

MARCO-BOLO WP2/WP5 joint workshop for common planning towards delivering on MSFD indicators

In preparation of the WP2/WP5 joint workshop, a Data Analysis Challenge was launched on the 23rd of August, inviting members of the eDNA community to participate in analysing phytoplankton and fish eDNA datasets using their preferred bioinformatic pipelines. A virtual workshop was held on the 7th and 8th of October 2024 with MARCO-BOLO WP2 and WP5 collaborators to process and analyse all the submitted data files and to discuss the next steps for calculating Essential Biodiversity Variables (EBV) & Essential Ocean Variables (EOV) metrics, as well as Marine Strategic Framework Directive (MSFD) indicators.

The organisation and preparation of the Data Analysis Challenge and the WP2/WP5 joint workshop was done by VLIZ, UNESCO and UiT. The following collaborating partners participated in the WP2/WP5 workshop: VLIZ, UNESCO, UiT, MBA, CNRS and MS.

The workshop included a brief overview of all the data that was submitted through the Data Challenge and a discussion on which metrics would be calculated for the data comparison and pipeline sensitivity. Four different metabarcoding sequence datasets were made available to all participants to run their bioinformatic pipelines on: one phytoplankton dataset (18S rRNA; Caracciolo et al., 2022) and three fish aquarium datasets (12S rRNA, 16S rRNA and COI; Lisbon Aquarium). The latter was provided by the A-Fish-DNA-Scan and ME-BARCODE group at the University of Minho. In total, 37 participants submitted data for the workshop, with most of them submitting multiple entries.

Dataset	#submission	#participants
Phytoplankton – 18S	18	14
Fish – 12S	29	21
Fish – 16S	25	20
Fish – COI	28	23

The following work was carried out during the workshop:

Task 1: Data formatting

Participants submitted their datasets in various formats, with some providing a single comprehensive file while others submitted three separate files including a taxa table, sequence table, and a table including either Operational Taxonomical Units (OTUs) or Amplicon Sequence Variants (ASVs). To standardise the data for analysis, all submissions were converted into a single long format file. All contributors of the workshop were assigned specific datasets to reformat, ensuring a collaborative approach to data harmonisation. Subsequently, a script was developed to calculate several key parameters from the reformatted data. These parameters included the total number of taxa and read counts at family, genus, and species levels, as well as the total number of target ASVs and species at



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 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).

each taxonomic level. Additionally, the number of non-target ASVs were also recorded. Finally, all these calculated parameters were consolidated into a single file, providing a comprehensive overview of the bioinformatics output data from all participants.

Task 2: Metadata pipeline

As part of the Data Analysis Challenge, participants were tasked with providing comprehensive bioinformatic metadata for their chosen sequence processing pipelines. This information included the pipeline used, whether it was an established pipeline or an in-house developed pipeline, and detailed descriptions of all bioinformatic tools employed along with the specific parameters used for each step. To facilitate standardised data collection, the workshop organizers provided participants with a metadata form for the Data Analysis challenge, which outlined the typical steps involved in bioinformatic processing of metabarcoding sequence data. During the workshop, all the submitted information was extracted and compiled into a master file. Importantly, separate entries were created for each dataset that participants chose to process with their respective pipelines, ensuring a detailed and organised record of the various bioinformatic approaches used across different datasets.

Task 3: Metric calculations

To compare the data results from the different bioinformatic pipelines used for each of the provided datasets, a number of metrics were discussed to be calculated: alpha and beta diversity indices, as well as a sensitivity index. The latter quantifies the number of target species detected by each pipeline in comparison to a reference list of species from the samples' origin.

Discussions were held regarding the future outcomes and commitments with the data compiled in this workshop. The results from participants who processed the 18S rRNA dataset will be used for the development and fine-tuning of the genetic indicator workflows for MSFD, within WP5. This will enable the assessment of how varying bioinformatic pipelines might impact the outputs of these indicators, providing valuable insights for standardisation efforts in marine environmental monitoring. Similarly, within WP2 efforts will be made to compare existing methods for calculating Essential Biodiversity Variables (EBVs) and Essential Ocean Variables (EOVs) using the Fish datasets results of the various pipelines.

Special thanks to all participants of the Data Analysis Challenge for providing us with their results: Adam Bennett, Alejandro Abdala, Anders Lanzén, Audrey Bourret, Cassandre Héritier--Tellier, Chris Hempel, Christian Atallah, Daniel Straub, Elisa Banchi, Emese Meglécz, Emma Strand, Evan Morien, Falk Hildebrand, Frank Panitz, Géraldine Pascal & Olivier Rué, Geslain Enora, Jeffrey Miller, Kingsly Chuo Beng, Luka Lenaroto, Luke Thompson, M. Pilar Cabezas, Mads Reinholdt Jensen, Magali Lescot, Marius Eisele, Mykle Hoban, Ngoc-Loi NGUYEN, Nicolas Henry, Oriol Canals, Onur Doğan, Pedro Junger, Saara Suominen, Sean Jungbluth, Sten Anslan, Tomáš Větrovský, Victor Pylro, Vitor Silva and Zachary Gold.



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 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).



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., Thiébaud, E., ... 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>



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 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).