ECOGRAPHY

Research

Estimating the extended and hidden species diversity from environmental DNA in hyper-diverse regions

Jean-Baptiste Juhel, Virginie Marques, Rizkie Satriya Utama, Indra Bayu Vimono, Hagi Yulia Sugeha, Kadarusman Kadarusman, Christophe Cochet, Tony Dejean, Andrew Hoey, David Mouillot*, Régis Hocdé* and Laurent Pouyaud

J.-B. Juhel (https://orcid.org/0000-0003-2627-394X) ✉ (jeanbaptiste.juhel@gmail.com), D. Mouillot and R. Hocdé (https://orcid.org/0000-0002-5794-2598), MARBEC, Univ. Montpellier, CNRS, Ifremer, IRD, Montpellier, France. – V. Marques (https://orcid.org/0000-0002-5142-4191), CEFE, PSL Research Univ., EPHE, CNRS, UM, UPV, IRD, Montpellier, France. – R. S. Utama, I. B. Vimono and H. Y. Sugeha, National Research and Innovation Agency Republic of Indonesia (BRIN), Pusat Penelitian Oseanografi (P2O), Ancol Timur-Jakarta, Indonesia. – K. Kadarusman, Politeknik Kelautan dan Perikanan Sorong, KKD BP Sumberdaya Genetik, Konservasi dan Domestikasi, Papua Barat, Indonesia. – C. Cochet and L. Pouyaud, Inst. des Sciences de l'Evolution de Montpellier, Montpellier, France. – T. Dejean, SPYGEN, Le Bourget-du-Lac, France. – A. Hoey, ARC, Centre of Excellence for Coral Reef Studies, James Cook Univ., Queensland, Australia.

Ecography
2022: e06299
doi: 10.1111/ecog.06299
Subject Editor: Simon Creer
Editor-in-Chief: Miguel Araújo
Accepted 14 June 2022

Species inventories are the building blocks of our assessment of biodiversity patterns and human impact. Yet, historical inventories based on visual observations are often incomplete, impairing subsequent analyses of ecological mechanisms, extinction risk and management success. Environmental DNA (eDNA) metabarcoding is an emerging tool that can provide wider biodiversity assessments than classical visual-based surveys. However, eDNA-based inventories remain limited by sampling effort and reference database incompleteness. In this study, we propose a new framework coupling eDNA surveys and sampling-theory methods to estimate species richness in under-sampled and hyper-diverse regions where some species remain absent from the checklist or undetected by visual surveys. We applied this framework to the coastal fish diversity in the heart of the coral triangle, the richest marine biodiversity hotspot worldwide. Combining data from 279 underwater visual censuses, 92 eDNA samples and an extensive custom genetic reference database, we show that eDNA metabarcoding recorded 196 putative species not detected by underwater visual census including 37 species absent from the regional checklist. We provide an updated checklist of marine fishes in the 'Raja Ampat Bird's Head Peninsula' ecoregion with 2534 species including 1761 confirmed and 773 highly probable presences. The Chao lower-bound diversity estimator, based on the incidence of rare species, shows that the region potentially hosts an additional 123 fish species, including pelagic, cryptobenthic and vulnerable species. The extended and hidden biodiversity along with their asymptotic estimates highlight the ability of eDNA to expand regional inventories and species distributions to better guide conservation strategies.

Keywords: chao estimator, checklist, coral triangle, dark diversity, eDNA, metabarcoding

NORDIC SOCIETY OIKOS

www.ecography.org

© 2022 The Authors. Ecography published by John Wiley & Sons Ltd on behalf of Nordic Society Oikos

*Shared leading authorship.

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

16000587, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecog.06299 by Cochrane France, Wiley Online Library on [28/10/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Introduction

Species inventories are the basic data in fields of ecology, biogeography and conservation (Menegotto and Rangel 2018). Such inventories are the building blocks of biodiversity patterns assessments (Oberdorff et al. 2019), conservation strategies (McGowan et al. 2020) and human impact assessments (Ceballos et al. 2017). Yet, historical inventories are often incomplete with a non-negligible proportion of species being undetected due to insufficient sampling effort, limitations of visual surveys or particular traits (e.g. elusive behavior, small body size) (Mora et al. 2008, Brandl et al. 2018). This negative bias in species richness estimates may impair the subsequent analyses of ecological mechanisms, extinction risks and conservation outcomes (Menegotto and Rangel 2018). Given the escalating impacts of climate change and local human stressors on biodiversity (O'Hara et al. 2021), more accurate and up-to-date species inventories are urgently needed, especially in biodiversity hotspots (Barlow et al. 2018).

Monitoring marine fishes is crucial to guide management and conservation strategies. As human-related biodiversity erosion is accelerating across the oceans (O'Hara et al. 2021), implementing cost and time efficient biodiversity censuses is increasingly important. Indonesia, within the center of the coral triangle, supports the highest number of reef fishers (Teh et al. 2013) and is the second largest fish producer globally (CEA 2018) with increasing illegal and destructive practices (Varkey et al. 2010). To address this issue, Indonesia is committed to implement protection measures on 32.5 million hectares (so 10% of its EEZ) by 2030 (Indonesia Ministry of National Development Planning 2019, Campbell et al. 2020). However, the evaluation of these management measures currently considers only conspicuous fish species that contribute to small-scale fisheries (Campbell et al. 2020). Yet, many species that are often missed by visual surveys are important for ecosystem functioning like cryptobenthic and pelagic fishes that fuel reef productivity (Brandl et al. 2019, Morais and Bellwood 2019).

Environmental DNA (eDNA) metabarcoding is an emerging tool that can provide wider biodiversity assessments than classical visual surveys particularly for rare and elusive species (Boussarie et al. 2018, Garlapati et al. 2019, Polanco Fernández et al. 2021). This non-invasive method is based on retrieving DNA naturally released by organisms in their environment, amplified by polymerase chain reaction (PCR) using universal primers, sequenced and taxonomically assigned using a genetic reference database (Ruppert et al. 2019, Polanco Fernández et al. 2021). Yet, species diversity using eDNA metabarcoding is often limited by the incompleteness of reference databases to accurately assign each sequence to a given species (Marques et al. 2020a, Polanco Fernández et al. 2021). Additionally, the detectability of species using eDNA is sensitive to environmental conditions such as sea current or temperature (Harrison et al. 2019), hence some species can be missed in particular habitats or simply by chance. To overcome such sampling incompleteness, a wide range of statistical methods have been proposed to estimate the 'true' species richness or the number of undetected species across samples based on the occurrence of rare species (Chao et al. 2017). The coupling of eDNA metabarcoding and sampling-theory-based methodology presents a promising approach to estimate biodiversity in under-sampled hyper-diverse regions but has never been tested.

Here we developed a quantitative framework to estimate the extended and hidden species diversity from eDNA samples, enhanced by an augmented reference database, in a region where visual surveys have been carried out and where an extensive species checklist has been assembled. We applied this framework to marine fishes of the Bird's Head Peninsula (West Papua, Indonesia) located in the center of the coral triangle, known as the richest marine biodiversity hotspot (Allen and Erdmann 2012, Mangubhai et al. 2012).

Methods

Study area

Indonesia is the world's largest archipelagic state. It hosts a large diversity of marine ecosystems such as estuarine beaches, mangroves, coral reefs, seagrass beds and algal beds (Mangubhai et al. 2012). The sampling area covered the southwest coast of the Bird's Head Peninsula between latitudes 00.953°S–04.337°S and longitudes 130.603°E–134.163°E, with a focus area on the Kaimana Regency coasts between the latitudes 02.991°S–04.337°S and longitudes 131.598°E–134.163°E, including Triton Bay and Lengguru seafront (Fig. 1, Hocdé et al. 2020). The Lengguru seafront consists of drowned karsts with several adjacent tropical marine ecosystems such as fringing coral reefs, small island ecosystems, large shallow inlets and bays or drowned river canyon, seagrass meadows, mangrove forests, submerged freshwater springs and wide stratified estuaries. The study area encompasses the 'SW coast of Papua', the 'Raja Ampat, Bird's Head Peninsula', the 'Cendrawasih Bay' and the eastern part of the 'Banda Sea and the Moluccas' ecoregions defined by Veron et al. (2009) (Supporting information). Even though a certain spatial mismatch between UVCs and eDNA samples can be noticed (Fig. 1), both sampling methods were performed in the same coastal reef habitats.

Updated marine fish checklist

We constructed an extensive species checklist of the 'Bird's Head Peninsula' (BHP) of West Papua Province' region based on historical fishing records and visual surveys (Kulbicki et al. 2013) including the ecoregions of the study area, extended with species occurring within and in the adjacent ecoregions with similar environments (Allen and Erdmann 2012, Froese and Pauly 2020), and with the specimen collected and observed during the 2017 survey. Species names were checked and updated using the authoritative reference and searchable on-line database Eschmeyer's Catalog of Fishes (<http://

16000587, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecog.06299 by Cochrane France, Wiley Online Library on [28/10/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

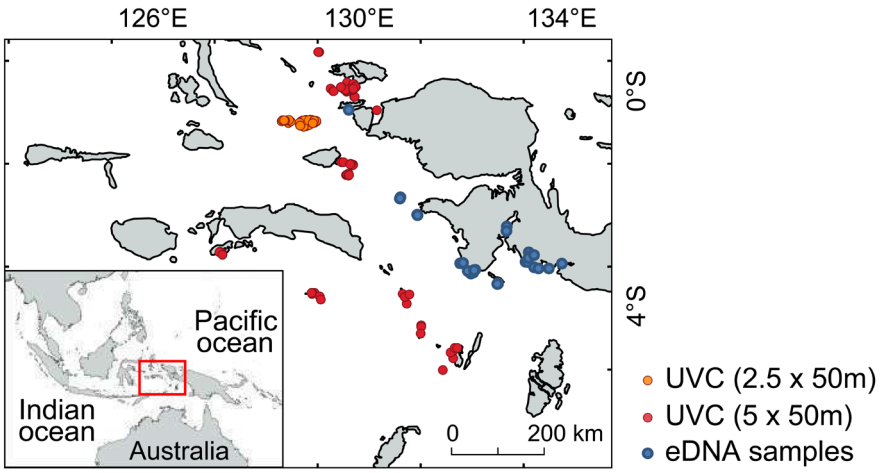

Figure 1. Map of the 'Raja Ampat Bird's Head Peninsula' region of West Papua (Indonesia) showing the location of environmental DNA (eDNA) samples and underwater visual censuses (UVC). Underwater visual censuses were retrieved from the Reef Life Survey initiative (<https://reeflifesurvey.com>) and Cinner et al. (2016).

researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp>, Fricke et al. 2020). This extensive checklist identifies 2534 marine fish species including 1761 species with confirmed occurrences belonging to 582 genera and 144 families; it also includes 736 species that are present in close regions and similar environments (Supporting information). This exceptional fish diversity is subject to a range of threats (Mangubhai et al. 2012, Campbell et al. 2020).

Underwater visual census

We retrieved data from 186 UVC transects performed during Aug–Sept 2014, Sept 2015 and Mar2018 from the Reef Life Survey initiative (<https://reeflifesurvey.com>). Additionally, we used data from 93 UVC transects performed between 2004 and 2013 in the region (Cinner et al. 2016, Fig. 1). All surveys used standardized protocols with two divers recording fish identity, abundance and size in 5 × 50 m, or 2.5 × 50 m for Cinner et al. (2016), blocks either side of the transect line. The two transect blocks include independent counts that are averaged to characterized the transect (Edgar et al. 2020).

Environmental DNA filtering and processing

We collected 92 water samples along the south coast of the BHP region of West Papua between Oct and Nov 2017 across different reef habitats (estuarine and brackish waters excluded) distributed over an area of 500 km from east to west, with a focus (80 of the 92 samples, or 87%) from the easternmost 210 km sector (Fig. 1). We collected the water samples in DNA-free plastic bags from a dinghy, during closed-circuit rebreather diving (depths between 10 and 100 m) as close as possible to the habitat or using Niskin water samplers (depths between 100 and 300 m) (Hocdé et al. 2020). Every water sampling session were performed before and never at the same time as fish collection to avoid in situ contamination. We coupled a pressure and temperature sensor to the Niskin bottle to control the sampling depth and characterize the water mass via the vertical temperature profile. For each sample, we filtered 2 l of seawater with sterile Sterivex filter capsules (Merck© Millipore; pore size 0.22 µm) and disposable sterile syringes. Immediately after, we filled the filter units with lysis conservation buffer (CL1 buffer SPYGEN©) and stored them in 50 ml screw-cap tubes at −20°C. The DNA extraction and amplification were performed following a modified protocol of Pont et al. (2018) including 12 separate PCR amplifications per sample. A teleost-specific 12S mitochondrial rDNA primer (teleo, forward primer-ACACCGCCCGTCACTCT, reverse primer-CTTCCGGTACACTTACCATG, Valentini et al. 2016) was used for the amplification of metabarcoding sequences (see Supporting information for laboratory analyses and bioinformatic analyses).

Among fish eDNA 12S primers, teleo provides a strong performance to detect fish diversity even in highly diverse ecosystems (Collins et al. 2019, Polanco Fernández et al. 2022). Although alternative fish eDNA primers might cover a larger proportion of fishes in the reference database and hence be more informative on species identification, there is currently no primers located outside the 12S with similar performance (Zhang et al. 2020).

We followed a contamination control protocol during both field and laboratory stages (Valentini et al. 2016). Water sample processing included the use of disposable gloves and single-use filtration equipment, and the bleaching (50% bleach) of Niskin bottles between samples. Staffs who performed eDNA filtration were not involved in tissue sampling of fish and used a dedicated workspace to avoid both contact and airborne contamination.

Genetic reference database completion

During the same survey along the south–western coast of the BHP in West Papua, we collected 1466 individuals from 413 species, 180 genera and 69 families of fishes along the shore.

16000587, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecog.06299 by Cochrane France, Wiley Online Library on [28/10/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The specimens were mainly collected by hand or with 4–8 m long bottom gillnets deployed by open-circuit and closed-circuit divers in the 0–100 m depth range (Hocdé et al. 2020). Some brackish and estuarine fishes were also collected with 10 m beach purse seines and pelagic fish with line fishing and spearfishing. We used morphological features and 652 bp CO1 (cytochrome oxidase 1) targeted genetic sequencing to identify the specimens. Then we amplified and sequenced the individuals on a large fraction of the 12S mitochondrial rDNA region (480 bp) with two distinct pairs of primers respectively designed for teleosts and elasmobranchs to improve sequencing results. Finally, the 12S teleo region defined in Valentini et al. (2016) was extracted from the obtained sequences to complement the EMBL genetic reference database (European Molecular Biology Laboratory, <www.ebi.ac.uk>, ver. 141, downloaded on January 2020, Baker et al. 2000) and improve taxonomic assignments (see Supporting information for the reference database and the methodological details of its completion).

To evaluate the completeness of the online database for the teleo region of the 12S mitochondrial DNA, we performed an in silico PCR on the EMBL database with ecoPCR (Ficetola et al. 2010) using the teleo primer sequences, allowing up to three mismatches. We compared the generated list of sequenced species to the extensive species checklist of the BHP ecoregion. Among the 1761 species of the Bird's Head Peninsula checklist for which presence is confirmed, only 496 species (28%) were sequenced in EMBL for the teleo region. The addition of sequences retrieved from our fish sampling increased this list to 762 sequenced species (43.4%). Additionally, 21 species absent from the historical checklist were collected, or observed and clearly identified, during the development of the genetic reference database (see Supporting information for the extensive checklist).

Taxonomic assignments

The metabarcoding workflow was based on the VSEARCH toolkit and the clustering algorithm SWARM that groups multiple sequence variants into MOTUs (molecular operational taxonomic units, Mahé et al. 2014) to clean PCR and sequencing errors. We performed taxonomic assignments using the ecotag program (lowest common ancestor algorithm) from the OBITOOLS toolkit (Boyer et al. 2016) against our custom reference database and the global public EMBL genetic database (release 141, downloaded on January 2020). For each MOTU, we chose the taxonomic assignment with the highest similarity from either the custom reference database or EMBL. We only retained the assignments with 100% similarity to either reference database so matching perfectly over the full length of the sequence (Supporting information). Some sequences could match at 100% but correspond to several species due to limited taxonomic resolution on our marker region, preventing a taxonomical assignment at the species level. For those sequences, we determined, if possible, the most probable species being detected based on the list of species corresponding to the sequence and the known spatial distribution of those species. For other sequences, it was not possible to narrow down the list of possible species if those are all known to occur in the region or in the vicinity of the region, so these sequences were tagged with a list of possible assignations (Supporting information) and removed from the analyses.

Fish traits

The extended fish diversity may be characterized by certain traits or behaviors which may limit the detection by classical (fishing or visual records) and eDNA surveys (Thalinger et al. 2021). To investigate this bias, we retrieved available data on habitat (reef or pelagic), diet, circadian activity, maximum body length, and IUCN (International Union for Conservation of Nature) conservation status for all the species detected by eDNA from Fishbase (<www.fishbase.org>) and compared them among the different sets of species.

Quantitative framework to estimate the extended and hidden species diversity

Regional species checklists catalogue all the species present in a given region based on the compilation of historical inventories and observations. Within this regional biodiversity, we define the visible diversity as the set of species that are detected by visual- or video-based sampling methods while the species only detected by eDNA represent the hidden diversity (Boussarie et al. 2018, Fig. 2). Further, we designate the set of species unique to eDNA samples but absent from the known regional diversity as the extended diversity. Species not detected by any of these sampling methods (visual-, video-, and eDNA-based surveys) but listed in the regional checklist and potentially present in the area make the dark diversity (Pärtel et al. 2011). The dark diversity thus represents the set of species that should be present in a certain region, based on their habitat requirements, their biogeographic distribution, their dispersal ability or their historical presence, yet not detected by any method (Moeslund et al. 2017).

Moreover, the detection potential of eDNA is conditional to fish behavior (Thalinger et al. 2021) and the persistence of eDNA in the environment (Harrison et al. 2019) so some species can be missed (Stat et al. 2019). In this case, asymptotic estimates can be inferred using species or MOTU accumulation curves (Juhel et al. 2020, Mathon et al. 2022). Yet, these asymptotes cannot provide the level of species dissimilarity between methods so cannot estimate the asymptotic extended or hidden species diversity. As an alternative, sampling-theory-based methodologies, based on the occurrence of rare species, can provide asymptotic estimates of both species richness within sites and species dissimilarity among sites (Chiu et al. 2014, Chao et al. 2017).

By analogy, to estimate the asymptotic visible, hidden and extended species diversity we used the lower bound estimation framework. More precisely, we estimated the asymptotic species richness sampled by each method using the bias-corrected lower bound estimator iChao2 (Chiu et al. 2014) based on incidence-raw data (Eq. 1):

16000587, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecog.06299 by Cochrane France, Wiley Online Library on [28/10/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

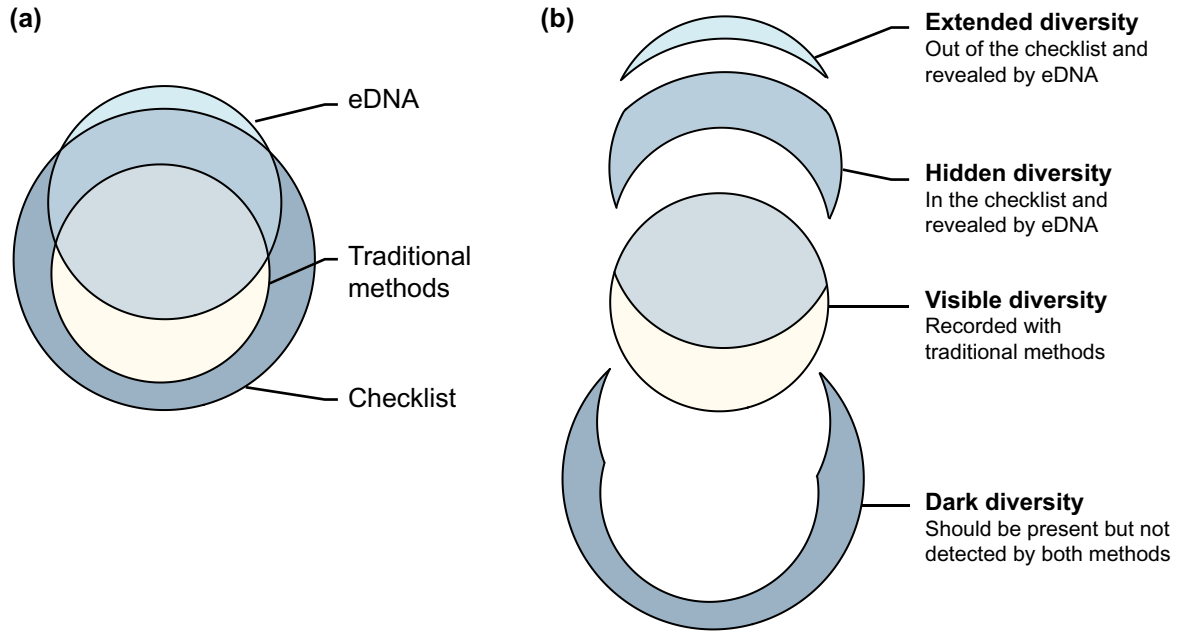

Figure 2. Conceptual diagrams illustrating the framework estimating visible, hidden, extended and dark species diversity in the context of partially known regional species checklist (a) and the description of each diversity portion (b).

$$\hat{S}_{\text{Chao2}} = \begin{cases} S_{\text{obs}} + [(N-1)/N]Q_1^2/(2Q_2), & \text{if } Q_2 > 0 \\ S_{\text{obs}} + [(N-1)/N]Q_1(Q_1-1)/2, & \text{if } Q_2 = 0 \end{cases} \quad (1)$$

where S_{obs} is the number of species observed in the N samples, Q_1 and Q_2 are the frequency counts of species found in one sample and two samples, respectively, for a given method.

Since the hidden diversity represents the number of species detected by eDNA but not by visual surveys we need an estimate of species diversity shared by the two methods. We used the bias-corrected Chao2-shared estimator (Pan et al. 2009) expressed as (Eq. 2):

$$\tilde{S}_{\text{shared}} = S_{\text{shared}} + K_{\text{UVC}}\frac{f_{+1}(f_{+1}-1)}{2(f_{+2}+1)} + K_{\text{eDNA}}\frac{f_{1+}(f_{1+}-1)}{2(f_{2+}+1)} + K_{\text{eDNA}}K_{\text{UVC}}\frac{f_{11}(f_{11}-1)}{4(f_{22}+1)} \quad (2)$$

Where $K_{\text{eDNA}} = (n_{\text{eDNA}}-1)/n_{\text{eDNA}}$

$K_{\text{UVC}} = (n_{\text{UVC}}-1)/n_{\text{UVC}}$

S_{shared} is the number of species observed by both methods across all the samples, n_{eDNA} is the number of eDNA samples, n_{UVC} is the number of visual surveys. Regarding the species recorded by each method, f_{+1} denotes the number of shared species that are recorded only once by visual surveys, f_{1+} denotes the number of shared species that are recorded only once by eDNA samples while f_{11} counts the number of shared species that are recorded only once by both methods. The same way, f_{+2} denotes the number of shared species that are recorded twice by visual surveys, f_{2+} denotes the number of shared species that are recorded twice by eDNA samples while f_{22} represents the number of shared species that are recorded twice by both methods.

We used the iChao2 index ($\hat{S}_{\text{Chao2}}$) to estimate the asymptotic diversity recorded by visual survey ($\hat{S}_{\text{visible}}$) and eDNA samples ($\hat{S}_{\text{eDNA}}$). To estimate the asymptotic hidden diversity recorded by eDNA ($\hat{S}_{\text{hidden}}$), we withdrew the estimated shared diversity ($\hat{S}_{\text{shared}}$) from the asymptotic diversity recorded by this method (S_{eDNA})(Eq. 3):

$$\hat{S}_{\text{hidden}} = \hat{S}_{e\ \text{DNA}} - \hat{S}_{\text{shared}} \quad (3)$$

We also used the iChao2 index to estimate the extended diversity ($\hat{S}_{\text{ext}}$) considering only the species out of the checklist sampled by eDNA. The estimators were computed using the SpadeR package in the R programming environment (Chao et al. 2015).

Results

Environmental DNA reveals hidden and extended fish diversity

The 279 UVC detected 725 species including 400 species previously sequenced for the teleo region, so potentially detectable with eDNA. These 400 species covered 52.6% of the sequenced checklist (Fig. 3, see Supporting information for the complete list of species).

From the 333 369 000 total initial reads of the 92 eDNA samples, 82 099 MOTUs were generated using the

16000587, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecog.06299 by Cochrane France, Wiley Online Library on [28/10/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

SWARM clustering algorithm from which 2576 MOTUs passed the bioinformatic filters and 506 MOTUs matched at 100% similarity to a given fish taxa. A total of 455 species, belonging to 232 genera and 87 families, were assigned. Among these species, 418 were referenced in the extended checklist of the region. Thus, with 92 water samples, 54.8% (417 out of 761) fish species in the checklist and sequenced for the teleo region were detected (Fig. 3) while UVCs detected 2.6% more (400) species with 4.3 times more surveys (i.e. 279 individual transects). UVCs detected 141 species that were not picked up by eDNA. Conversely, eDNA detected 159 species from the checklist that were not seen in UVCs, revealing a significant part of hidden diversity (Fig. 3). Additionally, eDNA detected 37 species that were absent from the historical checklist revealing a marked extended fish diversity in this region. Of these 37 species, 20 species were referenced in the checklist of Indonesia (Fishbase) and 12 species had not previously been recorded in Indonesia (Fig. 3, see Supporting information for the list of species).

Environmental DNA extends the regional checklist towards elusive species

Investigating fish species taxonomy and traits revealed that eDNA extended the checklist, i.e. extended diversity in the conceptual framework, mainly towards elasmobranch (+12%), pelagic (+57%), piscivorous (+3%), nocturnal (+18.4%) and vulnerable (+8%) species (Fig. 4). The maximum body length of species absent from the checklist was significantly larger (mean ± SD: 104 ± 138 cm) than those included in the checklist (50 ± 64 cm) (permutation t test, $t=-3.91$, p-value < 0.001) and species detected by UVCs (38.6 ± 42.1 cm) (permutation t test, $t=-3.6$, p-value < 0.001, Fig. 4f). Additionally, eDNA detected the smallest cryptobenthic species of the checklist, *Trimma xantochrum*, *Trimma halonevum* and *Trimma haimassum* with maximum body length ranging from 2.5 and 3.1 cm while UVCs missed them (Fig. 4f).

Asymptotic estimates of extended and hidden fish diversity

The bias-corrected iChao2 and Chao2-shared estimators provided asymptotic values of the visible, hidden and extended fish diversity. The overall fish diversity potentially revealed by both UVC and eDNA surveys was estimated at 700 species, adding 104 species to the regional pool. The visible diversity detected by UVCs ($\hat{S}_{visible}$) was estimated to increase from 400 to 502 species (67.1–71.7% of the pool) and the diversity recorded by eDNA samples was estimated to increase from 455 to 528 species so from 76.3% to 75.4% of the pool (Fig. 5a, b). The diversity recorded by both methods ($\tilde{S}_{shared}$) was estimated to increase from 255 to 330 species (43–47% of the pool) and the hidden diversity recorded by eDNA only was estimated to remain constant (196–198 species, 33–28% of the pool).

The asymptotic extended diversity ($\hat{S}_{ext}$) was estimated to be 3.3 times greater than the observed extended diversity, hence increasing from 37 to 123 missing species (Fig. 5c, d). The diversity recorded by eDNA within the checklist, thus including the hidden diversity and a part of the visible diversity, was estimated to increase from 418 to 455 species (55–60% of the checklist) leaving 306 fish species of the checklist undetected (i.e. dark diversity in the conceptual framework).

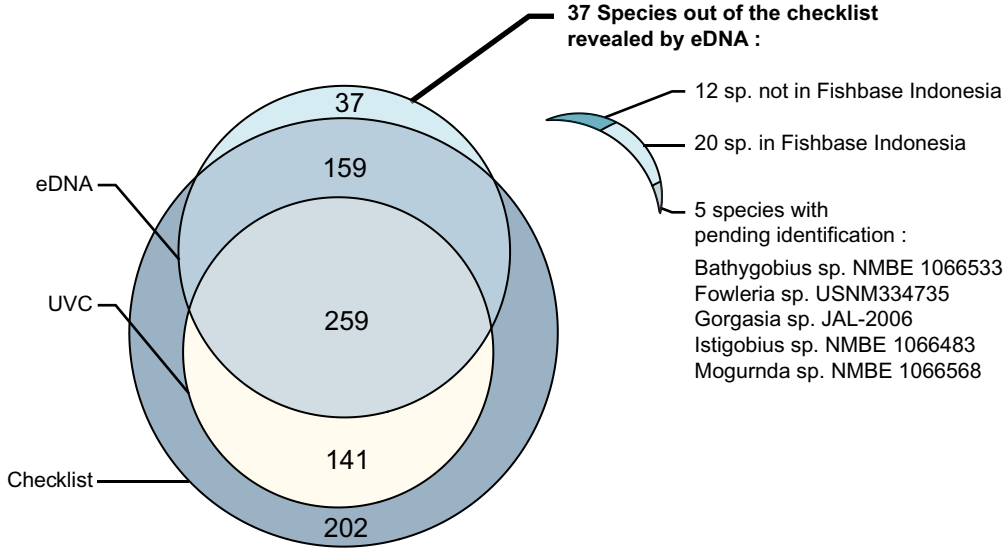

Figure 3. Venn diagrams showing species richness recorded using eDNA and underwater visual census (UVC) inside versus outside the historical fish checklist of the Bird's Head of Peninsula, Indonesia. Only species sequenced for the 12S mitochondrial rDNA region amplified by the teleo primer are considered. The validated species detection (green) included the catch, the morphological and sequencing identification. Data supporting this figure are available in the Supporting information.

16000587, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecog.06299 by Cochrane France, Wiley Online Library on [28/10/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

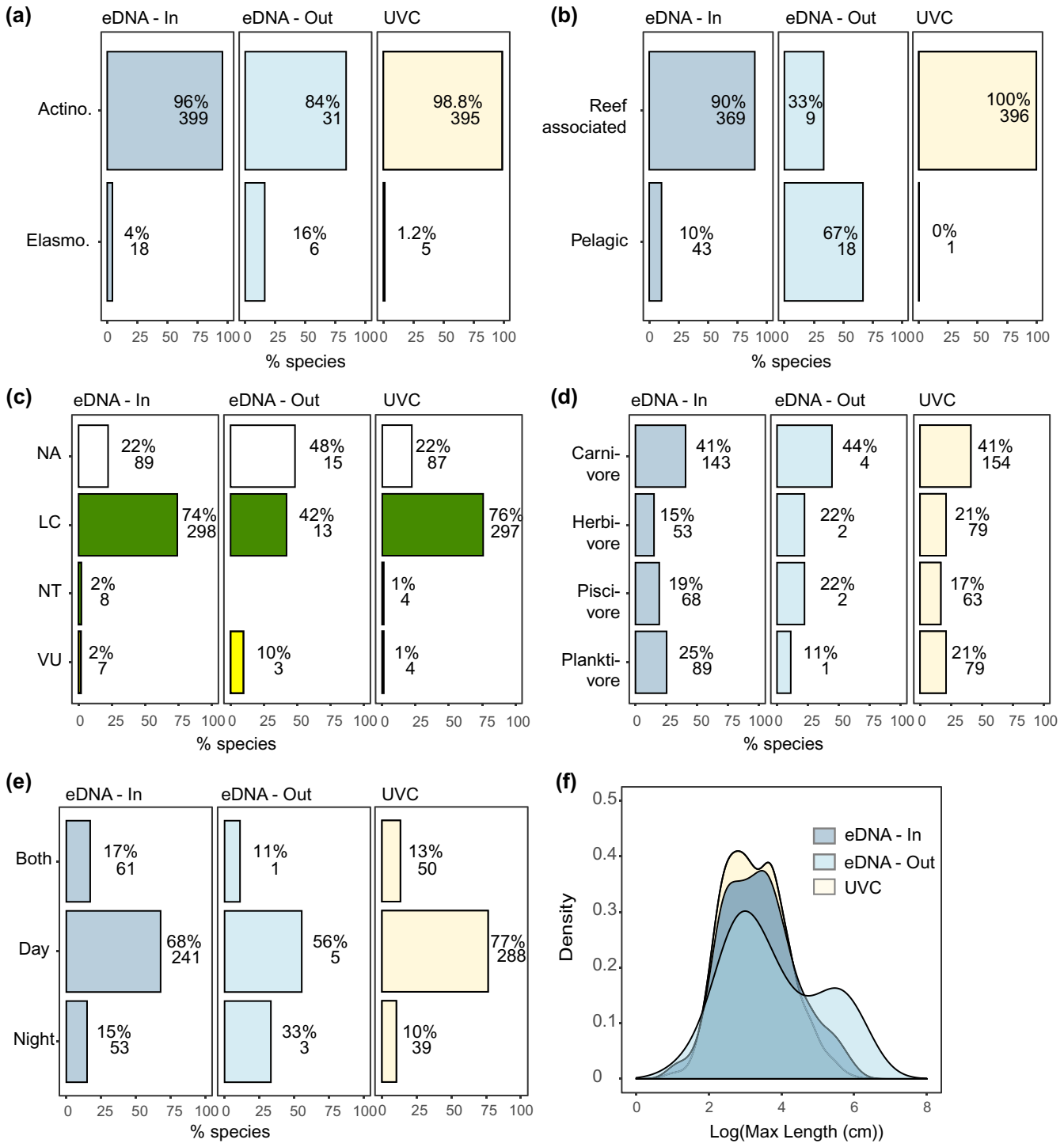

Figure 4. Class (a), habitat (b), IUCN conservation status (c), diet (d), circadian activity (e) and maximum length distribution (log transformed, f) of the fish species found in eDNA samples referenced in the checklist ('In') or absent from it ('Out'). The percentage and the number of species are given for each category. The initials of conservation status are standardized from the IUCN Red List (NA: unknown, LC: least concern, NT: near threatened, VU: vulnerable). Data supporting this figure are available in the Supporting information.

16000587, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecog.06299 by Cochrane France, Wiley Online Library on [28/10/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

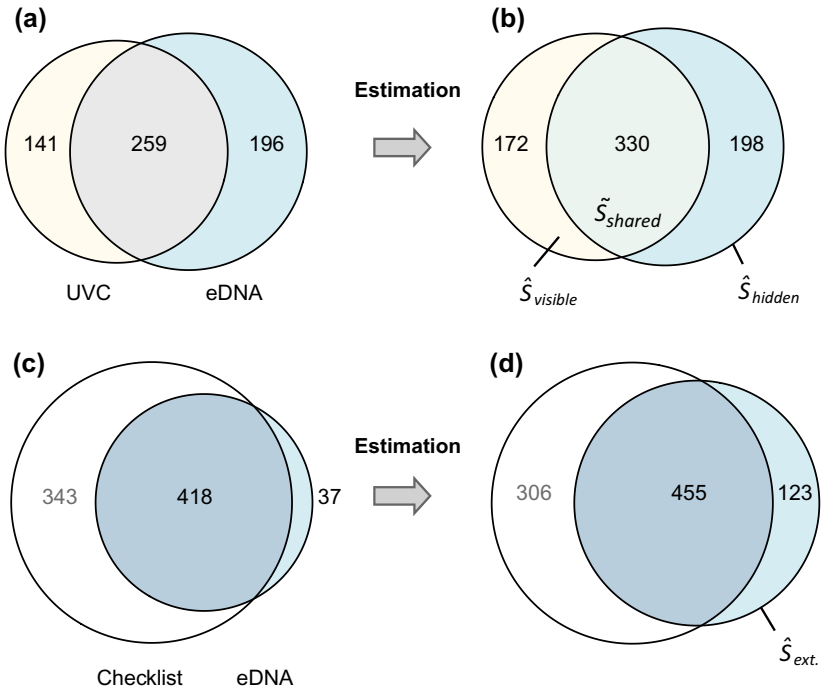

Figure 5. Venn diagrams showing the number of species found in environmental DNA (eDNA) samples and underwater visual censuses (UVC) with the estimates of the visible $\hat{S}_{visible}$ and hidden diversity $\hat{S}_{hidden}$ but also the diversity recorded by both methods $\tilde{S}_{shared}$ (a, b); and the number of species found by eDNA inside/outside the checklist and the extended diversity estimates $\hat{S}_{ext.}$ (c, d). The number of species in the sequenced checklist not detected by eDNA is indicated in grey.

Discussion

Revisiting the fish checklist in the coral triangle

With 92 water samples, eDNA captured 54.8% of the sequenced fish biodiversity in the extended checklist. This detection capacity outperforms that of traditional methods such as visual surveys that can miss elusive, highly mobile or cryptobenthic species (Boussarie et al. 2018, Mathon et al. 2022) or destructive fishing surveys that target restricted habitats and small sets of species. Unlike UVC, eDNA monitoring is less restricted by logistical constraints and can be performed at depths inaccessible to divers. Thus, eDNA metabarcoding allows a more efficient monitoring, especially in remote, difficult to access, highly diverse habitats. This result can be considered as conservative since the spatial extent of the eDNA sampling is smaller than that of UVC so we could expect even more fish diversity detected with a more widespread eDNA sampling. This result is also consistent with many studies comparing sampling methods (Polanco Fernández et al. 2021, Marques et al. 2021). However, the incompleteness of reference databases limits the taxonomic assignment of eDNA sequences (Marques et al. 2020a, Polanco Fernández et al. 2021). In our study, the custom genetic reference database increased the percentage of species sequenced in the checklist from 28.4% to 43.2%. However, a large part of MOTUs obtained by eDNA metabarcoding still could not be assigned suggesting a greater species detection potential (Juhel et al. 2020). The completion of reference genetic databases is thus crucial to provide extensive biodiversity assessment with eDNA. Given these limitations in the use of eDNA, well established survey methods such as UVC remain paramount to provide baseline monitoring information (Edgar et al. 2020). Additionally, traditional fishing techniques allow to collect tissue samples to fill the gaps in genetic reference databases. Thus, monitoring methods remain complementary and should be applied considering the aim of the sampling and methodological tradeoffs.

In our study, eDNA revealed 37 fish species absent from the historical checklist with a high proportion of pelagic piscivorous species. Such mobile and elusive species are often missed by traditional methods that involve divers and are dependent on water visibility. This result confirms the significant and valuable contribution eDNA metabarcoding can provide for the monitoring of such species (Mathon et al. 2022). Additionally, a large number of species were recorded with our fishing effort alone and added to the checklist during the development of the reference database. New species occurrences were also recorded using eDNA metabarcoding, although those remain putative at this stage. These results confirm that the checklist of the Bird's Head of Papua region is not fully known yet, and its true biodiversity remains substantially underestimated with more investigations needed particularly on mesophotic reefs (Andradi-Brown et al. 2021). It also shows that eDNA metabarcoding can help to extend species geographic distributions (West et al. 2020) which are critical for IUCN risk assessments (O'Hara et al. 2021).

16000587, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecog.06299 by Cochrane France, Wiley Online Library on [28/10/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Despite the potential increase in species diversity detected using eDNA metabarcoding, some caution is needed before adding them to regional inventories. Sequences present in the online reference database that were used for taxonomic assignment may have been collected from individuals outside the region of interest. This can induce assignment errors due to biogeographical related genetic variation (Wadrop et al. 2016). Additionally, the robustness of taxonomic assignment is dependent on the taxonomic resolution of the barcode. For example, two phylogenetically close species can share a similar sequence on the 12S gene leading to incorrect species assignment if one of them is absent from the reference database (Jackman et al. 2021). This bias can induce uncertainty on some species detection, which needs to be considered as 'putative' until all species from the same genus are sequenced. In this study, we defined a robust method to consolidate the identification of specimens and improve the quality of the sequences deposited in the genetic reference database (Supporting information).

The potential of eDNA to reveal hidden biodiversity

Visual- and video-based sampling methods have long been used for monitoring underwater biota. However, these traditional methods sample only conspicuous species, overlooking cryptic and elusive species that can constitute a large portion of overall fish diversity (Boussarie et al. 2018, Brandl et al. 2018). Environmental DNA can reveal this hidden diversity and thus holds great potential for the evaluation of human impacts and the success of restoration and protection measures (Zinger et al. 2020, Boulanger et al. 2021), to ultimately optimize and monitor conservation strategies.

Beyond taxonomic assignment, impaired by the actual incompleteness of genetic reference databases and the unperfect taxonomic resolution of genetic markers, eDNA metabarcoding can be analyzed by generating MOTUs that can be defined as distinct taxonomic entities and act as a proxy of taxonomic diversity following adequate curation (Marques et al. 2020b). They can provide a more exhaustive biodiversity estimation, albeit not taxonomically assigned, to revisit biogeographic patterns (Juhel et al. 2020, Mathon et al. 2022). Implementing comprehensive, large-scale and long-term biodiversity observatories will thus significantly complement or challenge many known ecological patterns and processes from the local to the global scale (Boulanger et al. 2021, Mathon et al. 2022).

The asymptotic estimation of biodiversity in hyper-diverse and under-sampled regions

Although widening biodiversity estimates and extending regional species checklists, eDNA metabarcoding cannot provide an exhaustive assessment of species diversity within a given area since eDNA samples are often limited by the volume of water filtered and the narrow range of habitats investigated (Bessey et al. 2020). In our study, 92 eDNA 2 l-samples can hardly detect the whole fish diversity of the entire region and reaching this full inventory in such isolated area would be very costly. In this case, coupling eDNA and sampling-theory-based methods to estimate the level of asymptotic diversity in a given area presents a novel and promising approach. Here we propose a new quantitative framework, based on the incidence of rare species, to estimate the level of extended and hidden diversity in regional checklists. We show that greater efforts using eDNA sampling could lead to a drastic increase in both species geographic ranges and regional species checklists.

With the diversification of biodiversity monitoring methods, our framework allows to estimate the asymptotic diversity shared by different surveys of the same area. In our study, the spatial and temporal discrepancy between UVC and eDNA surveys increases the uncertainty in the estimated hidden diversity ($\hat{S}_{\text{hidden}}$). Meanwhile, updating the checklist would reduce potential spatial biases for the estimated extended diversity ($\hat{S}_{\text{ext.}}$). Our framework can be applied to many taxa and ecosystems since eDNA metabarcoding is being increasingly used to detect species in marine, freshwater and terrestrial ecosystems (Sales et al. 2020). Moreover, using diversity estimators such as the Chao lower-bound framework (Pan et al. 2009, Chiu et al. 2014) will substantially improve estimates of regional biodiversity and fuel subsequent ecological analyses.

Acknowledgements – We thank the National Research and Innovation Agency (BRIN) for promoting our collaboration and the Sorong Polytechnic of Marine and Fisheries (Politeknik KP Sorong, West Papua) for providing the vessel Airaha 02 that we used in this campaign. We thank Julien Leblond, Ghofir Abdul and Amir Suruwaki for contributing to the reference database completion and the fish sampling. We thank the crew of the Aihara 02 for assisting us during the operations and SPYGEN staff for the technical support in the laboratory. Fieldwork and laboratory activities were supported by the Lengguru 2017 Project (<www.lengguru.org>), that is was based on a joint collaboration between Indonesian and European scientists and was covered by a MoU signed between LIPI and IRD on 5 Apr 2017. We thank the International Joint Laboratory 'SEntinel LAboratory of the Indonesian MArine BiodiversiTy' (IJL SELAMAT) for its support. The Lengguru fieldwork was conducted according to relevant guidelines by the government of the Republic of Indonesia and under research permit issued by RISTEK (Indonesia) (permit no. 3179/FRP/E5/Dit.KI/IX/2017) and relevant Indonesian government collecting permit.
Funding – Fieldwork and laboratory activities were supported by the Lengguru 2017 Project (<www.lengguru.org>), conducted by the French National Research Inst. for Sustainable Development (IRD), the National Research and Innovation Agency (BRIN) with the Research Center for Oceanography (RCO, the Politeknik KP Sorong), the Univ. of Papua (UNIPA) with the help of the Inst. Francais in Indonesia (IFI) and with corporate sponsorship from the Total Foundation and TIPCO company. The eDNA sequencing was funded by Monaco Explorations.
Conflict of interest – The authors declare no competing interests.

Author contributions

David Mouillot, Régis Hocdé and **Laurent Pouyaud** share leading authorship. **Jean-Baptiste Juhel**: Conceptualization (equal); Data curation (equal); Formal analysis (lead);

16000587, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecog.06299 by Cochrane France, Wiley Online Library on [28/10/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Investigation (lead); Methodology (equal); Supervision (equal); Visualization (lead); Writing – original draft (lead); Writing – review and editing (lead). **Virginie Marques**: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Validation (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Rizkie S. Utama**: Investigation (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Indra B. Vimono**: Investigation (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Hagi Y. Sugeha**: Investigation (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Kadarusman** : Investigation (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Christophe Cochet**: Investigation (supporting). **Tony Dejean**: Resources (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Andrew Hoey**: Resources (equal); Writing – original draft (supporting); Writing – review and editing (supporting). **David Mouillot**: Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Supervision (equal); Validation (equal); Writing – original draft (equal); Writing – review and editing (equal). **Régis Hocdé**: Conceptualization (supporting); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Project administration (equal); Resources (equal); Supervision (equal); Writing – original draft (equal); Writing – review and editing (equal). **Laurent Pouyaud**: Conceptualization (supporting); Data curation (equal); Funding acquisition (equal); Investigation (equal); Project administration (equal); Writing – original draft (equal); Writing – review and editing (equal).

Transparent peer review

The peer review history for this article is available at <https://publons.com/publon/10.1111/ecog.06299>.

Data availability statement

The data supporting this study, the custom reference database and the updated checklist are given in the Supporting information. The sequencing run is available from the Dryad Digital Repository <https://doi.org/10.5061/dryad.sqv9s4n6f> and the metabarcoding pipelines are available in GitLab (<https://gitlab.mbb.univ-montp2.fr/edna/snakemake_rapidrun_swarm>). Sequence data and associated collection information, corresponding to the custom 12S reference database (Supporting information), have been made available on the GenBank (NCBI).

Supporting information

The Supporting information associated with this article is available with the online version.

References

Allen, G. R. and Erdmann, M. V. 2012. Reef fishes of the East Indies, Vol. I–III. – Tropical Reef Research, 1292 p.

Andradi-Brown, D. A. et al. 2021. Highly diverse reef fish communities in Raja Ampat, West Papua. – Coral Reefs 40: 111–230.

Baker, W. et al. 2000. The EMBL nucleotide sequence database. – Nucleic Acids Res. 28: 19–23.

Barlow, J. et al. 2018. The future of hyperdiverse tropical ecosystems. – Nature 559: 517–526.

Bessey, C. et al. 2020. Maximizing fish detection with eDNA metabarcoding. – Environ. DNA 2: 493–504.

Boulanger, E. et al. 2021. Environmental DNA metabarcoding reveals and unpacks a biodiversity conservation paradox in Mediterranean marine reserves. – Proc. R. Soc. B 288: 20210112.

Boussarie, G. et al. 2018. Environmental DNA illuminates the dark diversity of sharks. – Sci. Adv. 4: eaap9661.

Boyer, F. et al. 2016. OBITOOLS: a UNIX-inspired software package for DNA metabarcoding. – Mol. Ecol. Res. 16: 176–182.

Brandl, S. J. et al. 2018. The hidden half: ecology and evolution of cryptobenthic fishes on coral reefs. – Biol. Rev. 93: 1846–1873.

Brandl, S. J. et al. 2019. Demographic dynamics of the smallest marine vertebrates fuel coral reef ecosystem functioning. – Science 364: 1189–1192.

Campbell, S. J. et al. 2020. Fishing restrictions and remoteness deliver conservation outcomes for Indonesia's coral reef fisheries. – Conserv. Lett. 13: e12698.

CEA, California Environmental Associates 2018. Trends in marine resources and fisheries management in Indonesia: a 2018 eeview. – <www.weirdesign.com>.

Ceballos, G. et al. 2017. Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. – Proc. Natl Acad. Sci. USA 30: E6089–E6096.

Chao, A. et al. 2015. SpadeR: species prediction and diversity estimation with R. – R package ver. 0.1.0, <http://chao.stat.nthu.edu.tw/blog/software-download/>.

Chao, A. et al. 2017. Seen once or more than once: applying good-turing theory to estimate species richness using only unique observations and a species list. – Methods Ecol. Evol. 8: 1221–1232.

Chiu, C. H. et al.2014. An improved non-parametric lower bound of species richness via the good-turing frequency formulas. – Biometrics 70: 671–682.

Cinner, J. E. et al. 2016. Bright spots among the world's coral reefs. – Nature 535: 416–419.

Collins, R. A. et al. 2019. Non-specific amplification compromises environmental DNA metabarcoding with COI. – Methods Ecol. Evol. 10: 1985–2001.

Edgar, G. J. et al. 2020. Reef life survey: establishing the ecological basis for conservation of shallow marine life. – Biol. Conserv. 252: 108855.

Ficetola, G. T. et al. 2010. An in silico approach for the evaluation of DNA barcodes. – BMC Genom. 11: 434.

Fricke, R. et al. (eds) 2021. Eschmeyer's catalog of fishes: genera, species, references. – <http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp>, accessed November 2020.

Froese, R. and Pauly, D. (eds) 2020. FishBase. – <www.fishbase.org>, accessed November 2020.

16000587, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecog.06299 by Cochrane France, Wiley Online Library on [28/10/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Garlapati, D. et al. 2019. A review on the applications and recent advances in environmental DNA (eDNA) metagenomics. – Rev. Environ. Sci. Biotechnol. 18: 389.
Harrison, J. B. et al. 2019. Predicting the fate of the eDNA in the environment and implications for studying biodiversity. – Proc. R. Soc. B 286: 20191409.
Hocdé, R. et al. 2020. Mission report: LENGGURU 2017 expedition 'biodiversity assessment in reef twilight zone and cloud forests', 1/10/2017–30/11/2017 – Kaimana Regency.
Indonesia Ministry of National Development Planning 2019. Roadmap of SDGs Indonesia towards 2030, 179 pp. – <www.sdg2030indonesia.org/>.
Jackman, J. M. et al. 2021. eDNA in a bottleneck: obstacles to fish metabarcoding studies in megadiverse freshwater systems. – Environ. DNA 3: 837–849.
Juhel, J.-B. et al. 2020. Accumulation curves of environmental DNA sequences predict coastal fish diversity in the coral triangle. – Proc. R. Soc. B 287: 20200248.
Kulbicki, M. et al. 2013. Global biogeography of reef fishes: a hierarchical quantitative delineation of regions. – PLoS One 8: e81847.
Juhel, J.-B. et al. 2022. Data from: Estimating the extended and hidden species diversity from environmental DNA in hyperdiverse regions. – Dryad Digital Repository <https://doi.org/10.5061/dryad.sqv9s4n6f>.
Mahé, F. et al. 2014. Swarm: robust and fast clustering method for amplicon-based studies. – PeerJ. 2: e593.
Mangubhai, S. et al. 2012. Papuan bird's head seascape: emerging threats and challenges in the global center of marine biodiversity. – Mar. Pollut. Bull. 64: 2279–2295.
Marques, V. et al. 2020a. GAPeDNA: assessing and mapping global species gaps in genetic databases for eDNA metabarcoding. – Divers. Distrib. 27: 1880–1892.
Marques, V. et al. 2020b. Blind assessment of vertebrate taxonomic diversity across spatial scales by clustering environmental DNA metabarcoding sequences. – Ecography 43: 1779–1790.
Marques, V. et al. 2021. Use of environmental DNA in assessment of fish functional and phylogenetic diversity. – Conserv. Biol. 35: 1944–1956.
Mathon, L. et al. 2022. Cross-ocean patterns and processes in fish biodiversity on coral reefs through the lens of eDNA metabarcoding. – Proc. R. Soc. B 289: 0220162.
McGowan, J. et al. 2020. Conservation prioritization can resolve the flagship species conundrum. – Nat. Commun. 11: 994.
Menegotto, A. and Rangel, T. F. 2018. Mapping knowledge gaps in marine diversity reveals a latitudinal gradient of missing species richness. – Nat. Commun. 9: 4713.
Moeslund, J. E. et al. 2017. Using dark diversity and plant characteristics to guide conservation and restoration. – J. Appl. Ecol. 54: 1730–1741.
Mora, C. et al. 2008. The completeness of taxonomic inventories for describing the global diversity and distribution of marine fishes. – Proc. R. Soc. B 275: 149–155.
Morais, R. A. and Bellwood, D. R. 2019. Pelagic subsidies underpin fish productivity on a degraded coral reef. – Curr. Biol. 29: 1521–1527.
O'Hara, C. C. et al. 2021. At-risk marine biodiversity faces extensive, expanding and intensifying human impacts. – Science 372: 84–87.
Oberdorff, T. et al. 2019. Unexpected fish diversity gradients in the Amazon basin. – Sci. Adv. 5: eaav8681.
Pan, H.-Y. et al. 2009. A nonparametric lower bound for the number of species shared by multiple communities. – J. Agric. Biol. Envir. Stat. 14: 452–468.
Pärtel, M. et al 2011. Dark diversity: shedding light on absent species. – Trends Ecol. Evol. 26: 124–128.
Polanco Fernández, A. et al. 2021. Comparing environmental DNA metabarcoding and underwater visual census to monitor tropical reef fishes. – Environ. DNA 3: 142–156.
Polanco Fernández, A. et al. 2022. Comparing the performance of 12S mitochondrial primers for fish environmental DNA across ecosystems. – Environ. DNA 3: 1113–1127.
Pont, D. et al. 2018. Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation. – Sci. Rep. 8: 10361.
Ruppert, K. M. et al. 2019 Past, present, and future of environmental DNA (eDNA) metabarcoding: a systematic review in methods, monitoring, and applications of global eDNA. – Global Ecol. Conserv. 17: e00547.
Sales, N. G. et al. 2020. Fishing for mammals: landscape-level monitoring of terrestrial and semi-aquatic communities using eDNA from riverine systems. – J. Appl. Ecol. 57: 707–716.
Stat, M. et al. 2019 Combined use of eDNA metabarcoding and video surveillance for the assessment of fish biodiversity. – Conserv. Biol. 33: 196–205.
Thalinger, B. et al. 2021. The effect of activity, energy use, and species identity on environmental DNA shedding of freshwater fish. – Front. Ecol. Evol. 9: 623718.
Teh, L. S. L. et al. 2013. A global estimate of the number of coral reef fishers. – PLoS One 8: e65397.
Valentini, A. et al. 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. – Mol. Ecol. 25: 929–942.
Varkey, D. A. et al. 2010. Illegal, unreported and unregulated fisheries catch in Raja Ampat Regency, eastern Indonesia. – Mar. Policy 34: 228–236.
Veron, J. E. N. et al. 2009. Delineating the coral triangle. – Galaxea 11: 91–100.
Wadrop, E. et al. 2016. Phylogeography, population structure and evolution of coral-eating butterflyfishes (family Chaetodontidae, genus Chaetodon, subgenus Corallochaetodon). – J. Biogeogr. 43: 1116–1129.
West, K. M. et al. 2020. eDNA metabarcoding survey reveals fine-scale coral reef community variation across a remote, tropical island ecosystem. – Mol. Ecol. 29: 1069–1086.
Zhang, S. et al. 2020. A comprehensive and comparative evaluation of primers for metabarcoding eDNA from fish. – Methods Ecol. Evol. 11: 1609–1625.
Zinger, L. et al. 2020. Advances and prospects of environmental DNA in neotropical rainforests. – Adv. Ecol. Res. 62: 331–373.

16000587, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ecog.06299 by Cochrane France, Wiley Online Library on [28/10/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License