



Operational Sensing Life Technologies for Marine Ecosystems

Deliverable D2.3 – Development of new DNA-based indicators for pelagic habitats environmental status

Lead Beneficiary: LifeWatch ERIC

Author/s: Arvanitidis C, Huertas C, Kasapidis P, Zaíño Rodríguez J, Paneque J, Obst M, Piera J.

30/07/2025



Funded by
the European Union

Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the EU nor the EC can be held responsible for them.

D2.3 Development of new DNA-based indicators for pelagic habitats environmental status monitoring; ANERIS #101094924

Prepared under contract from the European Commission

Grant agreement No. 101094924

EU Horizon Europe Research and Innovation action

Project acronym: **ANERIS**
Project full title: **operAtional seNsing liFE technologies for maRIne ecosystemS**
Start of the project: January 2023
Duration: 48 months
Project coordinator: Jaume Piera

Deliverable title: Development of new DNA-based indicators for pelagic habitats environmental status monitoring
Deliverable n°: D2.3
Nature of the deliverable: Other
Dissemination level: Public
WP responsible: WP2
Lead beneficiary: LifeWatch ERIC

Citation: Arvanitidis, C., Huertas, C., Kasapidis, P., Zaíño Rodríguez, J., Paneque, J., Lopez, J., Obst, M., Piera, J. (2025). Deliverable 2.3 Development of new DNA-based indicators for pelagic habitats environmental status monitoring. EU Horizon Europe ANERIS Project, Grant agreement No. 101094924

Due date of deliverable: Month n°28
Actual submission date: Month n°32

Deliverable status:

Version	Status	Date	Author(s)
1.0	Draft	20 July 2025	Arvanitidis Christos (LifeWatch ERIC)
2.0	Internal review	30 July 2025	Kotoulas Georgios (HCMR)
3.0	Review	30 July 2025	Hablutzel Pascal (VLIZ)
3.0	Final	30 July 2025	Arvanitidis Christos (LifeWatch ERIC)
3.0	Final review	31 July 2025	Jaume Piera (CSIC)

The content of this deliverable does not necessarily reflect the official opinions of the European Commission or other institutions of the European Union

Table of Contents

Preface	4
Summary	4
List of Abbreviations	4
1. Concept	5
1.1 Criteria adopted	5
1.2 Data availability	6
2. Approach	6
2.1 Review	6
2.2 Selection of the new indicators	6
2.3 Metadata	7
3. Workflows for the calculation of the indices	7
3.1 DNA-based trends for the arrival of new non-indigenous taxa	7
Figure 1: Workflow depicting the stages of the calculation of the newly proposed index DNA-based trends for the arrival of new non-indigenous taxa.	8
3.2 DNA-based trend in the distributional range of invasive taxa	8
Figure 3: Workflow output depicting the range expansion of the taxa identified by the workflow for the calculation of the newly proposed index DNA-based trend for the distributional range of invasive taxa.	9
3.3 DNA-based Taxonomic Relatedness	9
Figure 4: Workflow depicting the stages of the calculation of the newly proposed index DNA-based taxonomic relatedness.	10
Figure 5: Funnel-plot of the 95% confidence intervals and taxonomic distinctness values of the samples superimposed, as it is produced by the workflow on the calculation of the newly proposed index DNA-based taxonomic relatedness (Dplus is the calculated Taxonomic Distinctness values).	11
Acknowledgements	11
References	12
Annex I. Functionalities of the common services used in all workflows for the three newly DNA-based indices	14
Annex II. Functionalities of the specific services used in each of the workflows used for the three newly DNA-based indices	15

Preface

This Deliverable refers to a new research product developed by the ANERIS project. It provides the methodology and reports the development of three DNA-based indicators for pelagic environmental status including their application.

Summary

Deliverable 2.3: Arvanitidis, C., Huertas, C., Kasapidis, P., Zaíño Rodríguez, J., Paneque, J., Lopez, J., Obst, M., Piera, J. (2025). Deliverable 2.3 Development of new DNA-based indicators for pelagic habitats environmental status monitoring. EU Horizon Europe.

As stated in the Milestone 2.3, the methodology for the development of the new DNA-based indicators follows best practices established by relevant EU funded projects, such as DEVOTES (DEvelopment Of innovative Tools for understanding marine biodiversity and assessing good Environmental Status) and GES4SEAS (Achieving Good Environmental Status for maintaining ecosystem SERVICES, by ASsessing integrated impacts of cumulative pressures).

In the next sections, the concept and the entire approach for the development and testing of these ANERIS research products is documented. Each of the newly introduced DNA-based indicators can be calculated by using the three dedicated workflows, available by <https://my.lifewatch.eu/> portal. The potential users can now take advantage of these products in order to compute the newly developed indicators in a safe, documented and transparent environment.

List of Abbreviations

ANERIS: operAtional seNsing lifE technologies for maRIne ecosystemsS

ARMS: Autonomous Reef Monitoring Structures

DNA: Deoxyribonucleic Acid.

DEVOTES: DEvelopment Of innovative Tools for Understanding marine biodiversity and assessing good Environmental Status.

EMOBON: European Marine Omics Biodiversity Observation Network.

EU: European Union

GES4SEAS: Achieving Good Environmental Status for Maintaining Ecosystem Services by Assessing Integrated Impacts of Cumulative Pressures.

iBOL: International Barcode of Life.

NEAT: Nested Environmental Status Assessment Tool.

WoRMS: World Register of Marine Species.

1. Concept

The concept of the following proposed DNA-based indicators for monitoring the status of marine habitats, including the pelagic ones, is based on existing non-molecular ones. The existing indicators are those of the Taxonomic Distinctness (Warwick & Clarke 2001). The innovation of the newly developed indicators lies in the type of information to be used: in the existing indices, the data used to come from conventional measurements, such as those of organismal occurrence, abundance, biomass or productivity while in the proposed indices they come from DNA metabarcoding sequences. This approach overlooked microbial communities of bacteria and protists, which play essential roles in the structure, function, and element cycling of marine ecosystems. In addition, DNA metabarcoding type of data scales across spatial and temporal dimensions and has the potential to provide additional properties for assessing the ecological status of the habitats in question, particularly by capturing early signals of change.

1.1 Criteria adopted

The primary criteria set for the development of the DNA-based indices proposed are provided below:

Simplicity: The ANERIS Consortium has adopted this criterion for all its activities based on the principle that simplicity makes things easier to understand and implement, especially for non-scientists (e.g., policy-makers, industry, citizen scientists or the public). If the indices are, therefore, simple to implement, they will also be simple to interpret and establish value ranges that correspond to certain ecological status levels.

General use: The simpler an index is, the greater the chance that the types of data it requires can be easily collected from a wide range of habitats and ecosystems. The types of data needed here are simple taxon occurrences and their higher classification, which is easy to be retrieved. This is particularly useful for unpredictable habitats like the pelagic one, which is one of the largest marine habitats. Although the proposed DNA-based indices can be used for the pelagic habitat, there is nothing to prevent their use in other habitats when the conditions for their application are met.

Comparability: The results of the proposed DNA-based indices should be directly comparable to those obtained using conventional data. This allows for a different perspective on the ecological status of the habitats in question.

Low sampling effort: An indicator that achieves large sample sizes and fast processing with low sample size and effort, ensures that any initiative to collect data from any habitat, including pelagic habitats, can contribute to assessing their ecological status. Additionally, efforts from non-scientists, such as citizen scientists, can provide valuable data and information for ecosystem status assessments.

1.2 Data availability

An accessible way of collecting DNA-based data is metabarcoding, a multi-step process that includes isolating, amplifying, and reading specific DNA differing between taxa, which can resolve taxonomy down to the species level. This process was standardized a long time ago, and a global aggregator of such data, iBOL, is already operational. Accordingly, metabarcoding data were selected for calculating the proposed DNA-based indices. These indices can be used as raw numbers or slightly transformed through simple calculations whose meaning has already been tested with other types of data, such as species occurrence, abundance, biomass, and productivity. In Europe, metabarcoding data are being selected from a range of observation networks, such as EMOBON (Pavlouli *et al.* 2025).

2. Approach

The following paragraphs explain the approach used to identify the proposed DNA-based indices, as well as their properties for monitoring marine habitats. The corresponding workflows for their (semi)automated calculation is provided in the next section of the document.

2.1 Review

The ANERIS Consortium conducted a thorough review of relevant literature and online resources (e.g., observation networks, databases, global aggregators, existing software, web services, and workflows). The review provided substantial information that guided the selection of the proposed DNA-based indices to ensure their breadth of application and operability (Texteira *et al.* 2016). The review process was greatly facilitated by the results of two EU-funded projects: DEVOTES (2024) and GES4Seas (2024).

2.2 Selection of the new indicators

The indicator selection process included reviewing the ecological status indices provided by the DEVOTES project and, in particular, the NEAT software (Berg, 2016). Following the aforementioned primary criteria, the following DNA-based indices are proposed:

1. DNA-based trends for the arrival of new non-indigenous taxa.

This index applies to metabarcoding occurrence data. A standardized and operational workflow has been developed to identify new non-indigenous taxa in pelagic -and not only- samples and habitats.

2. DNA-based trends in the distributional range of invasive taxa.

This index is applied to metabarcoding occurrence data. A standardized and operational workflow has been developed that can calculate expansions in the distributional ranges of taxa in pelagic -and not only- habitats. These services can be transformed into proper workflows that are available online.

3. DNA-based taxonomic relatedness.

This index is applied to metabarcoding occurrence data. At the same time, it uses the higher classification of the taxa identified in a pelagic -and not only- sample. This process provides expected ranges of values for the index and/or departures from those expectations. There are already standardized and operational web services that can calculate expansions on distributional ranges of taxa in pelagic habitats, which can be transformed into proper workflows available online.

2.3 Metadata

The metadata used to officially describe and publish the above indices follows the standard introduced by the DEVOTES project and being available in the NEAT software.

3. Workflows for the calculation of the indices

The following paragraphs provide an overview of the workflows developed for each of the newly proposed indices.

3.1 DNA-based trends for the arrival of new non-indigenous taxa

This indicator aims to detect non-indigenous species entering pelagic habitats early on by using environmental DNA metabarcoding. Biological samples used in order to develop the index are those collected using Autonomous Reef Monitoring Structures (ARMS; Kuwahara et al. 2013) by the EMOBON network. The samples have been then processed through the PEMA pipeline (Zafeiropoulos et al 2020), which produces a metabarcoding matrix of operational taxonomic units (OTUs). The resulting sequences are lists for each sampling station. These lists are then cross-checked against the WoRMS (Costello et al. 2013) catalog of historical and regional occurrence records to pinpoint taxa absent from prior observations in the area. Newly detected non-indigenous species are flagged, yielding a time series that tracks the rate at which novel taxa appear. The workflow is available at:

<https://my.lifewatch.dev/workflow/execute-workflow-form>

D2.3 Development of new DNA-based indicators for pelagic habitats environmental status monitoring; ANERIS #101094924



Figure 1: Workflow depicting the stages of the calculation of the newly proposed index DNA-based trends for the arrival of new non-indigenous taxa.

3.2 DNA-based trend in the distributional range of invasive taxa

This indicator aims to quantify the geographic spread of invasive species by using spatial analyses to chart their expansion. Starting with the PEMA-derived species list from barcodes, records are filtered to include only recognized invasive taxa. Each occurrence is enriched with the sampling coordinates associated with ARMS deployments, and all points are pooled across temporal and spatial datasets. Geographic information services offered by MyLifeWatch are then used to compute spatial metrics for each invasive taxon. Comparing successive timeframes reveals the direction and magnitude of range expansion. The outputs include maps that visualize newly occupied regions. The workflow is available at:

<https://my.lifewatch.dev/workflow/execute-workflow-form>

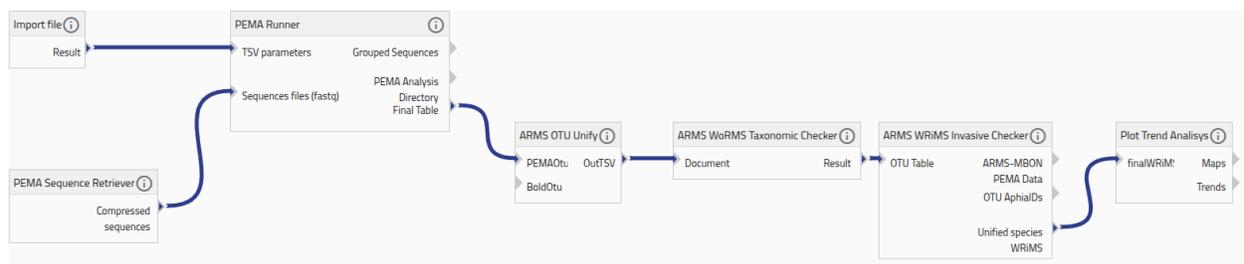


Figure 2: Workflow depicting the stages of the calculation of the newly proposed index DNA-based trend in the distributional range of invasive taxa.

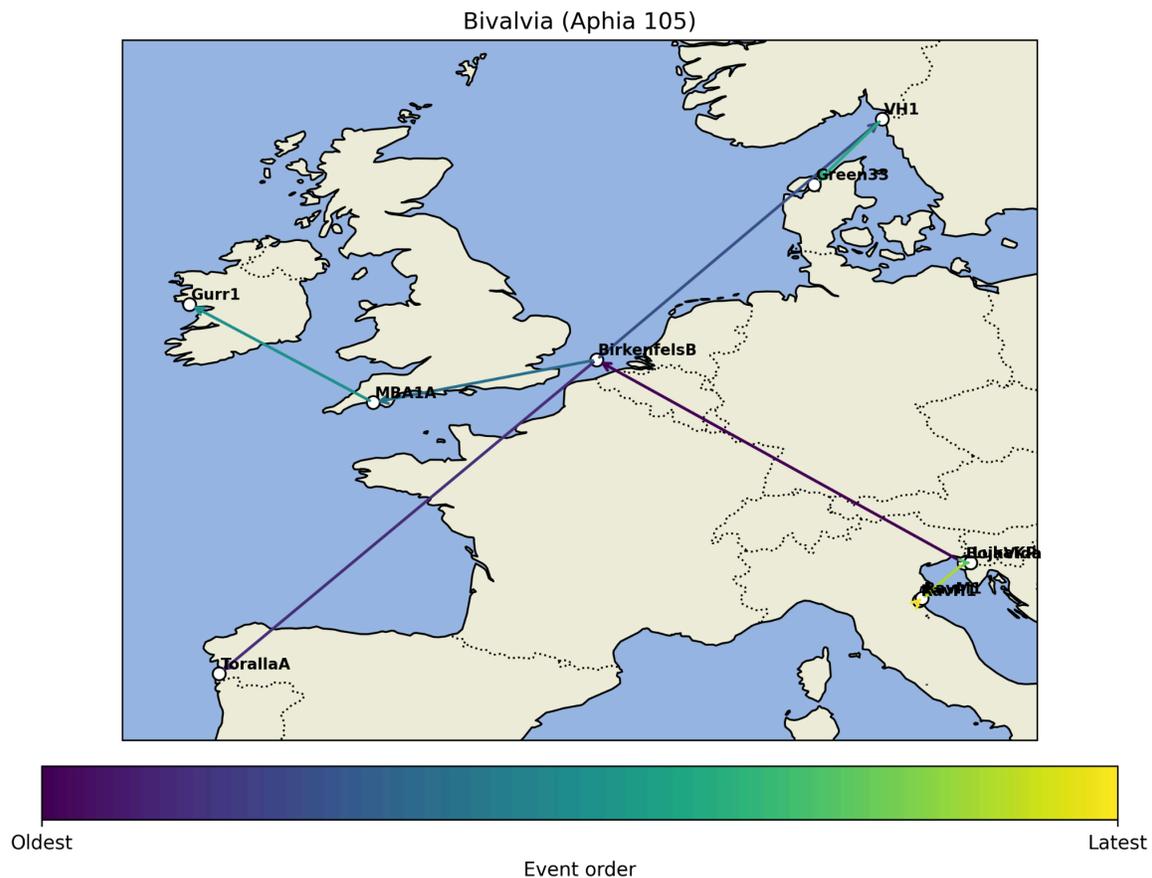


Figure 3: Workflow output depicting the range expansion of the taxa identified by the workflow for the calculation of the newly proposed index DNA-based trend for the distributional range of invasive taxa.

3.3 DNA-based Taxonomic Relatedness

This indicator assesses the taxonomic relationships among species in pelagic samples. It provides insight into community structure and the underlying ecological processes. First, species-level assignments from PEMA are harmonized against a standard taxonomy, such as WoRMS. Then, pairwise taxonomic distances are calculated using the Remote Virtual Laboratory for Taxonomic Computing (RvLab; Varsos *et al.* 2016), producing a distance matrix that captures distance among all detected species (taxa). This matrix can be normalized as needed and serves as the basis for diversity metrics, such as Faith's phylogenetic diversity, mean pairwise distance (MPD), or mean nearest taxon distance (MNTD). When relative abundance data are available, they can be incorporated to weight individual taxa. The resulting numerical index helps track shifts in community composition, detect homogenization, and

D2.3 Development of new DNA-based indicators for pelagic habitats environmental status monitoring; ANERIS #101094924

identify losses in functional diversity. The workflow is available at:
<https://my.lifewatch.eu/workflow/execute-workflow-form>

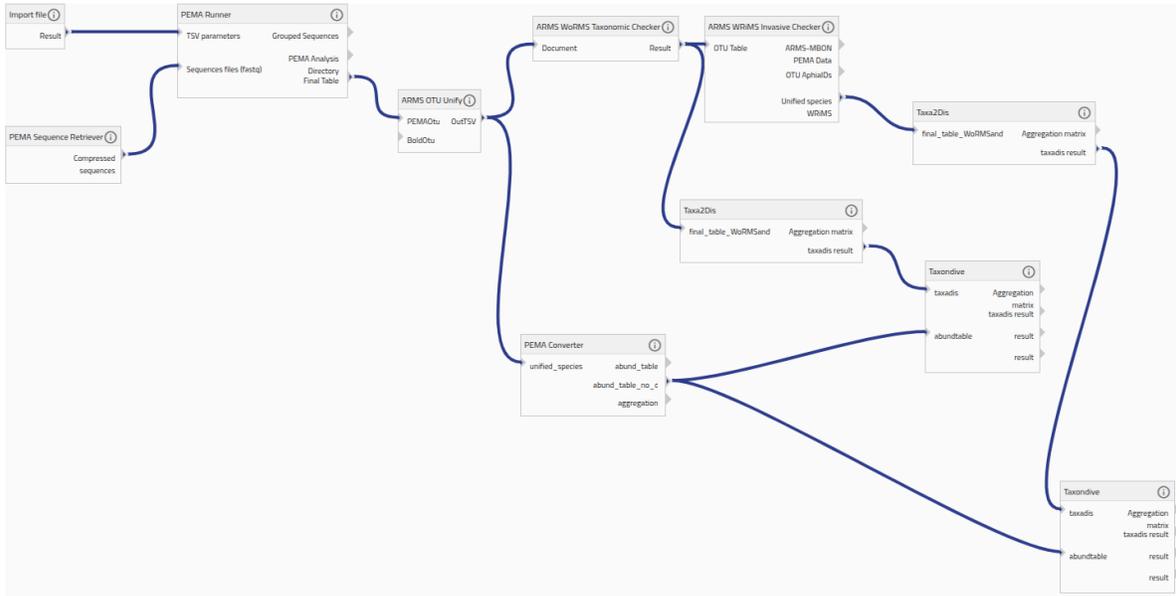


Figure 4: Workflow depicting the stages of the calculation of the newly proposed index DNA-based taxonomic relatedness.

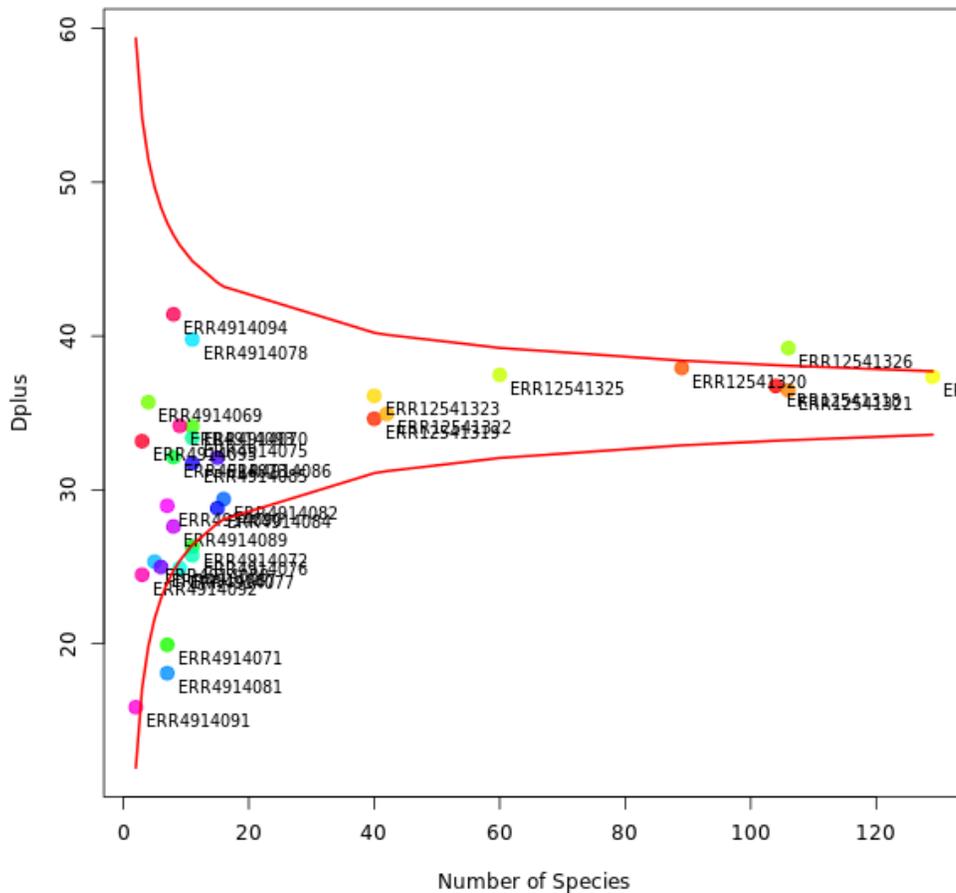


Figure 5: Funnel-plot of the 95% confidence intervals and taxonomic distinctness values of the samples superimposed, as it is produced by the workflow on the calculation of the newly proposed index DNA-based taxonomic relatedness (*Dplus* is the calculated Taxonomic Distinctness values).

Acknowledgements

The authors acknowledge the financial support provided by the European Union's Horizon Europe research and innovation programme under Grant Agreement No. [101094924], which has enabled the successful implementation of the ANERIS project. This funding has been essential for advancing our research and development activities in marine biodiversity assessment.

References

- Berg T, Murray, Ciarán M, J.Carstensen and J Andersen (2016). NEAT – Nested Environmental status Assessment Tool Manual. *Version 1 .1. 10.13140/RG.2.1.4144.6161*
- Costello MJ, Bouchet P, Boxshall G, Fauchald K, Gordon D, Hoeksema BW, Poore GCB, van Soest R, Stöhr S, Walter TC, Vanhoorne B, Decock W and W Appeltans (2013) Global Coordination and Standardisation in Marine Biodiversity through the World Register of Marine Species (WoRMS) and Related Databases. *PLOS One* 8: e51629.
<https://doi.org/10.1371/journal.pone.0051629>
- DEVOTES, 2024. <http://www.devotes-project.eu/>
- GES4SEAS, 2024. <https://www.ges4seas.eu/>
- Kuwahara A, Meyer C and A Collins (2013) Assessing Autonomous Reef Monitoring Structures (ARMS) as Biodiversity Monitors. *Integrative and Comparative Biology* 53: E119-E119
- Pavloudi C, Santi I, Azua I, Baña Z, Bastianini M, Belser C, Bilbao J, Bitz-Thorsen J, Broudin C, Camusat M, Cancio I, Caray-Counil L, Casotti R, Castel J, Comtet T, Cox CJ, Daguin C, Díaz de Cerio O, Exter K, Fauvelot C, Frada MJ, Galand PE, Garczarek L, González Fernández J, Guillou L, Hablützel PI, Heynderickx H, Houbin C, Kervella AE, Krystallas A, Lagaisse R, Laroquette A, Lescure L, Lopes E, Loulakaki M, Louro B, Magalhaes C, Maidanou M, Margiotta F, Montresor M, Not F, Paredes E, Percopo I, Péru E, Poulain J, Præbel K, Rigaut-Jalabert F, Romac S, Stavroulaki M, Souza Troncoso J, Thiébaud E, Thomas W, Tkacz A, Trano AC, Wincker P and N. Pade (2025) First release of the European marine omics biodiversity observation network (EMO BON) shotgun metagenomics data from water and sediment samples. *Biodiversity Data Journal*, 2025 Mar 12;13:e143585. doi: [10.3897/BDJ.13.e143585](https://doi.org/10.3897/BDJ.13.e143585)
- Teixeira H, Berg T, Uusitalo L, Fürhaupter K, Heiskanen A-S, Mazik K, Lynam CP, Neville S, Rodriguez JG, Papadopoulou N, Moncheva S, Churilova T, Kryvenko O, Krause-Jensen D, Zaiko A, Veríssimo H, Pantazi M, Carvalho S, Patrício J, Uyarra MC. À. Borja (2016) A Catalogue of Marine Biodiversity Indicators. *Frontiers in Marine Science* 3:207. Doi: 10.3389/fmars.2016.00207
- Varsos C, Patkos T, Oulas A, Pavloudi C, Gougousis A, Ijaz U, Filiopoulou I, Pattakos N, Vanden Berghe E, Fernández-Guerra A, Faulwetter S, Chatzinikolaou E, Pafilis E, Bekiari C, Doerr M, Arvanitidis C. (2016) Optimized R functions for analysis of ecological community data using the R virtual laboratory (RvLab). *Biodiversity Data Journal* 4: e8357.
<https://doi.org/10.3897/BDJ.4.e8357>
- Warwick R.M. and K.R. Clarke (2001) Practical measures of marine biodiversity based on relatedness of species. *Oceanography and Marine Biology: an Annual Review* 39: 207 - 231

D2.3 Development of new DNA-based indicators for pelagic habitats environmental status monitoring; ANERIS #101094924

Zafeiropoulos H, Viet HQ, Vasileiadou K, Potirakis A, Arvanitidis C, Topalis P, Pavludi C and E Pafilis (2020) PEMA: a flexible Pipeline for Environmental DNA Metabarcoding Analysis of the 16S/18S ribosomal RNA, ITS, and COI marker genes. *GigaScience* 9, gaa022, <https://doi.org/10.1093/gigascience/gaa022>

Annex I. Functionalities of the common services used in all workflows for the three newly DNA-based indices

PEMA Sequence Retriever: This service aims to download the sequences files (fastq files) that have been chosen by the user in order to run PEMA.

inputs	outputs
- None	- sequences_compressed.zip

PEMA Runner: This wrapper aims at running PEMA on all the sequences from the samples selected in the PEMA Sequence Retriever

inputs	outputs
- Parameters table in TSV format -sequences_compressed.zip (from PEMA Sequence Retriever)	- all_sequences_grouped.fa - final_table.tsv* - pema_analysis_dir.zip

*final_table.tsv: It is a TSV file with the number of columns that depends on the number of sequences/samples that are processed in the PEMA run. The first column is an ID. The final column contains the species information. The columns in between are for each sequence (each sample) that was processed and contain integer values; the title of these columns is the ENA code of the sequence (the same as in the MasterARMS csv file).

ARMS OTU Unify: This service aims at creating an OTU table with species information that can be processed by WoRMS and WRIMS.

inputs	outputs
- final_table (from PEMA Runner)	- otu.tsv

ARMS WoRMS Taxonomic Checker: This service takes the OTU tables from the previous step and check it using the WoRMS taxon match webservice.

inputs	outputs
--------	---------

- otu.tsv (from ARMS OTU Unify)	- worms.csv
---------------------------------	-------------

ARMS WRiMS Invasive Checker: This service takes the output from the previous step and uses the WRiMS taxon match webservice to check the known distribution of the species.

inputs	outputs
- worms.csv (from Worms)	- ARMS4Tesseract_PEMA_data.csv - OTU_withAphia.csv - final_table_WoRMSandWRiMS.csv*

*final_table_WoRMSandWRiMS.csv: table where each row records the presence/abundance of a specific OTU in a given sample (identified by the ENA accession in AccessionID). classification gives the full taxonomic path, OTU is the cluster ID, and Count is the integer read count for that OTU in that sample. Sampling metadata include latitude, longitude, and Sample_ID. Taxonomic crossreferences come via Aphia_ID with the WoRMS scientific name and rank.

Annex II. Functionalities of the specific services used in each of the workflows used for the three newly DNA-based indices

PEMA Converter: Used in DNA-Based Trend for the Arrival of New Non-Indigenous Taxa and DNA-Based Taxonomic Relatedness. This service aims to reformat and produce final output in various formats for human and machine2machine reading.

inputs	outputs
- otu.tsv (from ARMS OTU Unify)	- abund_table.csv - abund_table_no_dupls.csv* - aggregation.csv

*-abund_table_no_dupls.csv: output abundance table no duplicates filename.

Metamds Observations: Used in DNA-Based Trend for the Arrival of New Non-Indigenous Taxa. This service aims to reformat and produce final output in various formats for human and machine2machine reading.

D2.3 Development of new DNA-based indicators for pelagic habitats environmental status monitoring; ANERIS #101094924

inputs	outputs
- abund_table.csv (from PEMA Converter) - factorfile.csv	- mds_observations.png

Plot Trend Analysis: Used in DNA-Based Trend in the Distributional Range of Invasive Taxa. This service aims to reformat and produce final output in various formats for human and machine2machine reading.

inputs	outputs
- final_table_WoRMSandWRiMS.csv (from WRiMS)	- maps_by_species.zip - zone_groups_trends.zip

Taxa2Dis: Used in DNA-Based Taxonomic Relatedness. This service is used to transform the matrix of the WRiMS component output into a valid format (agg_matrix) for the taxa2dis function. This function will output the final distance matrix (taxadis).

inputs	outputs
- final_table_WoRMSandWRiMS.csv (from WRiMS)	- agg_matrix.csv - taxadis.csv

Taxondive: Used in DNA-Based Taxonomic Relatedness. This service is used to calculate taxonomic distinction metrics per sample (Δ^+ , expectation, and dispersion), export the results, and generate a graph with the results.

inputs	outputs
- taxadis.csv (from Taxa2Dis) - abund_table_no_dupls.csv (from PEMA Converter)	- rplot.png - legend.png - taxondive.csv - taxondive_table.csv