



NATIONAL BIODIVERSITY FUTURE CENTER

SPOKE 2 Linea 5 (A5) Develop innovative Multi-Omics based technologies to address emergent biodiversity threats.

D3.1 Report on omics SOPs in Marine Biodiversity science.

Standardised Operational Procedures (SOPs) for Omics analysis of
Marine Samples

**Università di Genova (UNIGE)
Consiglio Nazionale delle Ricerche (CNR)
Università Politecnica delle Marche (UNIVPM)
Stazione Zoologica Anton Dohrn (SZN)
Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (OGS)
Università di Bologna (UNIBO)**

List of contributors (in alphabetical order): Elisa Banchi, Emanuele Bosi, Alessia Cariani, Raffaella Casotti, Michela Castellano, Bruno Cataletto, Mauro Celussi, Manuela Coci, Cinzia Corinaldesi, Antonio Dell'Anno, Iole Di Capua, Simone Di Piazza, Cinzia Fabbro, Alice Ferrari, Carmelo Fruciano, Daniele Iudicone, Gian Marco Luna, Gabriella Luongo, Sarah Magozzi, Vincenzo Manna, Francesco Massa, Paolo Povero, Grazia Quero, Diana Sarno, Elisa Taviani, Anna Chiara Trano, Stefano Varrella, Paolo Vassallo, Luigi Vezzulli, Silvia Viaggi, Mirca Zotti

TABLE OF CONTENTS

Foreword	3
1. MARINE SAMPLE COLLECTION AND NUCLEIC ACIDS EXTRACTION	4
1.1 Seawater samples	4
1.2 Sediment samples	4
1.3 Plankton samples	6
1.4 Macrofauna	7
1.5 Corals samples	7
1.6 Fish samples	8
1.7 Exhaled breath condensate of cetaceans	8
2. METABARCODING OF MARINE SAMPLES	10
2.1 Metabarcoding of Marine Prokaryotes communities through 16S rRNA gene (V4-V5 region) sequencing	10
2.2 Metabarcoding of Marine Protist communities through 18S rRNA gene (V4-V9 region) gene sequencing	11
2.3 Metabarcoding of Marine Fungal assemblages through ITS2 sequencing	13
2.4 Metabarcoding of Marine Zooplankton communities through mitochondrial COI sequencing	15
2.5 Metabarcoding of Marine Fish communities	16
2.6 Metabarcoding of Marine Mammals	17
3. SHOTGUN METAGENOMIC ANALYSIS OF MARINE SAMPLES	17
4. VIRAL METAGENOMIC SEQUENCING	17
5. METATRANSCRIPTOMICS ANALYSIS OF MARINE SAMPLES	18
6. MARINE OMICS OBSERVATORIES	20

Foreword

Marine biodiversity is severely affected by multiple threats mostly resulting from human activities and climate driven unusual and catastrophic events. The impacts include deterioration of natural populations, invasion by exotic species, local extinction of native species, reduction of genetic diversity, loss and alteration of ecosystem dynamics and disease outbreaks linked to the introduction of pathogens and antibiotic resistant bacteria. To develop an effective response to these challenges new tools are needed that integrate science-based knowledge and offer the necessary information to formulate and implement effective action plans for assessing, monitoring, preserving and restoring Marine Mediterranean Biodiversity. To this end, Activity 5 will Implement last-generation technologies for the collection of Marine Omics data, their analysis and interpretation to implement Scalable, Fast and Cost-effective (SFC) response actions to Emergent Biodiversity threats. The activity will also establish an Italian Omics Observatory Network fostering the development and application of last generation biomolecular technologies for present and future Marine Biodiversity studies in the Mediterranean Sea. The network will be part of the National Marine Biodiversity Observatory System and it will be developed in close partnership with NBFC WP1-Spoke 1 in four main marine stations (Golfo di Napoli, Promontorio di Portofino, Golfo di Trieste and stazione Meda TeleSenigallia) where marine biodiversity and oceanographic monitoring programmes are well established.

This manual describes Standard Operating Procedures (SOPs) for omics analysis of marine samples including sample collection, preservation, nucleic acid extraction and analytical workflows. SOPs employed by the Italian Marine Omics Observatory are also described.

1. MARINE SAMPLE COLLECTION AND NUCLEIC ACIDS EXTRACTION

1.1 Seawater samples

For DNA analysis, as general procedure, the samples, whose volume varies with sites' characteristics (e.g, for Naples 5 L have proven to be satisfactory for both spring samples, which are very dense and for summer samples, which are the poorest in terms of biomass) are collected at selected depths by means of Niskin bottles, disposed into clean containers (e.g., HCl-washed plastic tanks) and filtered within 6 hours from collection. Three size fractions are collected: 0.2-3 μm , 3-20 μm , 20-200 μm . The protocol is a modification of the EMOBON and NEREA protocols (<https://repository.oceanbestpractices.org/handle/11329/1738> and www.nerea-observatory.org)

In detail:

Seawater samples are collected into clean containers (e.g. carboys) of appropriate volume. Ideally, seawater should be filter as soon as possible after collection, but if this is not possible, care should be taken to avoid to expose them to direct light or excess temperature. Once ready, seawater should be sieved through an acid-clean 200 μm metal sieve, and then sequentially filtered through 3 membrane filters (20 μm , Φ 47 mm; 3 μm and 0.2 μm , both Φ 142 mm) by means of appropriate filter rigs. The filters are to be handled only with forceps previously soaked in HCl and then rinsed in filtered seawater. Gloves should be worn at every time. Once finished the first round of filtration, extract the large filters (142 mm) from the filter holders and cut rapidly the filter into 4 visually equal pieces by using a sterile scalpel, after placing it on a clean, hard surface (e.g. in a Petri dish), possibly in a sterile environment (e.g. a hood). Immediately fold each piece by rolling it in a cylinder. Carefully place the filters into separate 5 ml cryotubes appropriately labelled. After the second seawater filtration, collect and store the negative control sample. Close the tripod with no filters and pass 2 L of MilliQ water (or Nanopure). Place new filters in the tripods and filter the same volume of MilliQ water as the seawater samples. e.g. if the observatory collects 5 L of seawater sample, filter 5 L of MilliQ. Extract the filters and process them as before.

At the end, the procedure will generate 6 tubes with half-cut filters of 0.22 μm (fraction 0.22-3); 6 tubes with half-cut filters of 3 μm (fraction 0.3-20); 2 tubes with 47 mm filters of 20 μm (fraction 20-200 μm) for a total of 18 tubes

The filters are stored at -80 °C until further processing. The DNeasy PowerWater Kit (QIAGEN) is recommended for DNA extraction according to the manufacturer's instructions. DNA's quality and quantity is assessed using NanoDrop or Qubit or Quantus). **For RNA analysis**, samples (volume depending on the sites' specifics) are collected at selected depths by means of Niskin bottles, disposed into clean containers (e.g., HCl-washed plastic tanks) through a 200 µm nylon mesh. Filtration is then performed onboard by means of portable peristaltic pumps onto Sterivex filters (0.22 µm, Millipore) to avoid external contamination. To minimize RNA transcriptional changes and degradation, filtration is carried out within 30 minutes from sampling or until filter clogs (if it occurs before this timeframe). The Sterivex cartridges are then filled (~1.8 mL) with DNA/RNA shield (Zymo Research), sealed, kept in a cold container until the home lab is reached and then stored at -80 °C until extraction. The RNeasy PowerWater Kit (QIAGEN) is recommended for RNA isolation according to the manufacturer's instructions.

References:

- Santi I, Beluche O, Beraud M, Buttigieg PL, Casotti R, Cox CJ, Cunliffe M, Davies N, de Cerio OD, Exter K, Kervella AE, Kotoulas G, Lagaisse R, Laroquette A, Louro B, Not F, Obst M, Pavloudi C, Poulain J, Præbel K, Vanaverbeke J and Pade N (2023) European marine omics biodiversity observation network: a strategic outline for the implementation of omics approaches in ocean observation. *Front. Mar. Sci.* 10:1118120. doi: 10.3389/fmars.2023.1118120
- Fanelli E., Aguzzi J., Casotti R., Conversano F., D'Alelio D., Iudicone D., Marini S., Stefanni S., 2019 NEREA, the Naples Ecological REsearch for Augmented observatories: towards an end-to-end transdisciplinary approach for the study of marine ecosystems. 2019 IMEKO TC-19 International Workshop on Metrology for the Sea Genoa, Italy, October 3-5, 2019 Proceedings: 103-107 (2019) <http://hdl.handle.net/10261/207800>

1.2 Sediment samples

Samples (top 1 cm) are collected in cryovial or sterile petri dishes and stored at -80 °C until further processing. If you use a Multi-corer or Van Veen grab, use sterile gloves to collect the surface sediment layer (top 1 cm) with a microspatula into a pre-labelled cryovial/petri dish. Microspatula should be sterilised with ethanol wipes between grabs and gloves changed between samples. If you collect while diving, use a 50 ml Flacon tubes to scoop up the top 1 cm of sediments. Falcon tubes are stored

at -80°C from collection until further processing. For collection of multiple sediment layers use a corer, a clean rod to push out sediment and a sterile microspatula to collect the different layers (e.g. 0–1/1–3 cm /3-5 cm depth).

Before total DNA extraction from the sediments, samples are pre-treated to remove as much extracellular DNA as possible and improve the success of DNA amplification by PCR (Danovaro et al., 2010). Sediments are resuspended with 5 ml of Wash Solution 1 (50 mM Tris-HCl, pH 8.3; 200 mM NaCl; 5 mM Na₂EDTA; 0.05% Triton X-100), vortexed and centrifuged at 6000x g for 2 mins. Once discharged the supernatants, the pellets are washed with 5ml of Wash Solution 2 (50 mM Tris-HCl, pH 8.3; 200 mM NaCl; 5 mM Na₂EDTA) and Wash Solution 3 (10 mM Tris-HCl, pH 8.3; 0.1 mM Na₂EDTA) and centrifuged as described above between each washing. Total DNA is then extracted from the resulting pellet using the DNeasy PowerSoil Kit (QIAGEN) following manufacturer instructions. The purity of DNA is checked by determining the ratio of absorbance at 260 nm to absorbance at 280 nm, and the DNA is quantified by NanoDrop ND-1000 Spectrophotometer or Qubit™ 3 Fluorometer or similar instruments.

For viral metagenomics, viral particles are extracted from marine sediment samples according to Corinaldesi et al. (2017). Briefly, 50 g of wet sediment is mixed with 50 ml of virus-free seawater (pre-filtered onto 0.02-µm-pore-size filters and autoclaved) into a sterile jar and aliquots of 2 ml of the sediment slurry were split into 50 sterile tubes to increase the viral recovery efficiency. Each sample is then supplemented with tetrasodium pyrophosphate solution (5 mM final concentration), incubated for 15 min s in the dark, and sonicated three times (Branson Sonifier 2200, 60W) for 1 min. After centrifugation (800 xg for 10 min), the supernatants are collected and the sediments are resuspended again with 4 ml of virus-free seawater, centrifuged (800 xg for 10 min), and the supernatant collected. This step is repeated three times. The supernatants are combined and incubated with DNase I (1 U/ml) in the dark at room temperature for 15 mins to remove extracellular DNA. The supernatants are pre-filtered through 0.2 µm pore size filters (Whatman Anopore) to remove cellular organisms before concentrating viral particles onto 0.02 µm pore size Al₂O₃ filters (Anodisc; diameter 47 mm). Filters containing the collected viruses are subjected to DNA extraction by using the DNeasy PowerSoil Kit (QIAGEN) following manufacturer instructions.

The DNA concentrations in all samples are quantified fluorometrically with the NanoDrop ND-3000 (ThermoFisher, Waltham, US) by using SYBR Gold or Qubit™ 3 Fluorometer by using Qubit 1X dsDNA HS Assay Kit. If the amount of viral DNA is not sufficient to be directly sequenced, viral DNA is subjected to amplification using Multiple Displacement Amplification (MDA) such as the highly processive, proof-reading proficient, Phi29 polymerase (GenomiPhi kit) and random hexamer primers, before further processing. Samples amplified with the GenomiPhi kit are then processed with the DNeasy PowerClean Cleanup Kit (QIAGEN) following manufacturer instructions.

References

- Corinaldesi, C., Tangherlini, M., & Dell'Anno, A. (2017). From virus isolation to metagenome generation for investigating viral diversity in deep-sea sediments. *Scientific Reports*, 7(1), 8355.
- Danovaro R. Methods for the Study of Deep-Sea Sediments, Their Functioning and Biodiversity. CRC Press 2010; p. 458

1.3 Plankton samples

- a) Phytoplankton samples are collected by a 20 µm mesh size plankton net while the boat proceeds at very low speed, towing the net horizontally at the surface for 3-7 minutes, depending on the transparency of the water. The sample collected in the cod end is sieved through a 200 µm metal sieve (or 200 µm net) and diluted to 1 L by adding filtered seawater. The sample is then filtered through 4 10 µm polycarbonate membrane filters (Φ 47 mm) in 250 ml aliquotes, generating 4 replicates. At the end of the filtrations, one filter rig is rinsed with MilliQ water and 250 ml of MilliQ water are filtered through the same filter rig. If multiple filtration rigs are used, the first one could be chosen for this blank. After filtration, membranes are folded, placed in single cryotubes appropriately labelled, preserved in DNA/RNA Shield (optional), and stored in liquid nitrogen or at -80 °C until further processing. 5 tubes are generated. The DNeasy PowerWater Kit (QIAGEN) is recommended for DNA extraction following the manufacturer's instructions. DNA's quality and quantity is assessed using NanoDrop, Qubit, or Quantifluor). The RNeasy Mini Kit (QIAGEN) is recommended for RNA extraction following the manufacturer's instructions. b) Zooplankton samples are collected by vertical tows using a 200

µm mesh size plankton net. The depth of the net towing depends on the characteristics of the station, e.g. from 50 m depth to the surface for deeper sites (> 50m depth) or from bottom to the surface in case of a shallower site (<50 m depth). The sample collected in the cod end is sieved through a 2000 µm metal sieve and diluted to 1 L. The sample is then filtered through 4 10 µm polycarbonate membrane filters (Φ 47 mm) in 250 ml aliquotes, generating 4 replicates. At the end of the filtrations, one filter rig is rinsed with MilliQ water and 250 ml of MilliQ water are filtered through the same filter rig. If multiple filtration rigs are used, the first one could be chosen for this blank. 5 tubes are generated. After filtration, membranes are folded, placed in single tubes, preserved in DNA/RNA Shield (TBD) (optional), and stored in liquid nitrogen or at -80 °C until further processing.

Optional (NEREA observatory protocol): The zooplankton sample collected in the cod end is gently mixed and filtered on a 200 µm nitex mesh filter. The bulk zooplankton (pellet) collected on the filter is manually divided into four subsamples by means of an acid-clean spatula, which are placed in 2 mL cryogenic vials and immediately frozen in liquid nitrogen onboard and later stored at -80 °C, or fixed in ethanol 95%.

1.4 Macrofauna

This includes molluscs, crustaceans, sponges, annelids, small tunicates, etc. Collect whole organisms (n=3) or cut specimens samples underwater into small pieces (~5 g) using sterile blades. Document each specimen with a photo. Transfer organism/pieces into a 50 ml Falcon tube. On the surface after the dive use sterile gloves to wash animal three times with sterile artificial seawater to detach loosely attached bacteria and stored at -80°C or DNA/RNA shield (TBD) until further processing. The DNeasy PowerSoil kit (QIAGEN) is used for DNA extraction following the manufacturer's recommendation. DNA's quality and quantity is assessed using spectrofluorimetric measurements (e.g., Quantifluor). The RNeasy Mini Kit (QIAGEN) is used for RNA extraction following the manufacturer's recommendation.

1.5 Coral samples

Coral mucus samples are collected from 3 colonies in the field with autoclaved cotton swabs placed in 2 ml Eppendorf tubes. Invert those tubes and open underwater next to the coral surface to minimize contamination with microbes from the surrounding seawater. Swabs should be gently rolled over the coral surface and then placed back in the tube. To detach mucus specimens from swabs, each sample will be repeatedly vortexed and sonicated immersed in sterile artificial seawater. Coral fragments are collected from the same colonies using a hammer and chisel. Document each specimen with a photo. For each colony coral tissue is separated from the carbonate skeletal matrix by mechanical fragmentation in an agate mortar. The holobiont homogenate added of sterile artificial seawater is incubated to allow skeletal fragments to settle. The tissue suspension is then centrifuged at 9,300g for 15 mins at 4°C to pellet coral tissue fraction. Coral skeleton fragments are then washed using sterile artificial seawater and transferred into a Eppendorf tube. Mucus cellular pellets, skeleton, and tissue are then stored frozen at -80°C or in DNA/RNA shield (TBD) until further processing. The DNeasy PowerSoil kit (QIAGEN) is used for DNA extraction following the manufacturer's recommendation. DNA's quality and quantity is assessed using spectrofluorimetric measurements (e.g., Quantifluor). The RNeasy Mini Kit (QIAGEN) is used for RNA extraction following the manufacturer's recommendation.

1.6 Fish samples

Fish samples for genomics and transcriptomics are obtained from tissue biopsies of fish mainly caught by collaborators (fishermen) through nets or spearfishing. Samples of approximate size of 0.15-0.25 cm³ are collected from muscles (for DNA and RNA extractions), liver, gonads and brain (for RNA extractions only), according to protocols of the European Reference Genome Atlas (ERGA) initiative ([https://www.erga-biodiversity.eu/following the protocol](https://www.erga-biodiversity.eu/following%20the%20protocol)). All samples are collected in at least two replicates, stored in RNeasy (Thermo Fisher) or equivalent or 95% ethanol (DNA samples only) and kept refrigerated (frozen in the case of RNeasy samples to be kept for more than 24 hours prior to RNA extraction). To avoid contamination between samples tools (razor blades, scissors) are flame-sterilised after each tissue sampling (see also sampling protocols from MED_UNITS project-

https://cinea.ec.europa.eu/publications/study-advancing-fisheries-assessment-and-management-advice-mediterranean-aligning-biological-and_en).

Total genomic DNA extraction is performed on muscle samples using commercial kits (e.g., PureLink™ Genomic DNA Mini Kit, Invitrogen™), while RNA is extracted from the relevant tissues using either standard extraction with TRIzol™ Reagent (Invitrogen™) or commercial kits. Both operations are conducted according to the manufacturer's protocol. DNA and RNA purity is ascertained by checking the ratio between absorbance at 260 nm and 280 nm using the NanoDrop ND-1000 spectrophotometer, while the Qubit™ 3 fluorometer is used to estimate the amount of the nucleic acids contained in every sample. Both DNA samples and RNA samples are sequenced by short-read Illumina technology at dedicated partner or commercial facilities.

1.7 Exhaled breath condensate of cetaceans

Exhaled Breath Condensate (EBC) samples and photogrammetry data (total length/body mass, needed to assess body condition/health state) will be collected - simultaneously - from large whales (fin whales and sperm whales) encountered in the Ligurian Sea during the dedicated at-sea campaigns carried out by CIMA. EBC and photogrammetry data will be obtained using two different UAVs (VTOL). The UAV dedicated to photogrammetry will be equipped with a 35mm camera and LIDAR (for centimetric resolution altitude). Photogrammetry data will be obtained collecting vertical (nadiral) images of the whale total body, taken @40m altitude. EBC dedicated UAV will be equipped with two sampling kits (SKs) mounted on the UAV bow and top surface respectively. EBC UAV will fly at about 2 m over the whale's blowhole. Each SKs will be composed by a 6 Well Non-Treated Sterile Flat Tissue Culture Plate. For each survey day, water and air controls will be obtained from the seawater and from a SK exposed to air for 30 mins, respectively. The sample in SKs will be stored with RNAlater in microtubes in a refrigerated container.

DNA concentration (qubit dsDNA HS Assay) will be determined, and samples stored at -80°C. Specific primers for V4-V5 hypervariable regions of the 16S rRNA gene (Parada et al., 2016), will be used to cover bacteria and archaea diversity. The V4 hypervariable region of the 18S rRNA gene will be amplified with the primer set described in Piredda et al. (2017) to cover fungi and protists diversity. The obtained amplicon will be further prepared for sequencing in an Illumina MiSeq platform, based

on V3 chemistry, delivering an expected ~100K reads per sample. EBC samples showing potential or known human/animal pathogens will be selected for in-deep shotgun metagenomics. Metagenomics shotgun libraries will be prepared and sequenced using paired-end sequencing V3 chemistry of Illumina MiSeq technology. Bioinformatic analysis of the genomic and metagenomic data generated by Illumina NGS will be performed based on standard methodologies and bioinformatic pipeline (de Sousa et al., 2019).

References:

- de Sousa AGG, Tomasino MP, Duarte P, Fernández-Méndez M, Assmy P, Ribeiro H et al (2019). Diversity and Composition of Pelagic Prokaryotic and Protist Communities in a Thin Arctic Sea-Ice Regime. *Microbial Ecology* (78): 388–408.
- Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol* 18, 1403–1414 (2016).
- Piredda R, Tomasino MP, D'Erchia AM, Manzari C, Pesole G, Montresor M et al (2017). Diversity and temporal patterns of planktonic protist assemblages at a Mediterranean Long Term Ecological Research site. *FEMS Microbiology Ecology* 93 (1), fiw200

2. METABARCODING OF MARINE SAMPLES

2.1 Metabarcoding of Marine Prokaryotes communities through 16S rRNA gene (V4-V5 region) sequencing

The prokaryotic (bacteria and archaea) communities in marine samples are investigated by amplification of the 16S rRNA gene (V4 and V5 hypervariable regions, amplicon size (~400bp) using the primer set 515F-Y and 926R described by Parada et al. (2016).

Amplicons and libraries are prepared according to 16S Metagenomic Sequencing Library preparation protocol (15044223 Rev. B) and run on an Illumina platform (MiSeq 2x300bp or Novaseq 2X250bp) to obtain 200K reads per sample. Bioinformatic analyses are performed with QIIME2 (Bolyen et al. 2019). Raw sequences are quality filtered and denoised with DADA2 (Callahan et al. 2016). Taxonomy is assigned to the amplicon sequence variant (ASVs) using the more

updated Silva reference database (Quast et al., 2013). Statistical analyses and alpha and beta diversity estimation are performed in the R environment (R Core Team, 2019). Briefly:

- 1st PCR Gen specific primers targeting 16S rRNA v4-v5 region:

16S Amplicon PCR Forward Primer = 5': 515F-Y (5'-GTGYCAGCMGCCGCGGTAA-3')

16S Amplicon PCR Reverse Primer = 5': 926R (5'-CCGYCAATTYMTTTRAGTTT-3')

- PCR reaction are run in 25 µl reactions using 2xKAPA HiFi Hotstart Ready Mix, 1µM of each primer and 5ng/µL DNA template
- PCR cycles: 95°C for 3 min s, 25 cycles of: 98°C for 10 sec s, 53°C for 30 sec s, 72°C for 30 sec s, and 72°C for 10 min s.
- 1st PCR Clean up PCR reactions is purified with AMPure XP beads
- 2nd PCR Index reactions are run in a 50 µl reaction with KAPA Hifi HotStart and Illumina Nextera indexing primers
- PCR cycles: 95°C for 3 min s, 8 cycles of: 95°C for 30 sec s, 55°C for 30 sec s, 72°C for 30 sec s. And 72°C for 5 min s.
- 2nd PCR reactions are purified with AMPure XP beads.

Samples are validated on Agilent Bioanalyser and quantified using the Qubit or Picogreen assay (Invitrogen).

- Concentrations are calculated, and samples are pooled to a final concentration of 4 nM and loaded on the MiSeq Reagent Kit v3, 600 cycles. Up to 125 barcodes will be processed in the same library in order to achieve a sufficient amount of high-quality sequencing reads (200,000 per sample) for downstream analysis.

For samples with low bacterial and archaeal fraction (e.g., plankton, coral, macrofauna and fish) a nested PCR is performed by applying a first full-length amplification using the 27 F/1492 R 16S universal primer set in order to increase the target DNA (

16S Amplicon PCR Forward Primer = 5': 27F 5'-AGAGTTTGATCMTGGCTCAG-3'

16S Amplicon PCR Forward Primer =5': 1492R 5'-

TACGGYTACCTTGTTACGACTT-3') and a second amplification using the 515F-Y/926 R primers (Caroline Belser et al., 2023)

References:

- Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology* 37, 852–857 (2019).
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nature Methods* 13, 581–583 (2016).
- Belser et al. (2023) Integrative omics framework for characterization of coral reef ecosystems from the Tara Pacific Expedition. *Scientific Data* 10:326
- Parada AE, Needham DM, Fuhrman JA (2016). Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology* 18, 1403–1414 (2016).
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P et al (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 41, 590–596.
- R Core Team. R: A language and environment for statistical computing [Computer software]. Vienna: R Foundation for Statistical Computing (2019).

2.2 Metabarcoding of Marine Protist communities through 18S rRNA gene (V4-V9 region) gene sequencing

The V4 and V9 regions of eukaryote SSU rDNA gene are amplified using primers by Stoeck et al. (2010), modified as in Piredda et al. (2017). The protocol is based on two amplification steps. In the first amplification, V4 and V9 regions of the 18S rRNA gene are amplified using the selected V4 and V9 universal primers having a 5' end overhang sequence, corresponding to Nextera transposase primer.

Primers used for the amplification of V4 18S rRNA:

V4 18S Next.For =

5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG[CCAGCASCYGCGGTAATTCC]*-3'

V4 18S Next.Rev =

5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG[ACTTTCGTTCTTGATYRATGA]*-3'

Primers used for the amplification of V9 18S rRNA:

V9 18S Next.For:

5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG[TTGTACACACCGCCCGTCGC]*-3'

V9 18S Next.Rev=

5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG[CCTTCYGCAGGTTCACCTAC]*-3'

Amplifications are performed in a reaction mixture containing 2.5 or 5 ng of extracted DNA (for V9 and V4 region, respectively), 1x Buffer HF, 0.2 mM dNTPs, 0.5 μ M of each primer, and 1 U of Phusion High-Fidelity DNA polymerase (New England Biolabs Inc, Ipswich, MA, USA) in a final volume of 25 μ l. The cycling parameters for PCR are: initial denaturation 98°C for 30 secs, followed by 10 cycles of denaturation at 98°C for 10 secs, annealing at 44 or 56°C (V4 and V9 region, respectively) for 30 secs, extension at 72°C for 15 secs, and subsequently 15 cycles of denaturation at 98°C for 10 secs, annealing at 62°C for 30 secs, extension at 72°C for 15 secs, with a final extension step of 7 mins at 72°C.

All PCRs are performed in the presence of a negative control (RNase/DNase-free water). The PCR products (~270 bp for V9 and ~470 bp for V4) are visualised on 1.2% agarose gel and purified using the AMPure XP Beads (Agencourt Bioscience Corp., Beverly, MA, USA), at a concentration of 1.2x vol/vol, according to manufacturer's instructions.

In the second PCR step the purified V4 and V9 amplicons are used as templates. This amplification is performed with the Nextera index primers and Illumina P5 and P7 primers as required by the Nextera dual index approach. The dual index strategy consists of incorporating unique indices into both ends of the library molecules to allow sample identification for the subsequent bioinformatics analysis. The 50 μ l reaction mixture is made up of the following reagents: template DNA (40 ng), 1x Buffer HF, dNTPs (0.1 mM), Nextera index primers (index 1 and 2) and 1 U Phusion DNA Polymerase and AccuStart II PCR ToughMix or Q5 High-Fidelity 2x Master Mix. The cycling parameters are those suggested by the Illumina Nextera protocol. The dual indexed amplicons obtained (~330 bp for V9 and ~530 bp for V4) are purified using AMPure XP Beads, at a concentration of 0.6x and 1x vol/vol (for V4 and V9 amplicons, respectively), then checked for quality control on 2100 Bioanalyzer

(Agilent Technologies, Santa Clara, CA, USA) and quantified by fluorimetry using the Quant-iT™ PicoGreen-dsDNA Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA) on a NanoDrop 3300 (Thermo Fisher Scientific). Equimolar quantities of V4 and V9 amplicons are pooled and subjected to 2X 250 bp sequencing on a Illumina platform. After read quality assessment with FastQC (Andrews, 2010), primers are trimmed from demultiplexed reads using cutadapt (Martin, 2011). Contig assembly and denoising are performed using the DADA2 R library (Callahan et al., 2016). The resulting ASVs (amplicon sequence variants) are classified against the PR2 v4 database (Guillou et al., 2013)

References:

- Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM (2009) A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes 4(7):e6372
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), 581–583. <https://doi.org/10.1038/nmeth.3869>.
- Martin M (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads *EMBnet.journal* 17:10-12.
- Piredda R, Tomasino MP, D'Erchia AM, Manzari C, Pesole G, Montresor M, Kooistra W. HCF, Sarno D, Zingone A. (2017). Diversity and temporal patterns of planktonic protist assemblages at a Mediterranean Long Term Ecological Research site, *FEMS Microbiology Ecology* 93, fiw200. <https://doi.org/10.1093/femsec/fiw200>
- Stoeck T, Hayward B, Taylor GT, Varela R, Epstein SS (2006) A multiple PCR-primer approach to access the microeukaryotic diversity in environmental samples. *Protist* 157, 31–43 (2006).

2.3 Metabarcoding of Marine Fungal assemblages through ITS2 sequencing

Fungal diversity is investigated by using the primer set ITS3f: 5'-GCATCGATGAAGAACGCAGC -3' and ITS4r: 5'- TCCTCCGCTTATTGATATGC -3' which amplify the internal transcribed spacer-2 (ITS2) region of the fungal rRNA gene (White et al., 1990, Gardes et al., 1993; Nilsson et al., 2019). The amplification is carried out through PCR in a 25 µL final volume using the MyTaq Reaction Buffer 5x, 0.25µM of primers ITS3f and ITS4r, the myTaq HS DNA Polymerase (Meridian

Bioscience), and 1 μ L of template DNA. The amplification consists of initial denaturation at 95 °C for 5 mins, followed by 32 cycles of 94 °C for 40 secs, annealing at 55 °C for 40 secs, and elongation at 72 °C for 90 secs with a final step at 72°C for 7 mins. Amplicons are sequenced on an Illumina MiSeq platform. Paired-end sequences are analyzed within the QIIME2 environment (Boyle et al., 2019). First, the ITSxpress plugin is used to trim sequences targeting the ITS2 region (Rivers et al., 2018); then, trimmed paired-end sequences are analyzed through the DADA2 procedure (Callahan et al., 2016), and the resulting biologically significant Amplicon Sequence Variants (ASVs) are compared against the UNITE database (updated to the last version) for taxonomic affiliation (Rivers et al., 2018). Taxonomic affiliations are performed through the USEARCH SINTAX procedure (Edgar et al., 2018) to evaluate potential distant affiliations. Downstream analyses of ASVs are carried out through packages included in RStudio software (e.g., phyloseq, microeco; McMurdie and Holmes, 2013; Liu et al., 2021).

References:

- Boyle E, Rideout RR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA et al (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2." *Nature biotechnology* 37.8 (2019): 852-857.
- Callahan, Benjamin J., et al. (2016) "DADA2: High-resolution sample inference from Illumina amplicon data." *Nature methods* 13.7: 581-583.
- Edgar RC (2018). "Accuracy of taxonomy prediction for 16S rRNA and fungal ITS sequences." *PeerJ* 6: e4652.
- Gardes M, Bruns TD (1993). ITS primers with enhanced specificity for basidiomycetes-application to the identification of mycorrhizae and rusts." *Molecular ecology* 2.2 (1993): 113-118.
- Liu C, Cui Y, Li X, Yao M (2021). Microeco: an R package for data mining in microbial community ecology." *FEMS microbiology ecology* 97.2: fiae255.
- Nilsson, RH, Anslan S, Bahram M, Wurzbacher C, Baldrian P, Tedersoo L (2019). Mycobiome diversity: high-throughput sequencing and identification of fungi. *Nature Reviews Microbiology*, 17(2), 95-109.
- McMurdie PJ, Holmes S (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*. 8(4):e61217

Rivers AR, Weber KC, Gardner TG, Liu S, Armstrong SD (2018). ITSxpress: Software to rapidly trim internally transcribed spacer sequences with quality scores for marker gene analysis." *F1000Research* 7.1418: 1418.

White TJ, Brun T, Lee S, Taylor J (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics." *PCR protocols: a guide to methods and applications* 18.1: 315-322.

2.4 Metabarcoding of Marine Zooplankton communities through mitochondrial COI sequencing

Zooplankton diversity is investigated by using the primers m1COLint: 5'-FGGWACWGGWTGAACWGTWTAYCCYCC-3' for the forward and the jgHCO2198 5'-TAIACYTCIGGRTGICCRARAAYCA-3' for the reverse (amplicon ~313 bp; Leray et al., 2013). The following procedures are applied in the NEREA project and used for the study by Di Capua et al. (in preparation). The extraction is performed on bulk zooplankton pellets using Cryogrinding followed by Nucleospin RNA/DNA co-extraction (Macherey Nagel).

PCR amplification is performed in 5 to 10 ng genomic DNA, and on three replicate for each sample. Because of the high level of degeneracy in primer sequences, a "touchdown" PCR profile is used to minimize the probability of non-specific amplifications. 16 initial cycles are carried out: denaturation for 10 secs at 95°C, annealing for 30 secs at 62°C (–1°C per cycle) and extension for 60 secs at 72°C, followed by 25 cycles at 46°C annealing temperature. DNA libraries are then prepared using 250 ng of DNA along with proper technical negative controls for each barcoding run according to an in-house "SLC protocol" derived from NEBNext kit for Illumina, NEB. Libraries are paired-end and sequenced with Illumina platform (2X 250 bp). Reads and their mates that map onto run quality control sequences (PhiX genome) are removed and a comparison between samples and their related negative controls (PCR and extraction) is produced. Samples are shipped to the sequencing company in dry ice. Illumina paired-end raw reads are processed to generate Amplicon Sequence Variants (ASVs) using the DADA2 R package (Callahan et al., 2016). The quality profile of the reads is inspected by the plot quality profile. The primers are removed and forward and reverse reads are truncated based on the Quality score plots. Filtered reads are used to train the error model using machine learning approach then forward and reverse reads are dereplicated to generate unique

sequences and denoised (collapsed) in amplicon sequence variants (ASVs) applying the trained error model with pseudo pooling option. The ASV table is further filtered based on expected sequence length and checked for presence of chimera (removeBimeraDenovo).

Taxonomic assignment of representative sequence for each ASV is performed using standalone blast in the blast+ suite against a custom version of COI reference database in preparation (Di Capua et al., in preparation).

References:

- Balech B, Sandionigi A, Marzano M, Pesole G, Santamaria M (2022). MetaCOXI: An integrated collection of metazoan mitochondrial cytochrome oxidase subunit-I DNA sequences. Database, 2022(December 2021), 1–6. <https://doi.org/10.1093/database/baab084>.
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods, 13(7), 581–583. <https://doi.org/10.1038/nmeth.3869>.
- Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V. et al (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. Frontiers in Zoology, 10(1), 34. doi:10.1186/1742-9994-10-34.

2.5 Metabarcoding of Marine Fish communities

MiFish primers (Miya, 2022) will be used to amplify the 12S gene, using published amplification conditions and sequencing on a Miseq platform. Elas02 primers will be used to specifically target elasmobranchs (Taberlet et al., 2018). Other primers and markers (Tele02, Euka2, 18SV4, COI) will also be used.

References

- Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V. et al (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. Frontiers in Zoology, 10(1), 34. doi:10.1186/1742-9994-10-34.

Miya, M. (2022). Environmental DNA Metabarcoding: A Novel Method for Biodiversity Monitoring of Marine Fish Communities. *Annual Review of Marine Science* 14(1): 161-185.

Taberlet P, Bonin A, Zinger L, Coissac E (2018). *Environmental DNA: For Biodiversity Research and Monitoring*. Oxford University Press, Oxford, DOI: 10.1093/oso/9780198767220.001.0001.

2.6 Metabarcoding of Marine Mammals

Primers MarVer1 (12S) and MarVer3 (16S) targeting the hypervariable regions of the vertebrate mitochondrial 12S and 16S rRNA genes will be used for amplification of cetacean DNA (Valsecchi et al., 2020)

References:

Valsecchi, E., Bylemans, J., Goodman, S. J., Lombardi, R., Carr, I., Castellano, L., ... & Galli, P. (2020). Novel universal primers for metabarcoding environmental DNA surveys of marine mammals and other marine vertebrates. *Environmental DNA*, 2(4), 460-476.

3. SHOTGUN METAGENOMIC ANALYSIS OF MARINE SAMPLES

Shotgun metagenomics is performed on indexed DNA libraries with approximate fragment length of 500 bp. Library preparation is performed with the Illumina DNA Prep workflow (Document # 1000000025416 v09) kit following the manufacturer's recommendation. Library sequencing will be performed using Illumina Novaseq SP-500 flow cell, 2X250bp in order to obtain 2x250 pair end reads and 70M reads (10 Gb) per sample.

4. VIRAL METAGENOMIC SEQUENCING

After the extraction of nucleic acids, potential contaminants from prokaryotic and eukaryotic DNA are checked by quantitative PCR (qPCR) amplification using primers and TaqMan probes targeting prokaryotic 16S rRNA and eukaryotic 18S rRNA genes (Corinaldesi et al., 2017). After this quality check, viral DNA is used to generate different libraries in order to study dsDNA and ssDNA viruses (Fang & Tolun, 2016). Libraries are generated and sequenced on high-throughput platforms (e.g., NovaSeq or NextSeq by Illumina). The obtained reads are further screened for the presence of non-viral products using taxonomic assignment tools such as Diamond (Buchfink et

al., 2021). Reads are subsequently assembled through specific tools (e.g., metaviralSPAdes; Antipov et al., 2020) aimed at reconstructing viral scaffolds, which are then checked for completion through the CheckV tool (Nayfach et al., 2021). The abundance of high-quality viral genomes is obtained by the CoverM tool by mapping raw reads against the viral scaffolds (<https://github.com/wwood/CoverM>). Taxonomic inference is carried out on high-quality genomes through a consensus approach using multiple reference databases (e.g., UniProt) and tools (e.g., MMSeqs2, VPF-class; Steinegger et al., 2017; Pons et al., 2021), and scaffolds longer than 10 Kb are analyzed through the vContact2 tool to obtain in-depth information on their genomic novelty (Turner et al., 2021) and then screened for the presence of potential AMGs with the DRAMv suite (Shaffer et al., 2020). Lastly, host inference for each viral genome is carried out with the iPHOP tool (Roux et al., 2022).

References:

- Antipov D, Raiko M, Lapidus A, Pevzner PA (2020). Metaviral SPAdes: assembly of viruses from metagenomic data. *Bioinformatics*, 36(14), 4126-4129.
- Nayfach S, Camargo AP, Schulz F, Eloie-Fadrosch E, Roux S, Kyrpides NC (2021). CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nature biotechnology*, 39(5), 578-585.
- Pons JC, Paez-Espino D, Riera G, Ivanova N, Kyrpides NC, Llabrés M (2021). VPF-Class: taxonomic assignment and host prediction of uncultivated viruses based on viral protein families. *Bioinformatics*, 37(13), 1805-1813.
- Roux S, Camargo AP, Coutinho FH, Dabdoub SM, Dutilh BE, Nayfach S, Tritt A (2022). iPHoP: an integrated machine-learning framework to maximize host prediction for metagenome-assembled virus genomes. *bioRxiv*, 2022-07.
- Steinegger M, Söding J (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature biotechnology*, 35(11), 1026-1028.
- Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa S L, Solden LM et al (2020). DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic acids research*, 48(16), 8883-8900.
- Turner D, Kropinski AM, Adriaenssens EM (2021). A roadmap for genome-based phage taxonomy. *Viruses*, 13(3), 506.

5. METATRANSCRIPTOMICS ANALYSIS OF MARINE SAMPLES

The libraries (including rRNA depletion) will be prepared following Illumina protocol (e.g., Illumina Stranded Total RNA Prep) and run on an Illumina System (e.g., NovaSeq 6000) for a sequencing effort > 70M PE reads per sample. The metatranscriptomic data will be used to quantify and measure differences in the microbial gene expression.

The reads will be quality checked with FastQC (Andrews, 2010) and assembled with MEGAHIT (Li et al., 2015). Gene prediction will be performed with Prodigal (Hyatt et al. 2010) and annotation will be executed with different databases to increase coverage and resolution of the results such as the SEED Subsystems (Overbeek et al., 2015), the InterPro (Mitchell et al., 2019), and the Kyoto Encyclopedia of Genes and Genomes (KEGG; Kanehisa and Goto, 2000). Differential expression analyses will be performed with DeSeq2 (Love et al., 2014).

References:

- Andrews S (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform. 8, 111–119.
- Kanehisa M, Goto S (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27–30 (2000).
- Li D, Liu CM, Luo R, Sadakane K, Lam TW et al (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 15, 1674–1676.
- Love MI, Huber W, Anders S et al (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology, 15, 1–21.
- Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A et al (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res. 47, 351-360.
- Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, et al (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res. 33, 5691–5702.

6. MARINE OMICS OBSERVATORIES

WATER COLUMN	NBFC Omics Observatory
Sampling stations	N=4 LTER Stations: Golfo di Napoli, Promontorio di Portofino, Golfo di Trieste, Stazione Meda Senigallia
Sampling Depth*	N=1 Shallow (-1 m required), Additional depth(s) (optional). e.g. two depths are to be sampled when the water column is stratified (typically March to September in Naples) -Niskin bottles
Seawater Size fractions and replicates (filtration only)	N=3: 0.2-3 μm, 3-20 μm, 20-200 μm size fractions The procedure is as follows: water from Niskin bottles (2 replicate sampling per sampling period, 5 L each, total 10 L) is sieved through a 200 μm metal sieve and disposed in a plastic carboy. From this, a swinnex holder with PC 20 μm filter Φ 47 mm is placed before 2 large filter rigs Φ 142 mm with 3 μm and 0.2 μm pore size filters in sequence, connected to a peristaltic pump. Volume to be filtered varies according to sampling site's characteristics (e.g. 5 L). The filtration is repeated twice. Larger filters are cut into four parts to fit cryovials (see also Annex 1 "graphical methods") at each filtration round, generating 8 replicates in total from each size fraction (0.2-3 μm and 3-20 μm). Instead, 2 replicates are generated for the fraction 20-200 μm. A total of 18 filter samples are generated for each sampling depth, 9 filters for analysis and 9 filters for biobanking. Filters are placed in cryovials and immediately frozen in liquid Nitrogen OR placed on ice and frozen at -80 °C upon return to the lab. Virus (optional) The water from all filtrations (10 L in total if 5 L are filtered for each round) is collected in a dedicated carboy and 1 mL of a FeCl₃ solution is added and left to flocculate for at least 24H. The flocculate is collected by filtering through 0.45 μm filters. Metatranscriptomics (optional) Seawater collected by Niskin bottles and pre-filtered through a 200 μm mesh. Filtration performed by means of portable peristaltic pumps onto Sterivex filters (minimise external contamination) for max 30 minutes after water collection or until filter clogging. The Sterivex cartridges are then filled with RNA later or DNA/RNA shield (TBD), kept in a cold container until the home lab is reached and then at -80°C until extraction.

Net fractions and replicates (net + filtration) >20-200 µm size fraction >200-2000 µm size fraction	N=2: 20-200 µm, 200-2000 µm 20 µm mesh size horizontal plankton net (only surface depth, 5-7 mins tow). The material collected is then sieved through 200 µm metal sieve, brought to volume to 1 L and separated into 4 replicates, each filtered through 10 µm polycarbonate membrane filters (Φ 47 mm). 4 replicates are generated. 200 µm mesh size zooplankton net from 50 m depth to surface, sieved through 2000 µm metal sieve, bring to volume to 1 L and separated into 4 replicates, each filtered through 10 µm polycarbonate membrane filters (Φ 47 mm). 4 replicates are generated. Optional: WP2 net (200 µm mesh size) towed vertically from 50 m depth to surface. The sample is filtered on 2000 µm, if needed to remove large organisms, and then on 200 µm. Four replicates of bulk zooplankton were immediately frozen in liquid nitrogen on board and later stored at -80 °C.
Sampling interval per year	N=12 Monthly for 1 year
Biobanking	2 replicates per sample for each size fraction are stored at -80°C in 5 ml cryotubes (EXPLORE ZYMO SHIELD FOR PRESERVATION). DNA not used for analysis is also stored at -80°C
Negative controls	Milli-Q or Nanopure – same volume as samples (only 2 replicates per month, every single 5 L filtration)
Sampling device	ATLANTECO kit (Peristaltic pump + 2 tripods for Φ 142 mm filters)
Standards	Mock communities/Intercalibration exercises
DNA extraction	DNeasy PowerWater Kit (QIAGEN)
Library preparation	Illumina DNA prep protocols (METAGENOME)
Metabarcodes	0.2-3 µm (16SV4V5, 18SV4, 18SV9, ITS2 optional), 3-20 µm (16SV4V5, 18SV4, 18SV9, ITS2 optional), 20-200 µm (16SV4V5, 18SV4, 18SV9), 200-2000 µm (16SV4V5 optional, 18SV4, 18SV9, COI) 16S V4V5 (Parada et al., 2016 Environ Microbiol 18 1403–1414); 18S V4 (Stoeck et al., 2006 Protist 157 31–43); 18S V9 (Amaral-Zettler et al., 2009 PLoS One 4 e6372); COI (Leray et al., 2013 Front Zool 10(1) 34). ITS2 -optional (White et al., 1990; Nilsson et al., 2019) 200K reads per sample per amplicon (Illumina pair-end MiSeq 2x300bp or Novaseq 2X250bp).
Metagenomes	NovaSeq S4 flow cell (150x1) pair end reads (~450 bp library fragments) -70M reads (10 Gbases) per sample
eDNA	In field eDNA pump 0.45 µm, protocols to be agreed with A1-SPOKE1 (optional)
NGS database	Each observatory should employ a local data storage system (e.g., NAS infrastructure).

N of samples x sampling station	1 sampling depth x 5 size fractions x 6 months x 2 replicates (2 are for biobanking)= 60 PLUS 2 sampling depths x 5 size fractions x 6 months x 2 replicates=120 TOTAL OF 180 filters per year
N of analyses x sampling station	METABARCODES: a TOTAL of 510 SEQUENCING, see Annex I – Graphical Methods METAGENOME: a TOTAL of 78 sequencing, see Annex I-Graphical Methods
ANCILLARY DATA	MANDATORY T, sal, chl, Dissolved Inorganic Nutrients (NO ₃ , NO ₂ , NH ₄ , PO ₄ , SiO ₃) Optional: O ₂ , Turbidity, pH, light, FCM counts, POC, DOC, CDOM, HPLC, phyto Counts, Zoo counts, Zoo biomass
LOGSHEETS AND SAMPLING CODING	https://docs.google.com/spreadsheets/d/1GkeGFpWAM7W_YSoMJov4E7CvTCwKQkYc/edit?usp=sharing&oid=117756666113384702583&rtpof=true&sd=true