WILEY

MOLECULAR ECOLOGY
RESOURCES

RESOURCE ARTICLE OPEN ACCESS

A Long-Term Ecological Research Data Set From the Marine Genetic Monitoring Program ARMS-MBON 2018–2020

Nauras Daraghmeh[1,2] | Katrina Exter[3] | Justine Pagnier[1,2,4] | Piotr Balazy[5] | Ibon Cancio[6] | Giorgos Chatzigeorgiou[7] | Eva Chatzinikolaou[7] | Maciej Chelchowski[5] | Nathan Alexis Mitchell Chrismas[8] | Thierry Comtet[9] | Thanos Dailianis[7] | Klaas Deneudt[3] | Oihane Diaz de Cerio[6] | Markos Digenis[7,10] | Vasilis Gerovasileiou[7,10] | José González[11] | Laura Kauppi[12] | Jon Bent Kristoffersen[7] | Piotr Kukliński[5] | Rafał Lasota[13] | Liraz Levy[14] | Magdalena Małachowicz[5] | Borut Mavrič[15] | Jonas Mortelmans[3] | Estefania Paredes[11] | Anita Poćwierz-Kotus[5] | Henning Reiss[16] | Ioulia Santi[7,17] | Georgia Sarafidou[7,18] | Grigorios Skouradakis[7] | Jostein Solbakken[12] | Peter A. U. Staehr[19] | Javier Tajadura[6] | Jakob Thyrring[19] | Jesus S. Troncoso[11] | Emmanouela Vernadou[7] | Frederique Viard[20] | Haris Zafeiropoulos[7,21] | Małgorzata Zbawicka[5] | Christina Pavloudi[7,17] | Matthias Obst[1,2]

[1]Department of Marine Sciences, University of Gothenburg, Gothenburg, Sweden | [2]Gothenburg Global Biodiversity Centre (GGBC), University of Gothenburg, Gothenburg, Sweden | [3]Flanders Marine Institute (VLIZ), Oostende, West-Vlaanderen, Belgium | [4]LifeWatch ERIC, Seville, Spain | [5]Institute of Oceanology, Polish Academy of Sciences (IOPAN), Sopot, Poland | [6]Plentzia Marine Station (PiE-UPV/EHU), University of the Basque Country, Plentzia-Bizkaia, Basque Country, Spain | [7]Institute of Marine Biology, Biotechnology and Aquaculture (IMBBC), Hellenic Center for Marine Research (HCMR), Heraklion, Crete, Greece | [8]Royal Botanic Garden Edinburgh, Edinburgh, UK | [9]Sorbonne Université, CNRS, Station Biologique de Roscoff, Place Georges Teissier, Roscoff, France | [10]Department of Environment, Faculty of Environment, Ionian University, Zakynthos, Greece | [11]Centro de Investigación Mariña, Universidade de Vigo, Estación de Ciencias Mariñas de Toralla, Vigo, Pontevedra, Spain | [12]Tvärminne Zoological Station, University of Helsinki, Hanko, Finland | [13]Faculty of Oceanography and Geography, University of Gdansk, Gdynia, Poland | [14]The Interuniversity Institute of Marine Sciences in Eilat, Eilat, Israel | [15]National Institute of Biology, Marine Biology Station Piran, Piran, Slovenia | [16]Faculty of Biosciences and Aquaculture, Nord University, Bodø, Norway | [17]European Marine Biological Resource Centre (EMBRC-ERIC), Paris, France | [18]Institute of Oceanography (IO), Hellenic Centre for Marine Research (HCMR), Anavyssos, Greece | [19]Department of Ecoscience, Aarhus University, Roskilde and Aarhus, Denmark | [20]ISEM, Université de Montpellier, CNRS, EPHE, IRD, Montpellier cedex 05, France | [21]KU Leuven, Department of Microbiology, Immunology and Transplantation, Rega Institute for Medical Research, Laboratory of Molecular Bacteriology, Leuven, Belgium

Correspondence: Matthias Obst (matthias.obst@marine.gu.se)

ABSTRACT

Molecular methods such as DNA/eDNA metabarcoding have emerged as useful tools to document the biodiversity of complex communities over large spatio-temporal scales. We established an international Marine Biodiversity Observation Network (ARMS-MBON) combining standardised sampling using autonomous reef monitoring structures (ARMS) with metabarcoding for genetic monitoring

Christina Pavloudi and Matthias Obst shared senior authorship.

of marine hard-bottom benthic communities. Here, we present the data of our first sampling campaign comprising 56 ARMS units deployed in 2018–2019 and retrieved in 2018–2020 across 15 observatories along the coasts of Europe and adjacent regions. We describe the open-access data set (image, genetic and metadata) and explore the genetic data to show its potential for marine biodiversity monitoring and ecological research. Our analysis shows that ARMS recovered more than 60 eukaryotic phyla capturing diversity of up to ~5500 amplicon sequence variants and ~1800 operational taxonomic units, and up to ~250 and ~50 species per observatory using the cytochrome *c* oxidase subunit I (COI) and 18S rRNA marker genes, respectively. Further, ARMS detected threatened, vulnerable and non-indigenous species often targeted in biological monitoring. We show that while deployment duration does not drive diversity estimates, sampling effort and sequencing depth across observatories do. We recommend that ARMS should be deployed for at least 3–6 months during the main growth season to use resources as efficiently as possible and that post-sequencing curation is applied to enable statistical comparison of spatio-temporal entities. We suggest that ARMS should be used in biological monitoring programs and long-term ecological research and encourage the adoption of our ARMS-MBON protocols.

1 | Introduction

The declining health of coastal ecosystems is a major concern for society as unsustainable use of marine resources and growing anthropogenic impacts continue to threaten marine biodiversity and ecosystem functions (Lotze et al. 2018; Smale et al. 2019; Worm et al. 2006). While the majority of coastal habitats are known to have deteriorated in the past century and to have experienced substantial declines in biodiversity (Lotze et al. 2018; Micheli et al. 2013; Obst et al. 2018), it still remains to be fully understood how the loss of species may affect the functioning of these ecosystems (Fields and Silbiger 2022; Gamfeldt et al. 2015; Narwani et al. 2019; Virta, Soininen, and Norkko 2021). Uncertainty also prevails about the significance of individual versus cumulative anthropogenic pressures for the declining biodiversity in coastal ecosystems (Andersen et al. 2015). To better understand these relationships, consistent and well-coordinated biodiversity monitoring efforts are essential (Muller-Karger et al. 2018). These can provide comparable information on ecological variability over time and thereby help identify the thresholds at which critical changes take place (Ducklow, Doney, and Steinberg 2009). More extensive biological monitoring can also be used to contextualise taxonomic information with environmental and socio-economic data to analyse causes and impacts of biodiversity decline as well as the recovery of ecosystems in response to management and restoration efforts (Elliott 2014; Heymans et al. 2020; Jacquemont et al. 2022) but also for following up sustainability goals set by the United Nations (UN) (e.g., Sustainable Development Goal 14: Life Below Water) and the Kunming-Montreal Global Biodiversity Framework.

To make biological monitoring programs more effective, several methods have been proposed recently (Danovaro et al. 2016), one of them being DNA metabarcoding. In principle, DNA-based techniques are capable of identifying biological communities at high temporal and spatial frequency, and with fine taxonomic resolution (Staehr et al. 2022). But despite the frequent application of metabarcoding in marine ecological research, only few marine surveillance programs have so far implemented genetic protocols for long-term monitoring (Hallam et al. 2023; Mathieu et al. 2020). Though it has been shown that DNA metabarcoding may enhance or even outperform traditional approaches (Capurso, Carroll, and Stewart 2023; Fediajevaite et al. 2021), its lack of application in large-scale spatio-temporal contexts may be due to its novelty as a biomonitoring tool (Hallam et al. 2023) or methodological biases associated with it (Capurso, Carroll,

and Stewart 2023; Mathieu et al. 2020; Zhang et al. 2023). Furthermore, data management can become a demanding task when trying to link raw data with different layers of downstream analytical information across a multitude of sample localities and project collaborators, potentially discouraging the use of genomic tools in large-scale monitoring initiatives. To accelerate the use of DNA metabarcoding for long-term and geographically widespread biomonitoring, it is therefore crucial for the scientific community to develop best practices for sample collection, processing and analysis as well as procedures and standards for data management and quality control (Santi et al. 2022).

Recently, Obst et al. (2020) established the Autonomous Reef Monitoring Structures Marine Biodiversity Observation Network (ARMS-MBON) for the genetic monitoring of hard-bottom communities, as part of the marine thematic node of the Group on Earth Observations Biodiversity Observation Network (GEO BON). This program deploys autonomous reef monitoring structures (ARMS) in ports, marinas and nature reserves along the European coastline, as well as at a number of locations in polar regions and the Red Sea, to capture and analyse the genetic diversity across latitudes, oceans and benthic habitats. The use of ARMS—simple-to-produce units of stacked PVC plates mimicking the three-dimensional complexity of benthic habitats—enables standardised and non-destructive sampling of complex benthic communities (Leray and Knowlton 2015; Ransome et al. 2017). The network maintains, to date, 25 observatories each deploying ARMS on a regular basis at one to seven sampling sites. ARMS-MBON has now become part of the European Marine Omics Biodiversity Observation Network (EMO BON) (Santi et al. 2023), a larger European initiative coordinated by the European Marine Biological Resource Centre (EMBRC) for the observation of genomic biodiversity. EMO BON includes marine biodiversity observatories from the Arctic to the Red Sea and investigates biological communities sampled from the water column and the benthic substrate using shared protocols, data and metadata standards. Furthermore, EMO BON contributes to the UN Decade of Ocean Science for Sustainable Development by participating in the global Ocean Biomolecular Observing Network (OBON) (Meyer et al. 2022) program, and, together with other connected projects, it aims for a worldwide coordinated biomolecular observation system (Santi et al. 2023).

Here, we present the first data release from the initial ARMS-MBON sampling campaign (i.e., from all ARMS units deployed in

2018 and 2019 and retrieved between 2018 and 2020), comprising genetic samples and image data collected by observatories from the Gulf of Finland in the Baltic Sea to the Spanish Atlantic coast, and from the northern Red Sea to the Svalbard archipelago. The data set adheres to the FAIR (Findable, Accessible, Interoperable, and Re-usable) Data Principles and contains references to material samples, metadata descriptions, images, sequence data, derived taxonomic observations and documentation of the analytical process. In this paper, we provide some brief exploration of the data set and give examples of potential applications, but the ARMS-MBON data can be used by any interested party for comparative studies and to support DNA-based monitoring of marine biodiversity over large spatio-temporal scales.

2 | Materials and Methods

2.1 | Field Sampling and Laboratory Protocols

The original descriptions of observatory design, fieldwork and sample processing procedures as well as instructions for biobanking and data management have been expanded from Obst et al. (2020). Fieldwork and sample processing for samples of this first data release followed the official ARMS-MBON Handbook v2.0, which is in parts based on the initial protocols developed by the Global ARMS Program (https://naturalhistory.si.edu/research/global-arms-program). We used molecular protocols available under the ARMS-MBON Molecular Standard Operating Procedure v1.0 (MSOP). The Handbook and MSOP can be accessed on the ARMS-MBON GitHub repository (see Table S1 for links). Preserved replicates of the material samples

are stored and catalogued at the ARMS-MBON network partner and EMBRC member institution, the Hellenic Centre for Marine Research (HCMR), Crete, Greece and at the individual ARMS-MBON network partner institutions.

2.2 | Access and Benefit Sharing (ABS)

All collections of ARMS samples were carried out with the necessary ABS national permits. These allow the relevant stations to collect the samples in order to utilise the genetic material for taxonomic identification purposes. If anyone wishes to obtain and process any replicate material as stored in the individual partner stations, please note that any reutilisation needs to be renegotiated with ABS competent authorities in the providing countries.

2.3 | General Description of the Data Set for ARMS-MBON Data Release 001

The first ARMS-MBON data release package comprises data of 56 individual ARMS units which were deployed, retrieved and processed by ARMS-MBON network partner institutions at 15 observatories across 12 countries along the European coastline as well as the northern Red Sea and Svalbard (Figure 1A). This covers the entirety of ARMS-MBON sampling events for which deployments took place in 2018 and/or 2019. Deployment periods ranged from 37 to 649 days during the period of April 2018 and December 2020, with the majority being deployed for around 3 months to approximately 1 year (Figure 1B). The 56 ARMS units represent 56
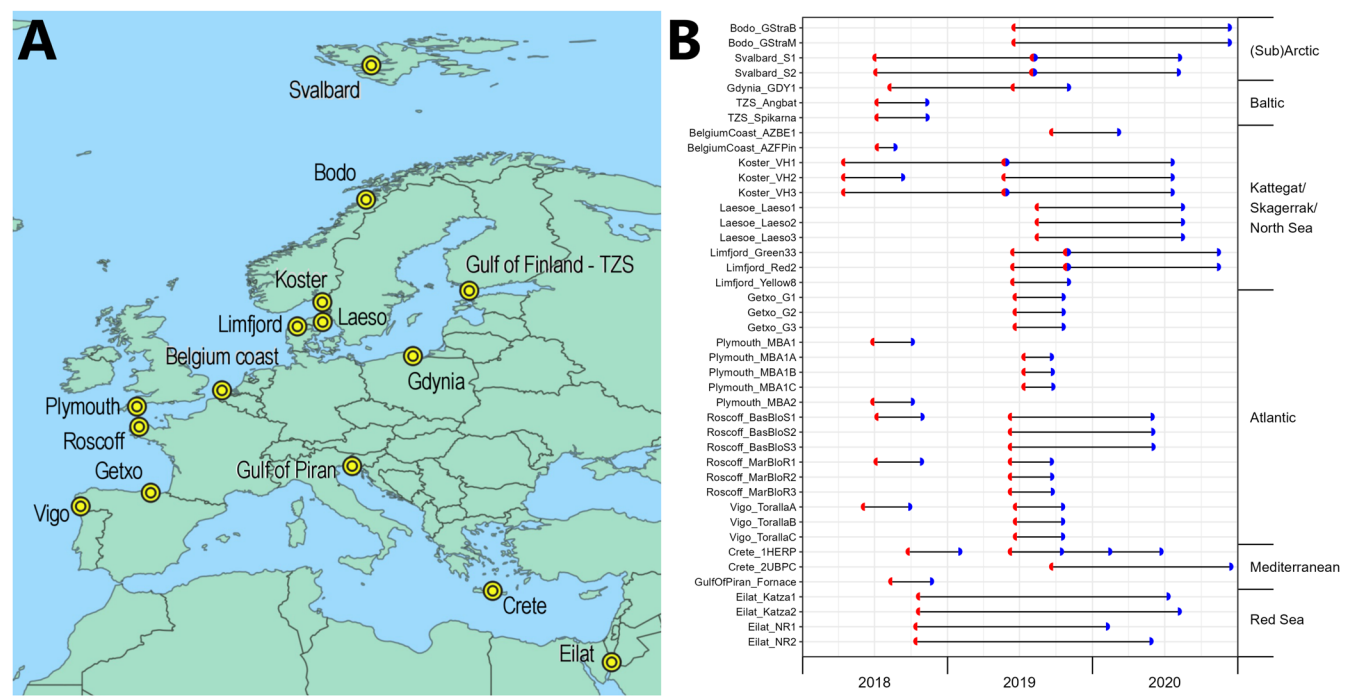
FIGURE 1 | (A) Locations of observatories that deployed ARMS units in 2018 and 2019 during the first ARMS-MBON sampling campaign. (B) Sampling events of the first ARMS-MBON sampling campaign. Axis on the left shows ObservatoryID_UnitID combinations, axis on the right shows groupings of observatories into larger regions. Red semicircle: Time of deployment. Blue semicircle: Time of retrieval. Where red and blue semicircles meet, a new ARMS unit was deployed for a consecutive period at the same spot upon retrieval of the first unit. Where lines contain more than two semicircles, either multiple units were deployed at the exact same spot at the same time but were retrieved at different time points (see Crete_1HERP), or an additional unit was placed directly next to an already deployed one later on and both units were retrieved at the same time (see Gdynia_GDY1).

individual sampling events and 42 unique Unit IDs, meaning where (i) upon retrieval a new ARMS unit was deployed at the same spot for a consecutive deployment period or (ii) for a certain period multiple units remained submerged simultaneously at the exact same spot but were either deployed or retrieved at different time points (i.e., to test the effect of deployment duration), this resulted in sampling events with the same Unit ID but separate Event IDs. General information on the observatories and ARMS deployments (e.g., coordinates, habitat type, deployment depth) and on the sampling events and their resulting material samples (e.g., date of deployment and retrieval, material sample IDs, preservative used) can be found on the GitHub repository (see Table S1 for links) and in Files S1 and S2, respectively.

In total, this first data release package comprises data of 190 material samples. For ARMS-MBON, three size fractions (one sessile and two motile) are processed for each ARMS unit, theoretically resulting in a set of 178 biological samples for the 56 units deployed here. However, during the initial phase of ARMS-MBON, various sampling and processing techniques were tested. This resulted in (i) some biological samples being stored at −20°C without any preservative or being preserved with ethanol (EtOH) instead of dimethyl sulfoxide (DMSO; as DESS buffer solution, see Handbook), which is now the standard preservative used in ARMS-MBON; (ii) some biological samples being processed as duplicates and preserved with DMSO as well as EtOH; (iii) three sediment and two plankton samples being collected and processed as part of this sampling campaign; (iv) some sample fractions being sieved with different mesh sizes; (v) two samples being processed as technical duplicates and (vi) only the two motile fractions being processed for one ARMS unit. Hence, a total material sample number of 190 resulted from this first campaign, including all biological and technical replicates (Table 1 and File S1).

2.4 | ARMS Plate Image Data

Images of each plate were recorded post-recovery and after disassembling the units, to visually document benthic communities according to the instructions in the ARMS-MBON Handbook v2.0. The image collection contains photographs of both sides of the settlement plates as well as close-ups of individual specimens or colonies. Download links for all images of this data release package are available via the GitHub repository (see Table S1 for link and File S1).

2.5 | Amplicon Sequencing

Molecular work for samples of the first data release package was carried out as detailed in the ARMS-MBON MSOP (see Table S1 for link) for DNA metabarcoding of the eukaryotic mitochondrial and nuclear marker genes cytochrome *c* oxidase subunit I (COI), 18S rRNA (18S) and internal transcribed spacer (ITS) region. Note that ITS was only targeted during the initial phase of ARMS-MBON and use of this marker gene has been discontinued. Negative controls were included during marker gene amplification. Each first-step polymerase chain reaction (PCR) included one to two blank samples to assess potential contamination qualitatively, which contained water instead of template DNA. These negative controls were subsequently included in the second-step PCR and sequenced along with actual samples to assess potential contamination quantitatively, too. All samples were subjected to amplicon sequencing by the network partner institute HCMR which also published all sequence data. Raw sequence files of the first data release package are available for download at the European Nucleotide Archive (ENA) (Yuan et al. 2023) through the accession numbers provided via the GitHub repository (see Table S1 for link and File S1 for accession numbers) and under the umbrella study PRJEB72316 (https://www.ebi.ac.uk/ena/browser/view/PRJEB72316). The accession numbers for the sequencing negative control samples are also included there. Information on the demultiplexing procedures applied on sequencing output data is provided as well, denoting if sequence reads deposited on ENA contain primer sequences or not (see Table S1 for link and File S1). Overall, the sequencing data comprise 200 samples (190 material samples plus 10 samples which were re-sequenced due to initially low read yield) for the COI and 18S marker genes, and 111 samples for ITS, plus four to nine negative control samples per marker gene (File S1).

TABLE 1 | Overview of the processed samples from the first ARMS-MBON sampling campaign.

Sample collection	Overview		
Number of ARMS units retrieved	56		
Number of derived material samples (i.e., biological samples and technical replicates; incl. three sediment and two plankton samples)	190		
Photographic images obtained from ARMS units	5258		
Marker gene sequencing	**COI**	**18S**	**ITS**
Sequencing batches	7	7	4[a]
Number of samples sequenced successfully (i.e., number of deposited ENA accessions for: material samples with successful PCR amplification + re-sequenced material samples + negative controls)	185 + 10 + 9	190 + 10 + 8	111[a] + 0 + 4

[a]ITS amplicon sequencing was discontinued during the first sampling campaign; 111 (out of 190) samples were processed for ITS sequencing.

2.6 | Bioinformatics Processing of Raw Sequence Data

To deposit taxonomic observations derived from the COI, 18S and ITS marker genes in the Ocean Biogeographic Information System (OBIS), sequence data were processed with the Pipeline for Environmental DNA Metabarcoding Analysis, PEMA v.2.1.4 (Zafeiropoulos et al. 2020; see Text S1 for details on bioinformatics processing). Given that different parameter settings can lead to rather different outcomes (Brandt et al. 2021; Zafeiropoulos et al. 2020), a fixed set of parameters was used for each marker gene and sequence data were processed separately for each sequencing run. Pre-processing (i.e., quality assessment, read/adapter/primer trimming, error correction, read merging and dereplication) were performed using FASTQC v0.11.8 (Andrews 2010), Trimmomatic v0.38 (Bolger, Lohse, and Usadel 2014), BayesHammer (Nikolenko, Korobeynikov, and Alekseyev 2013) as part of SPAdes v3.13.0 (Bankevich et al. 2012), PANDAseq v2.11 (Masella et al. 2012) and OBITools v1.2.12 (Boyer et al. 2016). All parameter files used for each PEMA run are available at the dedicated repository (see Table S1 for link).

For the 18S marker gene, sequences were clustered into operational taxonomic units (OTUs) using the VSEARCH v2.9.1 algorithm (Rognes et al. 2016) with a threshold of 0.97, while for ITS and COI, clustering was performed with Swarm v2 (Mahé et al. 2015), applying a threshold of $d = 10$ to infer amplicon sequence variants (ASVs). Note that PEMA initially defined the result of Swarm processing as inferred ASVs, that is, sequences which differ by one or more nucleotides, which is now corrected to swarm-clusters (Hakimzadeh et al. 2023). We here use PEMA's initial terminology for consistency reasons. Taxonomy was assigned to 18S OTUs and ITS ASVs with the CREST LCAClassifier v3.0 (Lanzé et al. 2012), using the PR2 v.4.13.0 (Guillou et al. 2013) and Unite v7.2 (Nilsson et al. 2018) databases, respectively. For COI sequences, taxonomic annotation was performed using the RDP classifier (Wang et al. 2007) with the MIDORI database v2.0 (Machida et al. 2017). Singletons and OTUs/ASVs unclassified at the domain level were removed. Abundances of OTUs/ASVs present in negative control samples were adjusted (see Text S1 for details). All bioinformatics analyses were supported by the High-Performance Computing System of the Institute of Marine Biology, Biotechnology and Aquaculture of HCMR (Crete, Greece) (Zafeiropoulos et al. 2021).

2.7 | Data Management, OBIS Submission and the ARMS-MBON GitHub Space

The data management proceeded in multiple stages: (i) collecting the (meta)data from the field scientists and from the processing of genetic data; (ii) harvesting and organising these in a public space, and quality controlling the metadata; (iii) adding semantics, organising the data in Research Object Crates (RO-Crates; https://www.researchobject.org/ro-crate/) (Peroni et al. 2022) and adding provenance and (iv) submitting the data to OBIS. These steps were performed as follows:

i. The types of (meta)data that were collected from the sampling teams of each observatory are outlined in Obst et al. (2020). The event metadata (observatory, event and sample metadata, ENA accession numbers for sequencing data) were collected via a project spreadsheet (Google Sheets) and via the ARMS-MBON project space on the PlutoF platform (https://www.plutof.ut.ee), and data (ARMS plate images and spreadsheets) were uploaded by each team to PlutoF. The PEMA parameter files and selected outputs (taxonomic classifications and fasta files for sequencing data) from the bioinformatics processing were uploaded to a Google Drive folder.

ii. The ARMS-MBON GitHub space (see Table S1 for links) was chosen for all subsequent steps of the data management. Particular reasons for choosing GitHub were its ease of access (within and external of the team) and the ability to track changes, to implement custom workflow support through actions and to build and host web-page-like 'landing pages'. The Google sheets and data on the Google Drive were harvested into respective GitHub repositories, as was the metadata from PlutoF (as Java Script Object Notation, JSON). All harvested data were quality controlled (i.e., consistency checks, correction of mistakes) and combined into a sampling event, an image, an omics (i.e., sequencing data) and an observatory spreadsheet (comma separated values, CSV); each data spreadsheet is accompanied by a metadata spreadsheet that defines and adds semantics to the column names. Due to their file size, the images themselves were not uploaded to GitHub, but their PlutoF (open access) URLs were, and these are included together with image metadata in a CSV file (an improved image database will be developed to better archive, annotate and serve the ARMS-MBON image data). The PEMA input and output files were uploaded to the *processing_batch1* repository. Due to their large combined size, the fasta files were instead uploaded to the Marine Data Archive (MDA) (https://www.mda.vliz.be) and their (open access) URLs are included in the relevant GitHub folder (see Table S1 for respective links to all here described data products).

iii. Once the data were organised into repositories in GitHub, we packed each repository as an RO-Crate. Within the RO-Crate JSON file, the contents of the repository are described and provided with provenance metadata using controlled vocabularies, thus making them machine-accessible and interoperable.

iv. The ARMS-MBON data release package from the first sampling campaign (i.e., all taxonomically classified occurrences from the marker gene analysis with a minimum of two sequence reads) were formatted into Darwin Core Archives (DwCA) for each marker gene for submission to OBIS. Metadata records for these data were created in the Integrated Marine Information System (IMIS; see Table S1 for OBIS links and IMIS DOIs) and all current and future links to the data are accessible there. The DwCAs, as well as all associated observatory and sampling event metadata, ENA accession numbers of amplicon sequencing data and links to image data specifically for the first data release package will be accessible via the *data_release_001* repository on GitHub (see Table S1 for link). The metadata files there represent subsets of the corresponding quality-controlled *combined*

metadata files which contain information on all observatories, sampling events, image data and genetic data of ARMS-MBON to date (i.e., not only for the sampling events described here). The *analysis_release_001* repository on GitHub comprises all relevant bioinformatics analysis data (i.e., parameter files and outputs of PEMA processing) associated with the OBIS submission. This repository is merely a duplicate of the *processing_batch1* repository. All code used for exploratory data analysis described below can be found in the *code_release_001* repository (see Table S1 for links to the repositories).

2.8 | Exploration of Sequencing Data

We explored the PEMA-processed sequencing data to present potential directions for utilising ARMS-MBON data sets. All following procedures are detailed in Text S2. Briefly, data from individual PEMA runs we provide on GitHub were merged for each marker gene and further curated to obtain a data set for visualisations and ecological assessments. As no confidence threshold was applied within PEMA for taxonomic assignments of COI ASVs (note that this is therefore also the case for the EurOBIS submission and users are urged to apply their own self-chosen cut-off), we excluded all rank assignments with a confidence value of below 0.8. We further removed certain samples and replicates to solely assess the ARMS mobile and sessile data and to reduce diversity inflation. Potential contaminant sequences were discarded.

We determined the number and/or abundance of unique phyla, species identified, ASVs/OTUs classified to species level and species shared between the data sets of the three marker genes. Where reference databases did not provide correct phylum level classification, we manually added these (see Text S2 for details). In terms of alpha diversity, we assessed the observed ASV/OTU richness and the number of identified species across observatories as well as frequency distributions of these two parameters (i.e., re-occurrence of ASVs/OTUs and species identified across observatories).

We tested for a significant monotonic relationship between sampling effort and diversity variables. Here, we computed Spearman's correlation of sequencing depth and ARMS deployment duration versus ASV/OTU richness and the number of species identified in each sample. Furthermore, we computed Spearman's correlation of the number of ARMS units deployed and the number of samples included in the analysis post-curation versus ASV/OTU richness and the number of species identified at each observatory. Where the correlation was statistically significant ($p < 0.05$) and moderate to strong (Spearman's $\rho > 0.5$), we performed analysis of simple linear regression to model the relationship between sampling effort predictor variables and dependent diversity variables. As no significant monotonic association was revealed between deployment duration and diversity variables (at a threshold of $p < 0.05$ for significance), we applied generalised additive mixed modelling (GAMM; with the factor *Observatory* as random effect, 15 levels) for this specific case to test if an alternative significant relation was present.

In order to test the application potential of the derived species observation data, we performed a scan against reference checklists for ecological key species. Species occurrences with at least two sequence reads were scanned against the following databases: (i) AZTI's Marine Biotic Index (AMBI) (Borja, Franco, and Pérez 2000; Borja, Chust, and Muxika 2019) for species very sensitive to disturbance; (ii) the World Register of Introduced Marine Species (WRiMS) (Costello et al. 2021, 2024) for species with alien status at the place of observation and (iii) the Red Lists of the International Union for Conservation of Nature (IUCN) and Baltic Marine Environment Protection Commission (Helsinki Commission, HELCOM) for species registered as Near Threatened, Vulnerable, Endangered or Critically Endangered. The AMBI and IUCN/HELCOM information were obtained using the World Register of Marine Species' (WoRMS) (Ahyong et al. 2024) REST services (https://www.marinespecies.org/rest/), whereas the WRiMS checks can be replicated using the Jupyter notebook on https://www.github.com/vliz-be-opsci/lw-iji-invasive-checker. Occurrences of red-listed species were confirmed by scanning against known distribution in WoRMS and corrected where necessary.

Samples were tested for differences in alpha diversity among locations with varying degrees of anthropogenic influence (i.e., industrial, semi-industrial, low human influence (LHI) and protected; see File S1). For statistical comparison, samples with < 5000 reads were removed and the remaining samples rarefied to an equal sequencing depth. Samples classified as 'industrial' were grouped into one category ('industrial/semi-industrial') with samples classified as 'semi-industrial'. Mean and standard deviation (SD) of the two alpha diversity measures were calculated for samples of each influence type. Data were subsequently checked for normality and $\log(1 + x)$-transformation applied where necessary. Subsequently, unidirectional analysis of variance (ANOVA) was performed to test for differences between influence types, with post hoc Tukey's test for pairwise comparisons. Where normality could not be achieved through transformation, non-parametric Kruskal–Wallis rank sum test was applied. All codes used for exploratory analysis can be found at the dedicated GitHub repository (see Table S1). Analyses and data visualisation were performed in R v4.1.0 (R Core Team 2021) via RStudio v2022.07.1 (RStudio Team 2022) using packages of the tidyverse v1.3.1 collection (Wickham et al. 2019) and the packages Biostrings v2.60.2 (Pagès et al. 2020), phyloseq v1.36.0 (McMurdie and Holmes 2013), vegan v2.6.2 (Oksanen et al. 2023), gamm4 v0.2.6 (Wood and Scheipl 2020), ggpubr v0.4.0 (Kassambara 2020), grafify v4.0 (Shenoy 2021), plyr v1.8.7 (Wickham 2011), scales v1.3.0 (Wickham, Pedersen, and Seidel 2023), egg v0.4.5 (Auguie 2019), UpSetR v1.4.0 (Conway, Lex, and Gehlenborg 2017), xlsx v0.6.5 (Dragulescu and Arendt 2020), writexl v1.5.0 (Ooms 2024) and openxlsx v4.2.5 (Schauberger and Walker 2021).

3 | Results

3.1 | Overall Description of the Data Set

Out of the 200 sequenced sample units (i.e., 190 material samples representing biological and technical replicates plus 10 re-sequenced samples) from the first ARMS-MBON sampling campaign, 195 and 200 were successfully sequenced (i.e., samples containing sequences that are deposited on ENA) using the

TABLE 2 | Overview of results from the sequence data processing using PEMA and resulting EurOBIS data sets.

	COI	18S	ITS
Overall number of unique ASVs/OTUs pre-curation	54,641	11,294	10,280
Number of PEMA-processed samples with classified ASVs/OTUs remaining after negative control correction (excl. negative controls)	189	189	42
Overall number of unique, classified, negative-control-corrected ASVs/OTUs	51,782	9,596	508
Sequencing depth (total read number of unique, classified, negative control-corrected ASVs/OTUs)	1,567,301	3,910,167	49,782
Number of occurrences deposited in EurOBIS	18,402	21,482	493

COI and 18S marker genes, respectively, whereas 111 samples (out of 111 samples included for this gene) were successfully sequenced using the ITS marker (Table 1). Sequence processing of these samples using PEMA resulted in 54,641, 11,294 and 10,280 unique ASVs/OTUs for COI, 18S and ITS, respectively (Table 2). After further curation and filtering (i.e., negative control correction and removal of unclassified sequences), 189 samples with 51,782 ASVs and 9,596 OTUs remained for the COI and 18S data sets, respectively. For ITS, 42 samples with 508 ASVs remained (note that for ITS, ASVs could be inferred from two sequencing runs only with the pipeline and parameters applied here, that is, for runs of July 2019 and April 2021). This corresponded to 1,567,301 sequence reads for COI, 3,910,167 sequence reads for 18S and 49,782 sequence reads for ITS (Table 2). All occurrence records with a minimum of two sequence reads of these classified sequences will be accessible through EurOBIS, as explained above. From the COI, 18S and ITS data sets, 18,402, 21,482 and 493 occurrences will be deposited, respectively (Table 2, see Table S1 for links to IMIS metadata and DOI records).

3.2 | Taxonomic Profiles of Three Marker Gene Data Sets

To explore the ecological properties of the processed sequencing data, we further curated the data set as described above (i.e., applying classification confidence threshold for COI, exclusion of some samples and removal of potentially spurious sequences). This curated data set—an illustration of how a potential user-curated data set may look—comprised a subset of 162 samples for COI and 18S, respectively, and 34 samples for ITS (Table 3). These samples contained 40,363, 8,700 and 372 ASVs/OTUs with 1,223,460, 2,875,245 and 24,978 sequence reads for the COI, 18S and ITS data sets, respectively. Application of these three marker genes led to the recovery of 65 eukaryotic phyla, of which 38, 57 and 9 were present in the COI (at the confidence threshold of 0.8 applied here), 18S and ITS data, respectively (Table 3, see File S2 for further details). In regards to relative read abundance, the COI data set classified to phylum level was dominated by metazoan phyla (i.e., eight out of the 10 most abundant phyla)—mainly Arthropoda, Annelida, Chordata, Bryozoa and Cnidaria—while almost half of all reads (~44%) belonged to sequences unclassified at phylum level (Figure 2A, Table 3 and File S2). For 18S, metazoans of the phyla Chordata, Mollusca, Arthropoda, Annelida and Cnidaria dominated the data set, but more non-metazoan phyla were among the most abundant taxa

compared to the COI data (Figure 2C and File S2). The majority of 18S reads could be classified to at least phylum level (~97%; Table 3). As expected, mainly fungal phyla were recovered using the ITS marker gene (Cnidaria was the only non-fungal phylum in the ITS data) and Ascomycota made up more than half of all reads in this gene's data set (Figure 2E and File S2). Approximately two-thirds (~61%) of ITS reads were classified to the phylum level (Table 3).

The use of the COI marker gene resulted in the most species identifications (at the confidence threshold of 0.8 applied here) and observations (occurrences with more than one read in a sample) compared to the use of 18S and ITS. We recovered 2,220 ASVs with species-level classification from the COI data, corresponding to 746 unique species and 2,772 observation records (i.e., occurrences of at least two reads; Table 3). For 18S and ITS, 399 and 92 OTUs/ASVs could be assigned a species name, which represented 135 and 82 unique species with 984 and 106 species observations, respectively (Table 3). In total, the species identifications represented 45 unique eukaryotic phyla: 35 for COI, 31 for 18S and four fungal phyla for ITS (Table 3, see File S2 for more details). In the case of COI, more than 100 species were identified for both the Arthropoda and Annelida phyla, with Molusca, Rhodophyta, Ochrophyta, Chordata and Cnidaria also representing a large share of the species-level identifications (Figure 2B and File S2). Contrarily, Annelida, Ciliophora and Myzozoa dominated species occurrences in the 18S data set (Figure 2D and File S2). Fungal species identifications retrieved from the ITS data mainly belonged to classes of Ascomycota (i.e., Eurotiomycetes, Dothideomycetes, Sordariomycetes, etc.) and Basidiomycota (i.e., Agaricomycetes, Tremellomycetes, etc.) (Figure 2F and File S2). The three marker genes recovered relatively distinct groups of species, as only one (i.e., ITS vs. 18S and ITS vs. COI) and 18 (i.e., COI vs. 18S) of the species recovered here overlapped between the three data sets (Figure S1), and 727, 116 and 80 identified species were unique to the COI, 18S and ITS data sets, respectively. No species occurred in all three marker gene data sets (Figure S1).

3.3 | Genetic and Species Diversity at ARMS-MBON Observatories

Alpha diversity of genetic units (i.e., ASVs and OTUs) and identified species varied between observatories. Between 60 (Eilat—Israel) and 246 species (Koster—Sweden) from COI data (at the

TABLE 3 | Overview of curated sequence data used for ecological analysis.

|	COI	18S	ITS
Number of sequencing samples	162	162	34
Overall number of ASVs/OTUs	40,363	8,700	372
Sequencing effort (total read number)	1,223,460	2,875,245	24,978
Number of phyla recovered	38	57	9
Percentage of reads classified at phylum level (with the confidence thresholds applied here)	56.39	96.65	61.15
ASV/OTUs with Linnean species name classification (with the confidence thresholds applied here)	2,220	399	92
Number of unique species identified with Linnean name	746	135	82
Derived species observation records (occurrences with a minimum of two sequence reads in a sample)	2,772	984	106
Number of phyla represented in species identifications	35	31	4 (15 classes)
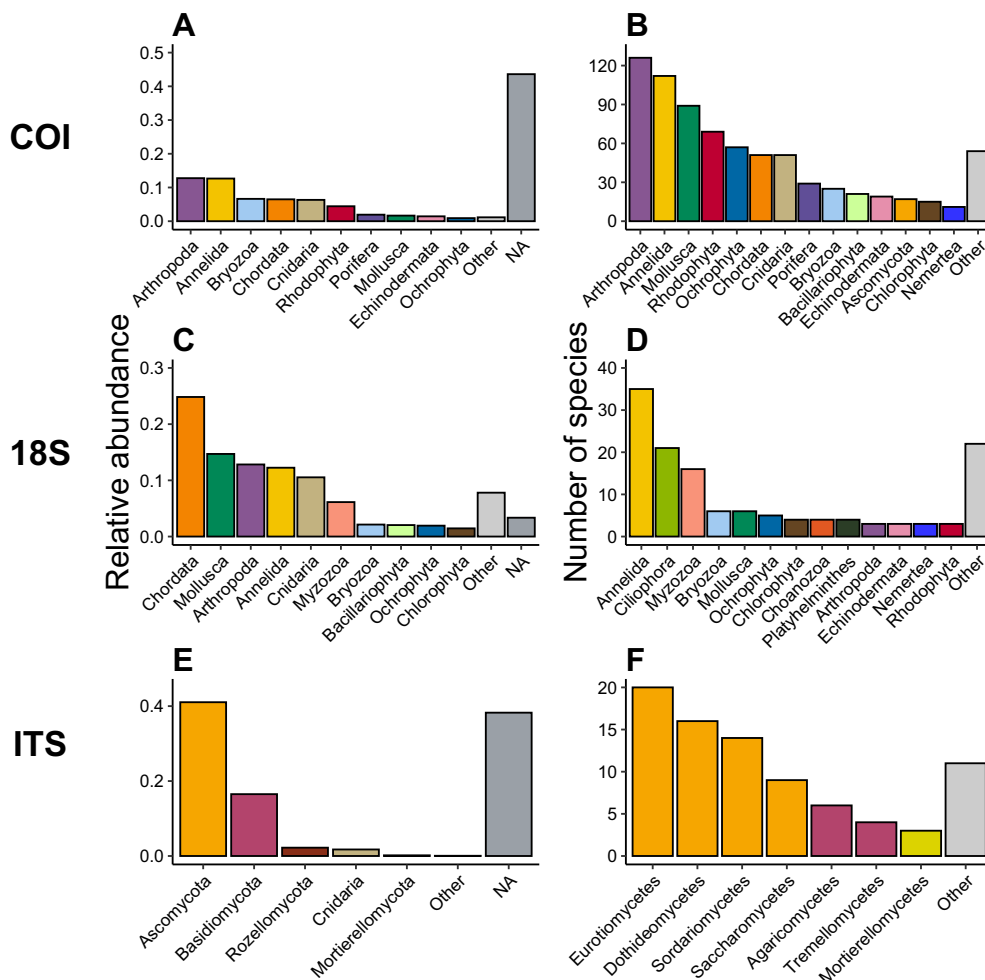
FIGURE 2 | Relative read abundance of the 10 most abundant phyla in the COI (A) and 18S (C) and five most abundant phyla in the ITS data sets (E). Less abundant phyla are grouped as *Other*, whereas relative abundance of sequence reads unclassified at phylum level are grouped as *NA*. Number of identified species within each phylum for COI (B) and 18S (D) and within each class for ITS (F). Phyla/classes with less than 10 (i.e., COI) or three (i.e., 18S and ITS) identified species are grouped as *Other*. Colours correspond to the same unique phyla across all plots. Class-level representation was chosen in (F) for better taxonomic resolution, and colours correspond to the fungal phylum each class belongs to.

confidence threshold applied here) and 2 (i.e., Getxo—Spain) and 53 species (Koster—Sweden) from 18S data were identified at the observatories (Figure 3; Table S2). Richness of ASVs/OTUs ranged between 365 (Læsø—Denmark) and 5,472 (TZS—Finland) for COI and between 253 (Getxo—Spain) and 2,136 (Koster—Sweden) for 18S (Figure 3; Table S2). While the high 18S OTU diversity at Koster on the Swedish west coast aligned with its high number of identified species, the high COI ASV richness at the TZS observatory on the Gulf of Finland contrasted its relatively low number of identified species ($n = 70$) compared to observatories with lower genetic diversity. In comparison, the two ARMS from the Norwegian coast at Bodø captured more than twice as many species ($n = 168$) as the TZS observatory but showed less than half the genetic diversity (2,275 ASVs) for COI. It should be noted that alpha diversity was driven by differences in sampling effort and sequencing depth to some extent (see below).

Analysis of frequency distributions across observatories showed that ~94% ($n = 38,090$) and ~4% ($n = 1,593$) of COI ASVs occurred at one or two observatories only, respectively (Figure 4A). Frequency distribution was less skewed for 18S data, with ~67% ($n = 5,794$) and ~18% ($n = 1,565$) of OTUs found at one or two observatories only, respectively (Figure 4B). Approximately 400 species (~53%) identified from the COI data occurred at one observatory only, whereas 160 species (~21%) occurred at two observatories (Figure 4C). A relatively low number of species identified from the COI data occurred at 10 or more observatories ($n = 13$, ~2%; Figure 4C). Half of all species ($n = 68$) identified using the 18S marker gene were found at one observatory only, and 17 species (~13%) occurred at only two observatories (Figure 4D). Two species identified from 18S data appeared at 10 or more observatories (Figure 4D).

3.4 | Influence of Sampling Effort and Deployment Duration on Diversity Measures

During the first ARMS-MBON sampling campaign, observatories deployed one (GulfOfPiran—Slovenia) to eight
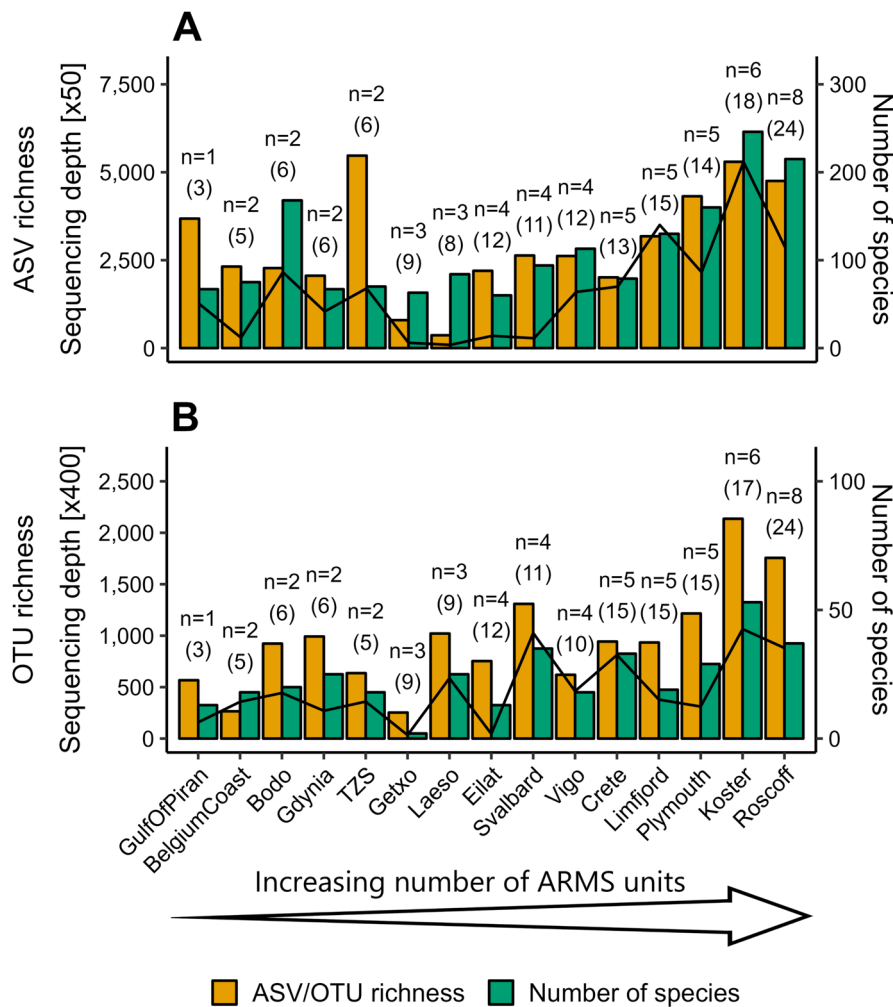

FIGURE 3 | Diversity observed for the COI (A) and 18S (B) data sets across the 15 observatories measured as ASV/OTU richness (yellow; shared axes on the left) and the number of identified species (green; axes on the right) at the classification confidence threshold applied here. The number of ARMS units deployed at the respective observatory is shown on top of the bars (n). Numbers in parentheses equal the number of genetic samples remaining for each observatory after curation and filtering for ecological analysis. The black line denotes sequencing depth (i.e., cumulative sequence read number of all genetic samples for each observatory; left shared axes). Observatories are ordered from left to right by an increasing number of ARMS units deployed.
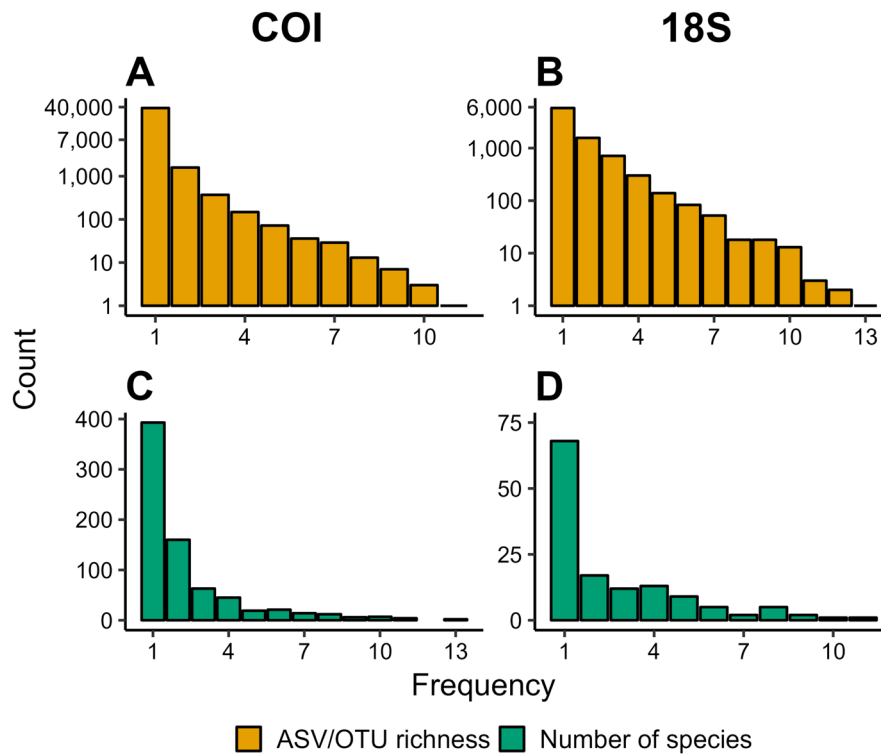
FIGURE 4 | Frequency distribution of ASVs for COI (A) and OTUs for 18S (B) and of species identified for COI (C) and 18S data (D). Counts represent the number of entities (i.e., ASVs/OTUs or species identified) occurring at a given frequency, in this case, the number of observatories. Note the $\log_2$-transformation of the y-axis in A and B.

(Roscoff—France) ARMS units, for periods ranging from 37 (BelgiumCoast—Belgium; UnitID AZFPin) to 649 days (Eilat—Israel; UnitID Katza2) (Figure 1B; see File S1 for further details). After data curation and filtering, this resulted in a data set of three (GulfOfPiran—Slovenia) to 24 samples (Roscoff—France) with at least one sequence read for each observatory for both the COI and 18S marker genes (Figure 3). Given differences in sequencing depth and quality, not all of the three fraction samples remained for each ARMS unit post-curation.

Computation of correlation indicated no significant monotonic association between deployment duration and the number of species identified in each sample for both marker genes (Figure 5A,B; see Table S3 for detailed results of correlation analysis). This was also the case for deployment duration vs. ASV/OTU richness (Figure 5A,B). Generalised additive mixed modelling (GAMM) indicated no significant non-linear effect of the smoother (i.e., deployment duration in days), except for a negative relation in the case of COI for ASV richness, although with high variability in response across observatories and in the effect of deployment duration (edf = 2.059, F = 11.18, p < 0.001; see Table S4 for detailed results of GAMM).

For both marker genes, sequencing depth (i.e., the number of sequence reads) significantly drove the number of species identified and the ASV/OTU richness observed in each sample, with moderate to strong correlation (p < 0.001 and Spearman's $\rho(160) \geq 0.60$ for all associations; Figure 5C,D). Analysis of linear regression indicated that sequencing depth accounted for around 23%–41% of variation in ASV/OTU richness and the

number of species identified in each sample for both marker genes. However, residual standard errors suggested the regression models did not fit the data sufficiently, although fit was more accurate for 18S compared to COI data (see Table S4 for detailed results of regression analysis).

For the COI marker gene, only the number of species identified at each observatory was significantly and moderately to strongly correlated to both the number of ARMS units deployed and the number of samples included in the analysis (p = 0.018, Spearman's $\rho(13)$ = 0.60; and p = 0.021, Spearman's $\rho(13) \geq 0.59$, respectively; see Figure S2A,C and Table S3). The observed COI ASV richness did not show a significant linear relationship with these two sampling effort parameters (Figure S2A,C and Table S3). In the case of 18S, both the observed OTU richness and the number of species identified at each observatory significantly and moderately correlated with both the number of ARMS units deployed and the resulting number of samples included in the analysis ($p \leq 0.012$ and Spearman's $\rho(13) \geq 0.64$ for all associations; see Figure S2B,D and Table S3). Analysis of linear regression indicated that the number of ARMS units deployed and the number of samples included accounted for around 37%–51% of variation in the number of species identified at each observatory for both marker genes. In the case of 18S, the number of ARMS units deployed and the number of samples included were both responsible for ~52% of variation in the OTU richness observed at the observatories. Given relatively high variation in the dependent variables and therefore considerable residual standard errors, observatory-wise regression models did not fit our
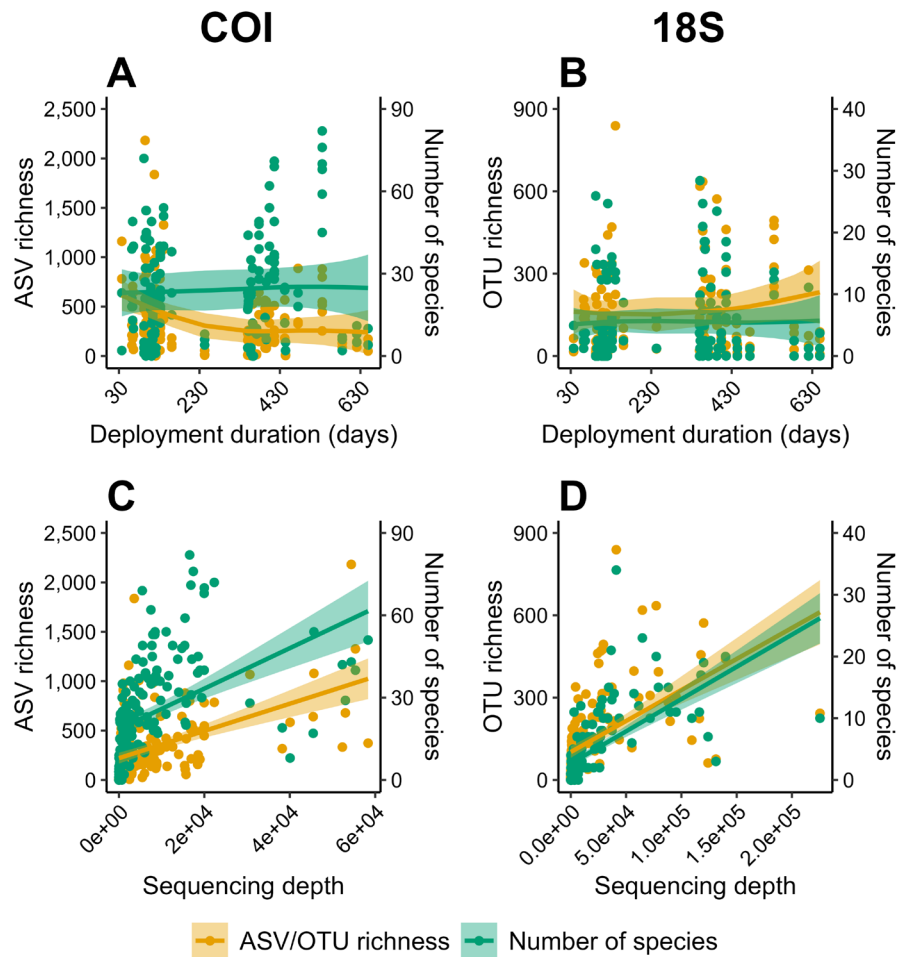
Molecular Ecology Resources, 2025

FIGURE 5 | Relationship of the observed ASV/OTU richness (yellow; axes on the left) and the number of species identified (green; axes on the right) in each sample versus the deployment duration in days of the respective ARMS units for the COI (A) and 18S (B) data sets; and relationship of ASV/OTU richness (yellow; axes on the left) and the number of species identified (green; axes on the right) in each sample versus sequencing depth (i.e., number of sequence reads in each sample) for the COI (C) and 18S (D) data sets. Solid lines represent the fixed effect smoother (deployment duration) based on GAMM (A, B) and linear regression (C, D) for ASV/OTU richness (yellow) and the number of species identified (green), shaded areas represent the corresponding 95% confidence intervals.

data accurately as described above for sample-wise models (Table S4).

3.5 | Identification of Ecological Key Species

The scan against databases for sensitive, non-indigenous and red-listed species resulted in observations of species in all three categories across the observatories (Figure 6). Overall, we observed 88 species registered in AMBI as sensitive to disturbance, 32 species listed as 'alien' in WRiMS at the location of occurrence, and 4 species registered as near threatened, vulnerable, endangered or critically endangered across all observatories (Table S5 and File S3). The observatory at Koster (Sweden) detected the highest number of sensitive species ($n = 37$), whereas Limfjord (Denmark) displayed the highest number of non-indigenous species (NIS). Red-listed species were only detected at the Plymouth (UK) (*Mya truncata*), Koster (Sweden) (*Cliona celata* and *Echinus esculentus*) and Roscoff (France) (*Cliona celata* and *Nucula nucleus*) observatories.

3.6 | Alpha Diversity Across Locations With Varying Degrees of Anthropogenic Influence

The number of identified species differed significantly among habitats with varying degrees of anthropogenic influence based on COI data (low human influence: 26 ± 8, industrial/semi-industrial: 29 ± 13, protected: 38 ± 15; ANOVA: $F_{2,60} = 5.043$; $p = 0.009$) (Figure 7A and Table S6) but based on 18S data they did not (industrial/semi-industrial: 5 ± 5, low human influence: 6 ± 4, protected: 7 ± 3; Kruskal-Wallis $\chi^2 = 2.543$, $p = 0.280$) (Figure 7B and Table S6). Mean ASV richness ranged from 216 ± 90 (low human influence) to 260 ± 200 (industrial/semi-industrial) and 260 ± 173 (protected) for COI data (Figure 7A) and no statistically significant difference was detected among habitat types given high within-group variability (ANOVA of $\log(1+x)$-transformed data: $F_{2,60} = 0.241$; $p = 0.787$; see Table S6 for details of statistical tests). The same was observed for 18S data (ANOVA: $F_{2,60} = 1.906$; $p = 0.157$), with OTU richness ranging from 123 ± 78 (industrial/semi-industrial) to 136 ± 53 (low human influence) and 165 ± 84 (protected) (Figure 7B and Table S6).
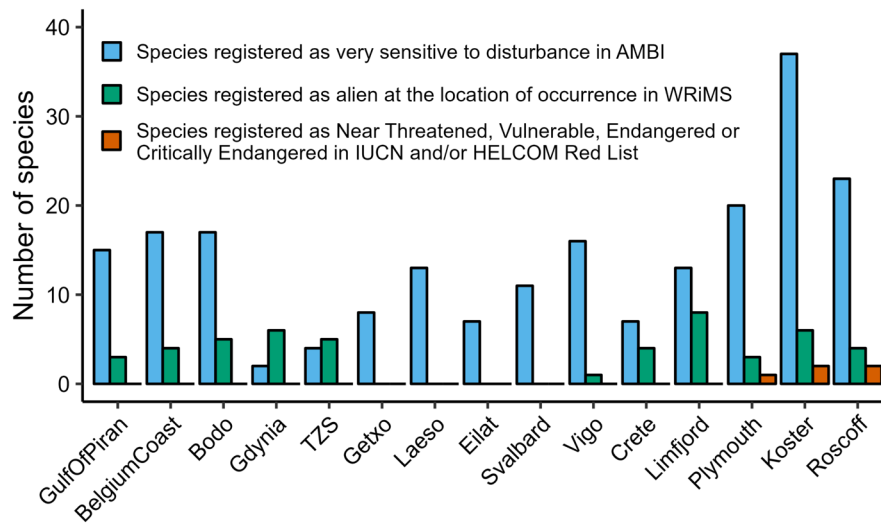
FIGURE 6 | Number of identified species (with occurrences of at least two sequence reads) across observatories listed in four different databases. Data of COI and 18S marker genes were pooled for each observatory. Blue—species registered as very sensitive to disturbance in AMBI; green—species registered as alien at the location of occurrence in WRiMS; orange—species registered as Near Threatened, Vulnerable, Endangered or Critically Endangered in the Red Lists of the IUCN and HELCOM. Observatories are ordered from left to right by increasing number of ARMS units deployed ($n = 1$ for GulfOfPiran, $n = 8$ for Roscoff).
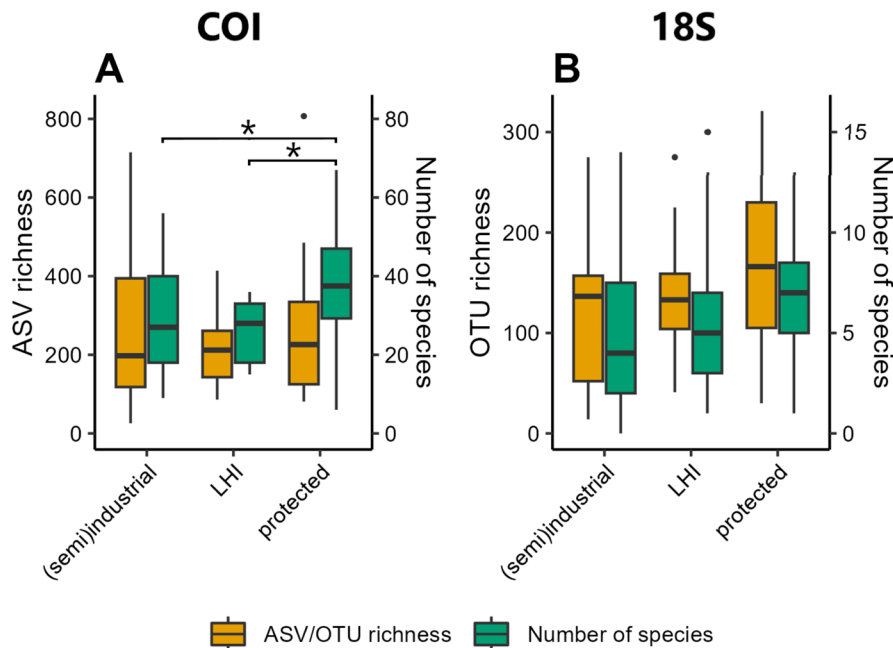

FIGURE 7 | Alpha diversity across locations with varying degrees of anthropogenic influence (i.e., industrial/semi-industrial, LHI—low human impact and marine protected area). Boxplots show alpha diversity measured as ASV/OTU richness (yellow; axes on the left) and the number of identified species (green; axes on the right) at the classification confidence threshold applied here for samples of each influence type for the COI (A) and 18S (B) data sets. Boxes span the interquartile range (IQR) from the first to the third quartile. Horizontal lines within boxes represent median values. Whiskers extend to minimum (lower) and maximum (upper) values up to 1.5*IQR beyond either side of the IQR. Values falling outside this range are outliers represented by single black dots. Significant differences in the number of identified species for the COI data set according to ANOVA are shown in (A) by horizontal bars with asterisks above boxplots (*$p < 0.05$).

4 | Discussion

Here we present the data set of raw and processed amplicon sequencing data from the first sampling campaign (2018–2020) of ARMS-MBON. To the best of our knowledge, this campaign represents the most geographically widespread sampling

initiative using ARMS to date. Raw sequencing (and image) data are open-access and, thus, they can be subjected to reprocessing by all interested users according to their needs. In addition, we processed sequencing data using a dedicated pipeline, and all taxonomic occurrences are published on EurOBIS. This processed data set was further filtered and curated,

representing a potential use case for ecological analysis or taxonomic screening. Analysis of this curated data set shows that ARMS are able to capture eukaryotic taxa belonging to more than 60 phyla, while composition of the hard-bottom benthic cryptofauna in terms of dominant phyla was comparable with those of earlier studies using ARMS (e.g., Carvalho et al. 2019; Ip et al. 2022; Pearman et al. 2018, 2020). Our data set also allows for comparisons of communities at all taxonomic levels (i.e., including genus, class, order level identifications) which are typically applied in local and regional community studies (Staehr et al. 2022). Although we do not present any prokaryotic data here, the preserved physical samples of ARMS-MBON sampling campaigns can also be used for comparative studies of non-eukaryotic microbial communities in the future (Ip et al. 2022; Pearman et al. 2019). Finally, the catalogue of ARMS plate images collected can become a valuable data resource in the future as it will allow for analysis of benthic community composition and growth dynamics with application of advanced image classification methods (Beijbom et al. 2015).

4.1 | Recovered Taxa Are Distinct Across Marker Genes and Observatories

The marginal overlap in identified species between the marker genes underlines the importance of applying multi-marker assays to increase taxonomic coverage in biomonitoring projects (e.g., da Silva et al. 2019; Duarte et al. 2023; Gibson et al. 2014; Gielings et al. 2021). The geographic distribution of genotypes and species identified across the ARMS-MBON network was relatively restricted with most observations of species and genotypes being unique to one or two observatories. Such pattern of sequences and taxa being unique to sample units or locations has been observed in many metabarcoding studies before (e.g., Carvalho et al. 2019; Villalobos et al. 2022) and in our case may be due to the still limited number of samples and observatories given the large geographic scale of the project. As samples and data accumulate over the coming years, it is likely that the partition with re-occurring observations will increase. In addition, the growing number of reference sequences (Porter and Hajibabaei 2018) and/or using customised reference databases (Mugnai et al. 2023) will likewise increase the taxonomic resolution and observation records derived from the samples.

4.2 | The Need to Account for Differences in Sequencing Depth and Sampling Effort

Deployment periods of ARMS units ranged from 37 to 649 days, with the majority being deployed for around 3 months to approximately 1 year. Our tests showed that in most cases deployment duration did neither have a significant effect on observed species nor on genetic diversity. The negative correlation between deployment duration and sequence diversity in the COI data set can most likely be attributed to the low sequencing depth in some samples of ARMS deployed for longer periods (i.e., units of observatories from Svalbard—Norway and Eilat—Israel and one unit of Crete—Greece). However, we want to stress the fact that a conclusion on the effect of deployment duration can ultimately only be achieved with dedicated testing, that is, by comparing the same location with both short- and long-term deployments

and equal sequencing depth. Previous studies have found contrasting results regarding the influence of deployment duration of sampling units. Using artificial substrate units (ASUs), Cahill et al. (2018) did not observe a significant effect of deployment duration on the number of specimens recovered and argued that overall recruitment patterns are predominantly driven by ecological and biogeographic conditions. In their comparative study, Leite et al. (2023) deployed ARMS and artificial seaweed monitoring systems (ASMS) for 6, 9 and 12 months and found that community composition changed over time given seasonal change. The authors indicate that maximum diversity can be recovered with sampling units deployed for < 12 months, which had also been shown by earlier studies (Leite et al. 2021).

In contrast to deployment duration, we showed that recovery of sequence and taxonomic diversity significantly depend on sequencing depth. This is a well-known problem in microbial amplicon sequencing (Cameron et al. 2021; McMurdie and Holmes 2014) and has also been shown in eukaryotic DNA metabarcoding studies (Alberdi et al. 2018; Grey et al. 2018; Shirazi, Meyer, and Shapiro 2021). For example, the substantial differences between observatories in species identified in our data sets can most likely be attributed to the differences in sequencing depth (see clear alignments of relatively low/high taxonomic diversity with relatively low-/high-sequencing depth in Figure 3, for example). This is a likely cause for some northern observatories (e.g., Koster—Sweden) showing higher diversity than some more southerly located (e.g., Getxo—Spain) and subtropical (e.g., Crete—Greece, and Eilat—Israel) observatories. As expected, the number of ARMS units deployed and the number of samples analysed for each observatory significantly drove the recovered alpha diversity, as well. Large variation in ASV richness across COI samples is likely the reason that the effect of sampling effort was not significant for this specific test case. Differences in sampling effort are known to drive observed alpha diversity measures in traditional ecological surveys (Gotelli and Colwell 2011) and, in particular, in metabarcoding studies (Evans et al. 2017; Grey et al. 2018). For the observatories in our data sets, there is a general trend towards higher diversity with increasing sampling effort. However, an increase in sampling effort only led to an increase in the diversity recovered if sequencing depth was sufficient in the additionally collected samples. Given these intrinsic data characteristics, we recommend the application of analytical tools to normalise for sequencing depth and to reduce sampling effort bias when analysing large data sets like the ones ARMS-MBON makes available.

Given that deployment duration did not influence alpha diversity, we suggest deploying ARMS for 6–12 months (during the season of most substantial growth), with preference for the second. We also urge for deployments of at least three ARMS units per site (i.e., with a distance of around 10 m from each other) as biological replicates to obtain a comprehensive representation of surrounding communities and to improve statistical power for comparative analysis. Given differences in sampling effort and in sequencing depth across samples—the latter being inherent to any amplicon sequencing study—we recommend that users apply statistical tools when using ARMS-MBON data to account for those during ecological analysis.

4.3 | ARMS Data Enable the Study of Large-Scale Biodiversity Patterns

We found indications that the alpha diversity of the hard-bottom benthos within marine protected areas (MPAs) is higher than in locations with more intense anthropogenic influence. Mean sequence and species richness were highest in samples from ARMS deployed in protected areas for both marker genes, but given high within-group variation across samples, this was only statistically significant for species richness of one marker gene. This can be an indication of the positive effect of protection measures (Edgar 2011; Edgar et al. 2014) but may also be related to the fact that MPAs are typically established in areas with already low anthropogenic influence where political implementation costs are low (Devillers et al. 2015; Stevenson et al. 2020). The fact that we found higher species diversity but not higher genetic diversity in MPAs remains unexplained at this point. A possible reason for this pattern could be that high genetic diversity reflects recruitment which may be high even in habitats with anthropogenic pressures, while MPAs support species beyond the initial recruitment stage. Further, propagules and larvae of opportunistic taxa, including NIS, may be particularly ubiquitous in harbours and marinas, but less so in stable ecosystems with high conservation status such as MPAs. In addition, small, heterogeneous but isolated MPAs may support high species but low genetic diversity due to reduced connectivity (Bell and Okamura 2005) and seascape and spatial factors may have a higher influence on genetic diversity than protection status (Benestan et al. 2023).

The fact that COI samples from sites with low human influence exhibited lower diversity compared to impacted sites may be attributed to stochastic factors. The majority of low-human-influence samples for this marker gene came from one observatory (Limfjord—Denmark), which may have driven richness parameters for this sample group. Such phenomena can be resolved in the future given the continuous sampling of ARMS-MBON on a large spatio-temporal scale. Data from the forthcoming sampling campaigns will help unravel differences in diversity more clearly in the future. Ultimately, ARMS data can be used to document biodiversity trends in benthic habitats in relation to human activities and eventually contribute to several essential biodiversity variables (EBVs) in the future (Kissling et al. 2018). Genetic and taxonomic data from ARMS-MBON may also be combined with open-access remote sensing data to link diversity patterns and taxonomic occurrences to environmental parameters (Pearman et al. 2019, 2020).

4.4 | ARMS Are Sensitive to Indicator, Non-Indigenous and Threatened Species

Our results show that ARMS can detect ecological indicator species sensitive to disturbance, which are typically used in environmental impact assessments (EIAs) (Bustos-Baez and Frid 2003; Dauvin 2005), environmental risk assessments (ERAs) (Kaikkonen et al. 2018) and national/regional monitoring programs (e.g., HELCOM 2013) to assess the health status of ecosystems. ARMS-MBON data are likely to improve such assessments because they provide information on species as well as genetic diversity, which can be analysed in relation to anthropogenic pressures. As such, ARMS should be deployed continuously in sites such as ports, marinas, wind farms or aquaculture facilities in order to assess impacts of human activities on marine biodiversity (Witalis et al. 2021). We also detected a (low) number of red-listed species, although the fact that this was the case only for observatories with the highest sampling effort underlines the rarity of such taxa and the need for continuous and considerable sampling to track and monitor them.

The data presented here can also be used for tracking the distribution and range shift of NIS as well as other taxa (Martaeng, Obst, and Kuklinski 2023; Wesselmann, Apostolaki, and Anton 2024) and may be applied in various monitoring programs and directives, such as European Union's Marine Strategy Framework Directive (MSFD) when assessing descriptors D1 on biological diversity and D2 on NIS (Bourlat et al. 2013) or the Water Framework Directive (WFD) (Duarte et al. 2023). Such investigations may be particularly enhanced by the intraspecific diversity that can be detected through large-scale genetic data sets such as the one of ARMS-MBON, which will allow for studies of population structure or connectivity of particular species with recently proposed metaphylogeographic analyses (Antich et al. 2023; Martaeng, Obst, and Kuklinski 2023; Turon et al. 2020). In addition, ARMS-MBON data are likewise useful for effective alien species matches between ports as part of same risk area assessments under the International Maritime Organisation's Ballast Water Management Convention (Stuer-Lauridsen et al. 2018). For the detection of NIS we relied on species listed as alien in WRiMS; however, this may be enhanced in the future by comparing occurrences in ARMS-MBON data sets with information in repositories such as the Global Biodiversity Information Facility (GBIF) to identify potential novel invasions or range extensions.

4.5 | How to Build a Successful, Data-Producing, Long-Term Genetic Observatory Network

Lessons learned from running the ARMS-MBON project are the following:

- Standardisation: it is vital to ensure standard protocols (collecting and processing of material, managing the data) are published, understood and followed so that the resulting data are comparable over the space and time of the observatories.

- Constant engagement: it is necessary to engage the hearts and minds of the participants for the exciting (e.g., field work) as well as the challenging and tedious (e.g., data management) parts, as otherwise the resulting (meta)data are insufficient to fulfil the project's potential.

- Strength in numbers: it is necessary to have a multidisciplinary core of experts taking control of the different parts of the project (e.g., sampling, sequencing, bioinformatics, data management, analysis), as no one scientist can do all of these with the quality and timeliness to allow for regular and trustworthy data releases.

5 | Conclusion

The ARMS-MBON initiative is a network of long-term ecological research (LTER) sites committed to the scientific exploration of hard-bottom benthic communities along the coasts of Europe and adjacent regions. Data resulting from the network's consecutive sampling campaigns will enable the study of coastal biodiversity over large temporal and spatial scales and may crucially enhance efforts to monitor marine habitats with varying degrees of anthropogenic influences. Such long-term sampling programs also enable improved early detection and monitoring of specific groups of taxa, such as NIS. The strength of ARMS-MBON lies in its continuous application of standardised protocols and operating procedures, as well as centralised molecular sample processing and sequencing. These measures help reduce biases potentially introduced due to the large-scale experimental set-up. As improvements to protocols and standard procedures are constantly underway, these will further enhance standardisation within the network. Importantly, the possibility of utilising either the raw data or the data processed with a standardised bioinformatics pipeline gives users the freedom to choose the data product best suited to their specific needs. As ARMS-MBON continues its sampling efforts through EMO BON, the open-access data it delivers provide an increasingly critical source of information in times of utmost urgency for large-scale environmental monitoring.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All data presented in this manuscript are publicly available, see main text and Supporting Information for detailed descriptions. Standard operating procedures and protocols are available on the dedicated ARMS-MBON GitHub repository (https://github.com/arms-mbon/documentation). All metadata and access to all image data generated by ARMS-MBON to date can be found on GitHub (https://github.com/arms-mbon/data_workspace). All genetic raw data generated by ARMS-MBON to date can be accessed on the European Nucleotide Archive (ENA) through the accession numbers provided via the GitHub repository (https://github.com/arms-mbon/data_workspace) and under the umbrella study PRJEB72316 (https://www.ebi.ac.uk/ena/browser/view/PRJEB72316). Metadata, access to image data and accession numbers for genetic data specifically for the data set presented in this manuscript are provided in the Supporting Information and are also available on the respective GitHub repository (https://github.com/arms-mbon/data_release_001). Output files of the PEMA pipeline used as the basis for OBIS submissions and for exploratory analysis as shown in this study are accessible at https://github.com/arms-mbon/analysis_release_001. Files part of OBIS submissions are available at https://github.com/arms-mbon/data_release_001. The taxonomically classified occurrence data for the COI (https://obis.org/dataset/066f002f-58d5-4687-bdb8-b39cdaef0c2b), 18S (https://obis.org/dataset/0ada9b0c-14f5-4247-881e-9f6f62b2c165) and ITS (https://obis.org/dataset/ddab58b2-0072-41b8-afc5-ac10d937247f) samples analysed in this study are published in OBIS under the respective links. The metadata records for the COI (https://doi.org/10.14284/620), 18S (https://doi.org/10.14284/705), and ITS (https://doi.org/10.14284/706) OBIS submissions are found on IMIS under the respective DOIs. All code and input files used for exploratory analysis presented in this manuscript are available or linked to on https://github.com/arms-mbon/code_release_001.

Benefit-Sharing Statement

Benefits Generated: ARMS-MBON represents a large-scale research collaboration with scientists from across Europe and beyond. All network partners of the observatories mentioned in this manuscript provided genetic samples and are included as co-authors. All continuously generated raw and processed data from this network is shared with the public and scientific community (see above). Our research addresses the urgent need for large-scale and long-term monitoring of marine biotic communities through extensive collaborative efforts.

References

Ahyong, S., C. B. Boyko, N. Bailly, et al. 2024. "World Register of Marine Species (WoRMS)." WoRMS Editorial Board.

Alberdi, A., O. Aizpurua, M. T. P. Gilbert, and K. Bohmann. 2018. "Scrutinizing Key Steps for Reliable Metabarcoding of Environmental Samples." *Methods in Ecology and Evolution* 9, no. 1: 134–147. https://doi.org/10.1111/2041-210X.12849.

Andersen, J. H., B. S. Halpern, S. Korpinen, C. Murray, and J. Reker. 2015. "Baltic Sea Biodiversity Status vs. Cumulative Human Pressures." *Estuarine, Coastal and Shelf Science* 161: 88–92. https://doi.org/10.1016/j.ecss.2015.05.002.

Andrews, S. 2010. "FastQC: A Quality Control Tool for High Throughput Sequence Data." http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Antich, A., C. Palacín, J. Zarcero, S. O. Wangensteen, and X. Turon. 2023. "Metabarcoding Reveals High-Resolution Biogeographical and Metaphylogeographical Patterns Through Marine Barriers." *Journal of Biogeography* 50: 515–527. https://doi.org/10.1111/jbi.14548.

Auguie, B. 2019. "egg: Extensions for "ggplot2": Custom Geom, Custom Themes, Plot Alignment, Labelled Panels, Symmetric Scales, and Fixed Panel Size."

Bankevich, A., S. Nurk, D. Antipov, et al. 2012. "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing." *Journal of Computational Biology* 19, no. 5: 455–477. https://doi.org/10.1089/CMB.2012.0021.

Beijbom, O., P. J. Edmunds, C. Roelfsema, et al. 2015. "Towards Automated Annotation of Benthic Survey Images: Variability of Human Experts and Operational Modes of Automation." *PLoS One* 10, no. 7: e0130312. https://doi.org/10.1371/JOURNAL.PONE.0130312.

Bell, J. J., and B. Okamura. 2005. "Low Genetic Diversity in a Marine Nature Reserve: Re-Evaluating Diversity Criteria in Reserve Design." *Proceedings of the Royal Society B: Biological Sciences* 272, no. 1567: 1067–1074. https://doi.org/10.1098/RSPB.2005.3051.

Benestan, L., N. Loiseau, P. E. Guérin, et al. 2023. "Contrasting Influence of Seascape, Space and Marine Reserves on Genomic Variation in Multiple Species." *Ecography* 2023, no. 1: e06127. https://doi.org/10.1111/ECOG.06127.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30, no. 15: 2114–2120. https://doi.org/10.1093/BIOINFORMATICS/BTU170.

Borja, A., G. Chust, and I. Muxika. 2019. "Forever Young: The Successful Story of a Marine Biotic Index." *Advances in Marine Biology* 82: 93–127. https://doi.org/10.1016/BS.AMB.2019.05.001.

Borja, A., J. Franco, and V. Pérez. 2000. "A Marine Biotic Index to Establish the Ecological Quality of Soft-Bottom Benthos Within European Estuarine and Coastal Environments." *Marine Pollution Bulletin* 40, no. 12: 1100–1114. https://doi.org/10.1016/S0025-326X(00)00061-8.

Bourlat, S. J., A. Borja, J. Gilbert, et al. 2013. "Genomics in Marine Monitoring: New Opportunities for Assessing Marine Health Status." *Marine Pollution Bulletin* 74, no. 1: 19–31. https://doi.org/10.1016/J.MARPOLBUL.2013.05.042.

Boyer, F., C. Mercier, A. Bonin, Y. Le Bras, P. Taberlet, and E. Coissac. 2016. "Obitools: A Unix-Inspired Software Package for DNA Metabarcoding." *Molecular Ecology Resources* 16, no. 1: 176–182. https://doi.org/10.1111/1755-0998.12428.

Brandt, M. I., B. Trouche, L. Quintric, et al. 2021. "Bioinformatic Pipelines Combining Denoising and Clustering Tools Allow for More Comprehensive Prokaryotic and Eukaryotic Metabarcoding." *Molecular Ecology Resources* 21, no. 6: 1904–1921. https://doi.org/10.1111/1755-0998.13398.

Bustos-Baez, S., and C. Frid. 2003. "Using Indicator Species to Assess the State of Macrobenthic Communities." *Advances in Polychaete Research* 496: 299–309. https://doi.org/10.1007/978-94-017-0655-1_28.

Cahill, A. E., J. K. Pearman, A. Borja, et al. 2018. "A Comparative Analysis of Metabarcoding and Morphology-Based Identification of Benthic Communities Across Different Regional Seas." *Ecology and Evolution* 8, no. 17: 8908–8920. https://doi.org/10.1002/ECE3.4283.

Cameron, E. S., P. J. Schmidt, B. J. M. Tremblay, M. B. Emelko, and K. M. Müller. 2021. "Enhancing Diversity Analysis by Repeatedly Rarefying Next Generation Sequencing Data Describing Microbial Communities." *Scientific Reports* 11, no. 1: 1–13. https://doi.org/10.1038/s41598-021-01636-1.

Capurso, G., B. Carroll, and K. A. Stewart. 2023. "Transforming Marine Monitoring: Using eDNA Metabarcoding to Improve the Monitoring of the Mediterranean Marine Protected Areas Network." *Marine Policy* 156: 105807. https://doi.org/10.1016/J.MARPOL.2023.105807.

Carvalho, S., E. Aylagas, R. Villalobos, Y. Kattan, M. Berumen, and J. K. Pearman. 2019. "Beyond the Visual: Using Metabarcoding to Characterize the Hidden Reef Cryptobiome." *Proceedings of the Royal Society B: Biological Sciences* 286, no. 1896: 20182697. https://doi.org/10.1098/rspb.2018.2697.

Conway, J. R., A. Lex, and N. Gehlenborg. 2017. "UpSetR: An R Package for the Visualization of Intersecting Sets and Their Properties." *Bioinformatics* 33, no. 18: 2938–2940. https://doi.org/10.1093/bioinformatics/btx364.

Costello, M. J., S. Ahyong, R. Bieler, et al. 2024. "World Register of Introduced Marine Species (WRiMS). WoRMS Editorial Board." https://www.marinespecies.org/introduced.

Costello, M. J., S. Dekeyzer, B. Galil, et al. 2021. "Introducing the World Register of Introduced Marine Species (WRiMS)." *Management of Biological Invasions* 12, no. 4: 792–811. https://doi.org/10.3391/mbi.2021.12.4.02.

da Silva, L. P., V. A. Mata, P. B. Lopes, et al. 2019. "Advancing the Integration of Multi-Marker Metabarcoding Data in Dietary Analysis of Trophic Generalists." *Molecular Ecology Resources* 19, no. 6: 1420–1432. https://doi.org/10.1111/1755-0998.13060.

Danovaro, R., L. Carugati, M. Berzano, et al. 2016. "Implementing and Innovating Marine Monitoring Approaches for Assessing Marine Environmental Status." *Frontiers in Marine Science* 3: 214343. https://doi.org/10.3389/fmars.2016.00213.

Dauvin, J. C. 2005. "Expertise in Coastal Zone Environmental Impact Assessments." *Marine Pollution Bulletin* 50, no. 1: 107–110. https://doi.org/10.1016/J.MARPOLBUL.2004.11.013.

Devillers, R., R. L. Pressey, A. Grech, et al. 2015. "Reinventing Residual Reserves in the Sea: Are We Favouring Ease of Establishment Over Need for Protection?" *Aquatic Conservation: Marine and Freshwater Ecosystems* 25, no. 4: 480–504. https://doi.org/10.1002/AQC.2445.

Dragulescu, A., and C. Arendt. 2020. "xlsx: Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files."

Duarte, S., P. E. Vieira, B. R. Leite, M. A. L. Teixeira, J. M. Neto, and F. O. Costa. 2023. "Macrozoobenthos Monitoring in Portuguese Transitional Waters in the Scope of the Water Framework Directive Using Morphology and DNA Metabarcoding." *Estuarine, Coastal and Shelf Science* 281: 108207. https://doi.org/10.1016/J.ECSS.2022.108207.

Ducklow, H. W., S. C. Doney, and D. K. Steinberg. 2009. "Contributions of Long-Term Research and Time-Series Observations to Marine Ecology and Biogeochemistry." *Annual Review of Marine Science* 1: 279–302. https://doi.org/10.1146/ANNUREV.MARINE.010908.163801.

Edgar, G. J. 2011. "Does the Global Network of Marine Protected Areas Provide an Adequate Safety Net for Marine Biodiversity?" *Aquatic Conservation: Marine and Freshwater Ecosystems* 21, no. 4: 313–316. https://doi.org/10.1002/AQC.1187.

Edgar, G. J., R. D. Stuart-Smith, T. J. Willis, et al. 2014. "Global Conservation Outcomes Depend on Marine Protected Areas With Five Key Features." *Nature* 506, no. 7487: 216–220. https://doi.org/10.1038/nature13022.

Elliott, M. 2014. "Integrated Marine Science and Management: Wading Through the Morass." *Marine Pollution Bulletin* 86, no. 1–2: 1–4. https://doi.org/10.1016/J.MARPOLBUL.2014.07.026.

Evans, N. T., Y. Li, M. A. Renshaw, et al. 2017. "Fish Community Assessment With eDNA Metabarcoding: Effects of Sampling Design and Bioinformatic Filtering." *Canadian Journal of Fisheries and*

Aquatic Sciences 74, no. 9: 1362–1374. https://doi.org/10.1139/CJFAS-2016-0306.

Fediajevaite, J., V. Priestley, R. Arnold, and V. Savolainen. 2021. "Meta-Analysis Shows That Environmental DNA Outperforms Traditional Surveys, but Warrants Better Reporting Standards." *Ecology and Evolution* 11, no. 9: 4803–4815. https://doi.org/10.1002/ECE3.7382.

Fields, J. B., and N. J. Silbiger. 2022. "Foundation Species Loss Alters Multiple Ecosystem Functions Within Temperate Tidepool Communities." *Marine Ecology Progress Series* 683: 1–19. https://doi.org/10.3354/MEPS13978.

Gamfeldt, L., J. S. Lefcheck, J. E. K. Byrnes, B. J. Cardinale, J. E. Duffy, and J. N. Griffin. 2015. "Marine Biodiversity and Ecosystem Functioning: What's Known and What's Next?" *Oikos* 124, no. 3: 252–265. https://doi.org/10.1111/OIK.01549.

Gibson, J., S. Shokralla, T. M. Porter, et al. 2014. "Simultaneous Assessment of the Macrobiome and Microbiome in a Bulk Sample of Tropical Arthropods Through DNA Metasystematics." *Proceedings of the National Academy of Sciences of the United States of America* 111, no. 22: 8007–8012. https://doi.org/10.1073/PNAS.1406468111.

Gielings, R., M. Fais, D. Fontaneto, et al. 2021. "DNA Metabarcoding Methods for the Study of Marine Benthic Meiofauna: A Review." *Frontiers in Marine Science* 8: 1429. https://doi.org/10.3389/FMARS.2021.730063.

Gotelli, N. J., and R. K. Colwell. 2011. "Estimating Species Richness." In *Biological Diversity: Frontiers in Measurement and Assessment*, edited by A. E. Magurran and B. J. McGill, 39–54. Oxford: Oxford University Press.

Grey, E. K., L. Bernatchez, P. Cassey, et al. 2018. "Effects of Sampling Effort on Biodiversity Patterns Estimated From Environmental DNA Metabarcoding Surveys." *Scientific Reports* 8, no. 1: 1–10. https://doi.org/10.1038/s41598-018-27048-2.

Guillou, L., D. Bachar, S. Audic, et al. 2013. "The Protist Ribosomal Reference Database (PR2): A Catalog of Unicellular Eukaryote Small Sub-Unit rRNA Sequences With Curated Taxonomy." *Nucleic Acids Research* 41, no. D1: D597–D604. https://doi.org/10.1093/NAR/GKS1160.

Hakimzadeh, A., A. Abdala Asbun, D. Albanese, et al. 2023. "A Pile of Pipelines: An Overview of the Bioinformatics Software for Metabarcoding Data Analyses." *Molecular Ecology Resources* 24: e13847. https://doi.org/10.1111/1755-0998.13847.

Hallam, J., E. L. Clare, J. I. Jones, and J. J. Day. 2023. "High Frequency Environmental DNA Metabarcoding Provides Rapid and Effective Monitoring of Fish Community Dynamics." *Environmental DNA* 5, no. 6: 1623–1640. https://doi.org/10.1002/EDN3.486.

HELCOM. 2013. "HELCOM Core Indicators: Final Report of the HELCOM CORESET Project." *Balt. Sea Environ. Proc.* No. 136.

Heymans, J. J., A. Bundy, V. Christensen, et al. 2020. "The Ocean Decade: A True Ecosystem Modeling Challenge." *Frontiers in Marine Science* 7: 554573. https://doi.org/10.3389/FMARS.2020.554573.

Ip, Y. C. A., J. J. M. Chang, R. M. Oh, et al. 2022. "Seq′ and ARMS Shall Find: DNA (Meta)barcoding of Autonomous Reef Monitoring Structures Across the Tree of Life Uncovers Hidden Cryptobiome of Tropical Urban Coral Reefs." *Molecular Ecology* 32, no. 23: 6223–6242. https://doi.org/10.1111/MEC.16568.

Jacquemont, J., R. Blasiak, C. Le Cam, M. Le Gouellec, and J. Claudet. 2022. "Ocean Conservation Boosts Climate Change Mitigation and Adaptation." *One Earth* 5, no. 10: 1126–1138. https://doi.org/10.1016/J.ONEEAR.2022.09.002.

Kaikkonen, L., R. Venesjärvi, H. Nygård, and S. Kuikka. 2018. "Assessing the Impacts of Seabed Mineral Extraction in the Deep Sea and Coastal Marine Environments: Current Methods and Recommendations for Environmental Risk Assessment." *Marine Pollution Bulletin* 135: 1183–1197. https://doi.org/10.1016/J.MARPOLBUL.2018.08.055.

Kassambara, A. 2020. "ggpubr: "ggplot2" Based Publication Ready Plots. R Package Version 0.4.0."

Kissling, W. D., J. A. Ahumada, A. Bowser, et al. 2018. "Building Essential Biodiversity Variables (EBVs) of Species Distribution and Abundance at a Global Scale." *Biological Reviews* 93, no. 1: 600–625. https://doi.org/10.1111/BRV.12359.

Lanzé, A., S. L. Jørgensen, D. H. Huson, et al. 2012. "CREST-Classification Resources for Environmental Sequence Tags." *PLoS One* 7, no. 11: e49334. https://doi.org/10.1371/journal.pone.0049334.

Leite, B. R., S. Duarte, J. S. Troncoso, and F. O. Costa. 2023. "Artificial Seaweed Substrates Complement ARMS in DNA Metabarcoding-Based Monitoring of Temperate Coastal Macrozoobenthos." *Diversity* 15, no. 5: 657. https://doi.org/10.3390/D15050657/S1.

Leite, B. R., P. E. Vieira, J. S. Troncoso, and F. O. Costa. 2021. "Comparing Species Detection Success Between Molecular Markers in DNA Metabarcoding of Coastal Macroinvertebrates." *Metabarcoding and Metagenomics* 5: e70063. https://doi.org/10.3897/MBMG.5.70063.

Leray, M., and N. Knowlton. 2015. "DNA Barcoding and Metabarcoding of Standardized Samples Reveal Patterns of Marine Benthic Diversity." *Proceedings of the National Academy of Sciences of the United States of America* 112, no. 7: 2076–2081. https://doi.org/10.1073/PNAS.1424997112.

Lotze, H. K., H. Guest, J. O'Leary, A. Tuda, and D. Wallace. 2018. "Public Perceptions of Marine Threats and Protection From Around the World." *Ocean and Coastal Management* 152: 14–22. https://doi.org/10.1016/J.OCECOAMAN.2017.11.004.

Machida, R. J., M. Leray, S. L. Ho, and N. Knowlton. 2017. "Metazoan Mitochondrial Gene Sequence Reference Datasets for Taxonomic Assignment of Environmental Samples." *Scientific Data* 4, no. 1: 1–7. https://doi.org/10.1038/sdata.2017.27.

Mahé, F., T. Rognes, C. Quince, C. de Vargas, and M. Dunthorn. 2015. "Swarmv2: Highly-Scalable and High-Resolution Amplicon Clustering." *PeerJ* 2015, no. 12: e1420. https://doi.org/10.7717/PEERJ.1420.

Martaeng, R., M. Obst, and P. Kuklinski. 2023. "Phylogeographic Study Using Autonomous Reef Monitoring Structures Indicates Fast Range Expansion of the Invasive Bryozoan Juxtacribrilina Mutabilis." *Hydrobiologia* 850, no. 19: 4115–4126. https://doi.org/10.1007/s10750-023-05184-9.

Masella, A. P., A. K. Bartram, J. M. Truszkowski, D. G. Brown, and J. D. Neufeld. 2012. "PANDAseq: Paired-End Assembler for Illumina Sequences." *BMC Bioinformatics* 13, no. 1: 1–7. https://doi.org/10.1186/1471-2105-13-31.

Mathieu, C., S. M. Hermans, G. Lear, T. R. Buckley, K. C. Lee, and H. L. Buckley. 2020. "A Systematic Review of Sources of Variability and Uncertainty in eDNA Data for Environmental Monitoring." *Frontiers in Ecology and Evolution* 8: 1–14. https://doi.org/10.3389/FEVO.2020.00135.

McMurdie, P. J., and S. Holmes. 2013. "Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data." *PLoS One* 8, no. 4: e61217. https://doi.org/10.1371/JOURNAL.PONE.0061217.

McMurdie, P. J., and S. Holmes. 2014. "Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible." *PLoS Computational Biology* 10, no. 4: e1003531. https://doi.org/10.1371/JOURNAL.PCBI.1003531.

Meyer, R., N. Davies, K. J. Pitz, et al. 2022. "The Founding Charter of the Omic Biodiversity Observation Network (Omic BON)." *GigaScience* 12: 1–6. https://doi.org/10.1093/GIGASCIENCE/GIAD068.

Micheli, F., B. S. Halpern, S. Walbridge, et al. 2013. "Cumulative Human Impacts on Mediterranean and Black Sea Marine Ecosystems: Assessing Current Pressures and Opportunities." *PLoS One* 8, no. 12: e79889. https://doi.org/10.1371/journal.pone.0079889.

Mugnai, F., F. Costantini, A. Chenuil, M. Leduc, J. M. G. Ortega, and E. Meglécz. 2023. "Be Positive: Customized Reference Databases and New, Local Barcodes Balance False Taxonomic Assignments in Metabarcoding Studies." *PeerJ* 11: e14616. https://doi.org/10.7717/PEERJ.14616.

Muller-Karger, F. E., P. Miloslavich, N. J. Bax, et al. 2018. "Advancing Marine Biological Observations and Data Requirements of the Complementary Essential Ocean Variables (EOVs) and Essential Biodiversity Variables (EBVs) Frameworks." *Frontiers in Marine Science* 5: 373058. https://doi.org/10.3389/FMARS.2018.00211.

Narwani, A., M. Reyes, A. L. Pereira, et al. 2019. "Interactive Effects of Foundation Species on Ecosystem Functioning and Stability in Response to Disturbance." *Proceedings of the Royal Society B: Biological Sciences* 286, no. 1913: 20191857. https://doi.org/10.1098/RSPB.2019.1857.

Nikolenko, S. I., A. I. Korobeynikov, and M. A. Alekseyev. 2013. "BayesHammer: Bayesian Clustering for Error Correction in Single-Cell Sequencing." *BMC Genomics* 14, no. 1: 1–11. https://doi.org/10.1186/1471-2164-14-S1-S7/TABLES/3.

Nilsson, R. H., K.-H. Larsson, A. F. S. Taylor, et al. 2018. "The UNITE Database for Molecular Identification of Fungi: Handling Dark Taxa and Parallel Taxonomic Classifications." *Nucleic Acids Research* 47: 259–264. https://doi.org/10.1093/nar/gky1022.

Obst, M., K. Exter, A. L. Allcock, et al. 2020. "A Marine Biodiversity Observation Network for Genetic Monitoring of Hard-Bottom Communities (ARMS-MBON)." *Frontiers in Marine Science* 7: 1031. https://doi.org/10.3389/FMARS.2020.572680.

Obst, M., S. Vicario, K. Lundin, et al. 2018. "Marine Long-Term Biodiversity Assessment Suggests Loss of Rare Species in the Skagerrak and Kattegat Region." *Marine Biodiversity* 48, no. 4: 2165–2176. https://doi.org/10.1007/s12526-017-0749-5.

Oksanen, J., G. L. Simpson, F. G. Blanchet, et al. 2023. "vegan: Community Ecology Package."

Ooms, J. 2024. "writexl: Export Data Frames to Excel "xlsx" Forma."

Pagès, H., P. Aboyoun, R. Gentleman, and S. DebRoy. 2020. "Biostrings: Efficient Manipulation of Biological Strings".

Pearman, J. K., E. Aylagas, C. R. Voolstra, H. Anlauf, R. Villalobos, and S. Carvalho. 2019. "Disentangling the Complex Microbial Community of Coral Reefs Using Standardized Autonomous Reef Monitoring Structures (ARMS)." *Molecular Ecology* 28, no. 15: 3496–3507. https://doi.org/10.1111/mec.15167.

Pearman, J. K., G. Chust, E. Aylagas, et al. 2020. "Pan-Regional Marine Benthic Cryptobiome Biodiversity Patterns Revealed by Metabarcoding Autonomous Reef Monitoring Structures." *Molecular Ecology* 29, no. 24: 4882–4897. https://doi.org/10.1111/MEC.15692.

Pearman, J. K., M. Leray, R. Villalobos, et al. 2018. "Cross-Shelf Investigation of Coral Reef Cryptic Benthic Organisms Reveals Diversity Patterns of the Hidden Majority." *Scientific Reports* 8, no. 1: 1–17. https://doi.org/10.1038/s41598-018-26332-5.

Peroni, S., S. Soiland-Reyes, P. Sefton, et al. 2022. "Packaging Research Artefacts With RO-Crate." *Data Science* 5, no. 2: 97–138. https://doi.org/10.3233/DS-210053.

Porter, T. M., and M. Hajibabaei. 2018. "Over 2.5 Million COI Sequences in GenBank and Growing." *PLoS One* 13, no. 9: e0200177. https://doi.org/10.1371/JOURNAL.PONE.0200177.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ransome, E., J. B. Geller, M. Timmers, et al. 2017. "The Importance of Standardization for Biodiversity Comparisons: A Case Study Using Autonomous Reef Monitoring Structures (ARMS) and Metabarcoding to Measure Cryptic Diversity on Mo'orea Coral Reefs, French Polynesia." *PLoS One* 12, no. 4: e0175066. https://doi.org/10.1371/journal.pone.0175066.

Rognes, T., T. Flouri, B. Nichols, C. Quince, and F. Mahé. 2016. "VSEARCH: A Versatile Open Source Tool for Metagenomics." *PeerJ* 2016, no. 10: e2584. https://doi.org/10.7717/PEERJ.2584.

RStudio Team. 2022. *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, PBC.

Santi, I., O. Beluche, M. Beraud, et al. 2023. "European Marine Omics Biodiversity Observation Network: A Strategic Outline for the Implementation of Omics Approaches in Ocean Observation." *Frontiers in Marine Science* 10: 1118120. https://doi.org/10.3389/FMARS.2023.1118120.

Santi, I., R. Casotti, T. Comtet, et al. 2022. "European Marine Omics Biodiversity Observation Network (EMO BON) Handbook (Version 1.0). 63." https://doi.org/10.25607/OBP-1653.

Schauberger, P., and A. Walker. 2021. "openxlsx: Read, Write and Edit xlsx Files."

Shenoy, A. R. 2021. "grafify: An R Package for Easy Graphs, ANOVAs and Post-Hoc Comparisons." https://doi.org/10.5281/zenodo.5136508.

Shirazi, S., R. S. Meyer, and B. Shapiro. 2021. "Revisiting the Effect of PCR Replication and Sequencing Depth on Biodiversity Metrics in Environmental DNA Metabarcoding." *Ecology and Evolution* 11, no. 22: 15766–15779. https://doi.org/10.1002/ECE3.8239.

Smale, D. A., T. Wernberg, E. C. J. Oliver, et al. 2019. "Marine Heatwaves Threaten Global Biodiversity and the Provision of Ecosystem Services." *Nature Climate Change* 9, no. 4: 306–312. https://doi.org/10.1038/s41558-019-0412-1.

Staehr, P. A. U., K. Dahl, H. Buur, et al. 2022. "Environmental DNA Monitoring of Biodiversity Hotspots in Danish Marine Waters." *Frontiers in Marine Science* 8: 800474. https://doi.org/10.3389/FMARS.2021.800474.

Stevenson, S. L., S. N. C. Woolley, J. Barnett, and P. Dunstan. 2020. "Testing the Presence of Marine Protected Areas Against Their Ability to Reduce Pressures on Biodiversity." *Conservation Biology* 34, no. 3: 622–631. https://doi.org/10.1111/COBI.13429.

Stuer-Lauridsen, F., G. Drillet, F. T. Hansen, and J. Saunders. 2018. "Same Risk Area: An Area-Based Approach for the Management of Bio-Invasion Risks From Ships' Ballast Water." *Marine Policy* 97: 147–155. https://doi.org/10.1016/J.MARPOL.2018.05.009.

Turon, X., A. Antich, C. Palacín, K. Præbel, and O. S. Wangensteen. 2020. "From Metabarcoding to Metaphylogeography: Separating the Wheat From the Chaff." *Ecological Applications* 30, no. 2: e02036. https://doi.org/10.1002/EAP.2036.

Villalobos, R., E. Aylagas, J. K. Pearman, et al. 2022. "Inter-Annual Variability Patterns of Reef Cryptobiota in the Central Red Sea Across a Shelf Gradient." *Scientific Reports* 12, no. 1: 1–17. https://doi.org/10.1038/s41598-022-21304-2.

Virta, L., J. Soininen, and A. Norkko. 2021. "Biodiversity Loss Threatens the Current Functional Similarity of Beta Diversity in Benthic Diatom Communities." *Microbial Ecology* 81, no. 2: 293–303. https://doi.org/10.1007/S00248-020-01576-9.

Wang, Q., G. M. Garrity, J. M. Tiedje, and J. R. Cole. 2007. "Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences Into the New Bacterial Taxonomy." *Applied and Environmental Microbiology* 73, no. 16: 5261–5267. https://doi.org/10.1128/AEM.00062-07.

Wesselmann, M., E. T. Apostolaki, and A. Anton. 2024. "Species Range Shifts, Biological Invasions and Ocean Warming." *Marine Ecology Progress Series* 728: 81–83. https://doi.org/10.3354/MEPS14544.

Wickham, H. 2011. "The Split-Apply-Combine Strategy for Data Analysis." *Journal of Statistical Software* 40, no. 1: 1–29.

Wickham, H., M. Averick, J. Bryan, et al. 2019. "Welcome to the Tidyverse." *Journal of Open Source Software* 4, no. 43: 1686. https://doi.org/10.21105/joss.01686.

Wickham, H., T. L. Pedersen, and D. Seidel. 2023. "scales: Scale Functions for Visualization."

Witalis, B., A. Iglikowska, M. Ronowicz, and P. Kukliński. 2021. "Biodiversity of Epifauna in the Ports of Southern Baltic Sea Revealed by Study of Recruitment and Succession on Artificial Panels." *Estuarine, Coastal and Shelf Science* 249: 107107. https://doi.org/10.1016/J.ECSS.2020.107107.

Wood, S., and F. Scheipl. 2020. "gamm4: Generalized Additive Mixed Models Using 'mgcv' and 'lme4'." https://CRAN.R-project.org/package=gamm4.

Worm, B., E. B. Barbier, N. Beaumont, et al. 2006. "Impacts of Biodiversity Loss on Ocean Ecosystem Services." *Science* 314, no. 5800: 787–790. https://doi.org/10.1126/SCIENCE.1132294.

Yuan, D., A. Ahamed, J. Burgin, et al. 2023. "The European Nucleotide Archive in 2023." *Nucleic Acids Research* 52, no. D1: D92–D97. https://doi.org/10.1093/nar/gkad1067.

Zafeiropoulos, H., A. Gioti, S. Ninidakis, et al. 2021. "0s and 1s in Marine Molecular Research: A Regional HPC Perspective." *GigaScience* 10, no. 8: 1–12. https://doi.org/10.1093/GIGASCIENCE/GIAB053.

Zafeiropoulos, H., H. Q. Viet, K. Vasileiadou, et al. 2020. "PEMA: A Flexible Pipeline for Environmental DNA Metabarcoding Analysis of the 16S/18S Ribosomal RNA, ITS, and COI Marker Genes." *GigaScience* 9, no. 3: 1–12. https://doi.org/10.1093/GIGASCIENCE/GIAA022.

Zhang, M., Y. Zou, S. Xiao, and J. Hou. 2023. "Environmental DNA Metabarcoding Serves as a Promising Method for Aquatic Species Monitoring and Management: A Review Focused on Its Workflow, Applications, Challenges and Prospects." *Marine Pollution Bulletin* 194: 115430. https://doi.org/10.1016/J.MARPOLBUL.2023.115430.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.