MARCO-BOLO
STRENGTHENING BIODIVERSITY OBSERVATION IN SUPPORT OF DECISION MAKING

Deliverable 2.2
Datasets, databases and softwares/pipelines facilitating the implementation of eDNA-based monitoring. Version 1.0

15/01/2026

Mads Reinholdt Jensen[1], Kim Præbel[1], Pedro Ciarlini Junger[2], Chris Bowler[2], Joanna Warwick-Dugdale[3], Michael Cunliffe[3], Jennifer Beatty[4], Fabrice Not[4], Aubrie Onoufriou[5], Iveta Matejusová[5], Daniele Ballardini[6], Domenico D'Alelio[6], Stephen Formel[7], Saara Suominen[7], Emilie Boulanger[7], Hanneloor Heynderickx[8], Daniel Kumazawa Morais[1]

[1]**UiT – The Arctic University of Norway,** [2]**CNRS-IBENS,** [3]**MBA:** Marine Biological Association, [4]**SU:** Sorbonne Université, [5]**MS:** Marine Directorate, Scottish Government, [6]**SZN:** Stazione Zoologica Anton Dohrn, [7]**UNESCO-OBIS:** United Nations Educational, Scientific and Cultural Organization - Ocean Biodiversity Information System, [8]**VLIZ:** Flanders Marine Institute

MARK IF PUBLIC / SENSITIVE / CLASSIFIED

Funded by the European Union

Funded by the European Union under the Horizon Europe Programme, Grant Agreement No. 101082021 (MARCO-BOLO). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

UK Research and Innovation

UK participants in MARCO-BOLO are supported by the UKRI's Horizon Europe Guarantee under the Grant No. 10068180 (MS); No. 10063994 (MBA); No. 10048178 (NOC).

Document Information

Grant Agreement	101082021
Project Acronym	MARCO-BOLO
Project Title	MARCO-BOLO will strengthen European marine, coastal and freshwater biodiversity observation to understand and restore ocean health.

Deliverable Number	D2.2
Work Package Number	WP2
Deliverable Title	Datasets, databases and softwares/pipelines facilitating the implementation of eDNA-based monitoring.
Lead Beneficiary	6. UiT
Author(s)	Mads Reinholdt Jensen (UiT), Kim Præbel (UiT), Pedro Ciarlini Junger (CNRS), Chris Bowler (CNRS), Joanna Warwick-Dugdale (MBA), Michael Cunliffe (MBA), Jennifer Beatty (SU), Fabrice Not (SU), Aubrie Onoufriou (MS), Iveta Matejusová (MS), Daniele Bellardini (SZN), Domenico D'Alelio (SZN), Stephen Formel (UNESCO-OBIS), Saara Suominen (UNESCO-OBIS), Emilie Boulanger (UNESCO-OBIS), Hanneloor Heynderickx (VLIZ), Daniel Kumazawa Morais (UiT)
Due Date	30.11.2025
Submission Date	15.01.2026
Dissemination Level	PU - Public[1]
Type of Deliverable	DATA – data sets, microdata, etc.

Version 1.0	15.01.2026, Mads Reinholdt Jensen

[1] Dissemination level: **PU:** Public, information as referred to in European Commission Decision 2015/844

Executive summary

Work package 2 (WP2) of the MARCO-BOLO project focused on validating and enabling environmental DNA (eDNA)-based approaches for biodiversity monitoring in aquatic and terrestrial systems. In task 2.2, these objectives were addressed through the exploration and comparison of datasets, databases, software, and bioinformatic pipelines to facilitate the implementation of eDNA-based monitoring. While this task was initially envisioned to build on existing infrastructure (specific pipelines, datasets, and customized databases), the departure of the creator of these preexisting tools, who was involved in the initial application, led to a shift in focus. This deliverable was likely originally intended to result in a single, standardized approach to working with eDNA-derived biomonitoring data. However, we collectively concluded that no single database, software, or pipeline can address the diverse practical use cases within eDNA. Therefore, this report provides a broader context on existing approaches, databases, and pipelines, their applications, and how they compare. The main body of work carried out under this task was a comparison of bioinformatic pipelines for two types of metabarcoding data (eukaryote 18S and 12S, 16S and COI for fishes), where we invited people around the world to contribute results from running their respective pipelines on the same datasets (Dataset 1). We also generated eDNA metabarcoding data for time series samples collected by institutions involved in this task. The statuses of Datasets 2-7, where new data was generated, are presented here, each accompanied by a "readme" in varying formats. These data products will contribute to deliverables D2.3 and D2.4 but are presented here alongside their metadata. As this is a data deliverable rather than a narrative report, we focused on implementing WP1's data model for reporting on data and metadata. While this model is not yet finalized, we here use Dataset 2 to demonstrate its potential, transforming information from Google Sheets into JSON format and generating standardized readme files using a large language model (LLM). Currently, each dataset has its own readme format, but will be standardized under WP1's model during the project's final year. Dataset 8-9 are based on data generated prior to MARCO-BOLO and here primarily serve to indicate the locations of all utilized data products, as results from these are presented under D2.3 and D2.4. Under this task, we also engaged with the existing literature and here provide a comprehensive overview of what we consider to be widely used primer sets for eDNA metabarcoding, pipelines run for different marker genes, and reference databases used to assign taxonomy (Dataset 10). This work complements the data analysis challenge, where we aimed to include as many pipelines as possible, and is here presented as a standalone data product. Finally, we showcase two custom-built reference databases tailored for Nordic eDNA metabarcoding applications (Dataset 11), targeting the "Leray" fragment (COI) and the "MiFish" fragment (12S rRNA). These databases integrate and harmonize data from multiple repositories while incorporating rigorous curation steps to ensure high-quality references for eDNA research. They remain a work in progress, with ongoing efforts to refine and expand their scope to support diverse research applications

Contents

Dataset 1: MARCO-BOLO Data Analysis Challenge

MARCO-BOLO Deliverable 2.2, Dataset 1

Grant Agreement: 101082021 | Work Package: WP2 | Deliverable: D2.2

Short description: In September 2024, MARCO-BOLO launched the Data Analysis Challenge to survey the global environmental DNA (eDNA) community about their use and development of bioinformatics pipelines to process eDNA samples, as well as invite them to contribute their time and methodology towards an extensive comparison of these pipelines (see Milestone 8). Registered participants were invited to download any of the four metabarcoding datasets provided, analyse these with their pipeline, and return the output ASV/OTU and taxonomy tables for comparison. In total, 14 participants submitted a total of 18 output datasets for the 18S protist observatory dataset, and 25 participants submitted a total of 101 output datasets for the 12S, 16S and COI fish aquarium datasets. The resulting tables were harmonised for joint analysis, and bioinformatic parameters were extracted from the different results files received. The current data submission consists of the harmonised tables of participant outputs, the extracted bioinformatics metadata, and some supplementary files, all detailed below.

Original data:

18S protist metabarcoding from the ASTAN observatory, published by Caracciolo *et al*. (2022)

- Metabarcoding raw reads available on the European Nucleotide Archive (ENA): PRJEB48571
- Sampling dates are provided here in the file 'metadata_astan.csv'.
- Microscopic counts file available on Zenodo (Rigaut-Jalabert *et al*., 2021): https://doi.org/10.5281/zenodo.5033180. These were restructured to fit the structure of the challenge datasets and are provided here as 'microscopy_harmonised.csv' and 'metadata_microscopy.csv'.

12S, 16S, COI fish metabarcoding from aquarium samples.

- The metabarcoding raw reads were communicated to participants directly. These are currently not yet published and owned by the A-Fish-DNA-Scan and ME-BARCODE group at the University of Minho (contact: Filipe Costa fcosta@bio.uminho.pt).
- The lists of fish, shark and ray species inhabiting the Lisbon aquarium at the time of sampling ("residents"), and species used in feed are available on Zenodo: doi:10.5281/zenodo.17739996.
    - aquarium_12s16sCOI_residents.csv
    - aquarium_12s16sCOI_feed.csv

Sampling area:

- 18S protist metabarcoding: The SOMLIT-Astan sampling station (http://marineregions.org/mrgid/5372) in the Western English Channel, France (48°46'18" N–3°58'6" W).
- 12s, 16s, COI metabarcoding: The Lisbon Aquarium, Portugal (38°45'48.671" N-9°5'37.377").

Instructions:

Instructions to participants were communicated and archived on the public repository https://github.com/marco-bolo/wp2-wp5-workshop.

Participant output data:

The resulting output tables from each participant, harmonised and merged by target group (protists or fish), are available on Zenodo: doi:10.5281/zenodo.17739996.

- challengedata_18s.csv
- challengedata_12s16sCOI.csv

Due to storage limitations, the participant output data contains 9 out of 18 submissions for the 18s metabarcoding dataset. The remaining datasets will be added by the end of December 2025.

Participant bioinformatics workflows:

Key bioinformatic steps and parameters were extracted for each of the submissions included in the challenge data. This harmonised bioinformatics metadata is available on Zenodo: doi:10.5281/zenodo.17739996.

- metadata_18s.csv: bioinformatics metadata accompanying the challengedata_18s.csv
- metadata_12s16sCOI.csv: bioinformatics metadata accompanying the challengedata_12s16sCOI.csv

The selected fields describing the bioinformatics steps and parameters were taken from the FAIR eDNA checklist (Takahashi *et al.*, 2025).

Output data analysis:

All scripts to compare the participant outputs, analyse patterns and generate figures are being published to the public repository https://github.com/marco-bolo/data-analysis-challenge.

Status:

Dataset is marked *incomplete* — analysis is ongoing but data challenge participation is complete (September 2024 – November 2025).

Currently, participants are indicated with an anonymized identifier (submitterID). Names will be released upon agreement of all participants in the following version.

Maintainer: Emilie Boulanger (OBIS/IOC-UNESCO)

Last updated: 2025-11-28

Dataset 2: EMO BON UiT Genomic Observatory 12S MiFish + COI Leray-XT

MARCO-BOLO Deliverable 2.2, Dataset 2

Grant Agreement: 101082021 | Work Package: WP2 | Deliverable: D2.2

Metadata: JSON-LD (schema.org) — https://lab.marcobolo-project.eu/csv-to-json-ld/schema-jsonld/mbo_wp2_d2_ds2.json

Sequences: PRJEB98105 (European Nucleotide Archive)

Short description: Analysis of 12S rRNA (MiFish) and COI (Leray-XT) markers in environmental DNA from seawater through amplicon sequencing. Samples were collected between August 2021 and August 2024 at the ESC68N genomic observatory within the European Marine Omics Biodiversity Observation Network (EMO BON). Seawater was filtered onto Sterivex filters, and environmental DNA was extracted and amplified by PCR for Illumina MiSeq sequencing. Sequences are archived at the European Nucleotide Archive (ENA).

Sampling area: Norwegian Sea, Norway (68.93°N, 17.13°E) — a site monitoring annual tide-driven flux of marine larvae through a fjord system.

Protocols

- **Sampling protocol** — doi:10.5281/zenodo.17590006

Field protocol for EMO BON sampling operations covering CTD deployment for metadata collection, Niskin bottle water sampling, and filtration procedures for metabarcoding using Sterivex and polycarbonate membrane filters.

eDNA analysis:

- **Clean lab routines** — doi:10.5281/zenodo.17552614

Routines for working in the clean labs of the Research Group for Genetics at UiT – The Arctic University of Norway.

- **DNA extraction** — doi:10.5281/zenodo.17552850

Extraction of eDNA from Sterivex filters using the Qiagen DNeasy Blood & Tissue Kit, including clean lab procedures, filter processing, lysis, and DNA purification over a 2-day workflow.

- **PCR (COI/16S/18S)** — doi:10.5281/zenodo.17590354

PCR setup for DNA amplification targeting COI, 16S, and 18S markers using AmpliTaq Gold MasterMix for MiSeq sequencing.

- **PCR (12S)** — doi:10.5281/zenodo.17590466

PCR setup targeting the 12S marker for fish detection in eDNA samples, optimised for MiSeq with samples run in triplicate.

- **Library preparation** — doi:10.5281/zenodo.17590611

Post-PCR processing including pooling, MinElute cleanup, Qubit quantification, and library preparation using the Qiaseq One Step Amplicon Library Preparation Kit with Ampure bead cleanup for Illumina sequencing.

- **Bioinformatics** — MetaBarFlow (doi:10.5281/zenodo.7023055)

Pipeline for QA/QC, denoising, and taxonomic assignment.

Usage Limitations

- **Status:** Dataset is marked *incomplete* — analysis is ongoing but data collection is complete (August 2021 – August 2024)
- **Protocols:** Supporting protocols are in revision prior to final publication
- **Taxonomic scope:** 12S primers optimised for fish; COI provides broader eukaryotic coverage but with known primer biases
- **Spatial scope:** Single Arctic observatory; regional generalisability requires caution

License and Access Conditions (FAIR Principles)

Principle	Implementation
Findable	Persistent w3id.org URIs; ENA accession PRJEB98105; Zenodo DOIs for protocols
Accessible	Sequence data openly available via ENA; metadata in machine-readable JSON-LD
Interoperable	Schema.org vocabulary; MarineRegions spatial references; ORCID identifiers
Reusable	**CC-BY-4.0** (attribution required); full provenance via linked protocols

Link with the Project Data Management Plan

This dataset is produced under MARCO-BOLO's Data Management Plan (D7.2, doi:10.5281/zenodo.8208410), which establishes FAIR data practices, Horizon Europe compliance, and co-designed data governance with stakeholders across the data pathway.

Maintainer: Mads Reinholdt Jensen (UiT – The Arctic University of Norway)

Metadata creator: Stephen Formel (OBIS)

Last updated: 2025-11-25

Dataset 3: L4 Timeseries – 2001-2023 - Metabarcoding Total Eukaryotes (18S V9 region rRNA gene)

Content summary and structure of the dataset

This dataset was generated from the Western Channel Observatory (WCO) coastal station 'L4' times series from the Western English Channel, years 2001 – 2023 (~weekly sampling of samples from ≤5 m depth). Metabarcoding of eukaryotic plankton was conducted via amplification of the 18S rRNA gene-V9 region (primers: 1391F and EukB) and sequenced as 250 bp paired-end reads by Northumbria University (NU-OMICS) on Illumina (Miseq), for the Marine Biological Association (MBA; Michael Cunliffe group). The dataset comprises 652 paired-end (i.e. forward and reverse) reads in the form of 1,304 individual sequences in FASTQ format; sequencing adapters and primer sequences have been removed.

Details of data source and methodology

Collection of samples: Samples were collected approximately once per week from the surface (≤5 m depth) of the Western English Channel coastal Western Channel Observatory (WCO: https://www.westernchannelobservatory.org.uk/) site 'L4' (50°15.00' N, 4°13.02' W; site depth ~55 m). From the years 2001-2019, seawater was collected via CTD Niskin-rosette from the Plymouth Marine Laboratory RV 'Plymouth Quest'; 1 L per sample was filtered onto 0.45 µm cellulose-nitrate or polycarbonate filters, and stored at −80°C. Samples obtained in the years 2022-2023 were collected from the Marine Biological Association RV 'Sepia', either via CTD Niskin rosette (2 L volume filtered per sample), or by the under-way sampling system (1 L volume filtered per sample). These samples were collected onto 0.22 µm cellulose-nitrate filters and stored as above.

DNA extraction, metabarcode amplification and sequencing: All samples were extracted from filters within a 12-month extraction campaign. Most samples were extracted with the ZymoBIOMICS DNA Miniprep Kit (Zymo Research); 12 samples were extracted with the ZymoBIOMICS DNA/RNA Miniprep Kit (Zymo Research). In short, filter sections were added to 700 µl of kit lysis solution in kit lysis tubes, and were bead-beaten three times for one minute duration at 10 m s-1, with five-minute intervals on ice. From here, DNA extraction followed manufacturer's protocols with 100 µl final elution volume. Possible kit/extraction associated contaminants were determined via the inclusion of one blank (i.e. no sample) sample per extraction batch of 50 samples. The 'Total-eukaryote' 18S-V9 region rRNA gene was amplified using PCR with the forward primer 1391f (5'-GTACACACCGCCCGTC -3') and the reverse primer EukBr (5'-TGATCCTTCTGCAGGTTCACCTAC -3') (based on Amaral-Zettler *et al.*, 2009, and Stoeck *et al.*, 2010), according to the Earth Microbiome Project (EMP) protocol (https://earthmicrobiome.org/protocols-and-standards/18s/; including reference to Caporaso *et al.*, 2012), except for use of KAPA2G Robust Master Mix (Roche Diagnostics). Amplicons were sequenced on the Illumina MiSeq platform by NU-OMICS at Northumbria University as 250 bp paired-end reads.

Sequence processing: In total, 20,941,365 18S-V9 rRNA reads were generated from 575 samples (plus 32 extraction kit blanks and 20 sequencing negative controls). Sequencing adapters were removed with CASAVA (Illumina), and primers were trimmed with CutAdapt v1.5 (Martin, 2011).

License/access conditions (FAIR principles)

This dataset is published in the European Nucleotide Archive (ENA) at EMBL-EBI under the project accession PRJEB98106 (current status: private; release date: 30th November 2026).

Dataset 4: L4 Timeseries – 2001-2023 - Metabarcoding of fishes (12S) and broad eukaryotes (COI)

This dataset will be generated from the same set of samples that were extracted in the production of Dataset 3. Unfortunately, due to a number of unsuccessful attempts at dry-ice sample transfer with well-established couriers, samples were transported in person for a WP2 event held at UiT in September 2025. Sample processing is underway and will be completed during the final year of Marco-Bolo.

Dataset 5: V9-18S rDNA amplicon sequencing of NEREA environmental DNA samples

Grant Agreement: 101082021

Project Acronym: MARCO-BOLO

Project Title: MARine COastal BiOdiversity Long-term Observations

Deliverable Number: D2.2

Work Package Number: WP2

Deliverable Title: Datasets, databases and softwares/pipelines facilitating the implementation of eDNA-based monitoring

Dataset name: V9-18S rDNA amplicon sequencing of NEREA environmental DNA samples

License/access conditions (FAIR principles): This dataset is published in the European Nucleotide Archive (ENA) at EMBL-EBI under the project accession PRJEB98195 (current status: private; release date: 30th June 2026).

Short description: Analysis of the V9 region of 18S rDNA in the environmental DNA from seawater through amplicon sequencing. Samples were collected monthly between September 2021 and December 2023, across different sampling sites within the context of the NEREA observatory in the Gulf of Naples (Italy). Seawater was collected at discrete depths, filtered through 0.45 µm filters, and the environmental DNA was extracted from and amplified by PCR.

Sampling area: Gulf of Naples (Campania, Italy)

Sampling protocol: https://doi.org/10.1038/s41597-024-03787-y

eDNA analysis: https://doi.org/10.1098/rstb.2023.0178

Metadata: https://doi.org/10.5281/zenodo.17669960

Dataset 6: 16S rDNA amplicon sequencing of NEREA environmental DNA samples

Grant Agreement: 101082021

Project Acronym: MARCO-BOLO

Project Title: MARine COastal BiOdiversity Long-term Observations

Deliverable Number: D2.2

Work Package Number: WP2

Deliverable Title: Datasets, databases and softwares/pipelines facilitating the implementation of eDNA-based monitoring

Dataset name: 16S rDNA amplicon sequencing of NEREA environmental DNA samples

License/access conditions (FAIR principles): This dataset is published in the European Nucleotide Archive (ENA) at EMBL-EBI under the project accession PRJEB100988 (current status: private; release date: 30th June 2026).

Short description: Analysis of the V4-V5 regions of 16S rDNA in the environmental DNA from seawater through amplicon sequencing. Samples were collected in the same period as the V9-18S samples, with monthly sampling between September 2021 and December 2023, across different sampling sites within the context of the NEREA observatory in the Gulf of Naples (Italy). Seawater was collected at discrete depths, filtered through 0.45 µm filters, and the environmental DNA was extracted and amplified by PCR.

Sampling area: Gulf of Naples (Campania, Italy)

Sampling protocol: https://doi.org/10.1038/s41597-024-03787-y

eDNA analysis: https://doi.org/10.1038/s41597-024-03787-y

Metadata: https://doi.org/10.5281/zenodo.17670063

Dataset 7: Droplet digital PCR (ddPCR) of NEREA environmental DNA samples

Grant Agreement: 101082021

Project Acronym: MARCO-BOLO

Project Title: MARine COastal BiOdiversity Long-term Observations

Deliverable Number: D2.2

Work Package Number: WP2

Deliverable Title: Datasets, databases and softwares/pipelines facilitating the implementation of eDNA-based monitoring

Dataset name: Droplet digital PCR (ddPCR) of NEREA environmental DNA samples

License/access conditions (FAIR principles): This dataset is published in Zenodo under the MARCO-BOLO community accession (current status: private; release date: 30th November 2026).

Short description: Detection of European anchovy (*Engraulis encrasicolus*) DNA in seawater using droplet digital PCR (ddPCR). Samples were collected between January 2020 and August 2021, across different sampling sites within the context of the NEREA observatory in the Gulf of Naples (Italy). Seawater was filtered to obtain eDNA, which was subsequently analysed by ddPCR using species-specific primers designed to target a fragment of the mitochondrial cytochrome b (MT-CytB) gene of *Engraulis encrasicolus*. This allowed sensitive quantification of the species, and the number of target DNA copies per µL in each sample was subsequently estimated.

Sampling area: Gulf of Naples (Campania, Italy)

Protocol: https://doi.org/10.5281/zenodo.17670154 (ddPCR protocol) - https://doi.org/10.1098/rstb.2023.0178 (sampling eDNA extraction protocol)

Data and Metadata: https://doi.org/10.5281/zenodo.17670287

Dataset 8: Tara Oceans Global Survey - Plankton and Microbial eDNA Variables

MARCO-BOLO Deliverable 2.2, Dataset 8

Grant Agreement: 101082021 | Work Package: WP2 | Deliverable: D2.2

Sequences: List of sub-datasets that make up the Tara Oceans Global Survey, deposited at the European Nucleotide Archive.

- Tara Oceans umbrella project accession number: PRJEB402
- Accession numbers of the 18S datasets: PRJEB6610, PRJEB9737
- Accession numbers 16S V4V5: PRJEB36282, PRJEB36283, PRJEB36284, PRJEB36285, PRJEB4357
- Accession numbers of the shotgun metagenomic data (used to obtain the "miTags"): PRJEB1788, PRJEB9691, PRJEB9740, PRJEB9742, PRJEB1787, PRJEB4352

Short description: This dataset was generated outside of the Marco-Bolo project, by the Tara Oceans expedition (2009-2013). The expedition surveyed 210 ecosystems in 20 biogeographic provinces, collecting over 35,000 samples of seawater and plankton. This dataset contains V4-18S, V9-18S, V4V5-16S markers and miTags. MiTags were obtained by mapping metagenomic sequences to 16S/18S reference sequences as described in Salazar et al., 2019. Sampling strategy details and protocols can be found in Pesant et al., 2015. The sequences from this dataset are archived in the European Nucleotide Archive (ENA).

Processed datasets

For generating eDNA based EOVs, abundance tables generated from the sequencing data were available in Zenodo and at the ocean microbiome website.

18S V4 datasets:

- 18S V4 Swarm (https://zenodo.org/records/7235995)
- 18S V4 dada2 (https://zenodo.org/records/13881376)

18S V9 datasets:

- 18S V9 Swarm (https://zenodo.org/records/7236051)
- 18S V9 dada2 (https://zenodo.org/records/13881418)

16S V4V5 dataset (to be released around January 2026):

- 16S V4V5 dada2 (https://zenodo.org/records/7551744)

miTags dataset: https://ocean-microbiome.org/

Dataset 9: SOMLIT-Astan Timeseries - 2009-2016 - Metabarcoding Total Eukaryotes (18S V4 region rRNA gene)

MARCO-BOLO Deliverable 2.2, Dataset 9

Grant Agreement: 101082021 | Work Package: WP2 | Deliverable: D2.2

Sequences: PRJEB48571 (European Nucleotide Archive)

Short description: This dataset was generated outside of the Marco-Bolo project, but we were kindly provided access by Nicolas Henry. This dataset contains information on the V4-18S rRNA marker as presented in Caracciolo *et al*. (2022), spanning the years 2009 to 2016. Samples were collected from surface waters at a depth of 1-2 meters through bimonthly sampling (twice a month). The dataset includes two different size classes: 3 µm and 0.22 µm (Sterivex). For the 3 µm size class, 5 liters of water were filtered. The sequences from this dataset are archived in the European Nucleotide Archive (ENA).

Sampling area: The offshore station SOMLIT-Astan, at 2 nautical miles off Roscoff (48° 46 '18”N, 3° 58’ 6”W) with a permanently mixed water column at around 60 m of depth. (https://www.seanoe.org/data/00854/96634/)

Protocols

- **Sampling protocol, DNA extraction, PCR (18S), library preparation.**

See Caracciolo *et al*. (2022).

- **Bioinformatics**

Pipeline for QA/QC, denoising, and taxonomic assignment. https://gitlab.sb-roscoff.fr/somlit-astan-metab/asv-tables/18sv4-ampliseq/-/tree/v1.1.0?ref_type=tags.

Usage Limitations

- **Status:** Dataset is marked partly published — raw sequencing data and an original ASV table were published in Caracciolo *et al*. (2022), but the specific dataset used here is an ASV table generated with v1.1.0 of the pipeline specified above. This ASV table will eventually be made available at SEANOE with the doi:10.17882/107693, but updates may be implemented prior to final publication(s). Previous version is available at doi:10.5281/zenodo.8392524.
- **Protocols:** Supporting protocols are published in Caracciolo *et al*. (2022).
- **Taxonomic scope:** 18S-V4 primers, broad target of eukaryotes
- **Spatial scope:** Single station observatory in the English Channel; regional generalisability requires caution

Maintainer: Nicolas Henry (Station Biologique de Roscoff: Roscoff, Bretagne, FR)

Last updated: 2025-11-28

Dataset 10: Auxiliary primer, reference database and pipeline tables

As part of WP2's work, three tables were generated to provide an overview of frequently used eDNA primers (Table 1), reference databases (Table 2), and pipelines (Table 3). These are available and described in greater detail on GitHub (https://github.com/MadsRJ/MBO_WP2_Tables/). At the time of writing these tables were living documents but will be submitted to Zenodo before the end of the project (Nov. 2026).

Dataset 11: Auxiliary reference database creation examples

While many eDNA practitioners rely on existing databases for inferring taxonomic identity of their sequencing data, we here provide two examples of custom reference database creation. Although not yet fully developed and tested, this WP2 endeavor has laid the groundwork for 'NordicRefDBs' (https://github.com/MadsRJ/NordicRefDBs/) – including code for building reference databases from scratch. The philosophy of this work is to curate GenBank/BOLD records of specimens from Nordic countries.

References

Amaral-Zettler, L. A., McCliment, E. A., Ducklow, H. W., Huse, S. M. (2009). A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. PLoS One 4(7), e6372. https://doi.org/10.1371/journal.pone.0006372.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S. M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J. A., Smith, G., Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME Journal 6(8), 1621–1624. http://doi.org/10.1038/ismej.2012.8.

Caracciolo, M., Rigaut-Jalabert, F., Romac, S., Mahé, F., Forsans, S., Gac, J. P., Arsenieff, L., Manno, M., Chaffron, S., Cariou, T., Hoebeke, M., Bozec, Y., Goberville, E., Le Gall, F., Guilloux, L., Baudoux, A.-C., de Vargas, C., Not, F., Thiébaut, E., Henry, N., Simon, N. (2022). Seasonal dynamics of marine protist communities in tidally mixed coastal waters. Molecular Ecology, 31(14), 3761–3783. https://doi.org/10.1111/mec.16539.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal, 17(1), 10. https://doi.org/10.14806/ej.17.1.200.

Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., Le Bescot, N., Gorsky, G., Iudicone, D., Karsenti, E., Speich, S., Troublé, R., Dimier, C., Searson S., Tara Oceans Consortium Coordinators. (2015). Open science resources for the discovery and analysis of Tara Oceans data. Scientific Data 2, 150023. https://doi.org/10.1038/sdata.2015.23.

Rigaut-Jalabert, F., Guilloux, L., Hoebeke, M., Forsans, S., Caracciolo, M., Simon, N. (2021). Morphological phytoplankton counts for the SOMLIT-Astan time-series (2007-2017) (Version 2) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.5033180.

Salazar, G., Paoli, L., Alberti, A., Huerta-Cepas, J., Ruscheweyh, H.-J., Cuenca, M., Field, C. M., Coelho, L. P., Cruaud, C., Engelen, S., Gregory, A. C., Labadie, K., Marec, C., Pelletier, E., Royo-Llonch, M., Roux, S., Sánchez, P., Uehara, H., Zayed, A. A., Zeller, G., Sunagawa, S. (2019). Gene expression changes and community turnover differentially shape the global ocean metatranscriptome. Cell 179(5), 1068-1083.e21. https://doi.org/10.1016/j.cell.2019.10.014.

Stoeck, T., Bass, D. Nebel, M., Christen, R., Jones, M. D. M., Breiner, H.-W., Richards, T. A. (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Molecular Ecology 19(s1), 21-31. https://doi.org/10.1111/j.1365-294X.2009.04480.x.

Takahashi, M., Frøslev, T. G., Paupério, J., Thalinger, B., Klymus, K., Helbing, C. C., Villacorta-Rath, C., Silliman, K., Thompson, L. R., Jungbluth, S., Yong, S. Y., Formel, S. K., Jenkins, G., Laporte, M., Deagle, B., Rajbhandari, S., Stjernegaard, T. J., Bissett, A., Jerde, C. L. Hahn, E. E., Schriml, L. M., Hunter, C.,

Newman, P., Woollard, P., Harper, L. R., Dunn, N., West, K., Haderlé, R., Wilkinson, S., Acharya-Patel, N., Lopez, M. L. D., Cochrane, G., Berry, O. (2025). A metadata checklist and data formatting guidelines to make eDNA FAIR (Findable, Accessible, Interoperable, and Reusable). Environmental DNA, 7(3). https://doi.org/10.1002/edn3.70100.


MARCO-BOLO
STRENGTHENING BIODIVERSITY OBSERVATION IN SUPPORT OF DECISION MAKING

Project Coordinator
Nicolas Pade | nicolas.pade@embrc.eu

Project Manager
Giulia Vecchi | giulia.vecchi@embrc.eu

Press and Communications
Mathilde Vidal | mathilde@erinn.eu
Website: MarcoBolo-Project.eu
Twitter: @MARCOBOLO_EU
LinkedIn: MARCO-BOLO

Funded by the European Union

Funded by the European Union under the Horizon Europe Programme, Grant Agreement No. 101082021 (MARCO-BOLO). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

UK Research and Innovation

UK participants in MARCO-BOLO are supported by the UKRI's Horizon Europe Guarantee under the Grant No. 10068180 (MS); No. 10063994 (MBA); No. 10048178 (NOC).