



HAL
open science

A long-term ecological research data set from the marine genetic monitoring programme ARMS-MBON 2018-2020

Nauras Daraghmeh, Katrina Exter, Justine Pagnier, Piotr Balazy, Ibon Cancio, Giorgos Chatzigeorgiou, Eva Chatzinikolaou, Maciej Chelchowski, Nathan Alexis Mitchell Christmas, Thierry Comtet, et al.

► To cite this version:

Nauras Daraghmeh, Katrina Exter, Justine Pagnier, Piotr Balazy, Ibon Cancio, et al.. A long-term ecological research data set from the marine genetic monitoring programme ARMS-MBON 2018-2020. 2024. hal-04734855

HAL Id: hal-04734855

<https://hal.science/hal-04734855v1>

Preprint submitted on 14 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License

A long-term ecological research data set from the marine genetic monitoring programme ARMS-MBON 2018-2020

Authors

Nauras Daraghmeh^{1,2}, Katrina Exter³, Justine Pagnier^{2,4}, Piotr Bałazy⁵, Ibon Cancio⁶, Giorgos Chatzigeorgiou⁷, Eva Chatzinikolaou⁷, Maciej Chełchowski⁵, Nathan Alexis Mitchell Christmas⁸, Thierry Comtet⁹, Thanos Dailianis⁷, Klaas Deneudt³, Oihane Diaz de Cerio⁶, Markos Digenis^{7,10}, Vasilis Gerovasileiou^{7,10}, José González¹¹, Laura Kauppi¹², Jon Bent Kristoffersen⁷, Piotr Kukliński⁵, Rafał Lasota¹³, Liraz Levy¹⁴, Magdalena Małachowicz⁵, Borut Mavrič¹⁵, Jonas Mortelmans³, Estefania Paredes¹¹, Anita Poćwierz-Kotus⁵, Henning Reiss¹⁶, Ioulia Santi^{7,17}, Georgia Sarafidou⁷, Grigorios Skouradakis⁷, Jostein Solbakken¹², Peter A.U. Staehr¹⁸, Javier Tajadura⁶, Jakob Thyrring¹⁸, Jesus S. Troncoso¹¹, Emmanouela Vernadou⁷, Frederique Viard¹⁹, Haris Zafeiropoulos^{7,20}, Małgorzata Zbawicka⁵, Christina Pavloudi^{7,17*}, Matthias Obst^{1,2*+}

* Shared Senior Authorship: Christina Pavloudi, Matthias Obst

+ Corresponding Author: Matthias Obst, matthias.obst@marine.gu.se

Affiliations

¹ Department of Marine Sciences, University of Gothenburg, Box 461, 405 30 Gothenburg, Sweden

² Gothenburg Global Biodiversity Centre (GGBC), University of Gothenburg, Box 463, 405 30 Gothenburg, Sweden

³ Flanders Marine Institute (VLIZ), Jacobsenstraat 1, 8400 Oostende, West-Vlaanderen, Belgium

⁴ LifeWatch ERIC, Sector II-III Plaza de España 41071, Seville, Spain

⁵ Institute of Oceanology of the Polish Academy of Sciences (IOPAN), Ul. Powstancow Warszawy 55, 81-712 Sopot, Poland

⁶ Plentzia Marine Station (PiE-UPV/EHU), University of the Basque Country, Areatza Pasealekua, 48620 Plentzia-Bizkaia, Basque Country, Spain

⁷ Institute of Marine Biology, Biotechnology and Aquaculture (IMBBC), Hellenic Center for Marine Research (HCMR), 71500, Heraklion, Crete, Greece

⁸ Marine Biological Association of the UK, Citadel Hill, Plymouth, PL12PB, UK

⁹ Sorbonne Université, CNRS, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France

¹⁰ Department of Environment, Faculty of Environment, Ionian University, M. Minotou-Giannopoulou str. Panagoula, 29100 Zakynthos, Corfu, Greece

¹¹ Centro de Investigación Mariña, Universidade de Vigo, Estación de Ciencias Mariñas de Toralla, Illa de Toralla s/n, Vigo 36331 (Pontevedra), Spain

¹² Tvärminne Zoological Station, University of Helsinki, J.A. Palménin tie 260, 10900 Hanko, Finland

¹³ University of Gdansk, Institute of Oceanography, Department of Marine Ecosystems Functioning, Al. M. Piłsudskiego 46, 81-372 Gdynia, Poland

¹⁴ The Interuniversity Institute of Marine Sciences in Eilat, Eilat 88103, Israel

¹⁵ National Institute of Biology, Marine Biology Station Piran, Fornace 41, 6330 Piran, Slovenia

¹⁶ Faculty of Biosciences and Aquaculture, Nord University, Postboks 1490, 8049 Bodø, Norway

¹⁷ EMBRC-ERIC Headquarters, Sorbonne Université, 4 Place Jussieu, 75005 Paris, France

¹⁸ Department of Bioscience, Aarhus University, Frederiksborgvej 399, 4000 Roskilde, Denmark

¹⁹ ISEM, Université de Montpellier, CNRS, EPHE, IRD, 34095 Montpellier cedex 05, France

²⁰ KU Leuven, Department of Microbiology, Immunology and Transplantation, Rega Institute for Medical Research, Laboratory of Molecular Bacteriology, Herestraat 49 - box 1030, 3000 Leuven, Belgium

ABSTRACT

Molecular methods such as DNA/eDNA metabarcoding have emerged as useful tools to document 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 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 ~5,500 amplicon sequence variants and ~1,800 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 three to six 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 programmes and long-term ecological research and encourage the adoption of our ARMS-MBON protocols.

KEY WORDS

genetic monitoring, European Marine Omics Biodiversity Observation Network (EMO BON), invasive species, COI, 18S rRNA, ITS

1 INTRODUCTION

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

21 To make biological monitoring programs more effective, several methods have been proposed
22 recently (Danovaro et al., 2016), one of them being DNA metabarcoding. In principle, DNA-
23 based techniques are capable of identifying biological communities at high temporal and spatial
24 frequency, and with fine taxonomic resolution (Staehr et al., 2022). But despite the frequent
25 application of metabarcoding in marine ecological research, only few marine surveillance
26 programs have so far implemented genetic protocols for long-term monitoring (Hallam et al.,
27 2023; Mathieu et al., 2020). Though it has been shown that DNA metabarcoding may enhance
28 or even outperform traditional approaches (Capurso et al., 2023; Fediajevaite et al., 2021), its
29 lack of application in large-scale spatio-temporal contexts may be due to its novelty as a
30 biomonitoring tool (Hallam et al., 2023) or methodological biases associated with it (Capurso
31 et al., 2023; Mathieu et al., 2020; Zhang et al., 2023). Furthermore, data management can
32 become a demanding task when trying to link raw data with different layers of downstream
33 analytical information across a multitude of sample localities and project collaborators,

34 potentially discouraging the use of genomic tools in large-scale monitoring initiatives. To
35 accelerate the use of DNA metabarcoding for long-term and geographically widespread
36 biomonitoring, it is therefore crucial for the scientific community to develop best practices for
37 sample collection, processing, and analysis as well as procedures and standards for data
38 management and quality control (Santi et al., 2022).

39 Recently, Obst et al. (2020) established the Autonomous Reef Monitoring Structures Marine
40 Biodiversity Observation Network (ARMS-MBON) for the genetic monitoring of hard-bottom
41 communities, as part of the marine thematic node of the Group on Earth Observations
42 Biodiversity Observation Network (GEO BON). This programme deploys autonomous reef
43 monitoring structures (ARMS) in ports, marinas, and nature reserves along the European
44 coastline, as well as at a number of locations in polar regions and the Red Sea, to capture and
45 analyse the genetic diversity across latitudes, oceans, and benthic habitats. The use of ARMS
46 - simple-to-produce units of stacked PVC plates mimicking the three-dimensional complexity
47 of benthic habitats - enables standardised and non-destructive sampling of complex benthic
48 communities (Leray & Knowlton, 2015; Ransome et al., 2017). The network maintains, to date,
49 25 observatories each deploying ARMS on a regular basis at one to seven sampling sites.
50 ARMS-MBON has now become part of the European Marine Omics Biodiversity Observation
51 Network (EMO BON) (Santi et al., 2023), a larger European initiative coordinated by the
52 European Marine Biological Resource Centre (EMBRC) for the observation of genomic
53 biodiversity. EMO BON includes marine biodiversity observatories from the Arctic to the Red
54 Sea and investigates biological communities sampled from the water column and the benthic
55 substrate using shared protocols, data, and metadata standards. Furthermore, EMO BON
56 contributes to the UN Decade of Ocean Science for Sustainable Development by participating
57 in the global Ocean Biomolecular Observing Network (OBON) (Meyer et al., 2022)
58 programme, and, together with other connected projects, it aims for a worldwide coordinated
59 biomolecular observation system (Santi et al., 2023).

60 Here, we present the first data release from the initial ARMS-MBON sampling campaign (i.e.,
61 from all ARMS units deployed in 2018 and 2019 and retrieved between 2018 and 2020),
62 comprising genetic samples and image data collected by observatories from the Gulf of Finland
63 in the Baltic Sea to the Spanish Atlantic coast, and from the northern Red Sea to the Svalbard
64 archipelago. The data set adheres to the FAIR (Findable, Accessible, Interoperable, and Re-
65 usable) Data Principles, and contains references to material samples, metadata descriptions,
66 images, sequence data, derived taxonomic observations, and documentation of the analytical
67 process. In this paper, we provide some brief exploration of the data set and give examples of

68 potential applications, but the ARMS-MBON data can be used by any interested party for
69 comparative studies and to support DNA-based monitoring of marine biodiversity over large
70 spatio-temporal scales.

71

72 **MATERIALS AND METHODS**

73 **Field sampling and laboratory protocols**

74 The original descriptions of observatory design, field work, and sample processing procedures
75 as well as instructions for biobanking and data management have been expanded from Obst et
76 al. (2020). Field work and sample processing for samples of this first data release followed the
77 official ARMS-MBON Handbook v2.0, which is in parts based on the initial protocols
78 developed by the Global ARMS Program ([https://naturalhistory.si.edu/research/global-arms-](https://naturalhistory.si.edu/research/global-arms-program)
79 [program](https://naturalhistory.si.edu/research/global-arms-program)). We used molecular protocols available under the ARMS-MBON Molecular
80 Standard Operating Procedure v1.0 (MSOP). The Handbook and MSOP can be accessed on
81 the ARMS-MBON GitHub repository (see Supplementary Table S1 for links). Preserved
82 replicates of the material samples are stored and catalogued at the ARMS-MBON network
83 partner and EMBRC member institution, the Hellenic Centre for Marine Research (HCMR),
84 Crete, Greece, and at the individual ARMS-MBON network partner institutions.

85

86 **Access and Benefit Sharing (ABS)**

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

92

93 **General description of the data set for ARMS-MBON data release 001**

94 The first ARMS-MBON data release package comprises data of 56 individual ARMS units
95 which were deployed, retrieved and processed by ARMS-MBON network partner institutions
96 at 15 observatories across 12 countries along the European coastline, as well as the northern
97 Red Sea and Svalbard (Figure 1A). This covers the entirety of ARMS-MBON sampling events
98 for which deployments took place in 2018 and/or 2019. Deployment periods ranged from 37
99 to 649 days during the period of April 2018 and December 2020, with the majority being
100 deployed for around three months to approximately one year (Figure 1B). The 56 ARMS units
101 represent 56 individual sampling events and 42 unique Unit IDs, meaning where (i) upon

102 retrieval a new ARMS unit was deployed at the same spot for a consecutive deployment period,
103 or (ii) for a certain period multiple units remained submerged simultaneously at the exact same
104 spot but were either deployed or retrieved at different time points (i.e., to test the effect of
105 deployment duration), this resulted in sampling events with the same Unit ID but separate
106 Event IDs. General information on the observatories and ARMS deployments (e.g.,
107 coordinates, habitat type, deployment depth) and on the sampling events and their resulting
108 material samples (e.g., date of deployment and retrieval, material sample IDs, preservative
109 used) can be found on the GitHub repository (see Supplementary Table S1 for links) and in
110 Supplementary Files 1 and 2, respectively.

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

125

126 **ARMS plate image data**

127 Images of each plate were recorded post-recovery and after disassembling the units, to visually
128 document benthic communities according to the instructions in the ARMS-MBON Handbook
129 v2.0. The image collection contains photographs of both sides of the settlement plates as well
130 as close-ups of individual specimens or colonies. Download links for all images of this data
131 release package are available via the GitHub repository (see Supplementary Table S1 for link
132 and Supplementary File 1).

133

134

135

136 **Amplicon sequencing**

137 Molecular work for samples of the first data release package was carried out as detailed in the
138 ARMS-MBON MSOP (see Supplementary Table S1 for link) for DNA metabarcoding of the
139 eukaryotic mitochondrial and nuclear marker genes cytochrome *c* oxidase subunit I (COI), 18S
140 rRNA (18S) and internal transcribed spacer (ITS) region. Note that ITS was only targeted
141 during the initial phase of ARMS-MBON and use of this marker gene has been discontinued.
142 All samples were subjected to amplicon sequencing by the network partner institute HCMR
143 who also published all sequence data. Raw sequence files of the first data release package are
144 available for download at the European Nucleotide Archive (ENA) (Yuan et al., 2023) through
145 the accession numbers provided via the GitHub repository (see Supplementary Table S1 for
146 link and Supplementary File 1 for accession numbers) and under the umbrella study
147 PRJEB72316 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB72316>). The accession
148 numbers for the sequencing negative control samples are also included there. Information on
149 the demultiplexing procedures applied on sequencing output data is provided as well, denoting
150 if sequence reads deposited on ENA contain primer sequences or not; see Supplementary Table
151 S1 for link and Supplementary File 1). Overall, the sequencing data comprises 200 samples
152 (190 material samples plus ten samples which were re-sequenced due to initially low read
153 yield) for the COI and 18S marker genes, and 111 samples for ITS, plus four to nine negative
154 control samples per marker gene (Supplementary File 1).

155

156 **Bioinformatics processing of raw sequence data**

157 To deposit taxonomic observations derived from the COI, 18S and ITS marker genes in the
158 European Ocean Biogeographic Information System (EurOBIS), sequence data were processed
159 with the Pipeline for Environmental DNA Metabarcoding Analysis, PEMA v.2.1.4
160 (Zafeiropoulos et al., 2020). See Supplementary Text S1 details on bioinformatics processing.
161 Given that different parameter settings can lead to rather different outcomes (Brandt et al.,
162 2021; Zafeiropoulos et al., 2020), a fixed set of parameters was used for each marker gene and
163 sequence data were processed separately for each sequencing run. All parameter files used for
164 each PEMA run are available at the dedicated repository (see Supplementary Table S1 for link).
165 For the 18S marker gene, sequences were clustered into operational taxonomic units (OTUs)
166 using the VSEARCH v2.9.1 algorithm (Rognes et al., 2016) with a threshold of 0.97, while for
167 ITS and COI, clustering was performed with Swarm v2 (Mahé et al., 2015), applying a
168 threshold of $d = 10$ to infer amplicon sequence variants (ASVs). Note that PEMA initially
169 defined the result of Swarm processing as inferred ASVs, i.e., sequences which differ by one

170 or more nucleotides, which is now corrected to swarm-clusters (Hakimzadeh et al., 2023). We
171 here use PEMA's initial terminology for consistency reasons. Taxonomy was assigned to 18S
172 OTUs and ITS ASVs with the CREST LCAClassifier v3.0 (Lanzé et al., 2012), using the PR2
173 v.4.13.0 (Guillou et al., 2013) and Unite v7.2 (Nilsson et al., 2018) databases, respectively. For
174 COI sequences, taxonomic annotation was performed using the RDP classifier (Wang et al.,
175 2007) with the MIDORI database v2.0 (Machida et al., 2017). Singletons and OTUs/ASVs
176 unclassified at domain level were removed. Abundances of OTUs/ASVs present in negative
177 control samples were adjusted. All bioinformatics analyses were supported by the High
178 Performance Computing system of the Institute of Marine Biology, Biotechnology and
179 Aquaculture of HCMR (Crete, Greece) (Zafeiropoulos et al., 2021).

180

181 **Data management, EurOBIS submission, and the ARMS-MBON GitHub space**

182 The data management proceeded in multiple stages: (i) collecting the (meta)data from the field
183 scientists and from the processing of genetic data, (ii) harvesting and organising these in a
184 public space, and quality controlling the metadata, (iii) adding semantics, organising the data
185 in Research Object Crates (RO-Crates; <https://www.researchobject.org/ro-crate/>) (Peroni et al.,
186 2022) and adding provenance, and (iv) submitting the data to EurOBIS. These steps were
187 performed as follows:

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

196 (ii) The ARMS-MBON GitHub space (see Supplementary Table S1 for links) was chosen for
197 all subsequent steps of the data management. Particular reasons for choosing GitHub were: its
198 ease of access (within and external of the team); the ability to track changes, to implement
199 custom workflow support through actions, and to build and host web-page-like "landing
200 pages". The Google sheets and data on the Google Drive were harvested into respective GitHub
201 repositories, as was the metadata from PlutoF (as Java Script Object Notation, JSON). All
202 harvested data were quality controlled (i.e., consistency checks, correction of mistakes) and

203 combined into a sampling event, an image, an omics (i.e., sequencing data), and an observatory
204 spreadsheet (comma separated-values, CSV); each data spreadsheet is accompanied by a
205 metadata spreadsheet that defines and adds semantics to the column names. Due to their file
206 size, the images themselves were not uploaded to GitHub, but their PlutoF (open access) URLs
207 were, and these are included together with image metadata in a CSV file (an improved image
208 database will be developed to better archive, annotate, and serve the ARMS-MBON image
209 data). The PEMA input and output files were uploaded to the *processing_batch1* repository.
210 Due to their large combined size, the fasta files were instead uploaded to the Marine Data
211 Archive (MDA) (<https://www.mda.vliz.be>) and their (open access) URLs included in the
212 relevant GitHub folder (see Supplementary Table S1 for respective links to all here described
213 data products).

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

218 (iv) The ARMS-MBON data release package from the first sampling campaign (i.e., all
219 taxonomically classified occurrences from the marker gene analysis with a minimum of two
220 sequence reads) has been formatted into Darwin Core Archives (DwCA) for each marker gene
221 for submission to EurOBIS. These data have been submitted using the relatively new DNA
222 extension, and EurOBIS has been working on the inclusion of these data in their database,
223 hence, these data will be published in EurOBIS in the autumn of 2024. A metadata record for
224 these data has been created in the Integrated Marine Information System (IMIS; see
225 Supplementary Table S1 for links) and all current and future links to the data are accessible
226 there. The DwCAs, as well as all associated observatory and sampling event metadata, ENA
227 accession numbers of amplicon sequencing data and links to image data specifically for the
228 first data release package are accessible via the *data_release_001* repository on GitHub (see
229 Supplementary Table S1 for link). The metadata files there represent subsets of the
230 corresponding quality-controlled *combined* metadata files which contain information on all
231 observatories, sampling events, image data and genetic data of ARMS-MBON to date (i.e., not
232 only for the sampling events described here). The *analysis_release_001* repository on GitHub
233 comprises all relevant bioinformatics analysis data (i.e., parameter files and outputs of PEMA
234 processing) associated with the EurOBIS submission. This repository is merely a duplicate of
235 the *processing_batch1* repository. All code used for exploratory data analysis described below

236 can be found in the *code_release_001* repository. See Supplementary Table S1 for links to the
237 repositories.

238 **Exploration of sequencing data**

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

248 We determined the number and/or abundance of unique phyla, species identified, ASVs/OTUs
249 classified to species level, and species shared between the data sets of the three marker genes.
250 Where reference databases did not provide correct phylum level classification, we manually
251 added these. In terms of alpha diversity, we assessed the observed ASV/OTU richness and the
252 number of identified species across observatories, as well as frequency distributions of these
253 two parameters (i.e., re-occurrence of ASVs/OTUs and species identified across observatories).
254 We also determined the influence of sampling effort on diversity variables. Here, we computed
255 Spearman's correlation of sequencing depth and ARMS deployment duration versus
256 ASV/OTU richness and the number of species identified in each sample. Furthermore, we
257 computed Spearman's correlation of the number of ARMS units deployed and the number of
258 samples included in the analysis post-curation versus ASV/OTU richness and the number of
259 species identified at each observatory. Where the correlation was statistically significant ($p <$
260 0.05) and moderate to strong (Spearman's $\rho > 0.5$), we performed analysis of simple linear
261 regression to model the relationship between sampling effort predictor variables and dependent
262 diversity variables.

263 In order to test the application potential of the derived species observation data, we performed
264 a scan against reference checklists for ecological key species. Species occurrences with at least
265 two sequence reads were scanned against the following databases: i) AZTI's Marine Biotic
266 Index (AMBI) (Borja et al., 2000, 2019) for species very sensitive to disturbance; ii) the World
267 Register of Introduced Marine Species (WRiMS) (Costello et al., 2021, 2024) for species with
268 alien status at the place of observation; and iii) the Red Lists of the International Union for

269 Conservation of Nature (IUCN) and Baltic Marine Environment Protection Commission
270 (Helsinki Commission, HELCOM) for species registered as Near Threatened, Vulnerable,
271 Endangered or Critically Endangered. The AMBI and IUCN/HELCOM information were
272 obtained using the World Register of Marine Species' (WoRMS) (Ahyong et al., 2024) REST
273 services (<https://www.marinespecies.org/rest/>), while the WRiMS checks can be replicated
274 using the Jupyter notebook on <https://www.github.com/vliz-be-opsci/lw-iji-invasive-checker>.
275 Occurrences of red-listed species were confirmed by scanning against known distribution in
276 WoRMS and corrected where necessary.

277 Samples were tested for differences in alpha diversity among locations with varying degrees
278 of anthropogenic influence (i.e., industrial, semi-industrial, low human influence (LHI), and
279 protected; see Supplementary File S1). For statistical comparison, samples with less than 5,000
280 reads were removed and the remaining samples rarefied to an equal sequencing depth. Samples
281 classified as “industrial” were grouped into one category (“industrial/semi-industrial”) with
282 samples classified as “semi-industrial”. Mean and standard deviation (SD) of the two alpha
283 diversity measures were calculated for samples of each influence type. Data was subsequently
284 checked for normality and log(1+x)-transformation applied where necessary. Subsequently,
285 unidirectional analysis of variance (ANOVA) was performed to test for differences between
286 influence types, with post-hoc Tukey’s test for pairwise comparisons. Where normality could
287 not be achieved through transformation, non-parametric Kruskal-Wallis rank sum test was
288 applied. All code used for exploratory analysis can be found at the dedicated GitHub repository
289 (see Supplementary Table S1). Analyses and data visualisation were performed in R v4.1.0 (R
290 Core Team, 2021) via RStudio v2022.07.1 (RStudio Team, 2022) using packages of the
291 tidyverse v1.3.1 collection (Wickham et al., 2019) and the packages Biostrings v2.60.2 (Pagès
292 et al., 2020), phyloseq v1.36.0 (McMurdie & Holmes, 2013), vegan v2.6.2 (Oksanen et al.,
293 2023), ggpubr v0.4.0 (Kassambara, 2020), grafify v4.0 (Shenoy, 2021), plyr v1.8.7 (Wickham,
294 2011), scales v1.3.0 (Wickham et al., 2023), egg v0.4.5 (Auguie, 2019), UpSetR v1.4.0
295 (Conway et al., 2017), xlsx v0.6.5 (Dragulescu & Arendt, 2020), writexl v1.5.0 (Ooms, 2024),
296 and openxlsx v4.2.5 (Schauberger & Walker, 2021).

297

298 **RESULTS**

299 **Overall description of the data set**

300 Out of the 200 sequenced sample units (i.e., 190 material samples representing biological and
301 technical replicates, plus 10 re-sequenced samples) from the first ARMS-MBON sampling

302 campaign, 195 and 200 were successfully sequenced (i.e., samples containing sequences that
303 are deposited on ENA) using the COI and 18S marker genes, respectively, while 111 samples
304 (out of 111 samples included for this gene) were successfully sequenced using the ITS marker
305 (Table 1). Sequence processing of these samples using PEMA resulted in 54,641, 11,294 and
306 10,280 unique ASVs/OTUs for COI, 18S and ITS, respectively (Table 2). After further curation
307 and filtering (i.e., negative-control-correction and removal of unclassified sequences), 189
308 samples with 51,782 ASVs and 9,596 OTUs remained for the COI and 18S data sets,
309 respectively. For ITS, 42 samples with 508 ASVs remained (note that for ITS, ASVs could be
310 inferred from two sequencing runs only with the pipeline and parameters applied here, i.e., for
311 runs of July 2019 and April 2021). This corresponded to 1,567,301 sequence reads for COI,
312 3,910,167 sequence reads for 18S and 49,782 sequence reads for ITS (Table 2). All occurrence
313 records with a minimum of two sequence reads of these classified sequences will be accessible
314 through EurOBIS, as explained above. From the COI, 18S and ITS data sets, 18,402, 21,482,
315 and 493 occurrences will be deposited, respectively (Table 2, see Supplementary Table S1 for
316 links to IMIS metadata and DOI records).

317

318 **Taxonomic profiles of three marker gene data sets**

319 To explore the ecological properties of the processed sequencing data, we further curated the
320 data set as described above (i.e., applying classification confidence threshold for COI,
321 exclusion of some samples and removal of potentially spurious sequences, etc.). This curated
322 data set – an illustration of how a potential user-curated data set may look – comprised a subset
323 of 162 samples for COI and 18S, respectively, and 34 samples for ITS (Table 3). These samples
324 contained 40,363, 8,700 and 372 ASVs/OTUs with 1,223,460, 2,875,245 and 24,978 sequence
325 reads for the COI, 18S and ITS data sets, respectively. Application of these three marker genes
326 led to the recovery of 65 eukaryotic phyla, of which 38, 57 and 9 were present in the COI (at
327 the confidence threshold of 0.8 applied here), 18S and ITS data, respectively (Table 3, see
328 Supplementary File 2 for further details). In regards to relative read abundance, the COI data
329 set classified to phylum level was dominated by metazoan phyla (i.e., eight out of the ten most
330 abundant phyla) – mainly Arthropoda, Annelida, Chordata, Bryozoa and Cnidaria – while
331 almost half of all reads (~44%) belonged to sequences unclassified at phylum level (Figure 2A,
332 Table 3 and Supplementary File 2). For 18S, metazoans of the phyla Chordata, Mollusca,
333 Arthropoda, Annelida and Cnidaria dominated the data set, but more non-metazoan phyla were
334 among the most abundant taxa compared to the COI data (Figure 2C and Supplementary File
335 2). The majority of 18S reads could be classified to at least phylum level (~97%; Table 3). As

336 expected, mainly fungal phyla were recovered using the ITS marker gene (Cnidaria was the
337 only non-fungal phylum in the ITS data) and Ascomycota made up more than half of all reads
338 in this gene's data set (Figure 2E and Supplementary File 2). Approximately two-thirds (~61%)
339 of ITS reads were classified to the phylum level (Table 3).

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

361

362 **Genetic and species diversity at ARMS-MBON observatories**

363 Alpha diversity of genetic units (i.e., ASVs and OTUs) and identified species varied between
364 observatories. Between 60 (Eilat - Israel) and 246 species (Koster - Sweden) from COI data (at
365 the confidence threshold applied here) and 2 (i.e., Getxo - Spain) and 53 species (Koster -
366 Sweden) from 18S data were identified at the observatories (Figure 3; Supplementary Table
367 S2). Richness of ASVs/OTUs ranged between 365 (Læsø - Denmark) and 5,472 (TZS -
368 Finland) for COI and between 253 (Getxo - Spain) and 2,136 (Koster - Sweden) for 18S (Figure
369 3; Supplementary Table S2). While the high 18S OTU diversity at Koster on the Swedish west

370 coast aligned with its high number of identified species, the high COI ASV richness at the TZS
371 observatory on the Gulf of Finland contrasted its relatively low number of identified species
372 ($n=70$) compared to observatories with lower genetic diversity. In comparison, the two ARMS
373 from the Norwegian coast at Bodø captured more than twice as many species ($n=168$) as the
374 TZS observatory but showed less than half the genetic diversity (2,275 ASVs) for COI. It
375 should be noted that alpha diversity was driven by differences in sampling effort and
376 sequencing depth to some extent (see below).

377 Analysis of frequency distributions across observatories showed that ~94% ($n = 38,090$) and
378 ~4% ($n = 1,593$) of COI ASVs occurred at one or two observatories only, respectively (Figure
379 4A). Frequency distribution was less skewed for 18S data, with ~67% ($n = 5,794$) and ~18%
380 ($n = 1,565$) of OTUs found at one or two observatories only, respectively (Figure 4B).
381 Approximately 400 species (~53%) identified from the COI data occurred at one observatory
382 only, while 160 species (~21%) occurred at two observatories (Figure 4C). A relatively low
383 number of species identified from the COI data occurred at 10 or more observatories ($n = 13$,
384 ~2%; Figure 4C). Half of all species ($n = 68$) identified using the 18S marker gene were found
385 at one observatory only, and 17 species (~13%) occurred at only two observatories (Figure
386 4D). Two species identified from 18S data appeared at ten or more observatories (Figure 4D).
387

388 **Influence of sampling effort and deployment duration on diversity measures**

389 During the first ARMS-MBON sampling campaign, observatories deployed one (GulfOfPiran
390 - Slovenia) to eight (Roscoff - France) ARMS units, for periods ranging from 37
391 (BelgiumCoast - Belgium; UnitID AZFPin) to 649 days (Eilat - Israel; UnitID Katza2) (Figure
392 1B; see Supplementary File 1 for further details). After data curation and filtering, this resulted
393 in a data set of three (GulfOfPiran - Slovenia) to 24 samples (Roscoff - France) with at least
394 one sequence read for each observatory for both the COI and 18S marker genes (Figure 3).
395 Given differences in sequencing depth and quality, not all of the three fraction samples
396 remained for each ARMS unit post-curation.

397 Computation of correlation indicated no significant linear association between deployment
398 duration and the number of species identified in each sample for both marker genes (Figure
399 5A, B; see Supplementary Table S3 for detailed results of correlation analysis). This was also
400 the case for deployment duration vs. ASV/OTU richness (Figure 5A, B). For both marker
401 genes, sequencing depth (i.e., the number of sequence reads) significantly drove the number of
402 species identified and the ASV/OTU richness observed in each sample, with moderate to strong
403 correlation ($p < 0.001$ and Spearman's $\rho(160) \geq 0.60$ for all associations; Figure 5C, D).

404 Analysis of linear regression indicated that sequencing depth accounted for around 23% to 41%
405 of variation in ASV/OTU richness and the number of species identified in each sample for both
406 marker genes. However, residual standard errors suggested the regression models did not fit
407 the data accurately, although fit was more accurate for 18S compared to COI data (see
408 Supplementary Table S4 for detailed results of regression analysis).
409 For the COI marker gene, only the number of species identified at each observatory was
410 significantly and moderately to strongly correlated to both the number of ARMS units deployed
411 and the number of samples included in the analysis ($p = 0.018$, Spearman's $\rho(13) = 0.60$; and
412 $p = 0.021$, Spearman's $\rho(13) \geq 0.59$, respectively; see Supplementary Figure S2A, C and
413 Supplementary Table S3). The observed COI ASV richness did not show a significant linear
414 relationship with these two sampling effort parameters (Figure S2A, C and Supplementary
415 Table S3). In the case of 18S, both the observed OTU richness and the number of species
416 identified at each observatory significantly and moderately correlated with both the number of
417 ARMS units deployed and the resulting number of samples included in the analysis ($p \leq 0.012$
418 and Spearman's $\rho(13) \geq 0.64$ for all associations; see Supplementary Figure S2B,D and
419 Supplementary Table S3). Analysis of linear regression indicated that the number of ARMS
420 units deployed and the number of samples included accounted for around 37% to 51% of
421 variation in the number of species identified at each observatory for both marker genes. In the
422 case of 18S, the number of ARMS units deployed and the number of samples included were
423 both responsible for ~52% of variation in the OTU richness observed at the observatories.
424 Given relatively high variation in the dependent variables and therefore considerable residual
425 standard errors, observatory-wise regression models did not fit our data accurately as described
426 above for sample-wise models (Supplementary Table S4).

427

428 **Identification of ecological key species**

429 The scan against databases for sensitive, non-indigenous, and red-listed species resulted in
430 observations of species in all three categories across the observatories (Figure 6). Overall, we
431 observed 88 species registered in AMBI as sensitive to disturbance, 32 species listed as “alien”
432 in WRiMS at the location of occurrence, and 4 species registered as near threatened, vulnerable,
433 endangered or critically endangered across all observatories (Supplementary Table S5 and
434 Supplementary File 3). The observatory at Koster (Sweden) detected the highest number of
435 sensitive species ($n = 37$), while Limfjord (Denmark) displayed the highest number of non-
436 indigenous species (NIS). Red-listed species were only detected at the Plymouth (UK) (*Mya*

437 *truncata*), Koster (Sweden) (*Cliona celata* and *Echinus esculentus*) and Roscoff (France)
438 (*Cliona celata* and *Nucula nucleus*) observatories.

439

440 **Alpha diversity across locations with varying degrees of anthropogenic influence**

441 The number of identified species differed significantly among habitats with varying degrees of
442 anthropogenic influence based on COI data (low human influence: 26 ± 8 , industrial/semi-
443 industrial: 29 ± 13 , protected: 38 ± 15 ; ANOVA: $F_{2,60} = 5.043$; $p = 0.009$) (Figure 7A and
444 Supplementary Table S6) but based on 18S data they did not (industrial/semi-industrial: 5 ± 5 ,
445 low human influence: 6 ± 4 , protected: 7 ± 3 ; Kruskal-Wallis $\chi^2 = 2.543$, $p = 0.280$) (Figure 7B
446 and Supplementary Table S6). Mean ASV richness ranged from 216 ± 90 (low human
447 influence) to 260 ± 200 (industrial/semi-industrial) and 260 ± 173 (protected) for COI data
448 (Figure 7A) and no statistically significant difference was detected among habitat types given
449 high within-group variability (ANOVA of $\log(1+x)$ -transformed data: $F_{2,60} = 0.241$; $p = 0.787$;
450 see Supplementary Table S6 for details of statistical tests). The same was observed for 18S
451 data (ANOVA: $F_{2,60} = 1.906$; $p = 0.157$), with OTU richness ranging from 123 ± 78
452 (industrial/semi-industrial) to 136 ± 53 (low human influence) and 165 ± 84 (protected) (Figure
453 7B and Supplementary Table S6).

454

455 **DISCUSSION**

456 Here we present the data set of raw and processed amplicon sequencing data from the first
457 sampling campaign (2018-2020) of ARMS-MBON. To the best of our knowledge, this
458 campaign represents the most geographically widespread sampling initiative using ARMS to
459 date. Raw sequencing (and image) data are open-access and, thus, they can be subjected to re-
460 processing by all interested users according to their needs. In addition, we processed
461 sequencing data using a dedicated pipeline, and all taxonomic occurrences will be published
462 on EurOBIS. This processed data set was further filtered and curated, representing a potential
463 use case for ecological analysis or taxonomic screening. Analysis of this curated data set shows
464 that ARMS are able to capture eukaryotic taxa belonging to more than 60 phyla, while
465 composition of the hard-bottom benthic cryptofauna in terms of dominant phyla was
466 comparable with those of earlier studies using ARMS (e.g., Carvalho et al., 2019; Ip et al.,
467 2022; Pearman et al., 2018, 2020). Our data set also allows for comparisons of communities at
468 all taxonomic levels (i.e., including genus, class, order level identifications) which are typically
469 applied in local and regional community studies (Staehr et al., 2022). Although we do not
470 present any prokaryotic data here, the preserved physical samples of ARMS-MBON sampling

471 campaigns can also be used for comparative studies of non-eukaryotic microbial communities
472 in the future (Ip et al., 2022; Pearman et al., 2019). Finally, the catalogue of ARMS plate images
473 collected can become a valuable data resource in the future as it will allow for analysis of
474 benthic community composition and growth dynamics with application of advanced image
475 classification methods (Beijbom et al., 2015).

476

477 **Recovered taxa are distinct across marker genes and observatories**

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

491

492 **The need to account for differences in sampling effort**

493 Deployment periods of ARMS units ranged from 37 to 649 days, with the majority being
494 deployed for around three months to approximately one year. Our tests showed that in most
495 cases deployment duration did neither have a significant effect on observed species nor on
496 genetic diversity. The weak negative correlation between deployment duration and sequence
497 diversity in the COI data set can most likely be attributed to the low sequencing depth in some
498 samples of ARMS deployed for longer periods (i.e., units of observatories from Svalbard -
499 Norway and Eilat - Israel and one unit of Crete - Greece). However, we want to stress the fact
500 that a conclusion on the effect of deployment duration can ultimately only be achieved with
501 dedicated testing, i.e., by comparing locations with both short- and long-term deployments and
502 equal sequencing depth. Previous studies have found contrasting results regarding the influence
503 of deployment duration of sampling units. Using artificial substrate units (ASUs), Cahill et al.
504 (2018) did not observe a significant effect of deployment duration on the number of specimens

505 recovered and argued that overall recruitment patterns are predominantly driven by ecological
506 and biogeographic conditions. In their comparative study, Leite et al. (2023) deployed ARMS
507 and artificial seaweed monitoring systems (ASMS) for six, nine, and twelve months and found
508 that community composition changed over time given ecological succession. The authors
509 indicate that maximum diversity can be recovered with sampling units deployed for less than
510 twelve months, which had also been shown by earlier studies (Leite et al., 2021). In contrast to
511 deployment duration, we showed that recovery of sequence and taxonomic diversity
512 significantly depend on sequencing depth. This is a well-known problem in microbial amplicon
513 sequencing (Cameron et al., 2021; McMurdie & Holmes, 2014) and has also been shown in
514 eukaryotic DNA metabarcoding studies (Alberdi et al., 2018; Grey et al., 2018; Shirazi et al.,
515 2021). As expected, the number of ARMS units deployed and the number of samples analysed
516 for each observatory significantly drove the recovered alpha diversity, as well. Large variation
517 in ASV richness across COI samples is likely the reason that the effect of sampling effort was
518 not significant for this specific test case. Differences in sampling effort are known to drive
519 observed alpha diversity measures in traditional ecological surveys (Gotelli & Colwell, 2011)
520 and, in particular, in metabarcoding studies (Evans et al., 2017; Grey et al., 2018). Hence,
521 analytical tools need to be applied to account for them.

522 Given that deployment duration did not influence alpha diversity, we suggest deploying ARMS
523 for at least three to six months during the season of most substantial growth (i.e., during spring
524 to summer/fall or summer to fall on the respective hemisphere) to capture a representative
525 snapshot of benthic communities and to use human, time, and material resources as efficiently
526 as possible. We also urge for deployments of at least three ARMS units per site (i.e., with a
527 distance of around 10 m from each other) as biological replicates to obtain a comprehensive
528 representation of surrounding communities and to improve statistical power for comparative
529 analysis. Given differences in sampling effort and in sequencing depth across samples – the
530 latter being inherent to any amplicon sequencing study – we recommend that users apply
531 statistical tools when using ARMS-MBON data to account for those during ecological analysis.

532

533 **ARMS data enable the study of large-scale biodiversity patterns**

534 We found indications that alpha diversity of the hard-bottom benthos within marine protected
535 areas (MPAs) is higher than in locations with more intense anthropogenic influence. Mean
536 sequence and species richness was highest in samples from ARMS deployed in protected areas
537 for both marker genes, but given high within-group variation across samples, this was only
538 statistically significant for species richness of one marker gene. This can be an indication of

539 the positive effect of protection measures (Edgar, 2011; Edgar et al., 2014) but may also be
540 related to the fact that MPAs are typically established in areas with already low anthropogenic
541 influence where political implementation costs are low (Devillers et al., 2015; Stevenson et al.,
542 2020). The fact that we found higher species diversity but not higher genetic diversity in MPAs
543 remains unexplained at this point. A possible reason for this pattern could be that high genetic
544 diversity reflects recruitment which may be high even in habitats with anthropogenic pressures,
545 while MPAs support species beyond the initial recruitment stage. Further, propagules and
546 larvae of opportunistic taxa, including NIS, may be particularly ubiquitous in harbours and
547 marinas, but less so in stable ecosystems with high conservation status such as MPAs. In
548 addition, small, heterogeneous but isolated MPAs may support high species but low genetic
549 diversity due to reduced connectivity (Bell & Okamura, 2005) and seascape and spatial factors
550 may have a higher influence on genetic diversity than protection status (Benestan et al., 2023).
551 Given the continuous sampling on a large spatio-temporal scale, data from the forthcoming
552 sampling campaigns of ARMS-MBON can help unravel differences in diversity more clearly.
553 In any case, ARMS data can be used to document biodiversity trends in benthic habitats in
554 relation to human activities and eventually contribute to several essential biodiversity variables
555 (EBVs) in the future (Kissling et al., 2018). Genetic and taxonomic data from ARMS-MBON
556 may also be combined with open-access remote sensing data to link diversity patterns and
557 taxonomic occurrences to environmental parameters (Pearman et al., 2019, 2020).

558

559 **ARMS are sensitive to indicator, non-indigenous, and threatened species**

560 Our results show that ARMS can detect ecological indicator species sensitive to disturbance,
561 which are typically used in environmental impact assessments (EIAs) (Bustos-Baez & Frid,
562 2003; Dauvin, 2005), environmental risk assessments (ERAs) (Kaikkonen et al., 2018), and
563 national / regional monitoring programmes (e.g., HELCOM, 2013) to assess the health status
564 of ecosystems. ARMS-MBON data are likely to improve such assessments since they provide
565 information on species as well as genetic diversity, which can be analysed in relation to
566 anthropogenic pressures. As such, ARMS should be deployed continuously in sites such as
567 ports, marinas, wind farms, or aquaculture facilities in order to assess impacts of human
568 activities on marine biodiversity (Witalis et al., 2021). We also detected a (low) number of red-
569 listed species, although the fact that this was the case only for observatories with the highest
570 sampling effort underlines the rarity of such taxa and the need for continuous and considerable
571 sampling to track and monitor them.

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

588

589 **How to build a successful, data-producing, long-term genetic observatory network**

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

- 591 ● Standardisation: it is vital to ensure standard protocols (collecting and processing of
592 material, managing the data) are published, understood, and followed so that the
593 resulting data is comparable over the space and time of the observatories.
- 594 ● Constant engagement: it is necessary to engage the hearts and minds of the participants
595 for the exciting (e.g., field work) as well as the challenging and tedious (e.g., data
596 management) parts, as otherwise the resulting (meta)data are insufficient to fulfil the
597 project's potential.
- 598 ● Strength in numbers: it is necessary to have multidisciplinary core of experts taking
599 control of the different parts of the project (e.g., sampling, sequencing, bioinformatics,
600 data management, analysis), as no one scientist can do all of these with the quality and
601 timeliness to allow for regular and trustworthy data releases.

602

603

604

605 **CONCLUSION**

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

623

ACKNOWLEDGEMENTS

The ARMS-MBON network was established under the infrastructure program ASSEMBLE Plus (grant no. 730984) and is currently financed by the European Marine Biological Resource Centre (EMBRC) under the program European Marine Omics Biodiversity Observation Network (EMO BON). Data publication and analysis is funded by the projects DTO-bioflow (grant agreement no. 101112823) and MARCO BOLO (grant agreement no. 101082021). Funding for individual ARMS observatories was provided by the INTERREG project GEANS (North Sea Program of the European Regional Development Fund of the European Union), the Swedish Agency for Marine and Water Management (grant no. 3181-2019), the Flanders LifeWatch contribution (Research Foundation Flanders grant I000819N), and the Aquanis 2.0 project (FONDATION Total). Data management and computational resources were provided by the Swedish Biodiversity Data Infrastructure (grant no. 2019-00242) and by IMBBC (Institute of Marine Biology, Biotechnology and Aquaculture) of the HCMR (Hellenic Centre of Marine Research). Funding for establishing the IMBBC HPC has been received by the MARBIGEN (EU Regpot) project, LifeWatchGreece RI and the CMBR (Center for the study

and sustainable exploitation of Marine Biological Resources) RI. Guiding documents to obtain ABS clearance for access to genetic resources were developed under the projects INTERREG EBB (EAPA_501/2016) and H2020 EOSC-Life (grant no. 824087). JT was supported by the Carlsberg Foundation (case no. CF21-0564) and the Aage V. Jensen Foundation.

BIBLIOGRAPHY

- Ahyong, S., Boyko, C. B., Bailly, N., Bernot, J., Bieler, R., Brandão, S. N., Daly, M., De Grave, S., Gofas, S., Hernandez, F., Hughes, L., Neubauer, T. A., Paulay, G., Boydens, B., Decock, W., Dekeyzer, S., Vandepitte, L., Vanhoorne, B., Adlard, R., ... Zullini, A. (2024). *World Register of Marine Species (WoRMS)*. WoRMS Editorial Board. <https://www.marinespecies.org>
- Alberdi, A., Aizpurua, O., Gilbert, M. T. P., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution*, 9(1), 134–147. <https://doi.org/10.1111/2041-210X.12849>
- Andersen, J. H., Halpern, B. S., Korpinen, S., Murray, C., & Reker, J. (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>
- Antich, A., Palacín, C., Zarcero, J., Owen, J., Wangensteen, S., & Turon, X. (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*. <https://cran.r-project.org/package=egg>
- Beijbom, O., Edmunds, P. J., Roelfsema, C., Smith, J., Kline, D. I., Neal, B. P., Dunlap, M. J., Moriarty, V., Fan, T. Y., Tan, C. J., Chan, S., Treibitz, T., Gamst, A., Mitchell, B. G., & Kriegman, D. (2015). Towards Automated Annotation of Benthic Survey Images: Variability of Human Experts and Operational Modes of Automation. *PLOS ONE*, 10(7), e0130312. <https://doi.org/10.1371/JOURNAL.PONE.0130312>
- Bell, J. J., & Okamura, B. (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(1567), 1067–1074. <https://doi.org/10.1098/RSPB.2005.3051>

- Benestan, L., Loiseau, N., Guérin, P. E., Pérez-Ruzafa, A., Forcada, A., Arcas, E., Lenfant, P., Mallol, S., Goñi, R., Velez, L., Mouillot, D., Puebla, O., & Manel, S. (2023). Contrasting influence of seascape, space and marine reserves on genomic variation in multiple species. *Ecography*, 2023(1), e06127. <https://doi.org/10.1111/ECOG.06127>
- Borja, A., Chust, G., & Muxika, I. (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., Franco, J., & Pérez, V. (2000). A Marine Biotic Index to Establish the Ecological Quality of Soft-Bottom Benthos Within European Estuarine and Coastal Environments. *Marine Pollution Bulletin*, 40(12), 1100–1114. [https://doi.org/10.1016/S0025-326X\(00\)00061-8](https://doi.org/10.1016/S0025-326X(00)00061-8)
- Bourlat, S. J., Borja, A., Gilbert, J., Taylor, M. I., Davies, N., Weisberg, S. B., Griffith, J. F., Lettieri, T., Field, D., Benzie, J., Glöckner, F. O., Rodríguez-Ezpeleta, N., Faith, D. P., Bean, T. P., & Obst, M. (2013). Genomics in marine monitoring: New opportunities for assessing marine health status. *Marine Pollution Bulletin*, 74(1), 19–31. <https://doi.org/10.1016/J.MARPOLBUL.2013.05.042>
- Brandt, M. I., Trouche, B., Quintric, L., Günther, B., Wincker, P., Poulain, J., & Arnaud-Haond, S. (2021). Bioinformatic pipelines combining denoising and clustering tools allow for more comprehensive prokaryotic and eukaryotic metabarcoding. *Molecular Ecology Resources*, 21(6), 1904–1921. <https://doi.org/10.1111/1755-0998.13398>
- Bustos-Baez, S., & Frid, C. (2003). Using indicator species to assess the state of macrobenthic communities. *Advances in Polychaete Research*, 299–309. https://doi.org/10.1007/978-94-017-0655-1_28
- Cahill, A. E., Pearman, J. K., Borja, A., Carugati, L., Carvalho, S., Danovaro, R., Dashfield, S., David, R., Féral, J. P., Olenin, S., Šiaulys, A., Somerfield, P. J., Trayanova, A., Uyarra, M. C., & Chenuil, A. (2018). A comparative analysis of metabarcoding and morphology-based identification of benthic communities across different regional seas. *Ecology and Evolution*, 8(17), 8908–8920. <https://doi.org/10.1002/ECE3.4283>
- Cameron, E. S., Schmidt, P. J., Tremblay, B. J. M., Emelko, M. B., & Müller, K. M. (2021). Enhancing diversity analysis by repeatedly rarefying next generation sequencing data describing microbial communities. *Scientific Reports 2021 11:1*, 11(1), 1–13. <https://doi.org/10.1038/s41598-021-01636-1>
- Capurso, G., Carroll, B., & Stewart, K. A. (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., Aylagas, E., Villalobos, R., Kattan, Y., Berumen, M., & Pearman, J. K. (2019).

Beyond the visual: Using metabarcoding to characterize the hidden reef cryptobiome.

Proceedings of the Royal Society B: Biological Sciences, 286(1896).

<https://doi.org/10.1098/rspb.2018.2697>

Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: an R package for the

visualization of intersecting sets and their properties. *Bioinformatics*, 33(18), 2938–

2940. <https://doi.org/10.1093/bioinformatics/btx364>

Costello, M. J., Ahyong, S., Bieler, R., Boudouresque, C., Desiderato, A., Downey, R., Galil,

B. S., Gollasch, S., Hutchings, P., Kamburska, L., Katsanevakis, S., Kupriyanova, E.,

Lejeusne, C., Ma, K. C. K., Marchini, A., Occhipinti, A., Pagad, S., Panov, V. E., Poore,

G. C. B., ... Zhan, A. (2024). *World Register of Introduced Marine Species (WRiMS)*.

WoRMS Editorial Board. <https://www.marinespecies.org/introduced>

Costello, M. J., Dekeyzer, S., Galil, B., Hutchings, P., Katsanevakis, S., Pagad, S., Robinson,

T., Turon, X., Vandepitte, L., Vanhoorne, B., Verfaille, K., Willan, R., & Rius, M.

(2021). Introducing the World Register of Introduced Marine Species (WRiMS).

Management of Biological Invasions, 12(4), 792–811.

<https://doi.org/10.3391/mbi.2021.12.4.02>

da Silva, L. P., Mata, V. A., Lopes, P. B., Pereira, P., Jarman, S. N., Lopes, R. J., & Beja, P.

(2019). Advancing the integration of multi-marker metabarcoding data in dietary

analysis of trophic generalists. *Molecular Ecology Resources*, 19(6), 1420–1432.

<https://doi.org/10.1111/1755-0998.13060>

Danovaro, R., Carugati, L., Berzano, M., Cahill, A. E., Carvalho, S., Chenuil, A.,

Corinaldesi, C., Cristina, S., David, R., Dell'Anno, A., Dzhembekova, N., Garcés, E.,

Gasol, J. M., Goela, P., Féral, J. P., Ferrera, I., Forster, R. M., Kurekin, A. A., Rastelli,

E., ... Borja, A. (2016). Implementing and innovating marine monitoring approaches for

assessing marine environmental status. *Frontiers in Marine Science*, 3(NOV), 214343.

<https://doi.org/https://doi.org/10.3389/fmars.2016.00213>

Dauvin, J. C. (2005). Expertise in coastal zone environmental impact assessments. *Marine*

Pollution Bulletin, 50(1), 107–110.

<https://doi.org/10.1016/J.MARPOLBUL.2004.11.013>

Devillers, R., Pressey, R. L., Grech, A., Kittinger, J. N., Edgar, G. J., Ward, T., & Watson, R.

(2015). Reinventing residual reserves in the sea: are we favouring ease of establishment

- over need for protection? *Aquatic Conservation: Marine and Freshwater Ecosystems*, 25(4), 480–504. <https://doi.org/10.1002/AQC.2445>
- Dragulescu, A., & Arendt, C. (2020). *xlsx: Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files*. <https://cran.r-project.org/package=xlsx>
- Duarte, S., Vieira, P. E., Leite, B. R., Teixeira, M. A. L., Neto, J. M., & Costa, F. O. (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., Doney, S. C., & Steinberg, D. K. (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(4), 313–316. <https://doi.org/10.1002/AQC.1187>
- Edgar, G. J., Stuart-Smith, R. D., Willis, T. J., Kininmonth, S., Baker, S. C., Banks, S., Barrett, N. S., Becerro, M. A., Bernard, A. T. F., Berkhout, J., Buxton, C. D., Campbell, S. J., Cooper, A. T., Davey, M., Edgar, S. C., Försterra, G., Galván, D. E., Irigoyen, A. J., Kushner, D. J., ... Thomson, R. J. (2014). Global conservation outcomes depend on marine protected areas with five key features. *Nature* 2014 506:7487, 506(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(1–2), 1–4. <https://doi.org/10.1016/J.MARPOLBUL.2014.07.026>
- Evans, N. T., Li, Y., Renshaw, M. A., Olds, B. P., Deiner, K., Turner, C. R., Jerde, C. L., Lodge, D. M., Lamberti, G. A., & Pfrender, M. E. (2017). Fish community assessment with eDNA metabarcoding: Effects of sampling design and bioinformatic filtering. *Canadian Journal of Fisheries and Aquatic Sciences*, 74(9), 1362–1374. <https://doi.org/10.1139/CJFAS-2016-0306>
- Fediajevaite, J., Priestley, V., Arnold, R., & Savolainen, V. (2021). Meta-analysis shows that environmental DNA outperforms traditional surveys, but warrants better reporting standards. *Ecology and Evolution*, 11(9), 4803–4815. <https://doi.org/10.1002/ECE3.7382>
- Fields, J. B., & Silbiger, N. J. (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., Lefcheck, J. S., Byrnes, J. E. K., Cardinale, B. J., Duffy, J. E., & Griffin, J. N. (2015). Marine biodiversity and ecosystem functioning: what's known and what's next? *Oikos*, 124(3), 252–265. <https://doi.org/10.1111/OIK.01549>
- Gibson, J., Shokralla, S., Porter, T. M., King, I., Van Konynenburg, S., Janzen, D. H., Hallwachs, W., & Hajibabaei, M. (2014). Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasytematics. *Proceedings of the National Academy of Sciences of the United States of America*, 111(22), 8007–8012. <https://doi.org/10.1073/PNAS.1406468111>
- Gielings, R., Fais, M., Fontaneto, D., Creer, S., Costa, F. O., Renema, W., & Macher, J. N. (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., & Colwell, R. K. (2011). Estimating species richness. In A. E. Magurran & B. J. McGill (Eds.), *Biological Diversity: Frontiers in Measurement and Assessment* (pp. 39–54). Oxford University Press.
- Grey, E. K., Bernatchez, L., Cassey, P., Deiner, K., Deveney, M., Howland, K. L., Lacoursière-Roussel, A., Leong, S. C. Y., Li, Y., Olds, B., Pfrender, M. E., Prowse, T. A. A., Renshaw, M. A., & Lodge, D. M. (2018). Effects of sampling effort on biodiversity patterns estimated from environmental DNA metabarcoding surveys. *Scientific Reports 2018 8:1*, 8(1), 1–10. <https://doi.org/10.1038/s41598-018-27048-2>
- Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., De Vargas, C., Decelle, J., Del Campo, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W. H. C. F., Lara, E., Le Bescot, N., Logares, R., ... Christen, R. (2013). The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, 41(D1), D597–D604. <https://doi.org/10.1093/NAR/GKS1160>
- Hakimzadeh, A., Abdala Asbun, A., Albanese, D., Bernard, M., Buchner, D., Callahan, B., Caporaso, J. G., Curd, E., Djemiel, C., Brandström Durling, M., Elbrecht, V., Gold, Z., Gweon, H. S., Hajibabaei, M., Hildebrand, F., Mikryukov, V., Normandeau, E., Özkurt, E., M. Palmer, J., ... Anslan, S. (2023). A pile of pipelines: An overview of the bioinformatics software for metabarcoding data analyses. *Molecular Ecology Resources*, 00, 1–17. <https://doi.org/10.1111/1755-0998.13847>

- Hallam, J., Clare, E. L., Jones, J. I., & Day, J. J. (2023). High frequency environmental DNA metabarcoding provides rapid and effective monitoring of fish community dynamics. *Environmental DNA*, 5(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., Bundy, A., Christensen, V., Coll, M., de Mutsert, K., Fulton, E. A., Piroddi, C., Shin, Y. J., Steenbeek, J., & Travers-Trolet, M. (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., Chang, J. J. M., Oh, R. M., Quek, Z. B. R., Chan, Y. K. S., Bauman, A. G., & Huang, D. (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*. <https://doi.org/10.1111/MEC.16568>
- Jacquemont, J., Blasiak, R., Le Cam, C., Le Gouellec, M., & Claudet, J. (2022). Ocean conservation boosts climate change mitigation and adaptation. *One Earth*, 5(10), 1126–1138. <https://doi.org/10.1016/J.ONEEAR.2022.09.002>
- Kaikkonen, L., Venesjärvi, R., Nygård, H., & Kuikka, S. (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. <https://cran.r-project.org/package=ggpubr>
- Kissling, W. D., Ahumada, J. A., Bowser, A., Fernandez, M., Fernández, N., García, E. A., Guralnick, R. P., Isaac, N. J. B., Kelling, S., Los, W., McRae, L., Mihoub, J. B., Obst, M., Santamaria, M., Skidmore, A. K., Williams, K. J., Agosti, D., Amariles, D., Arvanitidis, C., ... Hardisty, A. R. (2018). Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. *Biological Reviews*, 93(1), 600–625. <https://doi.org/10.1111/BRV.12359>
- Lanzé, A., Jørgensen, S. L., Huson, D. H., Gorfer, M., Grindhaug, S. H., Jonassen, I., Øvreås, L., & Urich, T. (2012). *CREST-Classification Resources for Environmental Sequence Tags*. <https://doi.org/10.1371/journal.pone.0049334>
- Leite, B. R., Duarte, S., Troncoso, J. S., & Costa, F. O. (2023). Artificial Seaweed Substrates Complement ARMS in DNA Metabarcoding-Based Monitoring of Temperate Coastal Macrozoobenthos. *Diversity*, 15(5), 657. <https://doi.org/10.3390/D15050657/S1>

- Leite, B. R., Vieira, P. E., Troncoso, J. S., & Costa, F. O. (2021). Comparing species detection success between molecular markers in DNA metabarcoding of coastal macroinvertebrates. *Metabarcoding and Metagenomics* 5: E70063, 5, e70063-. <https://doi.org/10.3897/MBMG.5.70063>
- Leray, M., & Knowlton, N. (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(7), 2076–2081. <https://doi.org/10.1073/PNAS.1424997112>
- Lotze, H. K., Guest, H., O’Leary, J., Tuda, A., & Wallace, D. (2018). Public perceptions of marine threats and protection from around the world. *Ocean & Coastal Management*, 152, 14–22. <https://doi.org/10.1016/J.OCECOAMAN.2017.11.004>
- Machida, R. J., Leray, M., Ho, S. L., & Knowlton, N. (2017). Metazoan mitochondrial gene sequence reference datasets for taxonomic assignment of environmental samples. *Scientific Data* 2017 4:1, 4(1), 1–7. <https://doi.org/10.1038/sdata.2017.27>
- Mahé, F., Rognes, T., Quince, C., de Vargas, C., & Dunthorn, M. (2015). Swarmv2: Highly-scalable and high-resolution amplicon clustering. *PeerJ*, 2015(12), e1420. <https://doi.org/10.7717/PEERJ.1420>
- Martaeng, R., Obst, M., & Kuklinski, P. (2023). Phylogeographic study using autonomous reef monitoring structures indicates fast range expansion of the invasive bryozoan *Juxtacribrilina mutabilis*. *Hydrobiologia*, 850(19), 4115–4126. <https://doi.org/10.1007/s10750-023-05184-9>
- Mathieu, C., Hermans, S. M., Lear, G., Buckley, T. R., Lee, K. C., & Buckley, H. L. (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., & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE*, 8(4), e61217. <https://doi.org/10.1371/JOURNAL.PONE.0061217>
- McMurdie, P. J., & Holmes, S. (2014). Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLOS Computational Biology*, 10(4), e1003531. <https://doi.org/10.1371/JOURNAL.PCBI.1003531>
- Meyer, R., Davies, N., Pitz, K. J., Meyer, C., Samuel, R., Anderson, J., Appeltans, W., Barker, K., Chavez, F. P., Duffy, J. E., Goodwin, K. D., Hudson, M., Hunter, M. E., Karstensen, J., Laney, C. M., Leinen, M., Mabee, P., MacKlin, J. A., Muller-Karger, F.,

- ... Buttigieg, P. L. (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., Halpern, B. S., Walbridge, S., Ciriaco, S., Ferretti, F., Fraschetti, S., Lewison, R., Nykjaer, L., & Rosenberg, A. A. (2013). Cumulative Human Impacts on Mediterranean and Black Sea Marine Ecosystems: Assessing Current Pressures and Opportunities. *PLoS ONE*, *8*(12), e79889. <https://doi.org/10.1371/journal.pone.0079889>
- Mugnai, F., Costantini, F., Chenuil, A., Leduc, M., Ortega, J. M. G., & Megléc, E. (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., Miloslavich, P., Bax, N. J., Simmons, S., Costello, M. J., Pinto, I. S., Canonico, G., Turner, W., Gill, M., Montes, E., Best, B. D., Pearlman, J., Halpin, P., Dunn, D., Benson, A., Martin, C. S., Weatherdon, L. V., Appeltans, W., Provoost, P., ... Geller, G. (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*(JUN), 373058.
<https://doi.org/10.3389/FMARS.2018.00211>
- Narwani, A., Reyes, M., Pereira, A. L., Penson, H., Dennis, S. R., Derrer, S., Spaak, P., & Matthews, B. (2019). Interactive effects of foundation species on ecosystem functioning and stability in response to disturbance. *Proceedings of the Royal Society B*, *286*(1913).
<https://doi.org/10.1098/RSPB.2019.1857>
- Nilsson, R. H., Larsson, K.-H., Taylor, A. F. S., Bengtsson-Palme, J., Jeppesen, T. S., Schigel, D., Kennedy, P., Picard, K., Gi Ockner 10, F. O., Tedersoo, L., Saar, I., & Abarenkov, K. (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., Exter, K., Allcock, A. L., Arvanitidis, C., Axberg, A., Bustamante, M., Cancio, I., Carreira-Flores, D., Chatzinikolaou, E., Chatzigeorgiou, G., Christmas, N., Clark, M. S., Comtet, T., Dailianis, T., Davies, N., Deneudt, K., de Cerio, O. D., Fortič, A., Gerovasileiou, V., ... Pavlouidi, C. (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., Vicario, S., Lundin, K., Berggren, M., Karlsson, A., Haines, R., Williams, A.,

- Goble, C., Mathew, C., & Güntsch, A. (2018). Marine long-term biodiversity assessment suggests loss of rare species in the Skagerrak and Kattegat region. *Marine Biodiversity*, 48(4), 2165–2176. <https://doi.org/10.1007/s12526-017-0749-5>
- Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O’Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., ... Weedon, J. (2023). *vegan: Community Ecology Package*. <https://github.com/vegandevs/vegan>
- Ooms, J. (2024). *writexl: Export Data Frames to Excel “xlsx” Format*. <https://cran.r-project.org/package=writexl>
- Pagès, H., Aboyoun, P., Gentleman, R., & DebRoy, S. (2020). *Biostrings: Efficient manipulation of biological strings*. <https://bioconductor.org/packages/Biostrings>
- Pearman, J. K., Aylagas, E., Voolstra, C. R., Anlauf, H., Villalobos, R., & Carvalho, S. (2019). Disentangling the complex microbial community of coral reefs using standardized Autonomous Reef Monitoring Structures (ARMS). *Molecular Ecology*, 28(15), 3496–3507. <https://doi.org/10.1111/mec.15167>
- Pearman, J. K., Chust, G., Aylagas, E., Villarino, E., Watson, J. R., Chenuil, A., Borja, A., Cahill, A. E., Carugati, L., Danovaro, R., David, R., Irigoien, X., Mendibil, I., Moncheva, S., Rodríguez-Ezpeleta, N., Uyarra, M. C., & Carvalho, S. (2020). Pan-regional marine benthic cryptobiome biodiversity patterns revealed by metabarcoding Autonomous Reef Monitoring Structures. *Molecular Ecology*, 29(24), 4882–4897. <https://doi.org/10.1111/MEC.15692>
- Pearman, J. K., Leray, M., Villalobos, R., Machida, R. J., Berumen, M. L., Knowlton, N., & Carvalho, S. (2018). Cross-shelf investigation of coral reef cryptic benthic organisms reveals diversity patterns of the hidden majority. *Scientific Reports 2018 8:1*, 8(1), 1–17. <https://doi.org/10.1038/s41598-018-26332-5>
- Peroni, S., Soiland-Reyes, S., Sefton, P., Crosas, M., Castro, L. J., Coppens, F., Fernández, J. M., Garijo, D., Grüning, B., La Rosa, M., Leo, S., Ó Carragáin, E., Portier, M., Trisovic, A., Groth, P., & Goble, C. (2022). Packaging research artefacts with RO-Crate. *Data Science*, 5(2), 97–138. <https://doi.org/10.3233/DS-210053>
- Porter, T. M., & Hajibabaei, M. (2018). Over 2.5 million COI sequences in GenBank and growing. *PLOS ONE*, 13(9), e0200177. <https://doi.org/10.1371/JOURNAL.PONE.0200177>
- R Core Team. (2021). *R: A language and environment for statistical computing*. R

- Foundation for Statistical Computing. <https://www.r-project.org/>
- Ransome, E., Geller, J. B., Timmers, M., Leray, M., Mahardini, A., Sembiring, A., Collins, A. G., & Meyer, C. P. (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*(4), e0175066. <https://doi.org/10.1371/journal.pone.0175066>
- Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, *2016*(10), e2584. <https://doi.org/10.7717/PEERJ.2584>
- RStudio Team. (2022). *RStudio: Integrated Development Environment for R*. RStudio, PBC. <http://www.rstudio.com/>
- Santi, I., Beluche, O., Beraud, M., Buttigieg, P. L., Casotti, R., Cox, C. J., Cunliffe, M., Davies, N., de Cerio, O. D., Exter, K., Kervella, A. E., Kotoulas, G., Lagaisse, R., Laroquette, A., Louro, B., Not, F., Obst, M., Pavloudi, C., Poulain, J., ... Pade, N. (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., Casotti, R., Comtet, T., Cunliffe, M., Koulouri, P. (Yolanda), Macheriotou, L., Not, F., Obst, M., Pavloudi, C., Romac, S., Thiebaut, E., Vanaverbeke, J., & Pade, N. (2022). *European Marine Omics Biodiversity Observation Network (EMO BON) Handbook (Version 1.0)*. 63. <https://doi.org/10.25607/OBP-1653>
- Schauberger, P., & Walker, A. (2021). *openxlsx: Read, Write and Edit xlsx Files*. <https://cran.r-project.org/package=openxlsx>
- 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., Meyer, R. S., & Shapiro, B. (2021). Revisiting the effect of PCR replication and sequencing depth on biodiversity metrics in environmental DNA metabarcoding. *Ecology and Evolution*, *11*(22), 15766–15779. <https://doi.org/10.1002/ECE3.8239>
- Smale, D. A., Wernberg, T., Oliver, E. C. J., Thomsen, M., Harvey, B. P., Straub, S. C., Burrows, M. T., Alexander, L. V., Benthuyssen, J. A., Donat, M. G., Feng, M., Hobday, A. J., Holbrook, N. J., Perkins-Kirkpatrick, S. E., Scannell, H. A., Sen Gupta, A., Payne, B. L., & Moore, P. J. (2019). Marine heatwaves threaten global biodiversity and the provision of ecosystem services. *Nature Climate Change* *2019* *9*:4, *9*(4), 306–312. <https://doi.org/10.1038/s41558-019-0412-1>

- Staeher, P. A. U., Dahl, K., Buur, H., Göke, C., Sapkota, R., Winding, A., Panova, M., Obst, M., & Sundberg, P. (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., Woolley, S. N. C., Barnett, J., & Dunstan, P. (2020). Testing the presence of marine protected areas against their ability to reduce pressures on biodiversity. *Conservation Biology*, 34(3), 622–631. <https://doi.org/10.1111/COBI.13429>
- Stuer-Lauridsen, F., Drillet, G., Hansen, F. T., & Saunders, J. (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., Antich, A., Palacín, C., Præbel, K., & Wangensteen, O. S. (2020). From metabarcoding to metaphylogeography: separating the wheat from the chaff. *Ecological Applications*, 30(2), e02036. <https://doi.org/10.1002/EAP.2036>
- Villalobos, R., Aylagas, E., Pearman, J. K., Curdia, J., Lozano-Cortés, D., Coker, D. J., Jones, B., Berumen, M. L., & Carvalho, S. (2022). Inter-annual variability patterns of reef cryptobiota in the central Red Sea across a shelf gradient. *Scientific Reports 2022 12:1*, 12(1), 1–17. <https://doi.org/10.1038/s41598-022-21304-2>
- Virta, L., Soininen, J., & Norkko, A. (2021). Biodiversity Loss Threatens the Current Functional Similarity of Beta Diversity in Benthic Diatom Communities. *Microbial Ecology*, 81(2), 293–303. <https://doi.org/10.1007/S00248-020-01576-9>
- Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, 73(16), 5261–5267. <https://doi.org/10.1128/AEM.00062-07>
- Wesselmann, M., Apostolaki, E. T., & Anton, A. (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(1), 1–29. <https://www.jstatsoft.org/v40/i01/>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Golemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686.
<https://doi.org/10.21105/joss.01686>

- Wickham, H., Pedersen, T. L., & Seidel, D. (2023). *scales: Scale Functions for Visualization*.
<https://cran.r-project.org/package=scales>
- Witalis, B., Iglíkowska, A., Ronowicz, M., & Kukliński, P. (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>
- Worm, B., Barbier, E. B., Beaumont, N., Duffy, J. E., Folke, C., Halpern, B. S., Jackson, J. B. C., Lotze, H. K., Micheli, F., Palumbi, S. R., Sala, E., Selkoe, K. A., Stachowicz, J. J., & Watson, R. (2006). Impacts of biodiversity loss on ocean ecosystem services. *Science*, 314(5800), 787–790. <https://doi.org/10.1126/SCIENCE.1132294>
- Zafeiropoulos, H., Gioti, A., Ninidakis, S., Potirakis, A., Paragkamian, S., Angelova, N., Antoniou, A., Danis, T., Kaitetzidou, E., Kasapidis, P., Kristoffersen, J. B., Papadogiannis, V., Pavludi, C., Ha, Q. V., Lagnel, J., Pattakos, N., Perantinos, G., Sidirokastritis, D., Vavilis, P., ... Pafilis, E. (2021). 0s and 1s in marine molecular research: a regional HPC perspective. *GigaScience*, 10(8), 1–12.
<https://doi.org/10.1093/GIGASCIENCE/GIAB053>
- Zafeiropoulos, H., Viet, H. Q., Vasileiadou, K., Potirakis, A., Arvanitidis, C., Topalis, P., Pavludi, C., & Pafilis, E. (2020). PEMA: a flexible Pipeline for Environmental DNA Metabarcoding Analysis of the 16S/18S ribosomal RNA, ITS, and COI marker genes. *GigaScience*, 9(3), 1–12. <https://doi.org/10.1093/GIGASCIENCE/GIAA022>
- Zhang, M., Zou, Y., Xiao, S., & Hou, J. (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>

DATA ACCESSIBILITY AND BENEFIT-SHARING

Data Accessibility Statement

All data presented in this manuscript are publicly available, see main text and Supplementary Material 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 Supplementary Files and are also available on the respective GitHub repository (https://github.com/arms-mbon/data_release_001). Files used for EurOBIS submissions will be available on https://github.com/arms-mbon/data_release_001). Output files of the PEMA pipeline used as basis for EurOBIS submissions and for exploratory analysis as shown in this manuscript are accessible on https://github.com/arms-mbon/analysis_release_001. The URLs to IMIS metadata records of EurOBIS submissions and the respective DOIs will be made available by the end of April 2024. 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.

AUTHOR CONTRIBUTIONS

All authors maintained the observatories, organised the sampling campaigns, and contributed to the development of the presented methods and standards. MO, CP, KE, IS established and maintained the network. IS and CP organised the central sample processing and sequencing. ND, CP, HZ did the majority of the analytical work. KE, CP, JP, ND performed the majority of the data management. JT and PS contributed to image data management. ND wrote the final manuscript with major contributions from MO, CP, KE, JP. All other authors revised the manuscript.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

SUPPORTING INFORMATION

Supporting Information for online publication.docx

Contains Supplementary Texts S1 – S2, Supplementary Tables S1 – S6, and Supplementary Figures S1 – S2, which are referred to in the main text.

Supplementary_File_1_Observatory_SamplingEvent_Image_Omics_DataMeta_data_release_001.xlsx

Is referred to in the main text and contains metadata on observatories and sampling events for data_release_001 of ARMS-MBON, as well as links to image data and ENA accession numbers of genetic data for sampling events presented in this manuscript. It also contains info for Omics data on how sequence reads were demultiplexed.

Supplementary_File_2_abundance_phyla_species_identified_data_release_001.xlsx

Is referred to in the main text. The first three sheets contain abundances based on relative read numbers of identified phyla for the filtered and curated exploratory data set of all three marker genes. The last three sheets contain the number of species identified for each phylum (and class in case of ITS) for the three marker gene data sets. For the COI data set, taxonomic assignments below a confidence level of 0.8 were set as NA.

Supplementary_File_3_Observatories_AMBI_RedList_WRiMS_species.xlsx

Is referred to in the main text. Each sheet represents one of the 15 observatories and contains the identified species (with occurrences of at least two sequence reads) at each observatory which are registered as: i) very sensitive to disturbance in AMBI; ii) alien at the location of occurrence in WRiMS; and iii) Near Threatened, Vulnerable, Endangered or Critically Endangered in the Red Lists of the IUCN and HELCOM. Data of COI and 18S marker genes were pooled for each observatory.

Supplementary_File_4_erroneus_sequences_removed.xlsx

Is referred to in Supplementary Text S2. It contains two sheets, one each for the COI and 18S. Each sheets contains the ASVs/OTUs which were removed as potential contaminants based on their taxonomic classification, and the samples they occurred in with the respective read number.

TABLES

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

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

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	5,258		
Marker gene sequencing	COI	18S	ITS
Sequencing batches	7	7	4 [†]
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 [†] + 0 + 4

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

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)

FIGURES

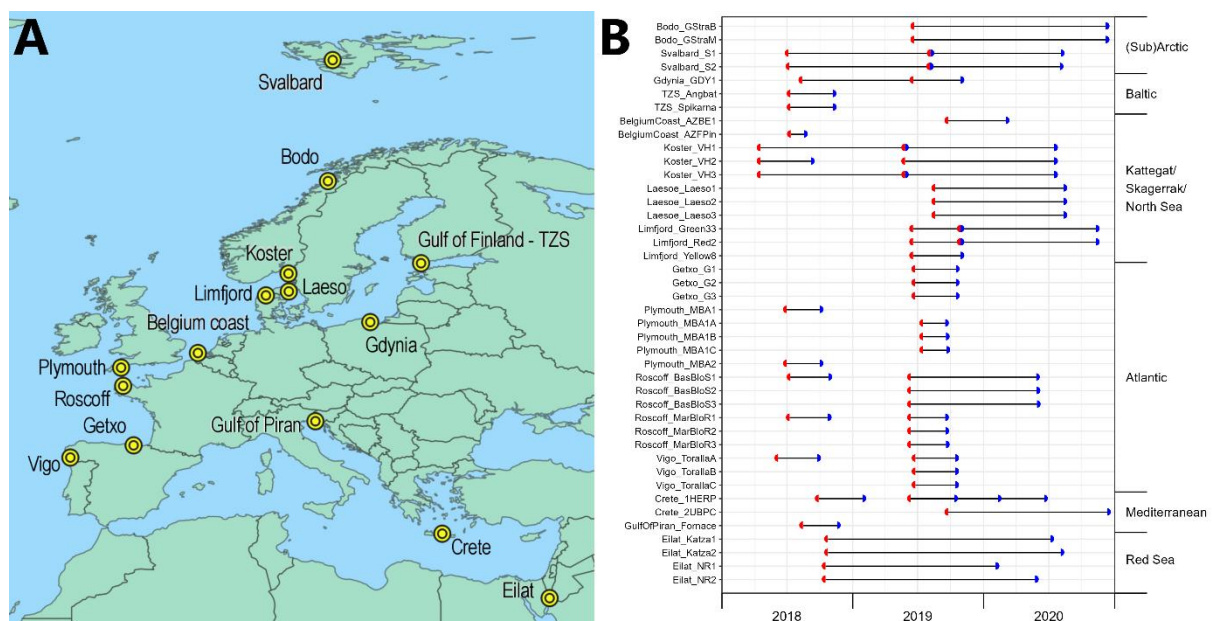


Figure 1. (A) Locations of observatories that deployed ARMS units in 2018 and 2019 during the first ARMS-MBON sampling campaign. (B) Sampling events of the first ARMS-MBON sampling campaign. Axis on the left shows ObservatoryID_UnitID combinations, axis on the right shows groupings of observatories into 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).

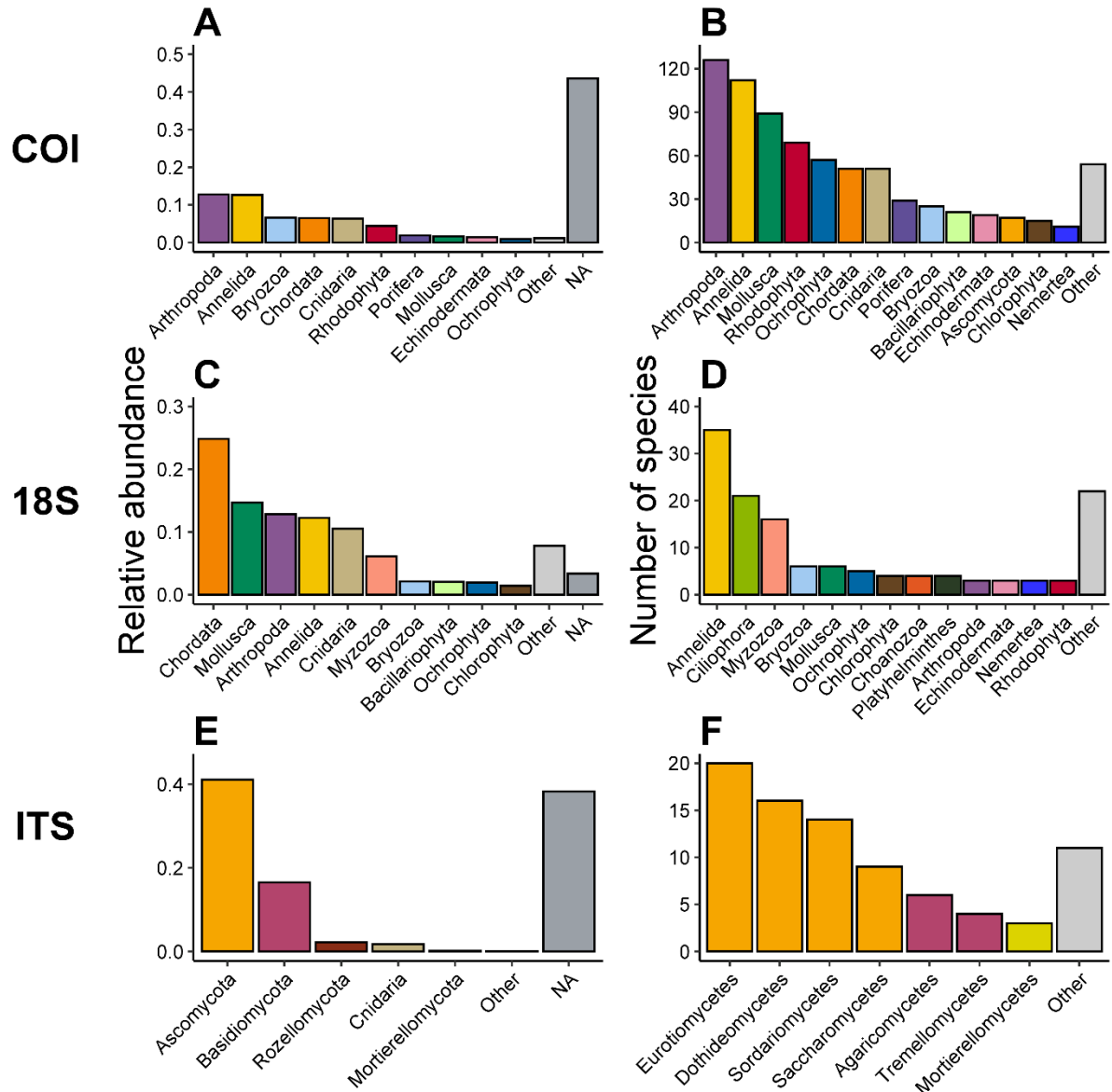


Figure 2. Relative read abundance of the ten most abundant phyla in the COI (A) and 18S (C) and five most abundant phyla in the ITS data sets (E). Less abundant phyla are grouped as *Other*, while relative abundance of sequence reads unclassified at phylum level are grouped as *NA*. Number of identified species within each phylum for COI (B) and 18S (D) and within each class for ITS (F). Phyla / classes with less than ten (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.

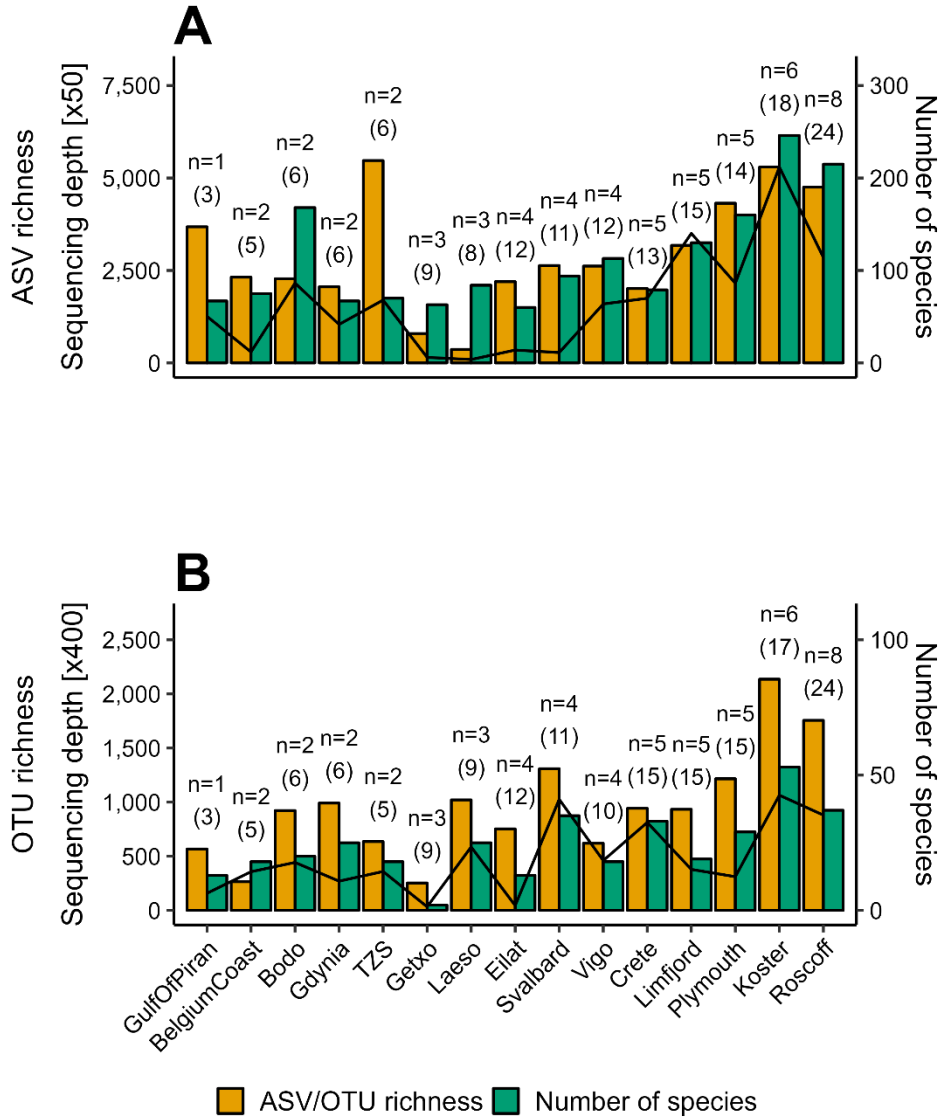


Figure 3. Diversity observed for the COI (A) and 18S (B) data sets across the 15 observatories measured as ASV/OTU richness (yellow; shared axes on the left) and the number of identified species (green; axes on the right) at the classification confidence threshold applied here. The number of ARMS units 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.

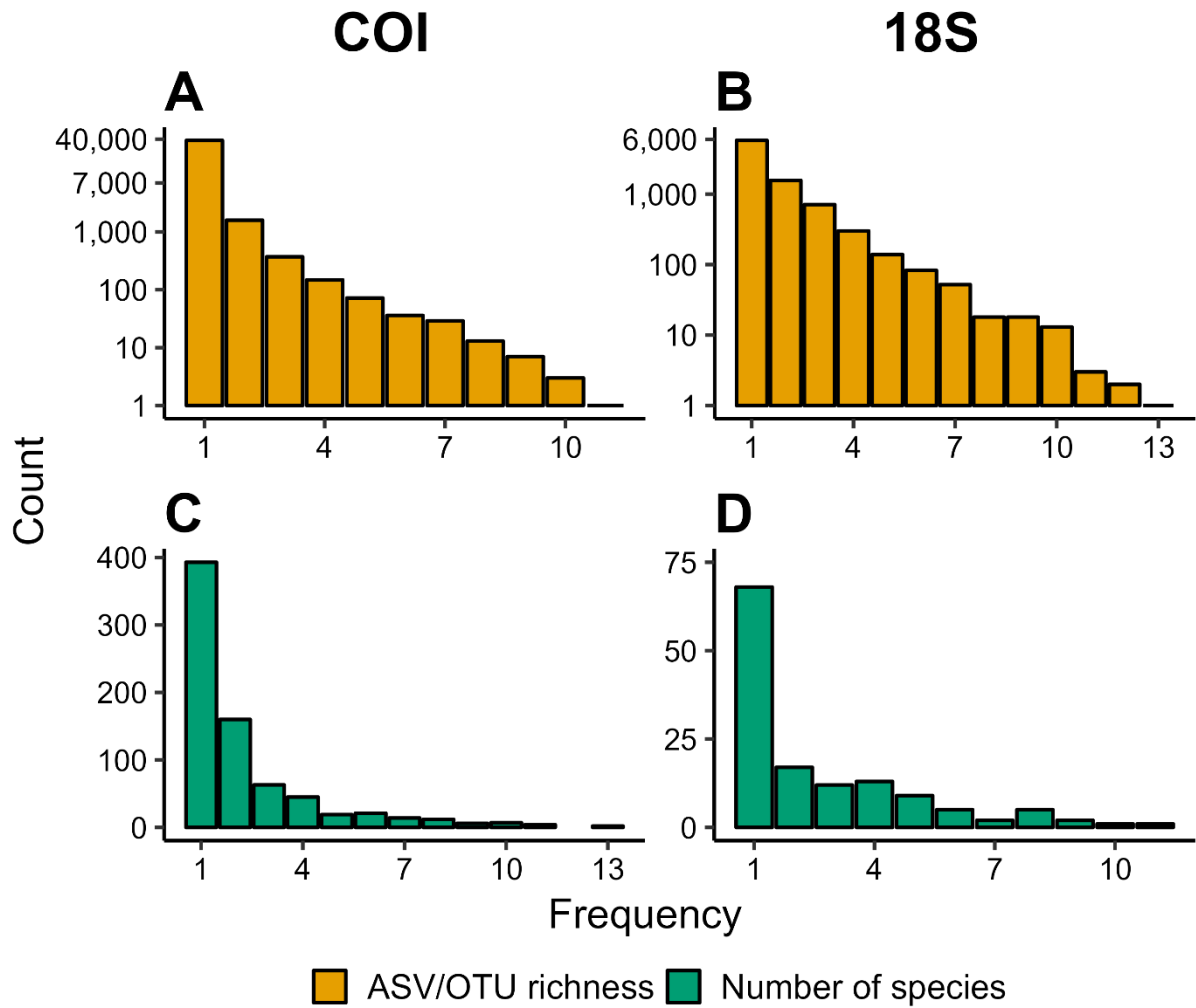


Figure 4. Frequency distribution of ASVs for COI (**A**) and OTUs for 18S (**B**) and of species identified for COI (**C**) and 18S data (**D**). Counts represent the number of entities (i.e., ASVs/OTUs or species identified) occurring at a given frequency, in this case the number of observatories. Note the log₂-transformation of the y-axis in **A** and **B**.

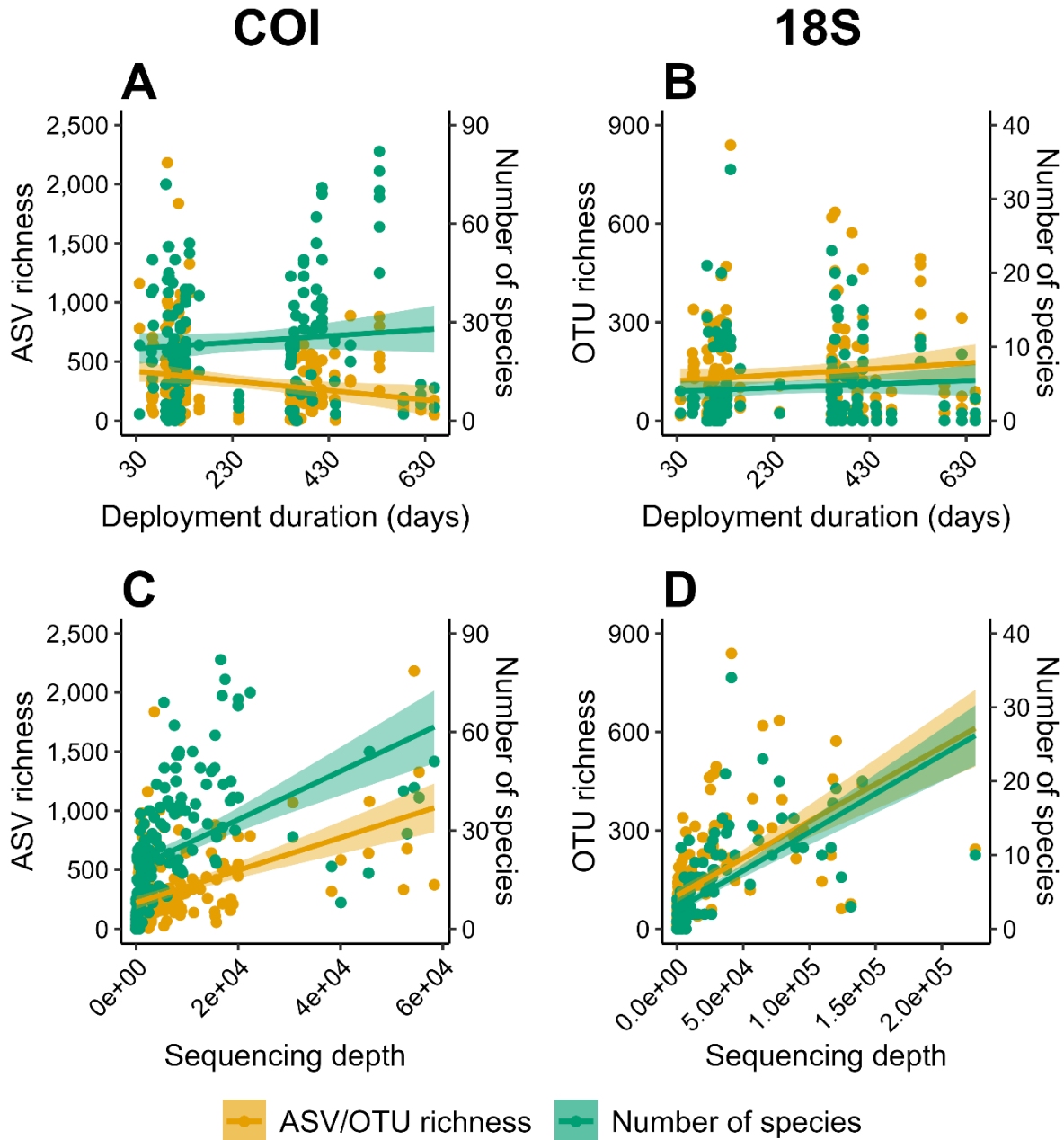


Figure 5. Relationship of the observed ASV/OTU richness (yellow; axes on the left) and the number of species identified (green; axes on the right) in each sample versus the deployment duration in days of the respective ARMS units for the COI (**A**) and 18S (**B**) data sets; and relationship of ASV/OTU 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 linear regression for ASV/OTU richness (yellow) and the number of species identified (green), shaded areas represent the corresponding 95% confidence intervals. No significant linear correlation was found for associations in **A** and **B**.

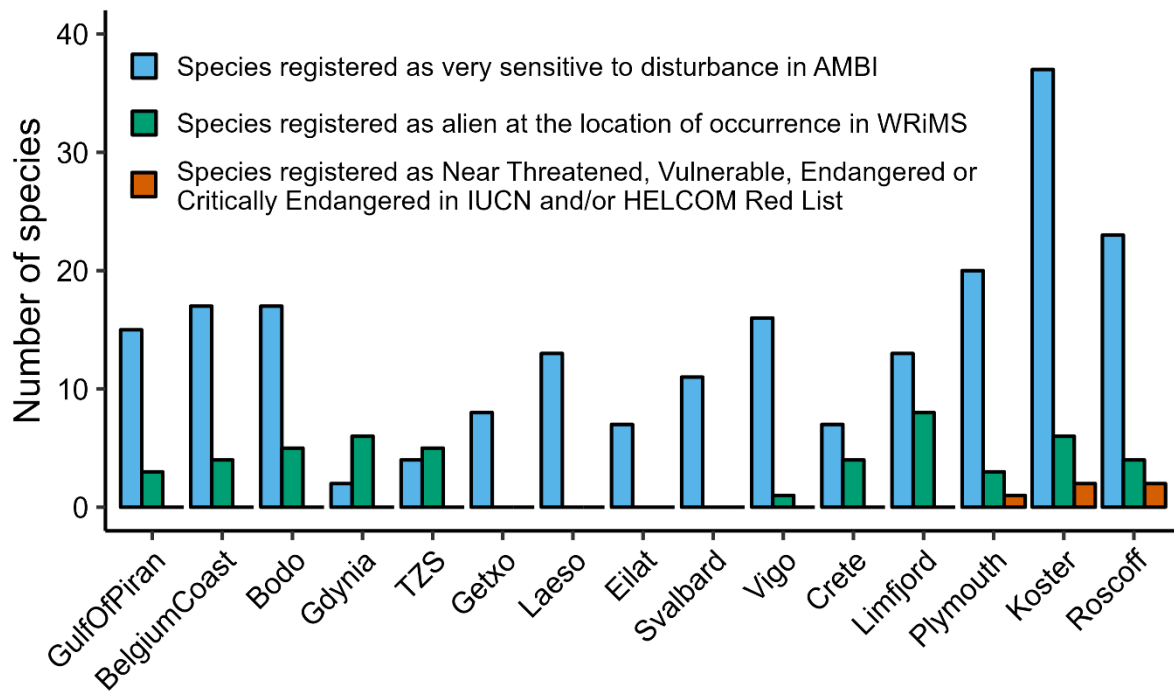


Figure 6. Number of identified species (with occurrences of at least two sequence reads) across observatories listed in four different databases. Data of COI and 18S marker genes were pooled for each observatory. Blue – species registered as very sensitive to disturbance in AMBI; green – species registered 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).

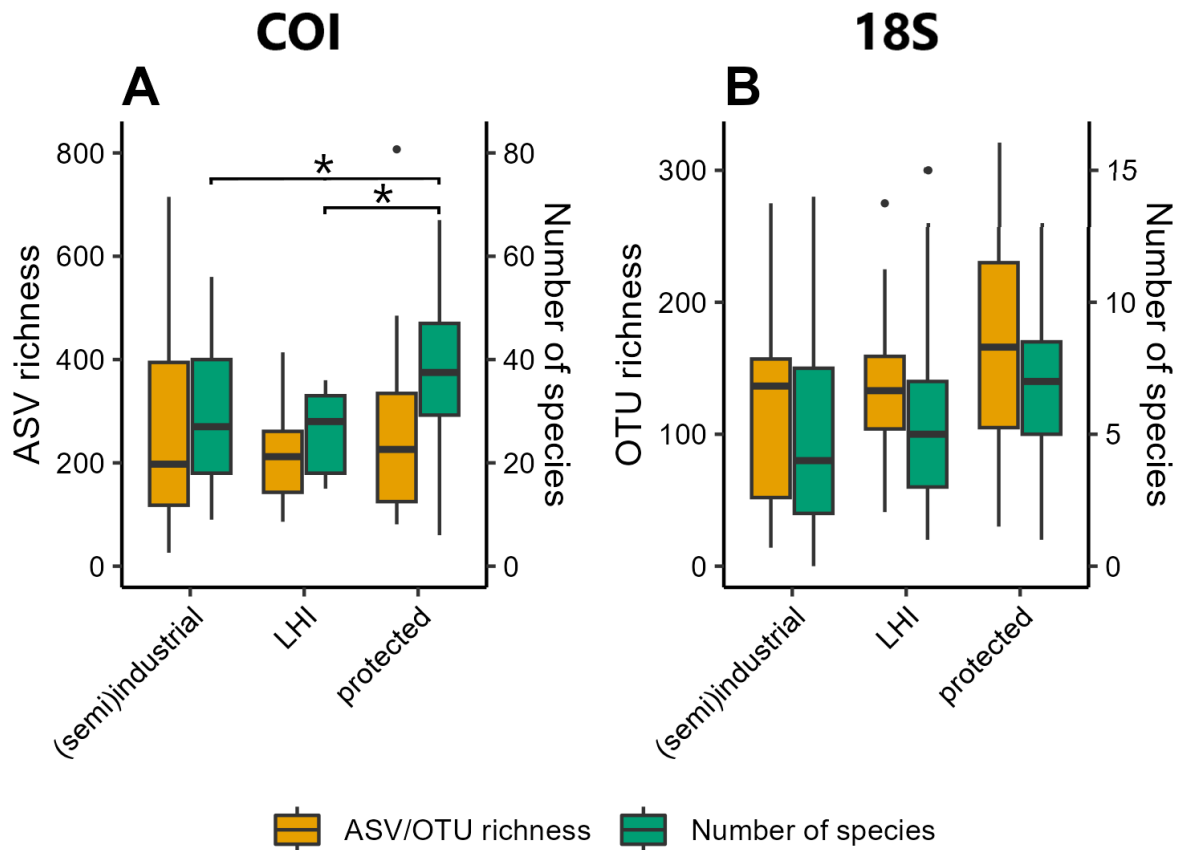


Figure 7. Alpha diversity across locations with varying degree of anthropogenic influence (i.e., industrial/semi-industrial, LHI - low human impact, and marine protected area). Boxplots show alpha diversity measured as ASV/OTU richness (yellow; axes on the left) and the number of identified species (green; 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 first to 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 (*: $p < 0.05$).