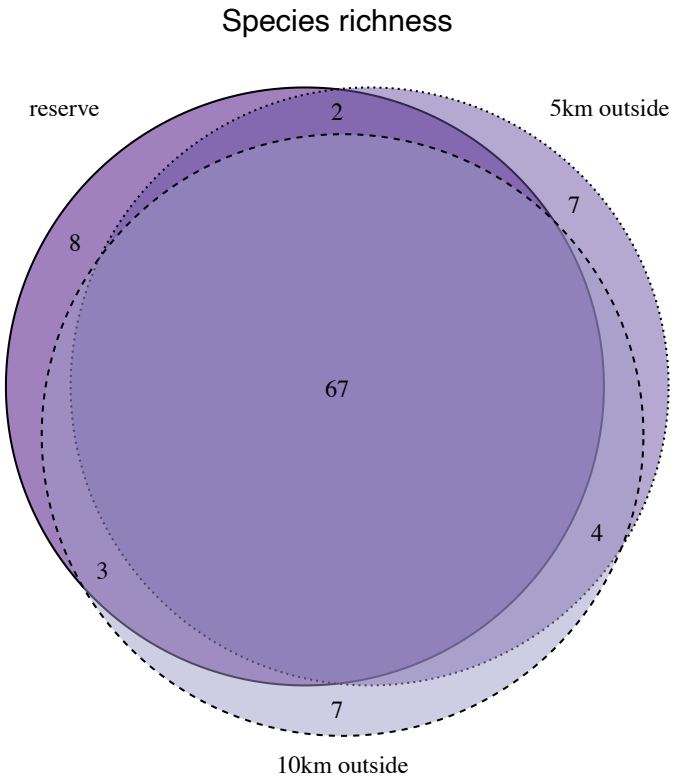
Supplementary Figure S1. Maps of the sampling transects in the studied reserves. (a) Banyuls, (b) Cap Roux, (c) Porquerolles, (d) Calvi and (e) Riou. The red lines represent the sampled transects, the dashed line represents the boundaries of the no-take reserves.



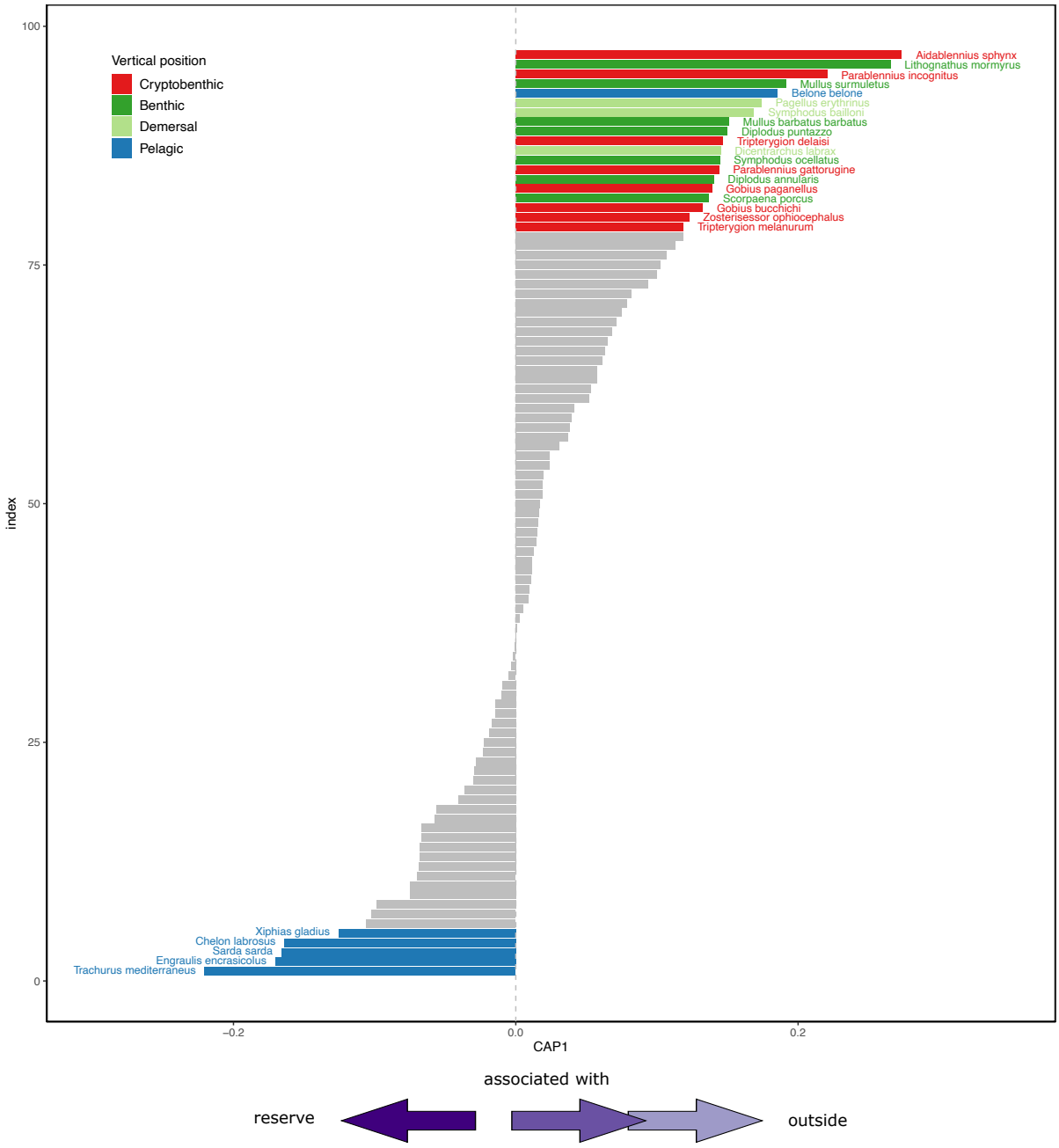
Supplementary Figure S2. Overview of fish taxa detected with the teleo marker. The bars display the number of species detected per family, coloured by species position in the water column.



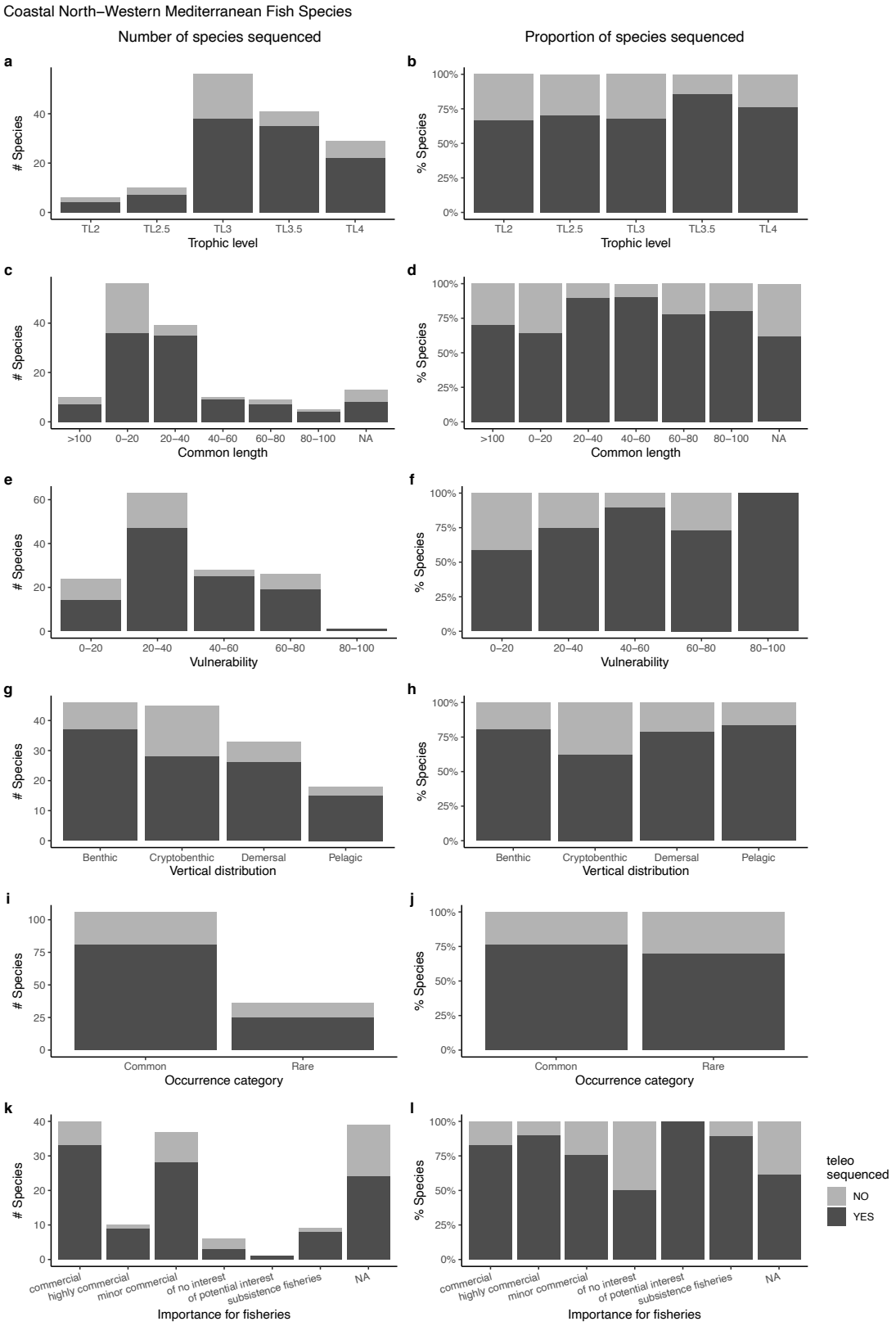
Supplementary Figure S3. Principal component analysis (PCA) on environmental variables. a) biplot of the first two PCA axis. b) biplot of the third and fourth PCA axis. Cumulatively these four axes explain 74.2% of the total environmental variance. The variables considered are, clockwise in panel (a), % rock or other hard substrate, mean distance to coast, % coarse and mixed sediment, % sandy mud, % muddy sand, temperature during sampling (temp), mean benthic chlorophyll a, % sand, mean depth (mean_bathy) and % seagrass *(Posidonia oceanica)* meadows. Variables (except for temperature and mean distance to coast) were calculated under a 500m buffer zone around each transect.



Supplementary Figure S4. Overlap of fish communities between no-take reserves, fished areas at 5km outside the reserve and fished areas at 10km outside the reserve. Numbers indicate the absolute number of species in each segment of the Venn diagram. See Supplementary Table 7.



Supplementary Figure S5. Species scores of the partial dbRDA along axis CAP1. The barplot represents the species with the greatest absolute score on the partial dbRDA axis CAP1, thus representing the species mostly associated to the reserves (left) or the fished areas outside (right). These are the same species scores as MS Fig. 4b but without the axis CAP2. Blue species are pelagic, red species are cryptobenthic, dark green are benthic, and light green are demersal.



Supplementary Figure S6. Representativeness of the genetic reference database for the regional species pool. We assessed the coverage by trophic level (a,b), common length (c,d), vulnerability to fishing (e,f), vertical distribution (g,h), rarity (i,j; Occurrence category) and commercial importance (k,l). We compared the number of species (left) as well as the proportion of species (right) sequenced (dark grey) versus not-sequenced (light grey) for our targeted 12S rRNA fragment (teleo). Fisher's exact tests showed that there was no bias between categories of each trait for the 25% missing species (p-value = 0.318, 0.0718, 0.117, 0.166, 0.385 and 0.160 respectively).

Supplementary Tables

Supplementary Table S1: Description of the six marine reserves included in our study: the Cerbère-Banyuls Natural Marine Reserve, the Carry reserve of the Côte Bleue Marine Park, the Riou island in the Calanques National Park, the Porquerolles no-take zone in the Port-Cros National Park, the Cap Roux Marine Protected Area (MPA) and the Calvi MPA in Corsica.

Marine reserve	Year of establishment	Size of the no-take zone (ha)	Management
(1) Cerbère-Banyuls	1974	65	Natural Marine reserve of Cerbère-Banyuls
(2) Carry-le-Rouet	1982	85	Côte Bleue Marine Park
(3) Riou	2012	1007	Calanques National Park
(4) Porquerolles	2007	152	Port-Cros National Park
(5) Cap Roux	2003	445	Cantonnement de pêche (/ established and management by fishermen)
(6) Calvi	1978	1074	Cantonnement de pêche (/ established and management by fishermen)

Supplementary Table S2. Estimated average marginal effect (AME) of the protection levels and habitat covariates for each richness model. The reserve level is taken as the intercept.

	Total richness					Cryptobenthic richness					Pelagic richness				
	AME	SE	p-value	2.5% CI	97.5% CI	AME	SE	p-value	2.5% CI	97.5% CI	AME	SE	p-value	2.5% CI	97.5% CI
PC1	-0.46	0.757	0.543	-1.943	1.023	0.233	0.258	0.367	-0.273	0.738	-0.109	0.171	0.523	-0.443	0.225
PC2	-0.959	0.774	0.215	-2.476	0.557	-0.016	0.264	0.951	-0.533	0.5	-0.65	0.174	0	-0.992	-0.308
PC3	-3.701	1.563	0.018	-6.764	-0.637	-1.082	0.532	0.042	-2.125	-0.038	-1.511	0.352	0	-2.201	-0.82
PC4	-0.335	1.523	0.826	-3.321	2.65	-0.006	0.519	0.991	-1.023	1.011	-0.853	0.343	0.013	-1.526	-0.18
Outside10	12.699	4.12	0.002	4.623	20.775	6.178	1.404	0	3.427	8.929	0.979	0.929	0.292	-0.841	2.799
Outside5	5.193	2.779	0.062	-0.253	10.64	3.002	0.947	0.002	1.147	4.858	-1.168	0.626	0.062	-2.395	0.06
	Vulnerable richness					Rare richness					Common richness				
	AME	SE	p-value	2.5% CI	97.5% CI	AME	SE	p-value	2.5% CI	97.5% CI	AME	SE	p-value	2.5% CI	97.5% CI
PC1	-0.15	0.119	0.209	-0.384	0.084	-0.623	0.418	0.136	-1.443	0.197	0.164	0.414	0.693	-0.648	0.976
PC2	0.021	0.105	0.84	-0.184	0.226	-0.939	0.428	0.028	-1.777	-0.1	-0.021	0.424	0.961	-0.851	0.809
PC3	-0.057	0.221	0.797	-0.49	0.376	-1.135	0.864	0.189	-2.829	0.559	-2.565	0.856	0.003	-4.243	-0.888
PC4	0.203	0.228	0.373	-0.244	0.65	0.446	0.842	0.596	-1.205	2.097	-0.782	0.834	0.349	-2.416	0.853
Outside10	0.214	1.039	0.837	-1.823	2.251	5.75	2.279	0.012	1.284	10.216	6.949	2.256	0.002	2.527	11.371
Outside5	0.836	1.491	0.575	-2.086	3.757	4.434	1.537	0.004	1.422	7.446	0.759	1.522	0.618	-2.223	3.741

Supplementary Table S3. Conditional predicted richness under three protection levels for each species category, predicted from their respective linear models in Supplementary Table 2. SE = standard error of fitted value. Data used for MS Fig. 6.

	Total richness				Cryptobenthic richness				Pelagic richness			
	yvals	SE	2.5% CI	97.5% CI	yvals	SE	2.5% CI	97.5% CI	yvals	SE	2.5% CI	97.5% CI
reserve	28.009	1.799	24.482	31.535	4.553	0.613	3.352	5.754	5.489	0.406	4.695	6.284
outside5	33.202	2.118	29.051	37.353	7.555	0.721	6.141	8.969	4.322	0.477	3.386	5.257
outside10	40.708	3.707	33.443	47.973	10.731	1.263	8.256	13.206	6.469	0.835	4.831	8.106
	Vulnerable Richness				Rare richness				Common richness			
	yvals	SE	2.5% CI	97.5% CI	yvals	SE	2.5% CI	97.5% CI	yvals	SE	2.5% CI	97.5% CI
reserve	1.039	0.231	0.586	1.492	8.39	0.995	6.44	10.34	19.618	0.985	17.687	21.549
outside5	1.106	0.288	0.541	1.67	12.825	1.171	10.529	15.12	20.377	1.16	18.105	22.65
outside10	1.215	0.554	0.13	2.301	14.14	2.05	10.123	18.158	26.568	2.03	22.59	30.545

Supplementary Table S4. dbRDA results (species ~ protection + PC1 + PC2 + PC3 + PC4). Permutation test with 9999 permutations.

Fixed variables	P model	R^2adj	Df	SumOfSqs	F marginal	P marginal
	0.001	0.129				
Protection			2	0.5935	1.7965	0.0009
PC1			1	0.5901	3.5723	0.0001
PC2			1	0.7206	4.3624	0.0001
PC3			1	0.3276	1.9831	0.0040
PC4			1	0.2943	1.7819	0.0105
Residual			65	10.7366		

Supplementary Table S5. dbRDA results of beta diversity partitions (turnover and nestedness) in function of the protection and habitat PC axes. Permutation test with 9999 permutations.

	Trunover						-	Nestedness					
Fixed variables	P model	R^2adj	Df	SumOfSqs	F marginal	P marginal		P model	R^2adj	Df	SumOfSqs	F marginal	P marginal
	0	0.252						0.286	0.153				
Protection			2	0.457	1.619	0.006				2	0.076	0.6425	0.812
PC1			1	0.616	4.366	0				1	0.052	0.8778	0.459
PC2			1	0.714	5.059	0				1	0.079	1.3226	0.225
PC3			1	0.266	1.889	0.009				1	0.111	1.8730	0.099
PC4			1	0.237	1.681	0.026				1	0.038	0.6389	0.670
Residual			65	9.168						65	3.86		

Supplementary Table S6. Summary of the reference database completion. References were extracted from the European Nucleotide Archive (ENA) or sampled and sequenced by us (completely new + ENA resampled).

	Number of fish species in list			
	Mediterranean sea		Coastal North-Western Mediterranean	
ENA at onset	240	31%	52	37%
(ENA resampled)	(35)	(5%)	(22)	(15%)
Completely NEW	80	10%	54	38%
Still missing	454	59%	36	25%
Total	774	100%	142	100%

Supplementary Table S7. Species corresponding to non-overlapping areas of the Venn diagram (Supplementary Fig. S4). Behaviour and Biology information come from FishBase.

a) Unique to reserves

Species	IUCN	Vertical	Behaviour	Biology / Diet	# sites
Atherina boyeri	LC	Pelagic	Gregarious	Feeding on […] fish larvae	1
Synodus saurus	LC	Benthic	NA	Feeds mainly on other fishes	2
Blennius ocellaris	LC	cryptobenthic	Solitary	Feeds on small invertebrates	1
Pseudaphya ferreri	LC	cryptobenthic	Gregarious	na	1
Sphyraena sphyraena	LC	Pelagic	Gregarious	Feeds mostly on fish	1
Xiphias gladius	LC	Pelagic	Gregarious	Feeds mostly on [pelagic] fishes	2
Chelidonichthys lucerna	LC	Benthic	NA	Feeds on fish	2
Pteroplatytrygon violacea	LC	Pelagic	Gregarious	Feeds on fish	1

b) Unique to fished areas (5km &10km outside)

Species	IUCN	Vertical	Behaviour	Biology / Diet	# sites
Micromesistius poutassou	LC	Pelagic	Gregarious	large individuals prey on […] small fish	1
Trisopterus capelanus	LC	NA	NA	Feeds on small fish	1
Lophius budegassa	LC	Benthic	Solitary	Feeds on nekton. mainly fishes	1
Corcyrogobius liechtensteini	LC	cryptobenthic	Solitary	na	2
Gammogobius steinitzi	LC	cryptobenthic	Solitary	na	1
Millerigobius macrocephalus	LC	cryptobenthic	Solitary	na	4
Odondebuenia balearica	LC	cryptobenthic	Solitary	na	1
Zosterisessor ophiocephalus	LC	cryptobenthic	Solitary	na	2
Ctenolabrus rupestris	LC	Benthic	Solitary	Feed on bryozoans. crustaceans and gastropods	1
Euthynnus alletteratus	LC	Pelagic	Gregarious	Opportunistic predator	2
Spicara flexuosa	LC	Demersal	Gregarious	Feed on zooplankton	1
Scophthalmus rhombus	LC	Benthic	Solitary	feeds on bottom-living fishes	1
Pegusa nasuta	DD	Benthic	Solitary	na	2
Macroramphosus scolopax	LC	Demersal	Gregarious	Feeds on invertebrates	1
Balistes capriscus	VU	Benthic	Both	Feeds on benthic invertebrates	1
Mustelus mustelus	VU	Demersal	Gregarious	Feeds also on bony fishes	1
Myliobatis aquila	VU	Demersal	Gregarious	Feeds on (benthic) fish	1
Raja asterias	NT	Benthic	NA	Feed on all kinds of benthic animals	1

Supplementary Methods

eDNA sample collection

Seawater samples were collected using a sampling protocol optimised for the monitoring of coastal marine communities using eDNA, based on four replicates per site. Sampling was conducted from May to June 2018 and in July 2019. For each replicate. 30 L of seawater was collected by continuously pumping water for 30 minutes along a 2 km transect. Transects were conducted by boat while navigating at 5 knots, and water was pumped from 1 m below the surface using sterile tubing and a peristaltic pump. Transects were conducted as close to the coastline as possible to sample close to the substrate and ensure the sampling of coastal communities. Seawater samples were filtered on site through a VigiDNA® 0.20-µM cross flow filtration capsule (SPYGEN, le Bourget du Lac, France) while changing gloves between each sample and each step to avoid (cross-) contamination. At the end of each filtration, the water inside the capsule was emptied and the capsule was filled with 80 mL of CL1 Conservation buffer (SPYGEN, le Bourget du Lac, France) and stored at room temperature until extraction.

eDNA extraction and sequencing

eDNA extraction was performed in a dedicated room for water DNA sample extraction, equipped with positive air pressure, UV treatment and frequent air renewal. Before entering this extraction room, personnel changed into full protective clothing comprising disposable body suit with hood, mask, laboratory shoes, overshoes and gloves in a connecting zone. All benches were decontaminated with 10% commercial bleach before and after each manipulation. For DNA extraction, each filtration 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 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, and two DNA extraction were carried per filtration capsule. The elution was performed by adding 100 µL of SE buffer twice. The two DNA samples were pooled before the amplification step.

After the DNA extraction the samples were tested for inhibition by qPCR[1]. If the sample was considered inhibited it was diluted 5-fold before the amplification. DNA amplifications 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 "teleo" primers[2], 4 µM human blocking primer for the "teleo" primers[2] and 0.2 µg/µL bovine serum albumin (BSA, Roche Diagnostic, Basel, Switzerland). 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. 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 and 1 min at 72 °C and a final elongation step at 72°C for 7 min in a room dedicated to amplified DNA with negative air pressure and physical separation from the DNA extraction rooms (with positive air pressure). Twelve replicate PCRs were run per sample.

After amplification, the samples were titrated using capillary electrophoresis (QIAxcel; Qiagen GmbH) and purified using the MinElute PCR purification kit (Qiagen GmbH). Before sequencing, purified DNA was titrated again using capillary electrophoresis.

The purified PCR products were pooled in equal volumes to achieve a theoretical sequencing depth of 1 000,000 reads per sample.

Reference database

Tissue samples were preserved in 96% ethanol and DNA was extracted using Qiagen's DNeasy Blood & Tissue kit. following the manufacturer's instructions with few modifications. PCR parameters were adapted from Thomsen et al. (2016)[3]. Reactions were performed with 1 - µl template DNA extract. 5-µl REDExtract-N-Amp™ PCR ReadyMix™ (Sigma Ref R4775). 2-µl ddH2O and 1-µl of each primer (at 2pM). Thermocycling parameters were: 94°C for 30 seconds 45 cycles of 94°C for 30 s. 45°C for 30 s. 72°C for 1 minute, and a final elongation step of 72°C for 5 minutes. Purification and sequencing of PCR products were carried out by Eurofins Genomics (Ebersberg. Germany). Electropherograms were checked using Codon code aligner version 4.2.7 for Windows (CodonCode Corporation. Dedham. MA. USA) and/or Geneious Prime 2019.0.4 (https://www.geneious.com).

The V05F_898 and teleo_R primer pair worked well for most of the species but produced too short or unsatisfactory sequences for 16 of them. We designed a new fish-specific primer pair targeting a 340 bp product at the end of the mitochondrial 12S rRNA gene region including the teleo eDNA metabarcode: the forward primer MF12S_F (5'-CTAGAGGAGCCTGTYVT) and the reverse primer MF12S_R (5'-GRHAAGTCGTAACATGGTA). PCR parameters were the same as above.

A total of 115 species were added to our local reference database. 80 of these are completely new, whereas 35 species were already present in public databases but sampled and sequenced to insure sequences match the local population. The final reference database used in this study contained sequences of 320 species corresponding to 41% of all Mediterranean fish species[4], but 75% of the coastal North-Western Mediterranean species[5] known to occur on our sampled habitat (table S6).

In order to assess the potential bias due to the 25% missing species in the reference database we investigated the number and proportion of species sequenced versus non-sequenced for the four traits considered in our study; vertical distribution, trophic level, vulnerability to fishing and common length, as well as for their regional rarity and commercial importance (Figure S6). The regional rarity was assessed based on the number of occurrences of each species within our

study area referenced in GBIF (https://www.gbif.org). Species were categorized using on the quartile definition of rarity of Gaston (1994)[6]: species in the lowest quartile of occurrences at the regional scale (i.e. the 25% of species with the least occurrences in our region) were categorized as rare. The remaining 75% of species were categorized as common. The commercial importance of each species was extracted from FishBase[4].

We tested whether there was any significant bias by conducting Fisher's exact tests for each trait. Fisher's exact tests showed that there was no coverage bias between categories of each trait (p-value = 0.166, 0.318, 0.117, 0.0718, 0.385 and 0.160 for vertical distribution, trophic level, vulnerability, length, regional rarity and commercial importance respectively). This absence of bias is also clear when looking at the proportion of species sequenced by category in Figure S6.

Removal of erroneous identifications and foreign species

The resulting dataset was manually inspected and corrected. One erroneous NCBI identification, *Istiophorus albicans*, was corrected as an *Engraulis* species. One taxa identified as *Clupeocephola* is a sequence intermediate to *Istiophorus albicans* and *Engraulis*, and was corrected as an *Engraulis* species. One freshwater species, *Gambusia holbrooki*, detected in a single sample was considered as a contamination and removed from the dataset. One brackish species, *Salmo salar*, is a commercially consumable fish and has previously been reported as a possible contaminant in marine eDNA studies[7] and was removed from the dataset. Five foreign species, *Cololabis saira*, *Encrasicholina punctifer*, *Schedophilus velaini*, *Dasyatis thetidis* and *Taeniurops meyeni*, were detected in multiple samples. These species do not occur in the Mediterranean but all have local relatives that could be genetically identical for the metabarcode used here but are not referenced in the local or public databases. As we cannot know which of the potential local relatives was truly present on our sites, all foreign detections were omitted from the final dataset.

Habitat and environmental data

We extracted the substrate data around each realized transect using the seabed habitat data derived from the 2016 EUSeaMap broad-scale predictive model, produced by EMODnet Seabed Habitats (http://www.emodnet-seabedhabitats.eu). Bathymetric data was downloaded from the EMODnet-bathymetry portal (http://portal.emodnet-bathymetry.eu/) and chlorophyll levels were downloaded from bio-ORACLE (http://www.bio-oracle.org).

In total, 11 environmental variables (Table S8) were calculated to be included in our models and account for the environmental variability between samples, sites and regions. Because our main focus is to test the effect of protection on species richness (GLM) and community dissimilarity (dbRDA), having too many covariables can lead to overfit models. In addition, many of the environmental variables covary amongst each other (figure S3). We thus conducted a Principal Component Analysis on the environmental variables (figure S3) which reduces the dimensionality of the dataset and yields a set of orthogonal axes. The first four axes explain most of the variance between our samples (74.2 %) and were used in the linear models of

species richness and the multivariate redundancy analyses to test and account for environmental variability.

Supplementary Table S8. List of the detailed variables calculated for each transect and included in the principal component analysis.

Category	Environmental variable
Substrate type (EUSeaMap)	*Posidonia oceanica* meadows
	Sand
	Sandy mud
	Muddy sand
	Coarse and mixed sediment
	Rock or other hard substrata
Bathymetry (EMODnet)	Mean bottom depth
Chlorophyll a (bio-ORACLE	Mean benthic chlorophyll
	Mean surface chlorophyll
Climate	Sea surface temperature
Location	Mean distance to coast

Supplementary References

1. Biggs, J. *et al.* Using eDNA to develop a national citizen science-based monitoring programme for the great crested newt (Triturus cristatus). *Biol. Conserv.* **183**, 19–28 (2015).
2. Valentini, A. *et al.* Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Mol. Ecol.* **25**, 929–942 (2016).
3. Thomsen, P. F. *et al.* Environmental DNA from seawater samples correlate with trawl catches of subarctic, deepwater fishes. *PLoS One* **11**, 1–22 (2016).
4. Froese, R. & Pauly, D. FishBase. (2019).
5. Charbonnel, E., Coudre, C. & Francour, P. *Inventaire des espèces de poissons rencontrées sur la Côte Bleue*. (2017).
6. Gaston, K. J. What is rarity? in *Rarity* 1–21 (Springer Netherlands, 1994). doi:10.1007/978-94-011-0701-3_1
7. Juhel, J.-B. *et al.* Accumulation curves of environmental DNA sequences predict coastal fish diversity in the coral triangle. *Proc. R. Soc. B Biol. Sci.* **287**, 20200248 (2020).