Operational Sensing Life Technologies for Marine Ecosystems

`

Milestone M2.4 – Workshop on the exploration of the available genomic workflows in use by the participating ERICs and associated parties

Lead Beneficiary: Flanders Marine Institute (VLIZ)

Authors: Pascal Hablützel, Marie-Catherine Bouqieaux, Hanneloor Heynderickx, Klaas Deneudt, Rune Lagaisse, Christos Arvanitidis, Cristina Huertas-Olivares, Catia Monteiro, Mar Humet Caballero, Filipa Martins, Rocio Nieto-Vilela, Ioulia Santi, Panagiotis Kasapidis, Jon Bent Kristoffersen

07/08/2023

Funded by
the European Union

Table of Contents

Preface & Summary

ANERIS aims to create and implement the next generation of scientific tools and methods for marine life-sensing and monitoring. In this framework, a workshop entitled “Workshop on genomic workflows” was held in Ostend, Belgium on the 30th and 31st of May 2023. The goals of the workshop were to explore the different available genomic workflows already in use by the participating Research Infrastructures (European Research Infrastructure Consortium; ERICs), partners, and associated parties of the project, to initiate more technical discussion and finally to set up an action plan and define tasks. The workshop consisted of 4 main themes: sampling design and protocols, wet lab protocols, bioinformatic workflows and connections to existing projects and initiatives.

Fig 1. Project partners attending the workshop on genomic workflows in Ostend.

List of Abbreviations

Abbreviation	
MARGENODAT	workflows for the MARine GENOmics DAta managemenT
NANOMICS	NAnopore sequeNcing for Operational Marine genomICS
OTU	Operational Taxonomic Unit
ASV	Amplicon Sequence Variant
ARMS	Autonomous Reef Monitoring Structures
NIS	Non-Indigenous Species
eDNA	Environmental DNA
CS	Case Study
OMB	Operational Marine Biology
SOP	Standard Operating Procedure
VRE	Virtual Research Environment

1. Workshop setup

The workshop was held in Ostend, Belgium on the 30th and 31st of May 2023 at VLIZ (InnovOcean Campus) and gathered partners of the WP2 from HCMR, LifeWatch, BIOPOLIS, EMBRC and VLIZ. The consortium was also joined by guests from the Open Science and EurOBIS teams at VLIZ. A participation list can be found in Annex 2.

The workshop had three main goals:

- Explore the available genomic workflows in use by the participating ERICs and associated parties
- Focus on technical discussions.
- Work out a plan of action for each theme discussed and define future tasks

The workshop was divided in four themes: sampling design and protocols, wet lab protocols, bioinformatic workflows and connections to existing projects or initiatives. For each of these themes, one or more ANERIS partners presented an introduction to the topic to serve as a basis for discussion. This brief introduction was composed of the following points: the presenters vision of how we can address the topics in ANERIS (keeping in mind the end-products required to address needs of Case Studies 2 and 4), their experience with the topic from previous or ongoing projects and how we can build on them or learn from them in ANERIS, and finally a list of questions that could be discussed to facilitate more technical discussions. The meeting agenda with all topics of presentations can be found in Annex 1 and the presentations of the workshop and meeting notes are available via the internal repository at https://aneris.eu/internal-repository/documents-meetings.

In the following points we summarize the discussion that took place for each theme, the decisions made and actions points defined.

2. Sampling design and protocols

The first session focussed on genomic sampling design and sampling protocols. We discussed how we can address the needs for Case Study 2 (CS 2): Improved spatial and temporal resolution of marine life monitoring based on genomics, and Case Study 4 (CS 4): Merging imaging and genomic information in different monitoring scenarios. We listed previous and ongoing initiatives and projects and discussed how we could learn from them and where we can connect to existing initiatives to join forces.

Under CS 2 we need to increase and complement the extent of current genomic monitoring using the ANERIS technologies by implementing low-cost methods. To improve the spatial and temporal coverage we identified two target groups, namely citizen scientists and existing

(genomic) research institutes in Europe not currently involved in routine genomic monitoring. As both groups have different capacities, we need to work out two sensible sets of sampling design and protocols. On one hand citizen scientists will need to be supplied with all sampling material, will require more training and will need to ship all samples for analysis; on the other hand, existing research institutes likely have more in-house knowledge and infrastructure, and therefore would be able to do more extensive genetic sampling and have the ability to combine this with imaging and bio-optic techniques under WP3 to deliver data to CS 4.

Citizen scientists could be engaged in the sampling of water for (e)DNA analysis or basic soft sediment sampling, because this is more feasible than extensive water filtrations for metagenomics for which we would need to supply dedicated, expensive sampling equipment like plankton nets, pumps or sediment samplers. This would also put a strain on the number of samples that can be collected. There is also the concern of getting sufficient sample material to study target taxa with unconcentrated water samples. Therefore, we should inquire with existing eDNA citizen science sampling efforts, like the UNESCO eDNA expeditions project [1], and see how they obtained desirable results while keeping sampling feasible. Under M2.1 we will develop this sampling kit for citizen scientists (lead: HCMR). Our aim is to keep sampling simple and feasible, and compile a cost-effective sampling kit, avoiding high investment costs and required expertise currently associated with water filtrations as required by the EMO BON plankton DNA filtration protocol. Another aspect is that citizen science campaigns have to be engaging and provide fast results to keep citizens motivated to participate and show them that they are making an impact. Some previous initiatives have shown that this type of campaign can have good results in terms of enthusiasm (eDNA expeditions project of UNESCO, BGE project [2] led by Naturalis, and the Ocean Sampling Day (OSD) worldwide research campaigns were positively received by the participants). This enthusiasm was achieved in one case by the use of QR codes that were used by citizens to register their samples and then to track and get updates about them (arrived at lab, sequenced, etc.). However, it has also been pointed out during our discussions that fast sequencing results could be challenging to achieve because sequencing can be slow and expensive, access to samples right away could be difficult in some cases and the quality and preservation of the samples could be concerning as well. These key issues will have to be addressed before setting up the citizen science genomic sampling campaign. For the organizational aspect of the campaign we could rely on the BioBlitzes. These events are a collaborative effort between scientists, naturalists, volunteers, and the general public to explore and document the biodiversity of a particular area within a specific timeframe. BioBlitzes provide a unique opportunity for participants to engage in hands-on fieldwork, learn from experts, and contribute to scientific research by recording species observations. Not only do they promote public awareness and appreciation of local ecosystems, but they also generate valuable data that can be used for conservation and environmental monitoring.

Research institutes could be involved for high quality water sampling and filtrations for (e)DNA analysis and in linking the DNA results with imaging data for the partners also involved in WP3. For sampling by research institutes, we could rely on EMO BON sampling protocols [3], as they

are well-established and standardised over Europe. By relying on the EMO BON protocols, we are adding to the existing 17 EMBRC genomic observatories over Europe that have been carrying out the bi-monhly genomic sampling of water column, soft sediment (microbial, macrobenthos and meiobenthos communities) and hard substrate fauna since June 2021 (Santi et al., 2023). However, not every aspiring local research partner will have the resources to carry out the full scale of sample types according to current SOPs or have the budget for these initial equipment investments. We will need to evaluate each station's capacities and possibly tailor protocols or decide on a specific set of feasible sample types, while still ensuring comparability with the EMO BON samples. Stations with limited genomic sampling experience will require initial training from EMBRC or experienced EMO BON observatories. Additionally, local environmental and biological conditions of new observatories might require tailored protocols for water filtrations.

When deciding on sample design and protocols we should also keep in mind the relevant Operational Marine Biology (OBM) products we aim to deliver under ANERIS. These will require specific sample types in terms of target taxa or habitats, which need dedicated sampling protocols. For the cross work-package work, we will also need to coordinate sampling effort to collect e.g. genomic and imaging samples in tandem and to target similar habitats or communities. Water samples of e.g. plankton communities are desirable for linking with imaging and bio-optic technologies under WP3 in ANERIS. Water samples taken during Bioblitz campaigns can be used for eDNA analysis of the specific target taxa or communities under investigation at a station during the BioBlitz events under WP4. Sediment sampling of macrobenthos, meiobenthos and microbial communities can be important in the framework of indicator taxa for e.g. pollution. We can explore existing biodiversity data here to test hypotheses regarding pollution gradients or environmental change. Some observatories focus their current research on rocky intertidal habitats, this will be challenging to harmonize with other ANERIS observatories sampling soft sediment and/or water column. We will look into using scrape samples on more exposed areas or deploying Autonomous Reef Monitoring structures (ARMS) [5, 6] in intertidal pools. The ARMS-MBON project uses stacked PVC plates to mimic reefs to sample colonising hard substrate fauna. It combines morphological identification of plate photography and metabarcoding of the sessile and motile taxa. ARMS-MBON protocols are standardised, and the network has a coverage of close to 20 observatories over Europe and the Arctic.

Sampling scale, frequency and timing is another topic for discussion. With decentralised sequencing, we need to identify how many samples each observatory is able to process within their allocated budgets. Depending on the target taxa for the OMB products, we will need dedicated sampling events in specific areas where current monitoring is missing or at certain times of the year, related to bloom patterns or spawning events. For each sample type, a sufficient number of replicates and negative controls need to be taken into account.

Sample preservation and shipment is to be discussed for both citizen science and research institute sampling efforts and will depend on sample type and target taxa. We will explore the use

of commercially available products like DNA/RNA shields vs. in-house produced products and aim for simple and more affordable room-temperature shipment.

To resume, two relatively independent sampling efforts are identified, one relying on feasible citizen science sampling of eDNA, and one more dedicated genomic sampling by research institutes to extend the current EMO BON observatories. They should be addressed individually with the local partners involved to fit in their in-house capacities. The sampling protocols should be adapted to fit the user's capacities and local environmental conditions as well as the desired sample type and community under study to supply data to the OMB products and Case Studies.

○ Future actions

Based on the discussion described above, the following points should be addressed in the future:

- Deliverable 2.1 is compiling a citizen science kit for genomic sampling, the content of this kit and the guidelines and training that needs to be provided is still to be decided (lead: HCMR)
- Discussions should be initiated with the people involved in (plankton) imaging technologies under WP3 regarding their proposed sampling design
- Discussion with people involved in BioBlitzes to ensure they could take meaningful eDNA samples and what communities they are targeting with the morphological identifications
- Contact people from eDNA expeditions of UNESCO to see if the DNA results they obtained were good (VLIZ)
- Intertidal observatories should discuss further regarding how they are going to take DNA samples, the use of Autonomous Reef Monitoring Structures (ARMS) to sample hard substrate fauna will be explored (BIOPOLIS)
- Feasibility of EMO BON protocols should be discussed with non-EMBRC genomic observatories, we should evaluate their in-house capacities and how we can sustainably support them, and discuss sample shipment and preservation (EMBRC)
- We need some preliminary data to see if we obtain the desired data for CS 2 and CS 4 and the OMB products we define, this data can be harvested from e.g. European monitoring under EMO BON, ARMS-MBON or OSD [7].

3. Wet lab protocols

Nanopore sequencing protocols will be developed, tested and optimised for different taxonomic groups and different sample types. Additionally, these protocols will be compared to the standard

Illumina metabarcoding data. Here, we discussed technical aspects of DNA extraction and PCR protocols as well as library preparation and sequencing.

○ DNA extractions and PCR

We will need to develop wet-lab protocols for different taxonomic groups, specifically focusing on DNA sequencing for fish, plankton, hard substrate and soft sediment taxa. In terms of extraction protocols we discussed two options; the use of commercially available extraction kits or developing a custom extraction protocol. The participants weighed the pros and cons of each option, considering factors such as cost, standardisation, and compatibility.

The commercial kits provide stable products, reduce the chance of cross-contamination and allow for better standardisation with existing projects. However, these kits are more costly, tend to have lower DNA yields and the exact content of the reagents is not disclosed by the manufacturers. With a custom in-house extraction protocol, the cost can be highly reduced, the DNA yield is typically higher, there is possibility for optimisation and the protocols are FAIR. On the other hand, this method is labor-intensive and thus requires more skills, there is a higher risk of cross-contamination and it does not allow for easy standardisation between projects.

The importance of including blank samples and mock communities in the sequencing process was highlighted. Here, we will rely on the use of existing mock communities for phytoplankton (from the EMO BON project), zooplankton (HCMR in-house), fish, and macrofauna sediment communities (from the MARBEFES [8] project). The protocols for metabarcoding will be specifically developed for Oxford Nanopore (Jain et al., 2016) sequencing which allows for the selection of longer reads. We have therefore decided to try and sequence the full 18S rRNA region when targeting phytoplankton, zooplankton and potentially even for extra-organismal DNA of macrofauna. The use of other primers, like COI and 16S, will also be expired for specific target taxa. In addition to the target region length, factors such as error rates, inhibitor presence, and cost were discussed for the choice of polymerase in the PCR step. The possibility of carrying out wet-lab protocols in-situ with Oxford Nanopore sequencing was considered. However, preliminary tests under other research projects proved that this is cumbersome and leads to inaccurate results. Further, the added benefit of demonstration to citizen scientists would be little, the full sequencing protocol still requires a full day, filled with detailed explanations while all the hands-on work would still have to be carried out by a trained scientist.

Overall, the participants emphasized the need for consensus on the target species and regions to sequence, and the importance of financial restrictions. To reduce the costs, the potential for collaboration in purchasing kits, reagents, and materials will be explored.

○ Library preparation and sequencing

Various topics related to sequencing technologies and library protocols were covered. Different nanopore devices and their features were discussed, including the Mk1b, Mk1C, and the upcoming Mk1D which will have improved temperature regulation and compatibility with iPad. GridIon and Flongle were also mentioned, with the latter being an adapter for smaller flow cells but which are more expensive. The PromethION versions were highlighted for their different flow cells, offering higher throughput and lower cost per base.

The discussion touched upon the error rate of nanopore sequencing, which was around 5% until 2022. Beginning this year (2023), Oxford Nanopore introduced their newest Q20+ platform with V14 kit chemistry ensuring min. 99% accuracy and the highest yield for all read lengths. However, not all users have been able to achieve this level of accuracy yet. Additionally, the use of the V14 kit chemistry allows duplex sequencing which can read both strands of DNA, instead of simplex sequencing, to increase the accuracy even further (up to 99.9%).

Two basic categories for the library protocols were compared, i.e. a ligation based method and a transposase method with click chemistry. Ligation was noted for its better yield and read length but it is also more work-intensive, time-consuming and costly. The transposase-based method is faster and more efficient (requires less template DNA), but it is not recommended for amplicon sequencing because it will disrupt the amplicons. The transposase method however could be used for metagenome shotgun-sequencing. In terms of amplicon protocol development the need for an efficient and short protocol was emphasized. Therefore the use of the rapid attachment chemistry and custom-made barcoding assays were suggested. Another option that was discussed is the use of rolling circle amplification. This would boost the consensus accuracy, but it is a complex process and it does not allow for an easy quality check of the amplification step. We agreed to look further into the different protocol steps associated with this method to further assess its potential. We also covered the possibility of using adaptive sampling. For this approach a reference library with sequences of interest is needed which requires more knowledge about the ecosystem you are studying beforehand and therefore might be of less interest here. Adaptive sampling can however be useful in the form of “barcode balancing”, where adaptive sampling helps to obtain equal numbers of reads for every sample. This removes the need for accurate quantification of each sample before making a pool of the samples. For basecalling, we discussed if it should be done in nodes or centrally. So far, Guppy basecalling (v4.3.4, Oxford Nanopore Technologies Ltd., UK), proposed by Oxford Nanopore, has been used by our participants. More advanced basecalling algorithms do exist but many also require a reference library. Since we are interested in identifying all organisms that are present, including ones we don’t know yet, we agree on using Guppy base calling in nodes for metabarcoding. A new and faster basecaller from Oxford Nanopore (Dorado) is under development and is expected to replace Guppy in the near future. When this happens we will explore a possible transition to Dorado for basecalling.

Overall, we would like to use and optimise a single amplicon protocol for different sample types (fish eDNA, plankton bulk DNA, and hard substrate and soft sediment eDNA or bulk DNA). To this end we will run a few test protocols, e.g. a four-primer amplicon protocol, ligation of native barcodes, PCR barcoding, rolling circle amplification, and explore the use of unique molecular identifiers. For these test-runs we agreed to re-use regular flow cells rather than buying new Flongles. The potential to update the chemistry and kit in the amplicon protocol at a later stage will also be considered here in order to match the rapidly evolving Nanopore technology. The protocol testing will be carried out with existing mock communities and can additionally be tested on samples from the EMO BON project. The latter project relies on Illumina sequencing technology, by processing these samples in ANERIS with Oxford Nanopore technology, which focuses on long read sequencing, a comparison between both technologies could be made.

○ Future actions

- Write down the pros and cons of the different extraction options and evaluate what is feasible in terms of budget. (HCMR, VLIZ, BIOPOLIS)
- Research the extraction protocols used by existing projects for compatibility (HCRM, VLIZ, BIOPOLIS)
- Establish a document with the desired protocol tests, their priority and their cost (HCRM, VLIZ, BIOPOLIS), a current version of the document is included in the Annex.
- Create or identify a repository to share final protocols (HCMR?)

4. Bioinformatic workflows

There are currently a number of bioinformatic workflows available to process raw sequencing output to quality controlled and taxonomically annotated sequences. To promote FAIR data principles (Findable, Accessible, Interoperable, Reusable) (Wilkinson et al., 2016), we need to address the challenges of data and code interoperability and reproducibility in bioinformatic workflows to facilitate (semi-)automated archiving of genomic data and metadata.

The final bioinformatic workflow should have the following qualities: accurate (scientifically correct results), reproducible, interoperable (input and output must be compatible with upstream and downstream analyses), adaptable (incorporate new knowledge and technologies if need be) and most of all, user friendly. Taking this into account, we should aim for a pipeline where each of its steps is like a single module that could be ignored if it doesn't suit the needs of the user.

Oxford Nanopore data brings some challenges to bioinformatic processing, as we will have to deal with low read accuracy, although we might see this level of sequencing errors improve over the upcoming years by continuous development and protocol optimisation. There are a number

of solutions proposed in literature to deal with this low read accuracy. A first option is to add Unique Molecular Identifiers (UMIs) during the laboratory processing, however from previous experiences we know that is costly, complex and has its limitations because it requires sufficient sequencing depth. Another option is the use of clustering algorithms. From previous Nanopore bioinformatic processing we learned that density-based clustering proved more favorable than greedy incremental clustering because the latter required a priori defined thresholds. However, not many people in the community have been using these clustering algorithms, and there is no consensus on how to run them (e.g. number of iterations). Testing this will require a large amount of data, which we can possibly source from existing datasets like from the EMO BON project.

We have for now, no clear overview of all pipelines existing to analyze sequences produced using the Oxford Nanopore technology. Using part of typical Illumina pipelines has been suggested, and while some tools from them could be useful, we need to keep in mind that there are differences between the two sequencing technologies (e.g. Oxford Nanopore errors contain more insertions than deletions). One tool developed and used before by partners to analyze Illumina data is PEMA (Pipeline for Environmental DNA Metabarcoding Analysis, Zafeiropoulos et al., 2020). An alternative to this tool for Oxford Nanopore sequencing analysis could be the use of SLIM 2.0 [12], for which the development has been planned during this project. SLIM 2.0 is a web application that provides a graphical user interface (GUI) and incorporates bioinformatics tools as modules. The pipeline that will be developed during this project should be ultimately incorporated into SLIM 2.0. However, as we want to also be able to run this pipeline locally or on a HPC, the workflow developed should be a standalone pipeline, which can later be integrated in SLIM 2.0

During the Interreg North Sea GEANS project, VLIZ developed a basic prototype pipeline for processing Nanopore sequencing data. The pipeline is currently available on request because it is unstable and needs further major developments in terms of clustering, taxonomic assignment and extracting raw number of reads before clustering. Additionally, as we optimise the pipeline, and for example test more iterations during the clustering, more computing power will be required to run the analysis. This computing power for the ANERIS project can be sourced under LifeWatch or by CPU and GPU hours available on HPC clusters on institute in-house servers.

The goal of MARGENODAT (workflows for the MARine GENOmics DAta managemenT) is not only to process raw sequencing data but also to provide a workflow that standardises data format and output to European Darwin Core Archive [13] and MIXS standards [14] which can be

used by non-experts. Currently no user friendly tools exist to reformat data to existing standards. As a result, we see that OTUs and ASVs are rarely deposited in data aggregators at the moment, hindering large scale biodiversity meta-analysis. Additionally, there is no convention on standards for archiving genetic result data or required metadata fields, with some data aggregators also lacking fields for depositing crucial workflow metadata. There currently are a number of data aggregators for biodiversity data, but they have no way of interlinking datasets and their ontology needs to be extended, their data standards are not fully interoperable and the underlying taxonomic backbones are often also not fully compatible. It will be important for the bioinformatic

community to collaborate, and involve key data aggregators and taxonomic backbones to work out standardized formats and protocols to improve data archiving and compatibility between tools and data systems.

As reference databases and genetic technologies are ever-evolving, rerunning previous workflows with new software or database versions will lead to improved results. Workflows should be developed so they can be easily rerun in the future, and there should be a standardized way to update previously archived result data. The integration of the workflow in a virtual research environment (VRE) could facilitate this.

We decided to focus on building a metabarcoding pipeline, which can later possibly be extended to a metagenomics workflow.

For the work at hand we need to create a team from people within our institute that can look into currently used internal workflows and that have connection to similar projects or initiatives dealing with bioinformatic workflows. While we should have enough resources at VLIZ to do the major development, BIOPOLIS and HCMR could also look into their institute if they have someone who could contribute if necessary. Each institute also has to evaluate their current in-house pipelines and see if we can take parts to the next level under the ANERIS pipeline.

○ Future actions

The following actions should be taken in the future:

- VLIZ to develop the general architecture of the bioinformatic workflow and report back to WP2 partners
- Look into the requirements for a possible integration into SLIM 2.0 (NORCE, VLIZ)
- Look also into the requirements to run the pipeline on the LifeWatch Virtual Research Infrastructure (VRE) (LifeWatch, NORCE, VLIZ)
- Have open discussion with stakeholders like data-aggregators and taxonomic backbones
- Evaluate in-house pipelines and identify in-house experts in partner institutes (VLIZ, HCMR, BIOPOLIS)

5. Connection to existing projects or initiatives

Our partners are distributed all across Europe, thereby covering a major part of the Atlantic coast and Mediterranean coast. By combining data from these major areas with data from existing projects, we want to provide large scale biodiversity assessments. Specifically, we would like to connect to projects such as EMO BON, MARBEFES, MARCO-BOLO [15], DTO-BioFlow [16], and eDNAqua-Plan [17]. From EMO BON we would like to explore the use of the sample biobank, developed mock communities, previously released European-wide data and established and standardized SOPs. The use of Autonomous Reef Monitoring Structures (ARMS) under ARMS-MBON will be explored for rocky intertidal observatories, and available data will be explored for OMB products. MARBEFES aims to standardize Illumina sequencing, they will sample soft sediment and eDNA and work on mock communities. MARCO-BOLO will work on developing indicator maps for genetic data under T5.2 application of integrative approaches on establishing ecological indicators. We also want to rely on citizen science to collect samples in coastal waters through regional and local initiatives, where we can connect to BioBlitzes and learn from the experiences of the eDNA expeditions project of UNESCO.

For the standardisation and centralisation of data archiving we will rely on existing formats like Darwin Core Archive and MiXS. We will explore cooperation with (Eur)OBIS and GBIF, as well as their underlying taxonomic backbones. The European node of OBIS (EurOBIS) and the World Register of Marine Species (WoRMS) are hosted at VLIZ, facilitating cooperation. There are a number of ongoing research projects and initiatives dealing with work similar to MARGENODAT. The eDNAqua-Plan project has a work package on Data Standards, Data Linking & Compatibility and will work on tasks standards for eDNA data repositories, DTO-BioFlow aims to enable sustained flows of biodiversity monitoring data into the Digital Twin of the Ocean (DTO), and will integrate tools into a VRE. It will be important to cooperate with these initiatives to harmonise our efforts and build sustainable solutions with applications beyond the scope of the developing project.

Acknowledgements

We want to thank all WP2 partners for their participation in the workshop and the fruitful discussions.

References

[1] https://www.unesco.org/en/edna-expeditions

[2] https://biodiversitygenomics.eu/citizen-science/?event-type=upcoming

[3] Ioulia Santi, Raffaella Casotti, Thierry Comtet, Michael Cunliffe, Yolanda Koulouri, Lara Macheriotou, Fabrice Not, Matthias Obst, Christina Pavloudi, Sarah Romac, Eric Thiebaut, Jan Vanaverbeke, Nicolas Pade. European Marine Omics Biodiversity Observation Network (EMO BON) – Handbook. Version 1.0, 2021. Available from: https://www.embrc.eu/newsroom/publications/european-marine-omics-biodiversity-observation-network-emo-bon-handbook

[4] 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. Available from:https://www.researchgate.net/publication/370889641_European_marine_omics_biodiversity_observation_network_a_strategic_outline_for_the_implementation_of_omics_approaches_in_ocean_observation [accessed Jul 13 2023].

[5] https://marinebon.org/arms-mbon/

[6] https://github.com/arms-mbon

[7] https://www.assembleplus.eu/research/ocean-sampling-day#:~:text=The%20Ocean%20Sampling%20Day%20(OSD,to%20produce%20contextual%20genomic%20data.

[8] https://marbefes.eu/

[9] Jain, M., Olsen, H.E., Paten, B. et al. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol 17, 239 (2016). https://doi.org/10.1186/s13059-016-1103-0

[10] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

[11] Haris Zafeiropoulos and others, PEMA: a flexible Pipeline for Environmental DNA Metabarcoding Analysis of the 16S/18S ribosomal RNA, ITS, and COI marker genes, *GigaScience*, Volume 9, Issue 3, March 2020, giaa022, https://doi.org/10.1093/gigascience/giaa022

[12] https://github.com/trtcrd/SLIM

[13] https://dwc.tdwg.org/terms/

[14] https://gensc.org/mixs/

[15] https://www.embrc.eu/newsroom/news/marco-bolo-start

[16] https://www.seascapebelgium.be/dto-bioflow

[17] https://seascapebelgium.be/ednaqua-plan

Annex

Annex 1: meeting agenda

Tuesday 30 May 2023	
09:00 - 09:30	Coffee
09:30 - 09:45	VLIZ - Marie-Catherine Bouquieaux / Pascal Hablützel
	Welcome and introduction
09:45 - 11:00	LifeWatch - Cristina Huertas / Christos Arvanitidis

	Study and sampling design: How can we address the needs for case study 2 and 4? What can we learn from previous initiatives (pan-European or local/regional)? Can we connect to other initiatives to join forces or to reach more local scientists?
11:00 - 11:30	Coffee break
11:30 - 12:30	EMBRC - Ioulia Santi
	Sampling protocol: What can we learn from previous or ongoing initiatives such as EMO BON, Ocean Sampling Day, etc.?
12:30 - 13:30	Lunch break
13:30 - 15:30	VLIZ - Hanneloor Heynderickx
	Wet lab protocols (DNA extraction and PCR): Lessons learned from previous eDNA projects using Nanopore sequencing.
15:30 - 16:00	Coffee break
16:00 - 18:00	HCMR - Jon Bent Kristoffersen
	Wet lab protocols (library preparation and sequencing): Lessons learned from previous eDNA projects using Nanopore sequencing.
19:00	Social dinner, Ostend

Wednesday 31 May 2023	
09:00 - 09:30	Coffee
09:30 - 11:00	VLIZ - Pascal Hablützel
	Bioinformatics
11:00 - 11:30	Coffee break

11:30 - 12:30	VLIZ - Pascal Hablützel
	MARGENODAT
12:30 - 13:30	Lunch break
13:30 - 15:30	VLIZ - Pascal Hablützel
	Connection to other projects or initiatives (EMO BON, EurOBIS, OBIS, etc.)
15:30 - 16:00	Coffee break
16:00 - 18:00	NORCE
	SLIM 2.0

Annex 2: List of participants (last updated on 24. May 2023)

Name	**Partner**	**Presence**
Marie-Catherine Bouquieaux	VLIZ	live
Rune Lagaisse	VLIZ	live
Pascal Hablützel	VLIZ	live
Hanneloor Heynderickx	VLIZ	live
Filipa Martins	BIOPOLIS	live
Cátia Monteiro	BIOPOLIS	live
Mar Humet	BiOPOLIS	online
Ioulia Santi	EMBRC ERIC	online
Christos Arvanitidis	LifeWatch ERIC	live

Jon Bent Kristoffersen	HCMR	live
Robío Nieto-Vilela	HCMR	online
Kasapidis Panagiotis	HCMR	online
Deneudt Klaas	VLIZ	live
Cristina HUERTAS-OLIVARES	LIFEWATCH ERIC	live

Annex 3: Small wet lab and bioinformatics projects (current export of list, to be extended and updated regularly)

Mock communities:

- EMO BON: Bacteria
- EMO BON: Phytoplankton
- HCMR: Zooplankton
- HCMR: macrofauna eDNA (in planning)
- HCMR: fish eDNA from local aquarium

In silico primer testing

- BIOPOLIS: CRABS for full 18S and 16S for eukaryotes
    - Primer bias
    - Taxonomic coverage of reference database
    - Taxonomic resolution

DNA extraction:

- Paramagnetic beads
    - BIOPOLIS: Manual protocol
        - 96 Pillar MagSleeve for 96 Well Plate (diagnocine.com) 1k
        - VP 407AM-N1 | V & P Scientific, Inc. (vp-sci.com) 3k+customs
    - BIOPOLIS: Kingfisher robot

PCR:

- VLIZ: feedback from eDNA expeditions on multiplexing of barcoding regions for eDNA samples

Library preparation:

- HCMR: 4 primer protocol
- VLIZ: review RCA protocol
- HCMR/VLIZ: test new Nanopore V14 kit chemistry
    - Duplex vs. simplex

Metagenomics:

- Ask Genoscope how far they are with protocol development
- Maybe develop our own protocol
- HCMR has a test-dataset that may be used for protocol development

Bioinformatics pipeline architecture

- VLIZ: develop the general architecture and report back to WP2
- VLIZ: Contact Tristan for requirements for integration into SLIM 2.0
- VLIZ: Checking the requirements to run the pipeline on the LifeWatch infrastructure