University of Southampton

Faculty of Environmental and Life Sciences

School of Ocean and Earth Science

Autonomous Biomolecular Sampling: Enhancing the Temporal and Spatial Resolution of Ocean Biodiversity Observations

by

Robyn Mairin Samuel

https://orcid.org/0000-0001-5989-4588

Thesis for the degree of Doctor of Philosophy

September 2025

University of Southampton

Abstract

Faculty of Environmental and Life Sciences

School of Ocean and Earth Science

Doctor of Philosophy

Autonomous Biomolecular Sampling: Enhancing the Temporal and Spatial Resolution of Ocean Biodiversity Observations

by

Robyn Mairin Samuel

On land, we can easily detect ecological changes through our senses, whether it's hearing a new bird species or noticing fewer insects, and these sensory cues drive scientific inquiry. In contrast, changes in marine ecosystems are far less perceptible, making it challenging to notice and investigate ecological trends. Environmental DNA (eDNA) monitoring provides a powerful biomolecular tool to reveal these hidden trends, offering critical insight into the impacts of anthropogenic activities on marine ecosystems.

Traditional approaches to understanding these impacts often rely on micro- and mesocosm experiments, which may not accurately reflect natural marine communities. Long-term place-based monitoring is essential to observe cumulative effects and support adaptive management strategies. However, conventional visual monitoring in marine environments is costly, logistically challenging, and sometimes hazardous, limiting its ability to deliver the temporal and spatial resolution required to detect change.

Autonomous eDNA sampling technologies, such as the Robotic Cartridge Sampling Instrument (RoCSI), offer a promising solution. Although the initial investment is substantial, a network of these autonomous samplers could provide consistent, low-cost biodiversity data, that is resilient to funding fluctuations, due to the option to store samples in biobanks for future analysis.

This doctoral research explores how biomolecular monitoring and autonomous technologies like RoCSI can enhance marine biodiversity monitoring. Chapter 2 demonstrates the RoCSI's ability to autonomously collect samples via a ship's underway system, increasing the spatial resolution of eDNA sampling. Chapter 3 focuses on the RoCSI's capability to capture high temporal resolution samples in a highly urbanised estuary. Chapter 4 optimises the RoCSI for long-term deployment by comparing different liquid nucleic acid preservatives. Chapter 5 discusses improvements in data management practices for biomolecular research, facilitating the integration of RoCSI data into a global/national scale observation network. Chapter 6 evaluates the Oxford Nanopore's portable MinION sequencer for its potential use alongside the RoCSI.

Overall, this thesis validates the RoCSI as an effective tool for high-resolution temporal and spatial biodiversity monitoring, optimises it use for long-term deployments, and advances data management practices for large-scale biomolecular observations. The findings provide foundational evidence for establishing a national-scale ocean biomolecular observatory that combines both autonomous and manual methods to deliver consistent, long-term ecological data needed to inform adaptive, place-based management of marine ecosystems.

Table of Contents

Table of Tables

Table of Figures

Table of Figures

Table of Figures

Table of Figures

Research Thesis: Declaration of Authorship

Print name: Robyn Mairin Samuel

Title of thesis: Autonomous Biomolecular Sampling: Enhancing the Temporal and Spatial Resolution of Ocean Biodiversity Observations

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:-

Samuel, R. M., Meyer, R., Buttigieg, P. L., Davies, N., Jeffery, N. W., Meyer, C., Pavloudi, C., Pitz, K. J., Sweetlove, M., Theroux, S., van de Kamp, J., & Watts, A. (2021). Toward a Global Public Repository of Community Protocols to Encourage Best Practices in Biomolecular Ocean Observing and Research. Frontiers in Marine Science, 8, 758694. https://doi.org/10.3389/FMARS.2021.758694/BIBTEX

Signature: ..... Date: 7th September 2025

Acknowledgements

Firstly, I would like to thank my supervisors for their guidance and inspiration throughout this PhD journey. My thanks also go to the Ocean Technology and Engineering group at the National Oceanography Centre. Your takeout Thursdays, breakfast clubs, international food days, and pub enforcements made coming to work fun, even when lab experiments were failing. A special thanks to my lab manager, Annika Simpson, for organising countless socials, keeping us on track in the lab, and being an all-round lovely person. I am grateful to all my friends in office 251/28 for all the coffee breaks and office shenanigans.

I would like to thank Pier Luigi Buttigieg, Neil Davies, Raïssa Meyer, Kathleen Johnson Pitz, all the Better Ocean Biomolecular Ocean Practices (BEBOP) group, the Omic Biodiversity Observation Network (Omic BON), and all the Ocean Biomolecular Observation Network (OBON), for their visionary thinking, which provided much-needed inspiration during the COVID -19 lockdowns. I extend my thanks to the DNA team and the marine Natural Capital and Ecosystem Assessment (mNCEA) programme at Natural England for helping me understand how eDNA analysis can be integrated into policy and management decision making.

Finally, I am immensely grateful to my friends for celebrating every PhD milestone and offering much-needed distractions. And to my family for all your support, I could probably write an entire PhD on everything I am grateful to you for. Above all, I want to thank my now-husband, Grant Young, and our wonderful son, Kye, for transforming tough times into the most magical moments.

Definitions and Abbreviations

AMS ..... Autonomous Microbial Sampler

Argo ..... A broad-scale global array of temperature/salinity profiling floats

AUV ..... Autonomous Underwater Vehicles

BeBOP ..... Better Biomolecular Ocean Practices

CAD ..... Computer-Aided Design

COI ..... mitochondrial gene cytochrome C Oxidase subunit I - A standardized molecular marker for classifying animal species

EBV ..... Essential Biodiversity Variable

eDNA ..... Environmental DeoxyriboNucleic Acids – The mixture of genomic material that can be extracted from an environmental sample (such as water, soil, faeces and air)

EOV ..... Essential Ocean Variable

eRNA ..... Environmental RiboNucleic Acids – The ribonucleic acids that can be extracted from an environmental sample (such as water, soil, faeces and air)

ESP ..... Environmental Sample Processor

FF ..... Flash Frozen

FISH ..... Fluorescent In Situ Hybridisation

GC Content ..... Guanine-Cytosine Content

GEO BON ..... Group on Earth Observations Biodiversity Observation Network

GLOMICON ..... Global Omics Observatory Network

GOOS ..... Global Ocean Observing Systems

GSC ..... Genomics Standard Consortium

IFFS ..... In-situ Filtration and Fixation Sampler

INSDC ..... International Nucleotide Sequence Database Collaboration

IOC ..... Intergovernmental Oceanographic Commissions

IODE ..... International Oceanographic Data and Information Exchange

KO ..... Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology

LAMP ..... Loop-mediated isothermal amplification

LR-AUV ..... Long Range-Autonomous Underwater Vehicle

MBON ..... Marine Biodiversity Observation Network

MIOP ..... Minimum Information about any Omic Protocol

MIxS ..... Minimum Information about any (x) Sequence

MS-SID ..... Microbial Sampler-Submersible Incubation Device

NAP ..... Nucleic Acid Preservative

NOC ..... National Oceanography Centre

OBIS ..... Ocean Biogeographic Information System

OBON ..... Ocean Biomolecular Observing Network

OBPS ..... Ocean Best Practice Systems

Omic BON ..... Omic Biodiversity Observation Network

OMZ ..... Oxygen Minimum Zone

OSD ..... Ocean Sampling Day

OTE ..... Ocean Technology and Engineering

PAP-SO ..... Porcupine Abyssal Plain Sustained Observatory

qPCR ..... Quantitative Polymerase Chain Reaction

RAS ..... Remote Access Sampler

RLT ..... RLT+ Buffer – The initial lysis buffer for the AllPrep extraction Kit (Qiagen)

RoCSI ..... Robotic Cartridge Sampling Instrument

ROV ..... Remote Operated Vehicles

SHA ..... Sandwich-Hybridization Assays

Shield ..... DNA/RNA Shield (Zymo Research)

SID ..... Submersible Incubation Device

SUPR ..... Suspended Particulate Rosette

UNESCO ..... United Nations Educational, Scientific and Cultural Organization

Chapter 1 Autonomous eDNA sampling with the Robotic Cartridge Sampling Instrument: Increasing the temporal and spatial resolution of marine biodiversity observations for the UN Ocean Decade for Sustainable Development

1.1 What is environmental DNA?

Environmental DNA (eDNA) is the mixture of genomic material that can be extracted from an environmental sample (such as water, soil, faeces and air). Potential sources of DNA within an environmental sample include cellular materials (e.g. broken cells, DNA molecules), pieces of tissue (mucus, scales, skin, etc.) and faeces that are sloughed off or excreted by organisms to the environment, as well as whole microorganisms and gametes (Pawlowski et al., 2020; Taberlet et al., 2012). The term eDNA is most commonly used for research focusing on macro-organisms. However, microbial life is always present in eDNA samples and increasingly research is focusing on both the microbial and macrobial life (Pawlowski et al., 2020). Therefore, in this thesis the term eDNA will be used to refer to all research that examines the nucleic acids (DNA and RNA) found in environmental samples, including all living organisms, from microbes to megafauna.

Samples of eDNA can be used to monitor target species, functional genes, and/or overall community composition (Bohmann et al., 2014). Targeted quantification of key taxa or functional genes is most commonly performed using quantitative PCR (qPCR), which provides quantitative data on the presence of DNA from the target taxa or gene within a sample (Knudsen et al., 2019; Tang et al., 2020). Metabarcoding is another widely used technique for determining overall community composition (Holman et al., 2019). Metabarcoding studies can provide relatively comprehensive data for the presence/absence of a wide range of taxa, however, abundance estimates from this method should be interpreted with caution (Bista et al., 2018; L. J. Clarke et al., 2017; Deiner, Renshaw, et al., 2017), due to interspecific differences in gene copy number, difference in DNA shedding rates, as well as biases introduced during PCR amplification. Shotgun sequencing or metagenomic approaches are less established methods for identifying community composition from eDNA samples, mainly used in microbial research

(Jo et al., 2020). However, metagenomics methods are rapidly developing and have the potential to provide overall community composition data with more reliable biomass estimates that are less impacted by amplification biases (Bista et al., 2018; Tessler et al., 2017; Venter et al., 2019). Metabarcoding, qPCR and metagenomics are the three approaches used and discussed in this thesis, but many other techniques such a fluorescent in situ hybridisation (FISH), or Loop-mediated isothermal amplification (LAMP) can also be used to analyse eDNA (Notomi et al., 2015; Williams et al., 2021; Zwirglmaier, 2005)

1.2 A brief history of eDNA and autonomous sampling in the marine environment

Techniques for capturing eDNA and identifying taxonomic diversity were first developed in microbiology. In this field, many microorganisms that cannot be cultured were instead identified using molecular techniques (Giovannoni et al., 1990; Ward et al., 1990). In 1987, Ogram et al. produced a protocol for the extraction of microbial DNA from sediment (Ogram et al., 1987). By 1990 the first metabarcoding studies were published, analysing the diversity of the 16S rRNA gene for bacterioplankton from the Sargasso Sea (Giovannoni et al., 1990) and the cyanobacterial mat from Octopus Spring, Yellowstone National Park (Ward et al., 1990).

The first articles using DNA barcodes to identify macro-organisms occurred in 2003, when Hebert et al. demonstrated that profiles of the mitochondrial gene cytochrome c oxidase I (COI), could be used to correctly distinguish between 200 closely allied lepidopteran species(Hebert, Cywinska, et al., 2003; Hebert, Ratnasingham, et al., 2003). Also in 2003, Willerslev et al. established ancient DNA (aDNA) analysis by revealing that permafrost and sediments contained preserved plant and animal DNA that could be used to characterize the taxonomic diversity of paleoenvironments(Willerslev et al., 2003).

Initial eDNA metabarcoding relied on the cloning of PCR products prior to Sanger sequencing. In 2004, this expensive and time-consuming cloning step was made unnecessary by the introduction of next generation sequencing (Shendure & Ji, 2008). At this point eDNA metabarcoding became a viable approach for a much broader range of research groups, starting within the field of microbiology (Galand et al., 2009; Petrosino et al., 2009; Roh et al., 2010) and soon extending to the study of macro-organisms (Ficetola et al., 2008; Thomsen et al., 2012; Valentini et al., 2009). Single species detection from environmental samples using qPCR also became commonplace around this time (Dejean et al., 2011; Erdner et al., 2010; Jerde et al., 2011).

As sequencing technologies developed, so did sampling technologies. Initial devices such as the Submersible Incubation Device (SID) and Remote Access Sampler (RAS) were designed to collect and preserve whole water samples (McKinney et al., 1997; Taylor & Doherty, 1990). Then, as molecular analysis became more accessible, devices were designed to filter water samples in-situ, such as the Autonomous Microbial Sampler (AMS) and the Suspended Particulate Rosette (SUPR; (Breier et al., 2009; Taylor et al., 2006). Further devices such as the In-situ Filtration and Fixation Sampler (IFFS) and Microbial Sampler-Submersible Incubation Device (MS-SID) were later designed to both filter and preserve samples in-situ (Bombar et al., 2015; Wurzbacher et al., 2012). Preserving samples in-situ allows for the preservation of DNA and RNA before exposing the water sample, and the microbial life within it, to the variable environmental conditions that occur during retrieval from depth, such as changes in pressure and temperature, which can affect gene expression profiles (Edgcomb et al., 2016). Now, large devices such as the Environmental Sample Processor (ESP) can filter, preserve and provide in-situ molecular analysis such as sandwich-hybridization assays (SHA) and qPCR (Preston et al., 2009; Robidart, Preston, et al., 2012; Robidart, Shilova, et al., 2012; Scholin et al., 2006, 2009).

In summary, over the last three decades, eDNA analysis has evolved from a tool primarily used to identify unculturable microbial life, to an effective approach for detecting taxa across the tree of life. Advances in technology now allow the nucleic acids within an environmental sample to reveal not only the presence of taxa but also the biological functioning of entire communities. Additionally, with the adoption of autonomous eDNA sampling technologies, biodiversity can now be observed at unprecedented spatial, temporal, and taxonomic resolutions.

1.3 The Robotic Cartridge Sampling Instrument

The Robotic Cartridge Sampling Instrument (RoCSI) is an autonomous eDNA sampler developed by the Ocean Technology and Engineering Group at the National Oceanography Centre, Southampton, UK. The RoCSI device filters seawater through commercially available Sterivex™ filters using a peristaltic pump. The Sterivex™ filters are loaded into specially adapted cartridges with luer-activated valves and stored in a bandolier. The Bandolier is mounted on a rotating Geneva wheel, which moves the samples from the initial filtration system to the secondary preservation system (Figure 1). In-line pressure sensors detect if the filters begin to clog; when this occurs, seawater filtration is halted, and samples are preserved. This mechanism allows the device to be function across a range of biomass levels.

The RoCSI has been designed for use with mass-produced, commercially available consumables, reducing reliance on bespoke components and minimising potential supply chain disruptions. Consumables for the RoCSI device include Sterivex™ cartridges and Flexboy®

blood bags to store the preservative. Both the filters and blood bags can be interchanged to meet the specific needs of the research; Sterivex™ filters come in either a 0.22 or 0.45µm pore size; Flexboy® blood bags come in a range of sizes; and the type of liquid preservative can be interchanged.

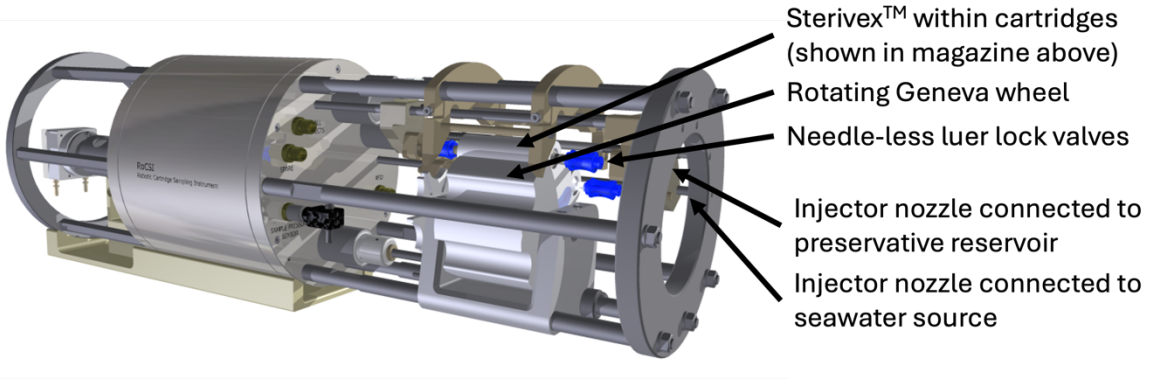


Figure 1 Computer-aided design (CAD) of the commercially available Robotic Cartridge Sampling instrument (RoCSI)

1.4 Autonomous eDNA sampling in the UN Decade of Ocean Science for Sustainable Development

A call for an international network of genomic observatories was first posited in 2012 by Davies et al. Since then, projects such as Tara Oceans and the Ocean Sampling Day (OSD) have begun tackling genomic observations at a global scale (Sunagawa et al., 2015; Tragin & Vaulot, 2018). However, individual projects typically focus sampling efforts along either temporal or spatial axis. For example, the TARA oceans expedition achieved vast spatial coverage by taking thousands of samples whilst circumnavigating the globe from 2009-2013, but without repeated temporal sampling at each location it is difficult to distinguish between temporal or spatial variability in biodiversity. The OSD, which began in 2014, has addressed this by coordinating across research groups for a yearly sampling event at midday of the northern hemisphere's summer solstice. This has enabled global comparisons of green microalgae using metabarcoding of the 18S V4 region and the repeated yearly sampling will help to identify decadal trends (Tragin & Vaulot, 2018). However, seasonal variation is not accounted for when sampling only once a year. Establishing a network of biodiversity observations using eDNA needs a coordinated effort much like the OSD but at a much larger scale, utilising a combination of data sources including autonomous technologies, citizen science and data contributions from research groups around the world.

1.4.1 The United Nations Decade of Ocean Science for Sustainable Development

The UN decade of ocean science for sustainable development was declared by the Intergovernmental Oceanographic Commissions (IOC) of the United Nations Educational, Scientific and Cultural Organization (UNESCO) to be held from 2021 -2030, as a way of building capacity across the UN member states to achieve the United Nations Sustainable Development Goal 14, 'to conserve and sustainably manage ocean and marine resources by 2030'. The decade has an overarching vision to 'develop scientific knowledge, build infrastructure and foster relationships for a sustainable and healthy ocean' and aims to achieve six societal outcomes: (1) A clean ocean, through identifying and removing sources of pollution; (2) A healthy and resilient ocean, with mapped and protected marine ecosystems; (3) A predicted ocean, enabling society to understand current and future ocean conditions; (4) A safe ocean, protecting people from ocean hazards; (5) A sustainably harvested ocean, providing food and resources for the blue economy; and (6) A transparent ocean, giving citizens equitable access to data, information and technologies (Heymans et al., 2020).

Environmental DNA has been highlighted as a key tool for achieving a quantitative understanding of ocean ecosystems and their functioning throughout the UN decade of ocean science (Ryabinin et al., 2019). This quantitative approach to assessing ecosystem variables is needed to advance ecosystem modelling and forecasting, which will contribute to outcome 3 of ocean decade 'a predicted ocean, enabling society to understand current and future ocean conditions'. An emphasis has been put on eDNA methods to assess ecological variables at temporal and spatial resolutions great enough for ecosystem modelling (Heymans et al., 2020). This is due to the relatively low cost of sampling and high taxonomic coverage of metabarcoding and metagenomic approaches.

Ocean observations of physical and chemical parameters have already come a long way through programs such as Argo, which collect temperature and salinity depth profiles from a fleet of over 4000 Argo floats covering the world's ocean (Roemmich et al., 2019). The monitoring of ocean chemistry (CO_2, Nutrients, trace metals, etc) has been vastly increased with the implementation of in-situ sensors deployed on mooring buoys, ship of opportunity, autonomous underwater vehicles (AUVs), and remote operated vehicles (ROVs)(Beaton et al., 2012; Rérolle et al., 2018). Collating these global observations in a format that is compatible for ecosystem modelling is a huge task requiring collaboration across regions and organisations. Therefore, a number of intergovernmental organisations, including the Ocean Biogeographic Information System (OBIS) of the IOC, the Global Ocean Observing Systems (GOOS), and the Marine Biodiversity Observation Network (MBON - the marine node of the Group on Earth Observations Biodiversity Observation Network (GEO BON)), are working together to establish a

network of integrated ocean observations using standardised data management that is accessible to all nations (Miloslavich et al., 2018).

GOOS has established a set of Essential Ocean Variables (EOVs) and GEOBON have outlined a set of Essential Biodiversity Variables (EBVs) that are needed for ecosystem modelling and feasible for cost effective global-scale observations (Muller-Karger et al., 2018). Satellite telemetry data, such as sea surface temperature, colour and altimetry along with in-situ sensors for subsurface temperature and salinity from Argo floats are already well-established data streams within the GOOS framework (Whitt et al., 2020). Collectively, these data have led to a greater understanding of ocean productivity, currents and improved climatic forecasting (Roemmich et al., 2019).

To increase the complexity and improve current oceanic modelling, future models need to include high temporal resolution biological variables beyond ocean colour measurements. Consequently, GOOS and GEOBON are now in the pilot stages of introducing molecular taxonomy of microbial and invertebrate diversity as both an EOV and an EBV (Goodwin et al., 2018; Muller-Karger et al., 2018). During this pilot stage small scale projects using eDNA methods will be expanded to a global scale. To do this a set of standard best practices is needed to integrate regional and national projects into a global effort. The Global Omics Observatory Network (GLOMICON) is currently running a project to compare best practices for the collection and extraction of seawater eDNA across research groups (Buttigieg et al., 2019). These cross-lab comparisons are beginning to reveal the extent of variation caused by differing protocols and will help to inform decisions on how to best align practices across research groups (Zaiko et al., 2022).

However, a careful balance will need to be struck between standardisation and innovation. Innovative methods and continuous optimisation of current methods can greatly facilitate eDNA research by reducing costs and/or biases associated with previous methods. Thus, standardisation risks impeding progress if measures aren't in place to review and update best practices (Hörstmann et al., 2020; Pearlman et al., 2019). This is particularly relevant to the field of eDNA, which has already been revolutionised with the introduction of high-throughput sequencing or next-generation sequencing (NGS) platforms and is likely to see further advances in sequencing technologies in the near future (Shendure et al., 2017).

Ocean Best Practice Systems (OBPS) is a project with links to GOOS and the International Oceanographic Data and Information Exchange (IODE), which is currently developing a system to record, update, disseminate and harmonise best practices within ocean research, with a working group specifically dedicated to omics & eDNA research (Hörstmann et al., 2020; Pearlman et al., 2019). The OBPS has been adopted by the IOC-UNESCO and as such is in a

unique position to coordinate a global community of omics and eDNA researchers. An OBPS omics/eDNA working group consisting of 65 participants from across the globe has been tasked with establishing a system for collating and categorising protocols, so they can be continuously tested and compared across research groups. This collaborative effort aims to facilitate the adoption of standard practices across regions and organisations. Establishing a global network of ocean eDNA observations is essential for identifying global scale patterns in marine biodiversity and improving ecosystem modelling and forecasting to support 'a predicted ocean'.

1.4.2 How will RoCSI sampling support the UN Ocean Decade

The RoCSI device will play a role in increasing the temporal and spatial resolution of ocean eDNA samples during the UN Decade. RoCSI can be deployed at fixed observatories to increase the temporal frequency of sampling or on ships of opportunity to increase the spatial coverage of eDNA sampling without the additional costs incurred from additional voyages using expensive research vessels. Integrating the RoCSI device with AUVs and ROVs will facilitate sampling at remote locations. Furthermore, integrating the RoCSI device with research vessels underway systems will enable high frequency sampling, which could be continued throughout adverse sea states, when sampling would otherwise be postponed for the health and safety of crew. Automated sampling will free up time for researchers, allowing time for more in-situ experimentation thus, maximising the total scientific output from a research cruise.

Customisable aspects of the RoCSI device such as interchangeable preservatives, filter types, sample volumes and sampling regimes increase the adaptability of RoCSI sampling for an array of best practice sampling protocols. For example, eDNA water sampling for fish biodiversity surveys often use 0.45µm filters and large volumes of water (>1L)(Mynott & Marsh, 2020; Thomsen et al., 2012). Whereas coastal multi-marker metabarcoding surveys aiming to capture bacterial, invertebrate and vertebrate biodiversity tend to use smaller pores size (0.22µm) and smaller water volumes (>400ml) (Djurhuus et al., 2018, 2020; Holman et al., 2019). This flexibility in sampling protocols with the RoCSI device, provides scope for the adaptation and adoption of standard practices that are established through the OBPS throughout the UN Ocean decade.

1.4.3 Key challenges for RoCSI ocean observations during UN Ocean Decade

1.4.3.1 Technology

During the UN Ocean Decade the RoCSI device will be developed for remote sampling with a LR-AUV. Systems will be developed for integration with other sensors, to enable smart sampling regimes based on the data from other remote sensors. For example, integration with a

fluorometer, could trigger sampling to begin when chlorophyll content crosses a predefined threshold. Targeted sampling like this enables remote sampling of temporary ocean features that are typically only sampled opportunistically, such as bloom conditions caused by ocean gyres or iron seeding events.

To progress toward fully automated genomic sensing the RoCSI device will need to be integrated with an autonomous nucleic acid extraction unit and an analysis unit. The OTE group at NOC are developing an extraction unit and the Amplitron device for in-situ quantitative amplification of target DNA. To increase the flexibility of sampling regimes these units will have a modular design, so the RoCSI device can be used either as a sampler or as part of a genomic sensor. Furthermore, the modular design will provide scope for integration with portable analysis devices, such as the MinION sequencer by Oxford Nanopore technologies, for in-situ metagenomic sequencing.

1.4.3.2 Data Management

A key challenge for coordinating a global scale network for marine biodiversity observations will be in managing the large quantities of sequencing data and metadata. High quality metadata will be needed for global meta-analysis to provide meaningful insights into data collected by multiple research organisations. So far, the omics/eDNA field has a good track record for making sequence data open access. It is commonplace for sequence data to be uploaded to the International Nucleotide Sequence Database Collaboration (INSDC) and many journals require the submission of data before publication (Karsch-Mizrachi et al. 2011). This means that a large amount of data is already freely accessible for cross comparisons (Mitchell et al., 2020). However, detailed and accurate metadata is also needed to increase the contextualisation and findability of this data to make it suitable for global scale comparisons.

Metadata standards, such as the 'Minimum Information about any (x) Sequence' (MIxS) form the Genomics Standard Consortium (GSC) provide metadata checklists to help to improve and standardise metadata (Bowers et al., 2017; Yilmaz et al., 2011). However, compliance to these standards is not universal and mistakes are often made which reduces the findability of the associated sequences. Platforms such as MGnify have been developed to facilitate the uploading of sufficient metadata for cross project comparisons of metagenomic data (Mitchell et al. 2020). The RoCSI automatically records sample metadata. Ensuring compliance with MGnify and MIxS standards will mean the RoCSI can further facilitate the process of uploading sequence data. This capability will allow sequences from RoCSI to be readily accessible for data mining and well-suited for future meta-analyses.

Currently, minimum metadata standards only require limited information on the methods used for eDNA collection, preservation, extraction and analysis. However, each aspect of the methods used can introduce different biases (Deiner et al., 2015; Gray et al., 2013; McCarthy et al., 2015; Zaiko et al., 2022). Therefore, metadata needs to include information on the types of methods used, so that they can be factored into future meta-analysis. This is especially relevant to the field of eDNA, where the rapid growth in eDNA research, over the last decade, has led to an array of different protocols being adopted and optimised by individual research groups.

Involvement in the OBPS community review process will help to align and improve eDNA methods across research groups. Furthermore, taking part in OBPS discussions and following the development of best practices will guide the recommendations for autonomous sampling of eDNA with the RoCSI device. Autonomous sampling devices that adhere to current best practices will enhance synergy between autonomous and manual sample collection efforts. Thus, increasing the quantity of samples that are directly comparable and suitable for a network of global observations.

Another key challenge will be in developing a system for integrating data from other sensors. Often, when sampling for eDNA, additional measurements are taken for variables such as temperature, salinity chlorophyll, etc. Many of these variables are also EOVs and/or EBVs. Therefore, forward planning to take these measurements using GOOS and GEOBON compliant protocols will make this data suitable for use by other ocean observation communities, thereby increasing the scientific value of the data. To facilitate this process autonomous systems could be developed to collate data from multiple sensors and output the data in an interoperable format.

1.4.3.3 Societal Impacts

Governmental organisations in the UK have shown interest in the adoption of autonomous eDNA sampling for routine marine biomonitoring, due to the significant decreases in costs when compared to other labour-intensive monitoring programmes involving morphological identification of species (Mynott & Marsh, 2020). The RoCSI device is designed to be easy to deploy and suitable for use in routine monitoring efforts, without the need for specialist technicians. However, clear and accessible resources will still be needed to explain how to use and adjust the device to suit the needs of governmental monitoring programs. Resources for the RoCSI should also clearly state the limitations of eDNA monitoring as opposed to current visual monitoring methods. So as not to falsely present eDNA as a cheaper replacement for visual surveys. Otherwise, key information on the size and life stage of taxa may be lost if visual surveys are replaced entirely by eDNA.

Ecological data is complex and cannot easily be interpreted to inform policy (Golumbeanu et al., 2014). To facilitate the interpretation of eDNA data, ongoing research is focused on developing standardised measures of ecosystem health through genomic biotic indices, which simplify complex ecological networks into a single value. (Aylagas et al., 2014, 2017; Cordier et al., 2017; Pawlowski et al., 2016). This can be done by scoring key indicator species by their sensitivity/tolerance to pollution then measuring occurrence and/or abundance of these species. More recently, supervised machine learning has been used to bypass the identification of taxa, instead identifying non-linear relationships between all ASVs and pollution, using training datasets which cover a known pollution gradient (Cordier et al., 2017; Cordier, Forster, et al., 2018; Cordier, Lanzén, et al., 2018; Frühe et al., 2020; Lanzén et al., 2020). These biotic indices provide a single value which is easily incorporated into legislative descriptions of 'High, Good, Moderate, Poor and Bad' environmental status (Golumbeanu et al., 2014).

While biotic indices are useful tools for the interpretation of ecological data, the use of autonomous devices, as well as machine learning, all contribute to the 'black box' effect. This distances end users from the processes involved in producing results and can lead to a mistrust of the end results. Providing outreach and education about eDNA for stakeholders and the general public is therefore crucial to shed light on eDNA processes. A general understanding of eDNA is needed to help to establish trust for legislative decisions made using eDNA data.

Citizen science projects that use eDNA not only support outreach and education but also generate robust scientific outcomes (Biggs et al., 2015; Deiner, Bik, et al., 2017; Larson et al., 2020; Miralles et al., 2016; Pocock et al., 2018; Schnetzer et al., 2016). Schnetzer et al. (2016) found that citizen scientists involved in OSD could collect high quality data comparable to scientific measurements, plus participants became more engaged in ocean issues showing more environmental awareness and ocean literacy (Schnetzer et al., 2016). Integrating a marine citizen science program with autonomous sampling initiatives would help engage and educate the public about marine biodiversity and eDNA research. Citizen sampling done in parallel with the RoCSI device, using the same filters, preservatives and sampling regime would generate vast quantities marine ecological data at a relatively low cost. Furthermore, citizens would have the opportunity to learn about the eDNA and establish trust for RoCSI derived eDNA data.

1.5 Aims and Objectives

The aim of this PhD thesis is to enhance the usability of the RoCSI through in-situ testing and methodological optimisation, thereby improving its application for marine biodiversity monitoring.

Objectives:

1. Compare RoCSI-collected samples to those obtained through traditional manual collection methods.
2. Assess the effectiveness of using the RoCSI for high spatial resolution of marine biodiversity monitoring.
3. Assess the effectiveness of using RoCSI for high temporal resolution of marine biodiversity monitoring.
4. Compare sample preservation methodologies for their suitability in long-term deployments of the RoCSI.
5. Investigate the potential integration of RoCSI with sample analysis technologies to enable near real-time in-situ monitoring.

1.6 Summary of Work Conducted:

Objectives 1 & 2:

Objectives 1 & 2 were addressed in Chapter 2, where the RoCSI device was configured to autonomously filter and preserve water samples every hour using the ship's underway system during the AE1714 research cruise from Bermuda. This enabled high spatial resolution sampling along the cruise track line (Objective 2). Additionally, manual samples were collected at various comparable time points to enable a direct comparison between autonomously collected RoCSI samples and those obtained through traditional manual collection methods (Objective 1).

Objective 3:

Chapter 3 addressed Objective 3 by deploying the RoCSI dockside in the highly urbanised Solent Estuary. Over a four day period, bi-hourly samples were collected as part of a pilot study designed to test the device configuration prior to its planned deployment at the Western Channel Observatory's (WCO) L4 buoy. However, due to COVID-19 lockdowns, the WCO L4 deployment was not feasible within the PhD timeframe. Although shorter in duration, the pilot study samples provided a higher temporal resolution dataset than originally planned for the WCO L4 deployment. This higher resolution dataset enabled the investigation of diel fluctuations in eDNA signals, which could influence the interpretation of studies with lower sampling frequencies.

Objective 4:

In Chapter 4, four liquid preservatives were compared for their suitability in the long-term (2 - month) preservation of both DNA and RNA. This research was also designed to inform the

planned L4 WCO deployment, where maintenance trips to the buoy can be delayed for up to two months due to adverse sea conditions. The preservatives included a laboratory-prepared Nucleic Acid Preservative, RNAlater® (Inivitrogen), RLT+ buffer (Qiagen), and DNA/RNA Shield (Zymo Research). This methodological comparison, alongside participation in the Ocean Best Practice System's Omics and eDNA task team, informed the perspectives on methodological sharing discussed in Chapter 5.

Objective 5:

Lastly, Objective 5 was addressed in Chapter 6, where the MinION desktop sequencer was tested aboard the RRS Discovery during a research cruise to the Porcupine Abyssal Plain. This work tested the MinION, exploring its suitability for enabling near real-time in-situ monitoring with the RoCSI.

1.7 Conclusion

The UN Ocean Decade presents a pivotal moment for advancing marine biodiversity research. Autonomous samplers, such as the RoCSI device, offer transformative potential by significantly expanding our capacity to collect high-resolution temporal and spatial data. This thesis outlines the steps taken to optimise the RoCSI, increasing its suitability for long-term marine biodiversity monitoring and demonstrating its capacity to generate valuable metadata aligned with emerging biomolecular standards. Integrating RoCSI sampling into a broader network of research projects, including citizen science initiatives, can unlock unprecedented insights into marine ecosystems. Such high-resolution data will be instrumental in informing ecological models and forecasting, ultimately enabling us to better understand and predict the ocean's future state. As the Ocean Decade progresses, continued innovation in autonomous sampling technologies like RoCSI will be crucial for achieving Outcome 3 of the Ocean Decade: “A predicted ocean, enabling society to understand current and future ocean conditions”.

1.8 References

Aylagas, E., Borja, A., & Rodríguez-Ezpeleta, N. (2014). Environmental status assessment using DNA metabarcoding: Towards a genetics based marine biotic index (gAMBI). *PLoS ONE*, *9*(3), 1–11. https://doi.org/10.1371/journal.pone.0090529

Aylagas, E., Borja, Á., Tangherlini, M., Dell’Anno, A., Corinaldesi, C., Michell, C. T., Irigoien, X., Danovaro, R., & Rodríguez-Ezpeleta, N. (2017). A bacterial community-based index to assess the ecological status of estuarine and coastal environments. *Marine Pollution Bulletin*, *114*(2), 679–688. https://doi.org/10.1016/j.marpolbul.2016.10.050

Beaton, A. D., Cardwell, C. L., Thomas, R. S., Sieben, V. J., Legiret, F. E., Waugh, E. M., Statham, P. J., Mowlem, M. C., & Morgan, H. (2012). Lab-on-chip measurement of nitrate and nitrite for in situ analysis of natural waters. *Environmental Science and Technology*, *46*(17), 9548–9556. https://doi.org/10.1021/es300419u

Biggs, J., Ewald, N., Valentini, A., Gaboriaud, C., Dejean, T., Griffiths, R. A., Foster, J., Wilkinson, J. W., Arnell, A., Brotherton, P., Williams, P., & Dunn, F. (2015). Using eDNA to develop a national citizen science-based monitoring programme for the great crested newt (Triturus cristatus). *Biological Conservation*, *183*, 19–28. https://doi.org/10.1016/j.biocon.2014.11.029

Bista, I., Carvalho, G. R., Tang, M., Walsh, K., Zhou, X., Hajibabaei, M., Shokralla, S., Seymour, M., Bradley, D., Liu, S., Christmas, M., & Creer, S. (2018). Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples. *Molecular Ecology Resources*, *18*(5), 1020–1034. https://doi.org/10.1111/1755-0998.12888

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., Yu, D. W., & de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology and Evolution*, *29*(6), 358–367. https://doi.org/10.1016/j.tree.2014.04.003

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology 2019 37:8*, *37*(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

Bombar, D., Taylor, C. D., Wilson, S. T., Robidart, J. C., Rabines, A., Turk-Kubo, K. A., Kemp, J. N., Karl, D. M., & Zehr, J. P. (2015). Measurements of nitrogen fixation in the oligotrophic North Pacific Subtropical Gyre using a free-drifting submersible incubation device. *Journal of Plankton Research*, *37*(4), 727–739. https://doi.org/10.1093/plankt/fbv049

Bowers, H. A., Marin, R., Birch, J. M., & Scholin, C. A. (2017). Sandwich hybridization probes for the detection of Pseudo-nitzschia (Bacillariophyceae) species: An update to existing probes and a description of new probes. *Harmful Algae*, *70*, 37–51. https://doi.org/10.1016/j.hal.2017.10.005

Breier, J. A., Rauch, C. G., McCartney, K., Toner, B. M., Fakra, S. C., White, S. N., & German, C. R. (2009). A suspended-particle rosette multi-sampler for discrete biogeochemical sampling in low-particle-density waters. *Deep-Sea Research Part I: Oceanographic Research Papers*, *56*(9), 1579–1589. https://doi.org/10.1016/j.dsr.2009.04.005

Buttigieg, P. L., Janssen, F., Macklin, J., & Pitz, K. (2019). The Global Omics Observatory Network: Shaping standards for long-term molecular observation. *Biodiversity Information Science and Standards*, *3*. https://doi.org/10.3897/BISS.3.36712

Chavez, F. P., Ryan, J., & Lluch-Cota, S. E. (2003). *From Anchovies to Sardines and Back: Multidecadal Change in the Pacific Ocean*. https://www.science.org

Clarke, G. L. (1940). Comparative Richness of Zooplankton in Coastal and Offshore Areas of the Atlantic. *The Biological Bulletin*, *78*(2), 226–255. https://doi.org/https://doi.org/10.2307/1537775

Clarke, L. J., Beard, J. M., Swadling, K. M., & Deagle, B. E. (2017). Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. *Ecology and Evolution*, *7*(3), 873–883. https://doi.org/10.1002/ece3.2667

Cordier, T., Esling, P., Lejzerowicz, F., Visco, J., Ouadahi, A., Martins, C., Cedhagen, T., & Pawlowski, J. (2017). Predicting the Ecological Quality Status of Marine Environments from eDNA Metabarcoding Data Using Supervised Machine Learning. *Environmental Science and Technology*, *51*(16), 9118–9126. https://doi.org/10.1021/acs.est.7b01518

Cordier, T., Forster, D., Dufresne, Y., Martins, C. I. M., Stoeck, T., & Pawlowski, J. (2018). Supervised machine learning outperforms taxonomy-based environmental DNA metabarcoding applied to biomonitoring. *Molecular Ecology Resources*, *18*(June), 1381–1391. https://doi.org/10.1111/1755-0998.12926

Cordier, T., Lanzén, A., Stoeck, T., Apothéloz-Perret-Gentil, L., & Pawlowski, J. (2018). Embracing Environmental Genomics and Machine Learning for Routine Biomonitoring. *Trends in Microbiology*, 1–11. https://doi.org/10.1016/j.tim.2018.10.012

Cristescu, M. E. (2019). Can Environmental RNA Revolutionize Biodiversity Science? *Trends in Ecology and Evolution*, *34*(8), 694–697. https://doi.org/10.1016/j.tree.2019.05.003

Dayton, P. K., Tegner, M. J., Parnell, P. E., & Edwards, P. B. (1992). Temporal and Spatial Patterns of Disturbance and Recovery in a Kelp Forest Community. *Ecological Monographs*, *62*(3), 421–445. https://doi.org/10.2307/2937118

Decelle, J., Martin, P., Paborstava, K., Pond, D. W., Tarling, G., Mahé, F., de Vargas, C., Lampitt, R., & Not, F. (2013). Diversity, Ecology and Biogeochemistry of Cyst-Forming Acantharia (Radiolaria) in the Oceans. *PLoS ONE*, *8*(1). https://doi.org/10.1371/journal.pone.0053598

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D. M., de Vere, N., Pfrender, M. E., & Bernatchez, L. (2017). Environmental DNA

metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, *26*(21), 5872–5895. https://doi.org/10.1111/mec.14350

Deiner, K., Renshaw, M. A., Li, Y., Olds, B. P., Lodge, D. M., & Pfrender, M. E. (2017). Long-range PCR allows sequencing of mitochondrial genomes from environmental DNA. *Methods in Ecology and Evolution*, *8*(12), 1888–1898. https://doi.org/10.1111/2041-210X.12836

Deiner, K., Walser, J. C., Mächler, E., & Altermatt, F. (2015). Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biological Conservation*, *183*, 53–63. https://doi.org/10.1016/j.biocon.2014.11.018

Dejean, T., Valentini, A., Duparc, A., Pellier-Cuit, S., Pompanon, F., Taberlet, P., & Miaud, C. (2011). Persistence of environmental DNA in freshwater ecosystems. *PLoS ONE*, *6*(8), 8–11. https://doi.org/10.1371/journal.pone.0023398

Djurhuus, A., Closek, C. J., Kelly, R. P., Pitz, K. J., Michisaki, R. P., Starks, H. A., Walz, K. R., Andruszkiewicz, E. A., Olesin, E., Hubbard, K., Montes, E., Otis, D., Muller-Karger, F. E., Chavez, F. P., Boehm, A. B., & Breitbart, M. (2020). Environmental DNA reveals seasonal shifts and potential interactions in a marine community. *Nature Communications*, *11*(1), 1–9. https://doi.org/10.1038/s41467-019-14105-1

Djurhuus, A., Pitz, K., Sawaya, N. A., Rojas-M?arquez, J., Michaud, B., Montes, E., Muller-Karger, F., & Breitbart, M. (2018). Evaluation of marine zooplankton community structure through environmental DNA metabarcoding Anni. *Limnology and Oceanography: Method*, *16*, 209–221. https://doi.org/10.1002/lom3.10237

Doherty, M., Tamura, M., Vriezen, J. A. C., McManus, G. B., & Katz, L. A. (2010). Diversity of oligotrichia and choreotrichia ciliates in coastal marine sediments and in overlying plankton. *Applied and Environmental Microbiology*, *76*(12), 3924–3935. https://doi.org/10.1128/AEM.01604-09

Edgcomb, V. P., Taylor, C., Pachiadaki, M. G., Honjo, S., Engstrom, I., & Yakimov, M. (2016). Comparison of Niskin vs. in situ approaches for analysis of gene expression in deep Mediterranean Sea water samples. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *129*, 213–222. https://doi.org/10.1016/j.dsr2.2014.10.020

Erdner, D. L., Percy, L., Keafer, B., Lewis, J., & Anderson, D. M. (2010). A quantitative real-time PCR assay for the identification and enumeration of Alexandrium cysts in marine sediments. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *57*(3–4), 279–287. https://doi.org/10.1016/j.dsr2.2009.09.006

Farnelid, H., Turk-Kubo, K., del Carmen Muñoz-Marín, M., & Zehr, J. P. (2016). New insights into the ecology of the globally significant uncultured nitrogen-fixing symbiont UCYN-A. *Aquatic Microbial Ecology*, *77*(3), 128–138. https://doi.org/10.3354/ame01794

Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using environmental DNA from water samples. *Biology Letters*, *4*(4), 423–425. https://doi.org/10.1098/rsbl.2008.0118

Foster, Z. S. L., Sharpton, T. J., & Grünwald, N. J. (2017). Metacoder: An R package for visualization and manipulation of community taxonomic diversity data. *PLoS Computational Biology*, *13*(2), 1–15. https://doi.org/10.1371/journal.pcbi.1005404

Frühe, L., Cordier, T., Dully, V., Breiner, H. W., Lentendu, G., Pawlowski, J., Martins, C., Wilding, T. A., & Stoeck, T. (2020). Supervised machine learning is superior to indicator value inference in monitoring the environmental impacts of salmon aquaculture using eDNA metabarcodes. *Molecular Ecology*, *October 2019*, 1–19. https://doi.org/10.1111/mec.15434

Galand, P. E., Casamayor, E. O., Kirchman, D. L., & Lovejoy, C. (2009). Ecology of the rare microbial biosphere of the Arctic Ocean. In *PNAS* (Vol. 106, pp. 22427–22432).

Gennip, S. van, Martin, A. P., Srokosz, M. A., Allen, J. T., Pidcock, R., Painter, S. C., Stinchcombe, M. C., & 1National. (2016). Plankton patchiness investigated using simultaneous nitrate and chlorophyll observations. *Journal of Geophysical Research: Oceans*, 4149–4156. https://doi.org/10.1002/2016JC011789.Received

Georges, A. A., El-Swais, H., Craig, S. E., Li, W. K. W., & Walsh, D. A. (2014). Metaproteomic analysis of a winter to spring succession in coastal northwest Atlantic Ocean microbial plankton. *ISME Journal*, *8*(6), 1301–1313. https://doi.org/10.1038/ismej.2013.234

Giovannoni, S. J., Britschgi, T. B., Moyer, C. L., & Field, K. G. (1990). Genetic diversity in Sargasso Sea bacterioplankton. *Nature*, *345*(6270), 60–63. https://doi.org/10.1038/345060a0

Gobler, C. J., Berry, D. L., Dyhrman, S. T., Wilhelm, S. W., Salamov, A., Lobanov, A. v., Zhang, Y., Collier, J. L., Wurch, L. L., Kustka, A. B., Dill, B. D., Shah, M., VerBerkmoes, N. C., Kuo, A., Terry, A., Pangilinan, J., Lindquist, E. A., Lucas, S., Paulsen, I. T., … Grigoriev, I. v. (2011). Niche of harmful alga Aureococcus anophagefferens revealed through ecogenomics. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(11), 4352–4357. https://doi.org/10.1073/PNAS.1016106108/SUPPL_FILE/SAPP.PDF

Golumbeanu, M., Oros, A., Nenciu, M., Zavatarelli, M., & Drago, A. (2014). Contribution of environmental indices in meeting the objectives and principles of the marine strategy

framework directive (MSFD). *Journal of Environmental Protection and Ecology*, *15*(3), 1130–1138.

Goodwin, K. D., Muller-Karger, F. E., Djurhuus, A., Allen, L. Z., Allen, A. E., McCrow, J. P., & Canonico Hyde, G. (2018). Molecular approaches for an operational marine biodiversity observation network. In *World Seas: An Environmental Evaluation Volume III: Ecological Issues and Environmental Impacts* (Second Edi). Elsevier Ltd. https://doi.org/10.1016/B978-0-12-805052-1.00032-2

Gray, M. A., Pratte, Z. A., & Kellogg, C. A. (2013). Comparison of DNA preservation methods for environmental bacterial community samples. *FEMS Microbiology Ecology*, *83*(2), 468–477. https://doi.org/10.1111/1574-6941.12008

Guardiola, M., Wangensteen, O. S., Taberlet, P., Coissac, E., Uriz, M. J., & Turon, X. (2016). Spatio-temporal monitoring of deep-sea communities using metabarcoding of sediment DNA and RNA. *PeerJ*, *2016*(12), 1–31. https://doi.org/10.7717/peerj.2807

Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E. M., & Troedsson, C. (2014). Characterization of the 18S rRNA Gene for Designing Universal Eukaryote Specific Primers. *PLOS ONE*, *9*(2), e87624. https://doi.org/10.1371/JOURNAL.PONE.0087624

Hagino, K., Onuma, R., Kawachi, M., & Horiguchi, T. (2013). Discovery of an endosymbiotic nitrogen-fixing cyanobacterium UCYN-A in Braarudosphaera bigelowii (Prymnesiophyceae). *PLoS ONE*, *8*(12). https://doi.org/10.1371/journal.pone.0081749

Harper, K. J., Goodwin, K. D., Harper, L. R., Lacasella, E. L., Frey, A., Dutton, P. H., Mayne, B., Simpson, T., & Harper, K. J. (2020). *Finding Crush : Environmental DNA Analysis as a Tool for Tracking the Green Sea Turtle Chelonia mydas in a Marine Estuary*. *6*(January), 1–13. https://doi.org/10.3389/fmars.2019.00810

Harrison, J. G., John Calder, W., Shuman, B., & Alex Buerkle, C. (2021). The quest for absolute abundance: The use of internal standards for DNA-based community ecology. *Molecular Ecology Resources*, *21*(1), 30–43. https://doi.org/10.1111/1755-0998.13247

Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1512), 313–321. https://doi.org/10.1098/rspb.2002.2218

Hebert, P. D. N., Ratnasingham, S., & DeWaard, J. R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B: Biological Sciences*, *270*(Suppl 1), S96. https://doi.org/10.1098/RSBL.2003.0025

Hernández-Becerril, D. U. (1995). Planktonic diatoms from the gulf of california and coasts off baja california: The genera rhizosolenia, proboscia, pseudosolenia, and former rhizosolenia species. *Diatom Research*, *10*(2), 251–267. https://doi.org/10.1080/0269249x.1995.9705348

Heymans, J. J., Bundy, A., Christensen, V., Coll, M., de Mutsert, K., Fulton, E. A., Piroddi, C., Shin, Y. J., Steenbeek, J., & Travers-Trolet, M. (2020). The Ocean Decade: A True Ecosystem Modeling Challenge. *Frontiers in Marine Science*, *7*(September 2020), 1–5. https://doi.org/10.3389/fmars.2020.554573

Holman, L. E., de Bruyn, M., Creer, S., Carvalho, G., Robidart, J., & Rius, M. (2019). Detection of introduced and resident marine species using environmental DNA metabarcoding of sediment and water. *Scientific Reports*, *9*(1), 1–10. https://doi.org/10.1038/s41598-019-47899-7

Holman, L. E., de Bruyn, M., Creer, S., Carvalho, G., Robidart, J., & Rius, M. (2021). Animals, protists and bacteria share marine biogeographic patterns. *Nature Ecology and Evolution*. https://doi.org/10.1038/s41559-021-01439-7

Hörstmann, C., Buttigieg, P. L., Simpson, P., Pearlman, J., & Waite, A. M. (2020). *A Best Practice for Developing Best Practices in Ocean Observation (BP4BP): Supporting Methodological Evolution through Actionable Documentation*. *April*. https://doi.org/10.25607/OBP-781

Jerde, C. L., Mahon, A. R., Chadderton, W. L., & Lodge, D. M. (2011). "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conservation Letters*, *4*(2), 150–157. https://doi.org/10.1111/j.1755-263X.2010.00158.x

Jo, J., Oh, J., & Park, C. (2020). Microbial community analysis using high-throughput sequencing technology: a beginner's guide for microbiologists. *Journal of Microbiology*, *58*(3), 176–192. https://doi.org/10.1007/s12275-020-9525-5

Kavanaugh, M. T., Hales, B., Saraceno, M., Spitz, Y. H., White, A. E., & Letelier, R. M. (2014). Hierarchical and dynamic seascapes: A quantitative framework for scaling pelagic biogeochemistry and ecology. *Progress in Oceanography*, *120*, 291–304. https://doi.org/10.1016/j.pocean.2013.10.013

Keeley, N., Wood, S. A., & Pochon, X. (2018). Development and preliminary validation of a multi-trophic metabarcoding biotic index for monitoring benthic organic enrichment. *Ecological Indicators*, *85*(January 2017), 1044–1057. https://doi.org/10.1016/j.ecolind.2017.11.014

Kelly, R. P., O'Donnell, J. L., Lowell, N. C., Shelton, A. O., Samhouri, J. F., Hennessey, S. M., Feist, B. E., & Williams, G. D. (2016). Genetic signatures of ecological diversity along an urbanization gradient. *PeerJ*, *4*, e2444. https://doi.org/10.7717/peerj.2444

Kitahashi, T., Sugime, S., Inomata, K., Nishijima, M., Kato, S., & Yamamoto, H. (2020). Meiofaunal diversity at a seamount in the pacific ocean: A comprehensive study using environmental DNA and RNA. *Deep-Sea Research Part I: Oceanographic Research Papers*, *160*(February), 103253. https://doi.org/10.1016/j.dsr.2020.103253

Knudsen, S. W., Ebert, R. B., Hesselsøe, M., Kuntke, F., Hassingboe, J., Mortensen, P. B., Thomsen, P. F., Sigsgaard, E. E., Hansen, B. K., Nielsen, E. E., & Møller, P. R. (2019). Species-specific detection and quantification of environmental DNA from marine fishes in the Baltic Sea. *Journal of Experimental Marine Biology and Ecology*, *510*(October 2018), 31–45. https://doi.org/10.1016/j.jembe.2018.09.004

Lanzén, A., Mendibil, I., Borja, Á., & Alonso-Sáez, L. (2020). A microbial mandala for environmental monitoring: Predicting multiple impacts on estuarine prokaryote communities of the Bay of Biscay . *Molecular Ecology*, *May*, 1–19. https://doi.org/10.1111/mec.15489

Laroche, O., Wood, S. A., Tremblay, L. A., Lear, G., Ellis, J. I., & Pochon, X. (2017). Metabarcoding monitoring analysis: the pros and cons of using co-extracted environmental DNA and RNA data to assess offshore oil production impacts on benthic communities. *PeerJ*, *5*, e3347. https://doi.org/10.7717/peerj.3347

Larson, E. R., Graham, B. M., Achury, R., Coon, J. J., Daniels, M. K., Gambrell, D. K., Jonasen, K. L., King, G. D., LaRacuente, N., Perrin-Stowe, T. I. N., Reed, E. M., Rice, C. J., Ruzi, S. A., Thairu, M. W., Wilson, J. C., & Suarez, A. v. (2020). From eDNA to citizen science: emerging tools for the early detection of invasive species. *Frontiers in Ecology and the Environment*, *18*(4), 194–202. https://doi.org/10.1002/fee.2162

Liu, J., & Zhang, H. (2021). Combining Multiple Markers in Environmental DNA Metabarcoding to Assess Deep-Sea Benthic Biodiversity. *Frontiers in Marine Science*, *8*. https://doi.org/10.3389/fmars.2021.684955

Marin, B., & Melkonian, M. (2010). Molecular Phylogeny and Classification of the Mamiellophyceae class. nov. (Chlorophyta) based on Sequence Comparisons of the Nuclear- and Plastid-encoded rRNA Operons. *Protist*, *161*(2), 304–336. https://doi.org/10.1016/j.protis.2009.10.002

McCarthy, A., Chiang, E., Schmidt, M. L., & Denef, V. J. (2015). RNA preservation agents and nucleic acid extraction method bias perceived bacterial community composition. *PLoS ONE*, *10*(3), 1–14. https://doi.org/10.1371/journal.pone.0121659

McGrath, K. C., Thomas-Hall, S. R., Cheng, C. T., Leo, L., Alexa, A., Schmidt, S., & Schenk, P. M. (2008). Isolation and analysis of mRNA from environmental microbial communities. *Journal of Microbiological Methods*, *75*(2), 172–176. https://doi.org/10.1016/j.mimet.2008.05.019

McKinney, E. S. A., Gibson, C. E., & Stewart, B. M. (1997). Planktonic diatoms in the north-west Irish sea: A study by automatic sampler. *Biology and Environment*, *97*(3), 197–202.

Miloslavich, P., Bax, N. J., Simmons, S. E., Klein, E., Appeltans, W., Aburto-Oropeza, O., Andersen Garcia, M., Batten, S. D., Benedetti-Cecchi, L., Checkley, D. M., Chiba, S., Duffy, J. E., Dunn, D. C., Fischer, A., Gunn, J., Kudela, R., Marsac, F., Muller-Karger, F. E., Obura, D., & Shin, Y. J. (2018). Essential ocean variables for global sustained observations of biodiversity and ecosystem changes. *Global Change Biology*, *24*(6), 2416–2433. https://doi.org/10.1111/gcb.14108

Miralles, L., Dopico, E., Devlo-Delva, F., & Garcia-Vazquez, E. (2016). Controlling populations of invasive pygmy mussel (Xenostrobus securis) through citizen science and environmental DNA. *Marine Pollution Bulletin*, *110*(1), 127–132. https://doi.org/10.1016/j.marpolbul.2016.06.072

Mitchell, A. L., Almeida, A., Beracochea, M., Boland, M., Burgin, J., Cochrane, G., Crusoe, M. R., Kale, V., Potter, S. C., Richardson, L. J., Sakharova, E., Scheremetjew, M., Korobeynikov, A., Shlemov, A., Kunyavskaya, O., Lapidus, A., & Finn, R. D. (2020). MGnify: The microbiome analysis resource in 2020. *Nucleic Acids Research*, *48*(D1), D570–D578. https://doi.org/10.1093/nar/gkz1035

Moisander, P. H., Beinart, R. A., Hewson, I., White, A. E., Johnson, K. S., Carlson, C. A., Montoya, J. P., & Zehr, J. P. (2010). Unicellular cyanobacterial distributions broaden the oceanic N2 fixation domain. *Science (New York, N.Y.)*, *327*(5972), 1512–1514. https://doi.org/10.1126/SCIENCE.1185468

Montoya, J. P., Holl, C. M., Zehr, J. P., Hansen, A., Villareal, T. A., & Capone, D. G. (2004). High rates of N2 fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean. *Nature*, *430*(7003), 1027–1031. https://doi.org/10.1038/nature02824

Moran, M. A. (2009). Metatranscriptomics: Eavesdropping on complex microbial communities. *Microbe*, *4*(7), 329–335. https://doi.org/10.1128/microbe.4.329.1

Morris, R. M., Longnecker, K., & Giovannoni, S. J. (2006). Pirellula and OM43 are among the dominant lineages identified in an Oregon coast diatom bloom. *Environmental Microbiology*, *8*(8), 1361–1370. https://doi.org/10.1111/j.1462-2920.2006.01029.x

Mulholland, M. R. (2007). The fate of nitrogen fixed by diazotrophs in the ocean. *Biogeosciences*, *4*, 37–51. www.biogeosciences.net/4/37/2007/

Muller-Karger, F. E., Miloslavich, P., Bax, N. J., Simmons, S., Costello, M. J., Pinto, I. S., Canonico, G., Turner, W., Gill, M., Montes, E., Best, B. D., Pearlman, J., Halpin, P., Dunn, D., Benson, A.,

Martin, C. S., Weatherdon, L. v., Appeltans, W., Provoost, P., … Geller, G. (2018). Advancing marine biological observations and data requirements of the complementary Essential Ocean Variables (EOVs) and Essential Biodiversity Variables (EBVs) frameworks. *Frontiers in Marine Science*, *5*(JUN), 1–15. https://doi.org/10.3389/fmars.2018.00211

Mynott, S., & Marsh, M. (2020). *Development of a Novel (DNA-based) Method for Monitoring Inshore Fish Communities Using a Programmable Large-Volume Marine eDNA Sampler* (Issue NERC330).

Notomi, T., Mori, Y., Tomita, N., & Kanda, H. (2015). Loop-mediated isothermal amplification (LAMP): principle, features, and future prospects. *Journal of Microbiology 2015 53:1*, *53*(1), 1–5. https://doi.org/10.1007/S12275-015-4656-9

Ogram, A., Sayler, G. S., & Barkay, T. (1987). The extraction and purification of microbial DNA from sediments. *Journal of Microbiological Methods*, *7*(2–3), 57–66. https://doi.org/10.1016/0167-7012(87)90025-X

Olejarz, J., Iwasa, Y., Knoll, A. H., & Nowak, M. A. (2021). The Great Oxygenation Event as a consequence of ecological dynamics modulated by planetary change. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-23286-7

Pawlowski, J., Apothéloz-Perret-Gentil, L., & Altermatt, F. (2020). Environmental DNA: What's behind the term? Clarifying the terminology and recommendations for its future use in biomonitoring. *Molecular Ecology*, *August*, 4258–4264. https://doi.org/10.1111/mec.15643

Pawlowski, J., Lejzerowicz, F., Apotheloz-Perret-Gentil, L., Visco, J., & Esling, P. (2016). Protist metabarcoding and environmental biomonitoring: Time for change. *European Journal of Protistology*, *55*, 12–25. https://doi.org/10.1016/j.ejop.2016.02.003

Pearlman, J., Bushnell, M., Coppola, L., Karstensen, J., Buttigieg, P. L., Pearlman, F., Simpson, P., Barbier, M., Muller-Karger, F. E., Munoz-Mas, C., Pissierssens, P., Chandler, C., Hermes, J., Heslop, E., Jenkyns, R., Achterberg, E. P., Bensi, M., Bittig, H. C., Blandin, J., … Whoriskey, F. (2019). Evolving and sustaining ocean best practices and standards for the next decade. *Frontiers in Marine Science*, *6*(JUN), 1–19. https://doi.org/10.3389/fmars.2019.00277

Pecl, G. T., Araújo, M. B., Bell, J. D., Blanchard, J., Bonebrake, T. C., Chen, I. C., Clark, T. D., Colwell, R. K., Danielsen, F., Evengård, B., Falconi, L., Ferrier, S., Frusher, S., Garcia, R. A., Griffis, R. B., Hobday, A. J., Janion-Scheepers, C., Jarzyna, M. A., Jennings, S., … Williams, S. E. (2017). Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. In *Science* (Vol. 355, Issue 6332). American Association for the Advancement of Science. https://doi.org/10.1126/science.aai9214

Petrosino, J. F., Highlander, S., Luna, R. A., Gibbs, R. A., & Versalovic, J. (2009). Metagenomic pyrosequencing and microbial identification. *Clinical Chemistry*, *55*(5), 856–866. https://doi.org/10.1373/clinchem.2008.107565

Pochon, X., Zaiko, A., Fletcher, L. M., Laroche, O., & Wood, S. A. (2017). Wanted dead or alive? Using metabarcoding of environmental DNA and RNA to distinguish living assemblages for biosecurity applications. *PLoS ONE*, *12*(11), 1–19. https://doi.org/10.1371/journal.pone.0187636

Pocock, M. J. O., Chandler, M., Bonney, R., Thornhill, I., Albin, A., August, T., Bachman, S., Brown, P. M. J., Cunha, D. G. F., Grez, A., Jackson, C., Peters, M., Rabarijaon, N. R., Roy, H. E., Zaviezo, T., & Danielsen, F. (2018). A Vision for Global Biodiversity Monitoring With Citizen Science. In *Advances in Ecological Research* (1st ed., Vol. 59, Issue January). Elsevier Ltd. https://doi.org/10.1016/bs.aecr.2018.06.003

Prasad, A. K. S. K., Nienow, J. A., & Hargraves, P. (2011). Plicate Species of the Diatom Genus Thalassiosira (Bacillariophyta) from the Atlantic and Gulf Coasts of Southeastern United States, with the Description of T. livingstoniorum sp. nov. . *Proceedings of the Academy of Natural Sciences of Philadelphia*, *161*(1), 1–34. https://doi.org/10.1635/053.161.0101

Preston, C. M., Marin, R., Jensen, S. D., Feldman, J., Birch, J. M., Massion, E. I., Delong, E. F., Suzuki, M., Wheeler, K., & Scholin, C. A. (2009). Near real-time, autonomous detection of marine bacterioplankton on a coastal mooring in Monterey Bay, California, using rRNA-targeted DNA probes. *Environmental Microbiology*, *11*(5), 1168–1180. https://doi.org/10.1111/j.1462-2920.2009.01848.x

Rérolle, V. M. C., Achterberg, E. P., Ribas-Ribas, M., Kitidis, V., Brown, I., Bakker, D. C. E., Lee, G. A., & Mowlem, M. C. (2018). High resolution ph measurements using a lab-on-chip sensor in surface waters of northwest European shelf seas. *Sensors (Switzerland)*, *18*(8), 1–23. https://doi.org/10.3390/s18082622

Robidart, J. C., Church, M. J., Ryan, J. P., Ascani, F., Wilson, S. T., Bombar, D., Marin, R., Richards, K. J., Karl, D. M., Scholin, C. A., & Zehr, J. P. (2014). Ecogenomic sensor reveals controls on N2-fixing microorganisms in the North Pacific Ocean. *ISME Journal*, *8*(6), 1175–1185. https://doi.org/10.1038/ismej.2013.244

Robidart, J. C., Magasin, J. D., Shilova, I. N., Turk-Kubo, K. A., Wilson, S. T., Karl, D. M., Scholin, C. A., & Zehr, J. P. (2019). Effects of nutrient enrichment on surface microbial community gene expression in the oligotrophic North Pacific Subtropical Gyre. *ISME Journal*, *13*(2), 374–387. https://doi.org/10.1038/s41396-018-0280-0

Robidart, J. C., Preston, C. M., Paerl, R. W., Turk, K. A., Mosier, A. C., Francis, C. A., Scholin, C. A., & Zehr, J. P. (2012). Seasonal Synechococcus and Thaumarchaeal population dynamics examined with high resolution with remote in situ instrumentation. *ISME Journal*, *6*(3), 513–523. https://doi.org/10.1038/ismej.2011.127

Robidart, J. C., Shilova, I. N., Zehr, J. P., Tiquia-Arashiro, S. M., Robidart, J. C., Shilova, I. N., & Zehr, J. P. (2012). *Molecular Biological Technologies for Ocean Sensing*. Springer. https://doi.org/10.1007/978-1-61779-915-0

Roemmich, D., Alford, M. H., Claustre, H., Johnson, K. S., King, B., Moum, J., Oke, P. R., Owens, W. B., Pouliquen, S., Purkey, S., Scanderbeg, M., Suga, T., Wijffels, S. E., Zilberman, N., Bakker, D., Baringer, M. O., Belbeoch, M., Bittig, H. C., Boss, E., … Yasuda, I. (2019). On the future of Argo: A global, full-depth, multi-disciplinary array. *Frontiers in Marine Science*, *6*(JUL), 1–28. https://doi.org/10.3389/fmars.2019.00439

Roh, S. W., Abell, G. C. J., Kim, K. H., Nam, Y. do, & Bae, J. W. (2010). Comparing microarrays and next-generation sequencing technologies for microbial ecology research. *Trends in Biotechnology*, *28*(6), 291–299. https://doi.org/10.1016/j.tibtech.2010.03.001

Ryabinin, V., Barbière, J., Haugan, P., Kullenberg, G., Smith, N., McLean, C., Troisi, A., Fischer, A. S., Aricò, S., Aarup, T., Pissierssens, P., Visbeck, M., Enevoldsen, H., & Rigaud, J. (2019). The UN decade of ocean science for sustainable development. *Frontiers in Marine Science*, *6*(JUL). https://doi.org/10.3389/fmars.2019.00470

Schnetzer, J., Kopf, A., Bietz, M. J., Buttigieg, P. L., Fernandez-Guerra, A., Ristov, A. P., Glöckner, F. O., Schnetzer, J., Kopf, A., Bietz, M. J., Kottmann, R., & Buttigieg, P. L. (2016). MyOSD 2014: Evaluating Oceanographic Measurements Contributed by Citizen Scientists in Support of Ocean Sampling Day †. *Journal of Microbiology & Biology Education*, *17*(1), 163–171. https://doi.org/10.1128/jmbe.v17i1.1001

Scholin, C., Everlove, C., Harris, A., Alvarado, N., Birch, J., Greenfield, D., Vrijenhoek, R., Mikulski, C., Jones, K., Doucette, G., Jensen, S., Roman, B., Pargett, D., Marin, R. I., Preston, C., Jones, W., & Feldman, J. (2009). Remote detection of marine microbes, small invertebrates, harmful algae, and biotoxins using the Environmental Sample Processor (ESP). *Oceanography*, *22*(2), 158–167. https://doi.org/10.5670/oceanog.2011.65

Scholin, C., Jensen, S., Roman, B., Massion, E., Marin, R., Preston, C., Greenfield, D., Jones, W., & Wheeler, K. (2006). The Environmental Sample Processor (ESP) - An autonomous robotic device for detecting microorganisms remotely using molecular probe technology. *Oceans 2006*, *April*, 3–6. https://doi.org/10.1109/OCEANS.2006.306885

Sepulveda, A. J., Hutchins, P. R., Forstchen, M., Mckeefry, M. N., & Swigris, A. M. (2020). The Elephant in the Lab (and Field): Contamination in Aquatic Environmental DNA Studies. *Frontiers in Ecology and Evolution*, *8*, 440. https://doi.org/10.3389/FEVO.2020.609973/BIBTEX

Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., & Waterston, R. H. (2017). DNA sequencing at 40: Past, present and future. *Nature*, *550*(7676). https://doi.org/10.1038/nature24286

Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology*, *26*(10), 1135–1145. https://doi.org/10.1038/nbt1486

Siegel, D. A., Behrenfeld, M. J., Maritorena, S., McClain, C. R., Antoine, D., Bailey, S. W., Bontempi, P. S., Boss, E. S., Dierssen, H. M., Doney, S. C., Eplee, R. E., Evans, R. H., Feldman, G. C., Fields, E., Franz, B. A., Kuring, N. A., Mengelt, C., Nelson, N. B., Patt, F. S., … Yoder, J. A. (2013). Regional to global assessments of phytoplankton dynamics from the SeaWiFS mission. *Remote Sensing of Environment*, *135*, 77–91. https://doi.org/10.1016/j.rse.2013.03.025

Stefanni, S., Stanković, D., Borme, D., de Olazabal, A., Juretić, T., Pallavicini, A., & Tirelli, V. (2018). Multi-marker metabarcoding approach to study mesozooplankton at basin scale. *Scientific Reports*, *8*(1), 1–13. https://doi.org/10.1038/s41598-018-30157-7

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D. R., Alberti, A., Cornejo-castillo, F. M., Costea, P. I., Cruaud, C., Ovidio, F., Engelen, S., Ferrera, I., Gasol, J. M., Guidi, L., Hildebrand, F., … Lepoivre, C. (2015). Structure and function of the global ocean microbiome. *Science*, *348*(6237), 1–10. https://doi.org/10.1126/science.1261359

Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. *Molecular Ecology*, *21*(8), 1789–1793. https://doi.org/10.1111/j.1365-294X.2012.05542.x

Takahashi, S., Tomita, J., Nishioka, K., Hisada, T., & Nishijima, M. (2014). Development of a prokaryotic universal primer for simultaneous analysis of Bacteria and Archaea using next-generation sequencing. *PLoS ONE*, *9*(8). https://doi.org/10.1371/journal.pone.0105592

Tang, W., Cerdán, E., Berthelot, H., Polyviou, D., Wang, S., Baylay, H., Whitby, H., Planquette, H., Mowlem, M., & Robidart, J. C. (2020). New insights into the distributions of nitrogen fixation and diazotrophs revealed by high-resolution sensing and sampling methods. *ISME Journal*. https://doi.org/10.1038/s41396-020-0703-6

Taylor, C. D., & Doherty, K. W. (1990). Submersible Incubation Device (SID), autonomous instrumentation for the in situ measurement of primary production and other microbial rate processes. *INSTRUMENTS AND METHODS*, *37*(2), 343–358.

Taylor, C. D., Doherty, K. W., Molyneaux, S. J., Morrison, A. T., Billings, J. D., Engstrom, I. B., Pfitsch, D. W., & Honjo, S. (2006). Autonomous Microbial Sampler (AMS), a device for the uncontaminated collection of multiple microbial samples from submarine hydrothermal vents and other aquatic environments. *Deep-Sea Research Part I: Oceanographic Research Papers*, *53*(5), 894–916. https://doi.org/10.1016/j.dsr.2006.01.009

Tessler, M., Neumann, J. S., Afshinnekoo, E., Pineda, M., Hersch, R., Velho, L. F. M., Segovia, B. T., Lansac-Toha, F. A., Lemke, M., Desalle, R., Mason, C. E., & Brugler, M. R. (2017). Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. *Scientific Reports*, *7*(1), 1–14. https://doi.org/10.1038/s41598-017-06665-3

Thompson, A. W., Foster, R. A., Krupke, A., Carter, B. J., Musat, N., Vaulot, D., Kuypers, M. M. M., & Zehr, J. P. (2012). Unicellular cyanobacterium symbiotic with a single-celled eukaryotic alga. *Science*, *337*(6101), 1546–1550. https://doi.org/10.1126/SCIENCE.1222700/SUPPL_FILE/THOMPSON.SM.PDF

Thomsen, P. F., Kielgast, J., Iversen, L. L., Møller, P. R., Rasmussen, M., & Willerslev, E. (2012). Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PloS One*, *7*(8), e41732. https://doi.org/10.1371/journal.pone.0041732

Tragin, M., & Vaulot, D. (2018). Green microalgae in marine coastal waters: The Ocean Sampling Day (OSD) dataset. *Scientific Reports*, *8*(1), 1–12. https://doi.org/10.1038/s41598-018-32338-w

Valentini, A., Pompanon, F., & Taberlet, P. (2009). DNA barcoding for ecologists. *Trends in Ecology and Evolution*, *24*(2), 110–117. https://doi.org/10.1016/j.tree.2008.09.011

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., … Smith, H. O. (2019). Sequencing of the Sargasso Sea. *Science*, *304*(5667), 66–75. https://doi.org/10.1126/science.1093857

Wang, S., Tang, W., Delage, E., Gifford, S., Whitby, H., González, A. G., Eveillard, D., Planquette, H., & Cassar, N. (2021). Investigating the microbial ecology of coastal hotspots of marine nitrogen fixation in the western North Atlantic. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-84969-1

Ward, D. M., Weller, R., & Bateson, M. M. (1990). 16S rRNA sequences reveal numerous uncultured inhabitants in a well-studied natural community. *Nature (London)*, *345*(6270), 63–65. http://dx.doi.org/10.1038/345063a0

Whitt, C., Pearlman, J., Polagye, B., Caimi, F., Muller-Karger, F., Copping, A., Spence, H., Madhusudhana, S., Kirkwood, W., Grosjean, L., Fiaz, B. M., Singh, S., Singh, S., Manalang, D., Gupta, A. sen, Maguer, A., Buck, J. J. H., Marouchos, A., Atmanand, M. A., … Khalsa, S. J. (2020). Future Vision for Autonomous Ocean Observations. *Frontiers in Marine Science*, *7*(September). https://doi.org/10.3389/fmars.2020.00697

Willerslev, E., Hansen, A. J., Binladen, J., Brand, T. B., Gilbert, M. T. P., Shapiro, B., Bunce, M., Wiuf, C., Gilichinsky, D. A., & Cooper, A. (2003). Diverse plant and animal genetic records from holocene and pleistocene sediments. *Science*, *300*(5620), 791–795. https://doi.org/10.1126/science.1084114

Williams, M. A., Hernandez, C., O'Sullivan, A. M., April, J., Regan, F., Bernatchez, L., & Parle-McDermott, A. (2021). Comparing CRISPR-Cas and qPCR eDNA assays for the detection of Atlantic salmon (Salmo salar L.). *Environmental DNA*, *3*(1), 297–304. https://doi.org/10.1002/EDN3.174

Wood, S. A., Biessy, L., Latchford, J. L., Zaiko, A., von Ammon, U., Audrezet, F., Cristescu, M. E., & Pochon, X. (2020). Release and degradation of environmental DNA and RNA in a marine system. *Science of the Total Environment*, *704*, 135314. https://doi.org/10.1016/j.scitotenv.2019.135314

Wurzbacher, C., Salka, I., & Grossart, H. P. (2012). Environmental actinorhodopsin expression revealed by a new in situ filtration and fixation sampler. *Environmental Microbiology Reports*, *4*(5), 491–497. https://doi.org/10.1111/j.1758-2229.2012.00350.x

Xu, G., Dong, C., Liu, Y., Gaube, P., & Yang, J. (2019). Chlorophyll Rings around Ocean Eddies in the North Pacific. *Scientific Reports*, *9*(1), 1–8. https://doi.org/10.1038/s41598-018-38457-8

Yamahara, K. M., Preston, C. M., Birch, J., Walz, K., Marin, R., Jensen, S., Pargett, D., Roman, B., Ussler, W., Zhang, Y., Ryan, J., Hobson, B., Kieft, B., Raanan, B., Goodwin, K. D., Chavez, F. P., & Scholin, C. (2019). In situ Autonomous Acquisition and Preservation of Marine Environmental DNA Using an Autonomous Underwater Vehicle. *Frontiers in Marine Science*, *6*(July), 1–14. https://doi.org/10.3389/fmars.2019.00373

Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J. R., Amaral-Zettler, L., Gilbert, J. A., Karsch-Mizrachi, I., Johnston, A., Cochrane, G., Vaughan, R., Hunter, C., Park, J., Morrison, N., Rocca-Serra, P., Sterk, P., Arumugam, M., Bailey, M., Baumgartner, L., … Glöckner, F. O. (2011). Minimum information about a marker gene sequence (MIMARKS) and minimum information

about any (x) sequence (MIxS) specifications. *Nature Biotechnology*, *29*(5), 415–420. https://doi.org/10.1038/nbt.1823

Zaiko, A., Greenfield, P., Abbott, C., von Ammon, U., Bilewitch, J., Bunce, M., Cristescu, M. E., Chariton, A., Dowle, E., Geller, J., Ardura Gutierrez, A., Hajibabaei, M., Haggard, E., Inglis, G. J., Lavery, S. D., Samuiloviene, A., Simpson, T., Stat, M., Stephenson, S., … Pochon, X. (2022). Towards reproducible metabarcoding data: Lessons from an international cross-laboratory experiment. *Molecular Ecology Resources*, *22*(2), 519–538. https://doi.org/10.1111/1755-0998.13485

Zaiko, A., Pochon, X., Garcia-Vazquez, E., Olenin, S., & Wood, S. A. (2018). Advantages and Limitations of Environmental DNA/RNA Tools for Marine Biosecurity: Management and Surveillance of Non-indigenous Species. *Frontiers in Marine Science*, *5*(September). https://doi.org/10.3389/fmars.2018.00322

Zehr, J. P. (2011). Nitrogen fixation by marine cyanobacteria. In *Trends in Microbiology* (Vol. 19, Issue 4, pp. 162–173). https://doi.org/10.1016/j.tim.2010.12.004

Zemb, O., Achard, C. S., Hamelin, J., de Almeida, M. L., Gabinaud, B., Cauquil, L., Verschuren, L. M. G., & Godon, J. J. (2020). Absolute quantitation of microbes using 16S rRNA gene metabarcoding: A rapid normalization of relative abundances by quantitative PCR targeting a 16S rRNA gene spike-in standard. *MicrobiologyOpen*, *9*(3). https://doi.org/10.1002/mbo3.977

Zhang, Y., Pavlovska, M., Stoica, E., Prekrasna, I., Yang, J., Slobodnik, J., Zhang, X., & Dykyi, E. (2020). Holistic pelagic biodiversity monitoring of the Black Sea via eDNA metabarcoding approach: From bacteria to marine mammals. *Environment International*, *135*(October 2019), 105307. https://doi.org/10.1016/j.envint.2019.105307

Zorz, J., Willis, C., Comeau, A. M., Langille, M. G. I., Johnson, C. L., Li, W. K. W., & LaRoche, J. (2019). Drivers of regional bacterial community structure and diversity in the northwest atlantic ocean. *Frontiers in Microbiology*, *10*(FEB), 1–24. https://doi.org/10.3389/fmicb.2019.00281

Zwirglmaier, K. (2005). Fluorescence in situ hybridisation (FISH) - The next generation. *FEMS Microbiology Letters*, *246*(2), 151–158. https://doi.org/10.1016/j.femsle.2005.04.015

Chapter 2 The Robotic Cartridge Sampling Instrument (RoCSI): Proof of concept for an autonomous biomolecular sampling device

Robyn Samuel[ab], Robin Brown[a], James Wyatt[a], John Walk[a], Richard Pinnell[a], Ali Baylay[b], Weiyi Tang[c], Matt Wilson[d], Luke E. Holman[e], Susan Evans[a], Claire Widdicombe[f], Marc Rius[gh], Matt Mowlem[a] and Julie Robidart[a].

[a]Ocean Technology and Engineering Group, National Oceanography Centre Southampton, United Kingdom

[b]School of Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton, United Kingdom

[c]Department of Geosciences, Princeton University, Princeton, NJ, United States

[d]Clinical and Experimental Sciences, University of Southampton, Southampton, UK

[e]Section for Molecular Ecology and Evolution, Globe Institute, University of Copenhagen, Copenhagen, Denmark

[f]Plymouth Marine Laboratory, Prospect Place, Plymouth, United Kingdom

[g]Department of Marine Ecology, Centre for Advanced Studies of Blanes (CEAB), CSIC, 17300 Blanes, Catalonia, Spain

[h]Centre for Ecological Genomics and Wildlife Conservation, Department of Zoology, University of Johannesburg, South Africa

2.1 Abstract

Amid the escalating biodiversity crisis, understanding the biodiversity of Anthropocene oceans is critical for tracking change and evaluating the success of restoration programs. Current biodiversity surveys frequently rely on visual observations, which are costly and logistically challenging. This study presents the Robotic Cartridge Sampling Instrument (RoCSI), an autonomous device designed to automate the collection and preservation of environmental biomolecules, such as DNA and RNA, in aquatic ecosystems. RoCSI enables high-resolution biodiversity monitoring and provides insights into how functional traits shape species distribution.

An initial deployment in the western North Atlantic Ocean evaluated RoCSI's performance against standard biomolecular manual sampling methods using qPCR and multi-marker

metabarcoding. RoCSI successfully sampled across a wide range of biomass conditions without disruption and achieved a highly significant correlation with manually collected samples (1:0.998, p-value <0.001) for the nifH gene abundances. No significant differences were detected in microbial and eukaryotic taxa between RoCSI and manual samples.

These results indicate the RoCSI's potential as a valuable tool for marine research, enabling the autonomous collection of biomolecular samples even in challenging conditions. By supporting the exploration of Essential Biodiversity Variables, such as phylogenetic and trait diversity, the RoCSI enhances our ability to study the often-overlooked variability within marine ecosystems, thereby advancing ecological monitoring and biodiversity research.

2.2 Introduction

Aquatic ecosystems have played a crucial role in making Earth habitable over geological timescales (Olejarz et al., 2021). They are vital for oxygen production, carbon sequestration, nutrient cycling, and provide significant value to national economies and food security for human populations (Pecl et al., 2017; Zehr, 2011). Despite their importance, our knowledge of aquatic ecosystem biodiversity remains limited due to the challenges in obtaining representative ecological samples. While satellite ocean colour imagery has shed light on the structure of phytoplankton communities (Siegel et al., 2013), in-situ sampling is necessary to address many Essential Biodiversity Variables (EBVs) and Essential Ocean Variables (EOVs) such as taxonomic diversity, trait diversity, and species distributions (Miloslavich et al., 2018; Muller-Karger et al., 2018). Obtaining representative in-situ ecological samples requires high-resolution sampling that can be prohibitively expensive and logistically challenging. As a result, long-term (>5 year) biological ocean observing programmes are estimated to only cover approximately 7% of the ocean surface area, mostly in coastal regions (Satterthwaite et al., 2021).

Molecular techniques that capture and identify biomolecules from environmental samples (i.e., environmental DNA/RNA or eDNA/eRNA) can provide rapid insight into the presence of keystone species and functional genes (Robidart et al., 2019), community composition (Holman et al., 2019), biogeographic patterns (Holman et al., 2021), and the abundance of key taxa (Salter et al., 2019; Harper et al., 2020). Furthermore, diversity and biotic indices can also be derived from environmental DNA (eDNA) analyses (Aylagas et al., 2014; Kelly et al., 2016), enabling us to measure whole-community changes and compare the change of such metrics between methods. An array of molecular techniques can be employed to assess biological diversity and functioning when DNA and RNA are simultaneously extracted from an environmental sample (Kitahashi et al., 2020; Laroche et al., 2017; Pochon et al., 2017; Zaiko et al., 2018). Since very

low volumes of nucleic acids are used in molecular analyses, a single environmental sample can potentially generate multiple DNA and RNA datasets (Ficetola & Taberlet, 2023).

Quantitative PCR (qPCR) and metabarcoding are commonly used analysis tools in eDNA studies (Taberlet et al., 2018). Quantitative data on target taxa and genes can be acquired through qPCR, providing valuable information on trends in the abundance/biomass of key taxa and functional genes (Farnelid et al., 2016; Harper et al., 2020; Robidart et al., 2014). Metabarcoding studies use universal primer sets to amplify highly variable gene regions (barcodes), enabling the resolution of taxonomic diversity in broad groups such as microbes (16S rRNA) or eukaryotic phytoplankton (18S rRNA). The vast species lists generated from metabarcoding data provide deeper insight into the range of species present, making eDNA metabarcoding a valuable tool for the early detection of invasive and cryptic species (Holman et al., 2019; Rey et al., 2020). Furthermore, using more than one primer set on a single eDNA sample allows for simultaneous multi trophic level assessments, which provide insight into cross-trophic level interactions (Liu & Zhang, 2021; Zhang et al., 2020; Ficetola & Taberlet, 2023).

The utility of biomolecular analyses in providing early warnings of biological hazards has driven the development of in situ analytical instruments, such as the Environmental Sample Processor (ESP). The ESP can filter and preserve biomolecular samples and perform in situ analytics (Scholin et al., 2009; Yamahara et al., 2019). Some samplers are designed to collect and filter large volumes of water (tens to hundreds of litres) but are limited in the number of samples they can handle (Govindarajan et al., 2022; Honda & Watanabe, 2007; Winslow et al., 2014). Others are developed specifically for functional research, with capabilities for in situ incubation (Taylor and Doherty, 1990; Bombar et al., 2015; Edgcomb et al., 2016). However, there remains a need for compact, easy-to-deploy automated sampling devices that enable continuous, discrete sampling for broad-scale biomolecular monitoring without requiring additional *in situ* analytics (McQuillan & Robidart, 2017).

This study describes the proof of concept for the latest Robotic Cartridge Sampling Instrument (RoCSI) comparing amplicon sequencing and qPCR results from field samples collected and preserved both using the prototype RoCSI and manually. The RoCSI is a user-friendly biomolecular sampling and preservation device, which can be utilised alongside standard physical and chemical sensors without the need for specialist technicians. Sampling procedures are easily customised via USB connection to any laptop with the custom software and graphical user interface. The RoCSI is designed to use commercially available consumables for medical and water quality applications, such as Sterivex™ (Millipore®) cartridges and standard Flexboy® blood bags (Sartorius AG), to avoid manufacturing bottlenecks that can occur with custom made consumables. We discuss how the results from this study have led to

design improvements in the commercially available RoCSI, making it a valuable tool for monitoring aquatic ecosystems, reducing the reliance on ship-based sampling and helping achieve net zero.

2.3 Materials & Methods

2.3.1 The Robotic Cartridge Sampling Instrument (RoCSI)

The RoCSI is a fully automated water filtration and biomolecule preservation device capable of collecting discrete samples continuously throughout a range of aquatic systems, from low particulate concentrations in oligotrophic waters to high particulate concentrations in coastal bloom conditions. There are few moving parts, and it is smaller than other commercially available options that provide more than 30 samples per deployment (Edgcomb et al. 2016; Scholin et al. 2009; Winslow et al. 2014; Figure 2B & C). It connects to a laptop via USB connection, allowing users to programme specific missions with software that monitors both the system's operation and the working environment. The software logs sample collection data, such as sample time/volume/duration and uses integrated pressure sensors to recognise clogging and stop sampling, thus mitigating faults or damages to the samples or system.

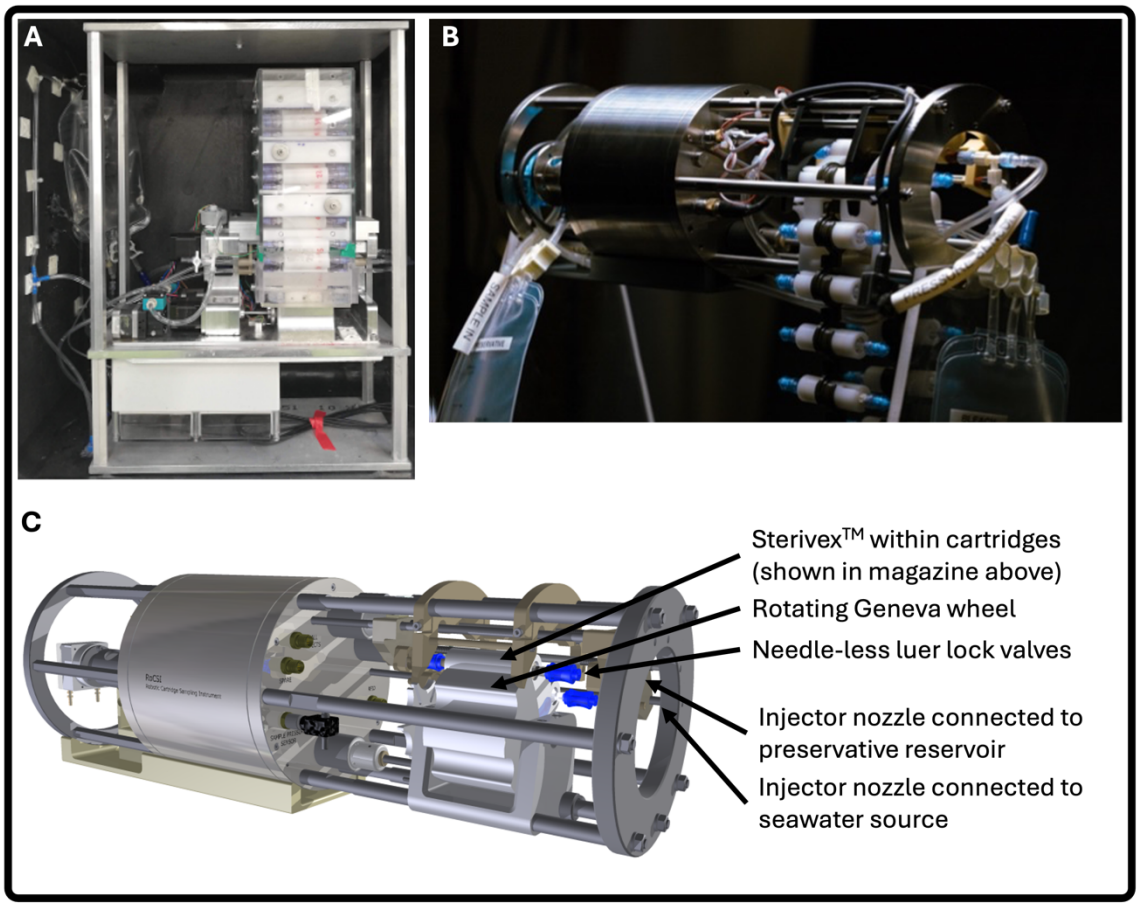


Figure 2 **A** Photograph of the prototype RoCSI used in this study. **B** Photograph of the commercially available RoCSI. **C** Labelled CAD diagram of the commercially available RoCSI.

RoCSI filters water through commercially available Sterivex™ Cartridges (Millipore), which are loaded into sample units on a magazine (Figure 2B). The magazine rotates around a geneva wheel driven by a motor directly coupled to the magazine or remotely using magnetic couplings to enable underwater operations. The magazine accurately positions a discrete sample unit in the correct position for the sampler injection system to connect (Figure 2C).

The injection system is driven using a motor directly coupled to the leadscrew, or remotely using a magnetic coupling for underwater use. The leadscrew has both left-hand and right-hand threads to allow the simultaneous injection of nozzles to both ends of the sampling unit. The sampling units have needle free valves fitted to both ends to seal the sample, thus reducing sample contamination before and after filtration. Once the injection system is engaged the sample pump is activated by the software and it is driven directly by a motor or via a magnetic coupling to enable underwater use. The pump activates for a time or volume set by the software and filtration happens according to the mission directive. After the mission parameter is met, the injection system is disengaged. The needle free valves ensure the sample is not lost due to leakage and decreases the risk of contamination (Figure 2C).

The Geneva wheel activates and places a new sample unit in position to begin sampling, at the same time the unit containing a sample is indexed to the station that provides an injection of preservative. An injection system at this station engages and provides a dosage of preservation fluid (e.g., RNAlater or any equivalent liquid fixative) to preserve the sample. The operation is repeated for the duration of the mission until the sample quantity is acquired. The Geneva wheel rotates and can therefore operate continuously by feeding filters into the sampling position. The number of samples that can be run is therefore unlimited and solely dependent on the space available to hold the sample units (Figure 2).

During all operations, the system software monitors the actions of the system and the environmental parameters it operates in. It includes sensors that track magazine and injection system engagement and continuously measure flow rate and pressure/differential pressure within the system. The software logs any errors that occur during these operations and can stop sampling if parameter thresholds are exceeded (e.g., over pressuring during sample filtration). Sample and system data are stored on a memory card, with the option for RS-232 telemetry for near real-time data acquisition. The commercially available RoCSI (McLane Research Laboratories) has been modified to support underwater operation up 6000m depth and has an additional bleach flushing step for automated decontamination (Table 1).

Table 1 List of RoCSI specifications compared with the prototype RoCSI specifications

	RoCSI prototype	RoCSI
Functional		
Depth rating	Surface for benchtop use	6000m
Samples per mission	Continuous	Continuous/Dependent on platform capacity
Sample types	Filtered water ≤2 L per sample In-situ preserved Sterivex	Filtered water ≤2 L per sample In-situ preserved Sterivex
Pore size	0.22µm/0.45µm	0.22µm/0.45µm
Pumping system	Peristaltic pump	Peristaltic pump
Decontamination system	Inlet valves for manual bleach flushing	Automated in-line bleach flushing
Weight	10 kg	15.5 kg; 10 kg wet weight
Size	140 mm x 150 mm x 300 mm	170 mm x 180 mm x 530 mm
Materials		
RoCSI System	Titanium	Titanium
Electronics housing	Plastic	Titanium
Electrical		
Power	12V DC / 2A nominal; 16V max 0.4 – 0.7A current at 12 V	12V DC / 2A nominal; 16V max 0.4 – 0.7A current at 12 V
Communication	RS-232 with cross-platform console and graphical user interfaces	RS-232 with cross-platform console and graphical user interfaces
Control		
Operations	Set mission prior to operations; modify as needed by starting a new mission	Set mission prior to operations; additional capacity to add decontamination runs between sampling events; sample termination based on individual thresholds (i.e. volume, time, pressure)

2.3.2 Samples collection

The survey was conducted as part of the AE1714 cruise from Bermuda, which took place from 29th July to 7th August 2017. The cruise route covered the waters around Bermuda and the Northwest Atlantic, moving from Bermuda to Nova Scotia, Canada, then along the United States coastline to New Jersey, before returning Southeast to Bermuda (Figure 4).

During the cruise, trace metal free surface seawater (~5 m) was continuously pumped from an in-situ TowFish (GeoFish, manufactured by University of California Santa Cruz) to the laboratory aboard the R/V *Atlantic Explorer* (Tang et al., 2020). A portion of the seawater flow was redirected to a 100ml flow-through seawater reservoir dedicated to the RoCSI, while another portion was sampled in parallel using a Masterflex L/S Digital Precision peristaltic pump (Cole-Parmer, Saint Neots, UK). To decontaminate the equipment, all instruments were cleaned with 10% bleach and rinsed with Milli-Q water. Sample blanks (i.e., samples taken using only Milli-Q water) were taken at the beginning of the cruise and at regular intervals throughout the cruise.

2.3.2.1 RoCSI sample collection

The RoCSI was configured to filter seawater from the TowFish for a maximum of 40 minutes hourly, throughout the cruise (Figure 3). The seawater was filtered through 0.22µm PES Sterivex filters with RNAlater preservative added immediately after each filtration cycle. To prevent damage to the filter membrane and to avoid cell lysis in high biomass regions, an overpressure limit of 800 mBar for over 10 seconds was set. If this limit was exceeded, the filtration would stop and RNAlater would be added to the sample. In high-biomass portions of the transect, the pressure sensor was turned off and a 30-minute limit for filtration was imposed (Supplementary Table 1).

The RoCSI sampled hourly for 24 hours. At the end of each day, the RoCSI samples were collected, the RNAlater was removed from the cartridges via the outlet, and the samples were frozen at -80ºC. Three RoCSI blanks using a Milli-Q water supply were included in the analysis. Blank 1 was taken before any seawater sampling, while Blanks 2 and 3 were collected between seawater samples.

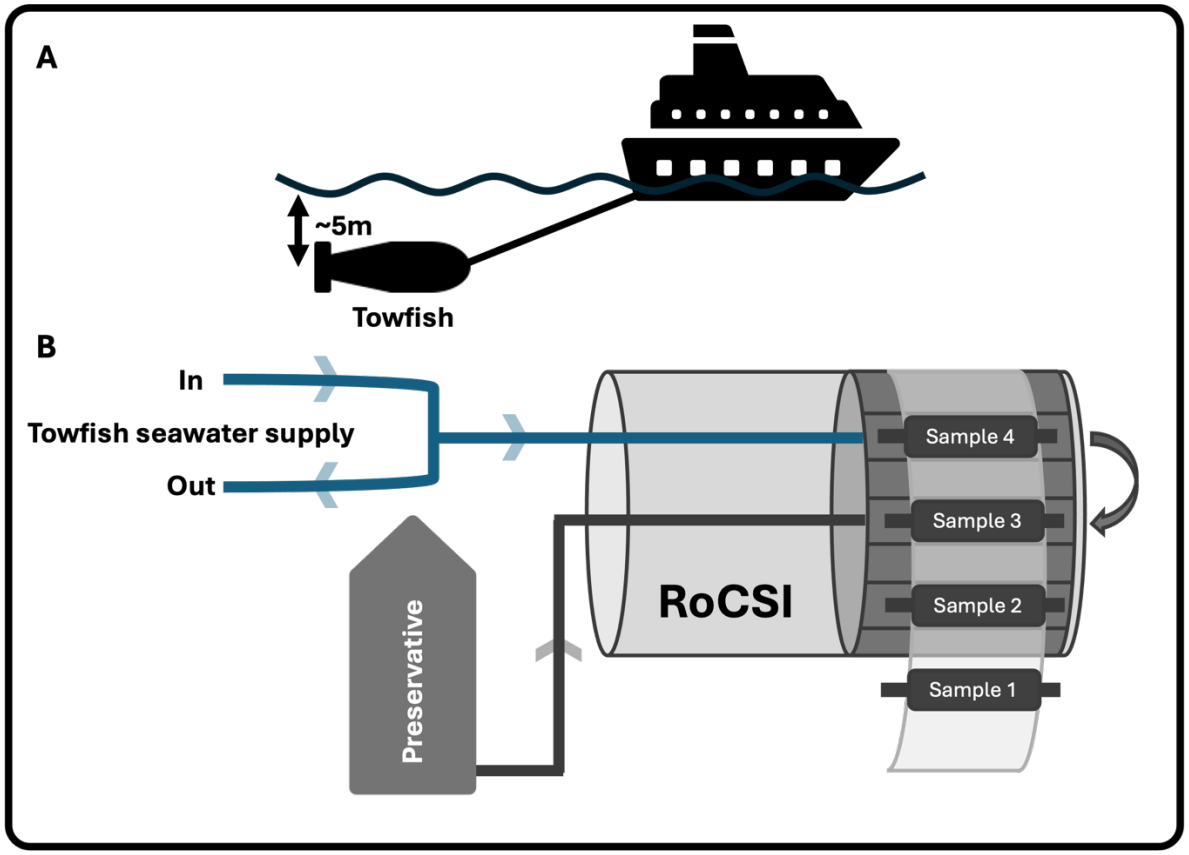


Figure 3 **A** Towfish seawater supply to ships underway system. **B** Simplified schematic of the RoCSI, showing seawater and preservative supply.

2.3.2.2 Manual sample collection

Three times daily, samples were collected from the Towfish flow and filtered manually using a Masterflex peristaltic pump and 0.22µm Sterivex filters. Samples were taken to approximately correspond with RoCSI samples, however filtering speed for the Masterflex pump was approximately twice as fast as the RoCSI, resulting in some minor differences in the sample time (Supplementary Table 1). After filtering, excess water was removed, the Sterivex filters were sealed with inert clay putty, and samples were flash frozen in liquid nitrogen before being stored at -80°C. Three manual blanks using the Milli-Q water supply were included in the analysis, each taken after the standard decontamination procedure between samples.

2.3.3 Nucleic acid purification & quality control

All nucleic acid isolation was performed in a dedicated clean lab, free from cultures and amplicons. Surfaces and instruments were wiped down with 5% bleach and RNase away for RNA purification. DNA was extracted using a modified version of the Qiagen All Prep protocol. Briefly, Sterivex filters were halved and added to autoclaved bead beater tubes, with 0.5 and 0.1mm glass beads and 800µl Qiagen RLT+ buffer for 2 minutes of bead beating at 30Hz before

on column extraction following the Qiagen All Prep protocol. DNA was eluted in 50µl Qiagen AE buffer.

DNA, for metabarcoding, was extracted using a modified DNeasy Plant Mini kit (Qiagen). In brief, the Sterivex filters were halved and added to separate autoclaved bead beater tubes as previously described. AP1 buffer (400 µl) was added, and the samples underwent three freeze-thaw cycles: 30 seconds in liquid nitrogen, followed by 3 minutes in a 65°C dry heat block. This was followed by 2 minutes of bead beating at 30 Hz using the Vortex-Genie (Scientific Industries) with a horizontal bead-beating attachment. Proteinase-k (45µl) was added to each sample and incubated sideways in an Eppendorf Thermomixer Compact, at 55°C for one hour at 400rpm. Samples were then processed following the DNeasy Plant kit manufacturer's protocol, with final elution in 50µl AE buffer.

RNA, for concentration comparisons, was extracted using the RNeasy Mini Kit (Qiagen) protocol, eluting into 50µl RNase-free water. Followed by RNase-free DNase (Qiagen) treatment and clean-up using RNA Clean and Concentrator (ZYMO), with final elution into 50µl DNase/RNase free water. Negative control DNA and RNA extraction blanks were processed in parallel. For all samples, DNA and RNA concentrations were quantified with Qubit fluorometer (Thermofisher) and quality was assessed using BioAnalyzer electropherograms (Agilent).

2.3.3.1 Quantitative PCR

Quantitative PCR (qPCR) reactions targeting the *nifH* gene, coding for a subunit of the key enzyme for nitrogen fixation (nitrogenase) from *Atlelocyanobacteria thalassa* (hereafter referred to as UCYN-A1 for unicellular cyanobacteria group A1), were run in triplicate on each sample. 30 µl reaction mixes contained 1x Accuprime PCR master mix (Invitrogen, CA, USA), 2.5mM MgCl2, 0.4µM UCYN-A *nifH* F 5'-AGCTATAACAACGTTTTATGCGTTGA-3' and UCYN-A *nifH* R 5'-CGGCCAGCGCAACCTA- 3', and 0.2µM FAM-labeled probe 5'- TCTGGTGGTCCTGAGCCTGGA-3', as described (Robidart et al., 2014; Tang et al., 2020). qPCR reactions, including no template controls (NTCs), were run in 96-well plates with optical tape (Applied Biosystems, CA, USA) on the Roche LightCycler 96 (Germany) according to the following conditions: 94°C for 75s, with 45 cycles of 95°C for 15s, 59°C for 30s. Standards and samples were run in triplicate and qPCR efficiencies calculated from the standard curves ranged from 98.6% to 107.2%.

2.3.3.2 Metabarcoding

The Illumina 16S library preparation protocol was followed for both 16S and 18S rRNA gene sequencing (Illumina 2013). The forward primer pro341F (5'- CCTACGGGNBGCASCAG - 3') and reverse primer pro805R (5' - GACTACNVGGGTATCTAATCC - 3') were used to amplify the 16S

rRNA V3-V4 gene region with an amplicon of 464 base pairs (bp), specific to Bacteria and Archaea (Takahashi et al., 2014). The 18S rRNA V4 gene region was targeted using forward primer F-574 (5' - GCGGTAATTCCAGCTCCAA - 3') and reverse primer R-952 (5' - TTGGCAAATGCTTTCGC - 3') for a 378bp amplicon, specific to Eukaryota (Hadziavdic et al., 2014). Primers were modified with overhang sequences and a second, limited cycle PCR was used to add barcodes and sequencing adaptors, as specified in the Illumina sequencing library preparation protocol (Illumina, 2013.). Samples were sequenced using the Miseq-Illumina platform at the University of Southampton's Environmental Sequencing Facility.

The raw sequencing data was initially processed by removing adaptor sequences and sequences with less than 250 bp and deposited at European Nucleotide Archive (accession number: PRJEB37775). Paired-end joining, denoising and taxonomic assignment was performed using QIIME2 release 2019.7 (Bolyen et al., 2019). In brief, denoising and pair-end joining were performed using DADA2 to produce amplicon sequence variants (ASVs). Phylogenetic trees were built using fasttree and mafft alignment. Taxonomic classification was performed with Naïve Bayes classifier and classify-sklearn method with the SILVA version 132 database, clustered at 99% sequence similarity as a reference (for full scripts see GitHub rms1u18/RoCSI_Bermuda_PoC).

2.3.4 Data analysis

Statistical tests and data visualisation for DNA and RNA yield were performed in R v3.6.3, with the ggpubr v0.2.5 package. Data was tested for normality with the Shapiro-Wilk test. Pearson's correlation was used for normally distributed data and Spearman's Rank correlation was used for non-parametric data.

Multivariate analysis and visual comparisons were completed using Primer7 with PERMANOVA+ add-on. Bray-Curtis similarity matrices were constructed using presence/absence ASV data. Data were visualised using non-metric multidimensional scaling (NMDS) and group average hierarchical cluster analysis. A two factor PERMANOVA was performed using the sampling method as a fixed factor and sample region as a random factor.

Taxonomic heat trees were plotted in R using metacoder v0.3.3 (Foster et al., 2017). In brief, taxa with less than 10 reads or unassigned were removed. Pairwise comparison of taxa in coastal and offshore groups were calculated using the Wilcoxon test. Node colour was set to the log ratio of mean taxon abundance in both coastal and offshore groups. The log ratio of mean taxon abundance was set to 0 when no significant difference was found between coastal and offshore samples (for full scripts see GitHub rms1u18/RoCSI_Bermuda_PoC).

2.4 Results

2.4.1 RoCSI performance

The RoCSI sampled as expected at hourly intervals in the 10-day cruise, collecting 178 Sterivex cartridge samples across a range of biomass, from oligotrophic open ocean to coastal bloom conditions (Figure 4; Supplementary Table 4). The volume of seawater filtered through each cartridge ranged from 200-2646ml, with an average sample volume of 1246ml (±478SD; Supplementary Table 4). Nearshore samples had lower filtration volumes due to the termination of filtration at maximum pressure limits set by the user, in a region where biomass was high and filtration was more prone to clogging.

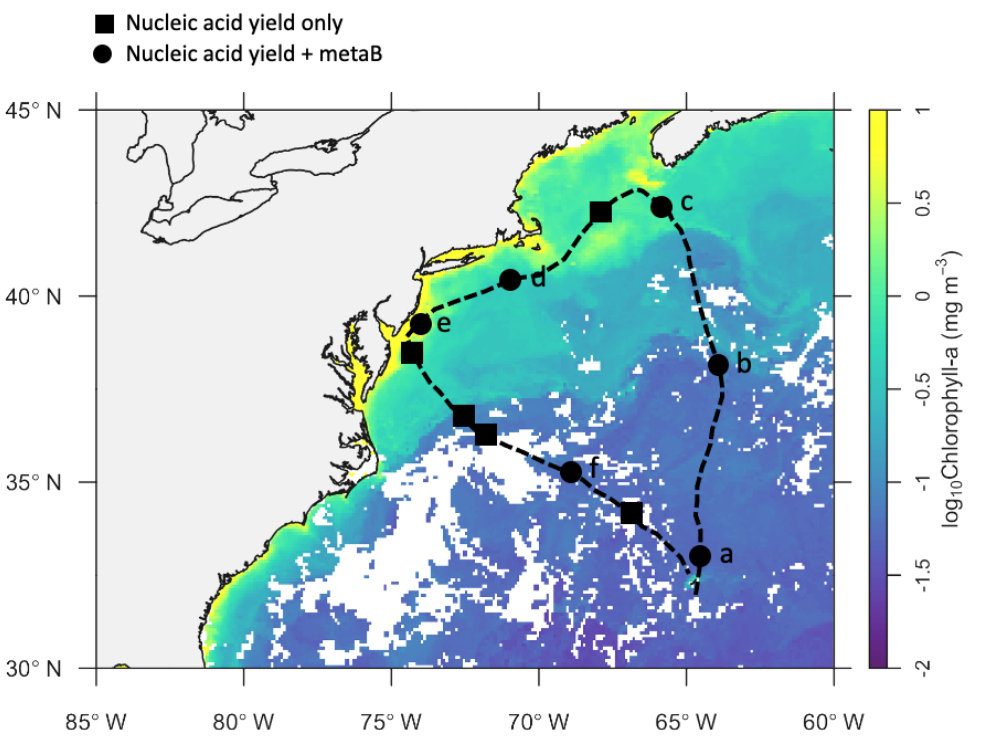


Figure 4 Map showing the AE1714 cruise track (black dashed line). The locations of the sample collection sites are indicated by labelled circles (a-f), where sample pairs were obtained for both DNA/RNA yield and metabarcoding analyses, and squares, where samples were only collected for DNA/RNA yield analysis. The map also includes log10-scaled chlorophyll-a concentrations, derived from 10-day composite satellite images for chlorophyll-a during the cruise (as described in Tang et al., 2019, 2020).

2.4.2 Nucleic acid yield

DNA and RNA yield from RoCSI samples were significantly correlated to manually collected samples (t = 3.276, df = 7, p-value = 0.014 and t = 4.3, df = 8, p-value = 0.003, respectively). DNA yields were on average 1.62 times higher in manually collected samples than in ROCSI samples whereas RNA yields were 1.71 times higher in ROCSI samples than manually collected samples (Figure 5).

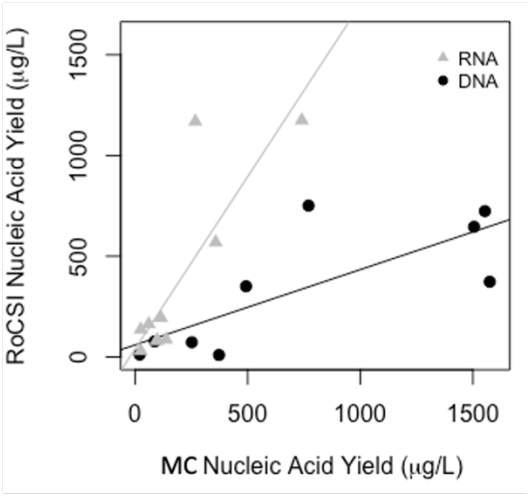


Figure 5 Scatterplot showing DNA (black) and RNA (grey) yield from RoCSI and manually collected samples with linear trendline. Significant correlation between RoCSI and manually collected (MC) DNA yield (R^2 = 0.778, p-value = 0.014, y = 0.3741x + 59.514). Significant correlation between RoCSI and MC RNA yield (R^2 = 0. 835, p-value = 0. 003, y = 1.7102 x + 39.998)

2.4.3 qPCR

UCYN-A1 *nifH* gene abundances were highly correlated between RoCSI-preserved samples and manually collected samples (R^2 = 0.9601, p-value <0.001; Figure 6), indicating quantitative recovery across 6 orders of magnitude of UCYN-A1 abundance relative to traditional sampling. UCYN-A1 *nifH* abundance from blank samples collected from RoCSI and on the benchtop filtration system, as well as qPCR NTCs, were below the limit of detection.

UCYN-A1 were ubiquitous throughout the transect, with higher *nifH* gene counts near the coast, reaching record abundances in this region (Tang et al., 2019, 2020). There was a strong gradient in UCYN-A1 abundances in the coastal-to-offshore transit, sharply declining greater than 3 orders of magnitude, coinciding with chlorophyll concentrations in the transition to the oligotrophic open ocean (Tang et al., 2019, 2020).

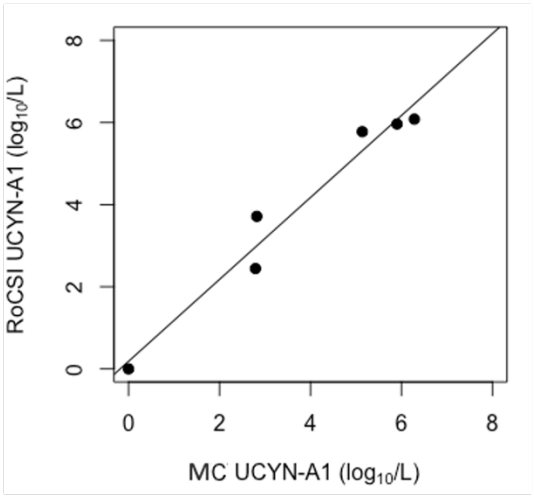


Figure 6 Significant correlation between UCYN-A1 *nifH* quantification from the qPCR of ROCSI and manually collected samples with linear trendline (R^2 = 0.9601, y = 0.9976 x + 0.183, p-value <0.001).

2.4.4 Metabarcoding

Across the 12 samples and 6 blank samples, 16S and 18S rRNA gene sequencing produced over 6.2 million reads, with approximately 3.5 million reads remaining after read merging and quality filtering. The average number of sequences per sample post-filtering was 75,828 (±49,480 SD) and 141,856 (±78,954 SD), for 16S and 18S rRNA genes, respectively. Blank field samples had an average of 34,049 (±42,463 SD) and 85,421 (±104,008 SD) sequences, for 16S and 18S rRNA genes respectively. DADA2 denoising resulted in 3642 ASVs for 16S rRNA gene and 4845 ASVs for 18S rRNA gene, from which, 1007 total taxa were identified. Of these, 184 Bacteria and Archaea were assigned from 16S rRNA gene sequences and 823 Eukaryotes were assigned from 18S rRNA gene sequences.

Samples underwent rarefaction to 36,325 sequences for 16S and 46,142 sequences for 18S, achieving saturation for both (see Supplementary Figure 11 and Figure 12 for the rarefaction curves). PERMANOVA revealed no significant differences in community composition for samples collected with the RoCSI, when compared with manually collected samples in either 18S (p-value = 0.488) or 16S rRNA gene ASVs (p-value = 0.252; see Table 2).

Both RoCSI and manually collected samples were pooled in to ‘Coastal’ or ‘Offshore’ regions. Coastal regions were defined as less than 200 km from the North American mainland coastline and Offshore as more than 200km from the mainland coastline. A significant difference was found between samples collected in the coastal region and open ocean samples, for both 18S (P = 0.011) and 16S rRNA genes (P = 0.019; Table 2).

Table 2 PERMANOVA results based on Bray-Curtis similarities using presence/absence data for 16S and 18S rRNA gene ASV composition comparing samples collected by different methods and in different regions.

	Factor	Df	Sum Sq	Pseudo-F	P-value
18S rRNA Metazoan	Method	1	3560.1	1.147	0.252
	Region	1	7212.7	1.797	**0.019**
	Residuals	8	32110		
	Total	11	45987		
16S rRNA Microbial	Method	1	1380.1	0.813	0.488
	Region	1	8131.6	3.333	**0.011**
	Residuals	8	19517		
	Total	11	30725		

Df - degrees of freedom; Sum Sq - sum of squares; Pseudo -F - F value by permutation, boldface indicates statistical significance with P<0.05, P-values based on 9999 permutations

Group average cluster analysis revealed a 4.1% average similarity for 16S rRNA gene and 15.8% for 18S rRNA gene, excluding NTCs. At 5% average group similarity, 16S rRNA gene ASVs cluster according to region sampled whereas at 25% average group similarity, 18S rRNA gene ASVs cluster according to the region sampled (orange coastal vs. blue offshore in Figure 7). In both 16S and 18S rRNA gene datasets, one manually collected sample (pair b) is clustered with the coastal samples rather than the offshore samples. This sample was taken 5 minutes after the RoCSI sample (Supplementary Table 4), whilst the ship was underway, and was taken from a site that was on the border of a chlorophyll front (Figure 4). The 5-minute delay in sample collection may have coincided with crossing the coastal front.

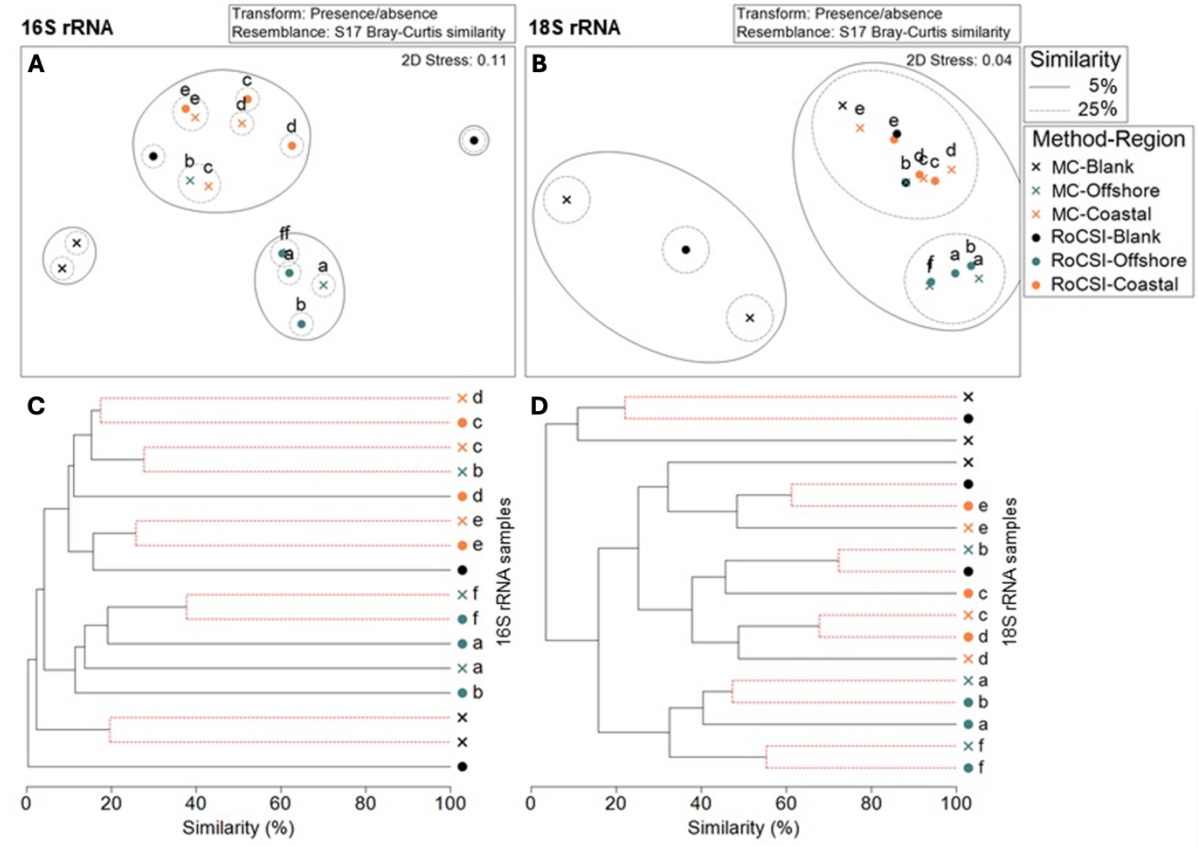


Figure 7 Ordination Based on Bray-Curtis Similarities of ASV Presence/Absence. Labels a-f indicate pairs of RoCSI and MC samples and correspond with the map labels in Figure 2. **A.** Non-metric multidimensional scaling (nMDS) plot of Bacteria and Archaea ASVs based on 16S rRNA gene metabarcoding, with similarity contours from group average cluster analysis. **B.** nMDS plot of Eukaryota ASVs based on 18S rRNA gene metabarcoding. **C.** Cluster analysis of Bacteria and Archaea ASVs from 16S rRNA gene metabarcoding with SIMPROF test for significant similarities between samples (red dashed lines indicate significant similarities). **D**. Cluster analysis of Eukaryota ASVs from 18S rRNA gene metabarcoding.

There were no significant differences in the alpha diversity indices for RoCSI and MC samples ($p > 0.05$). There were notable differences between coastal and offshore diversity indices, with a small but significant increase in Pielou's evenness for coastal samples compared to offshore samples when examining Bacterial and Archaeal (16S rRNA gene) communities (W=4.34, p-value=0.037; Table 3; Figure 8). Whereas for Eukaryotes (18S rRNA gene), Pielou's evenness was significantly greater in offshore communities compared to coastal communities (W= 4.83, p-value = 0.028) and Shannon's diversity was significantly greater in offshore (t= -2.68, p-value = 0.019; Table 3; Figure 8).

Table 3 Comparison of diversity indices in coastal or offshore regions.

		Coastal		Offshore				
	Index	Average	SD	Average	SD	t	W	p-value
16S rRNA	No. of ASVs	250.86	124.40	230.88	104.11	0.33		0.744
	Evenness (J')	0.94	0.01	0.93	0.01		4.34	0.037
	PD	22.14	6.78	19.61	6.78	0.69		0.501
	Diversity (H')	7.38	0.71	7.21	0.64	0.48		0.641
18S rRNA	No. of ASVs	463.14	185.62	633.63	380.03		0.33	0.563
	Evenness (J')	0.63	0.06	0.75	0.09		4.83	**0.028**
	PD	30.69	9.07	42.26	18.98		2.26	0.133
	Diversity (H')	5.49	0.86	6.83	1.06	-2.68		**0.019**

Significant differences <0.05 in bold. Abbreviations: SD – Standard Deviation; t = student's t-test; W – Wilcoxon test; J' – Pielou's evenness; PD – Faith's phylogenetic diversity; H' – Shannon's diversity

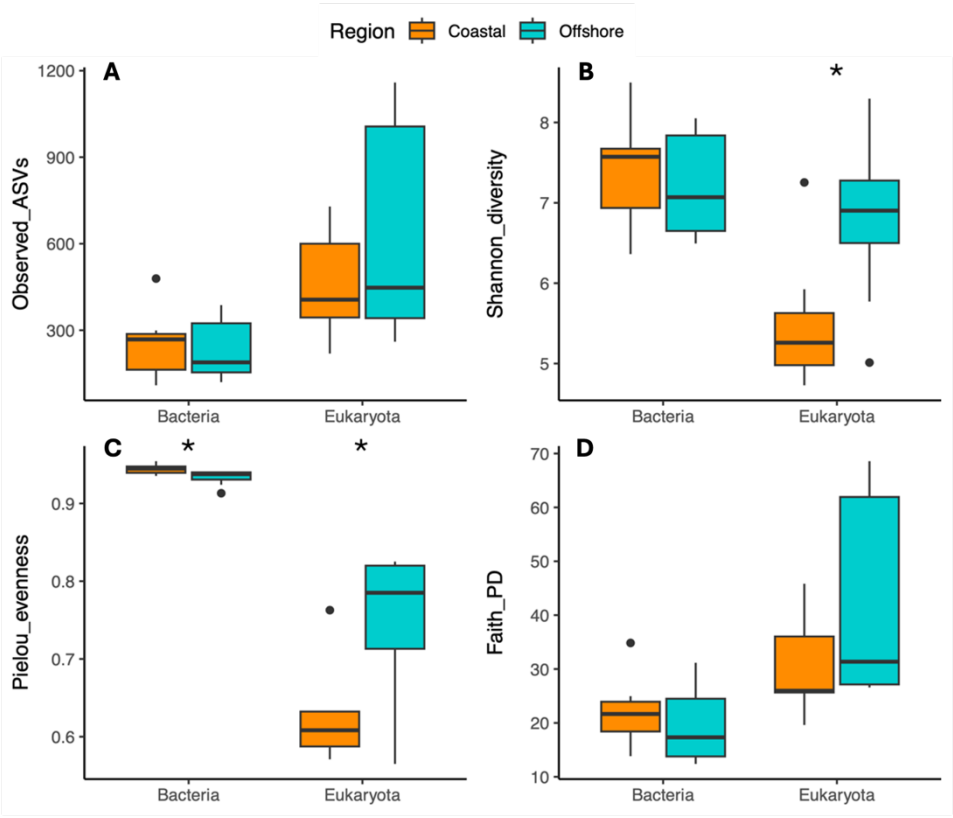


Figure 8 Comparison of coastal and offshore diversity indices from Bacteria 16S and Eukaryote 18S rRNA gene metabarcoding. Panel **A** showing the observed Amplicon Sequence Variants (ASVs), **B** showing Shannon's diversity indices, **C** showing Pielou's evenness indices, and **D** showing Faith's Phylogenetic Diversity (PD). * indicates significant differences $p<0.05$.

Heatmaps generated from 16S and 18S rRNA gene metabarcoding reveal the distribution of the 40 most abundant taxa (Figure 9). In coastal samples, 16S rRNA gene metabarcoding revealed a high abundance of sequences assigned to *Luteolibacter, Pseudomonas, Alphaproteobacteria*, *Mesonia, Bacteriodia* and *Actinobacteria*. Whereas offshore samples had a high abundance of sequences assigned *Halomonas, Halomonadacaea, Neosaia chiangmaiensis* and Firmicutes taxa. *Rhizobiales* were ubiquitous across samples for 16S rRNA gene metabarcoding (Figure 9).

The 18S rRNA metabarcoding revealed a high abundance of sequences assigned to Cyclopodia, *Centropages typicus*, Pseudoperkinsidae, Prymnesiales, *Braarudosphaera bigelowii*, Bathycoccus, Mamiellales. Whereas Syndiniales group V was the only taxa consistently in high abundance in the offshore samples. *Emiliania huxleyi* CCMP1516 and Syndiniales group III were ubiquitous across 18S rRNA metabarcodes (Figure 9).

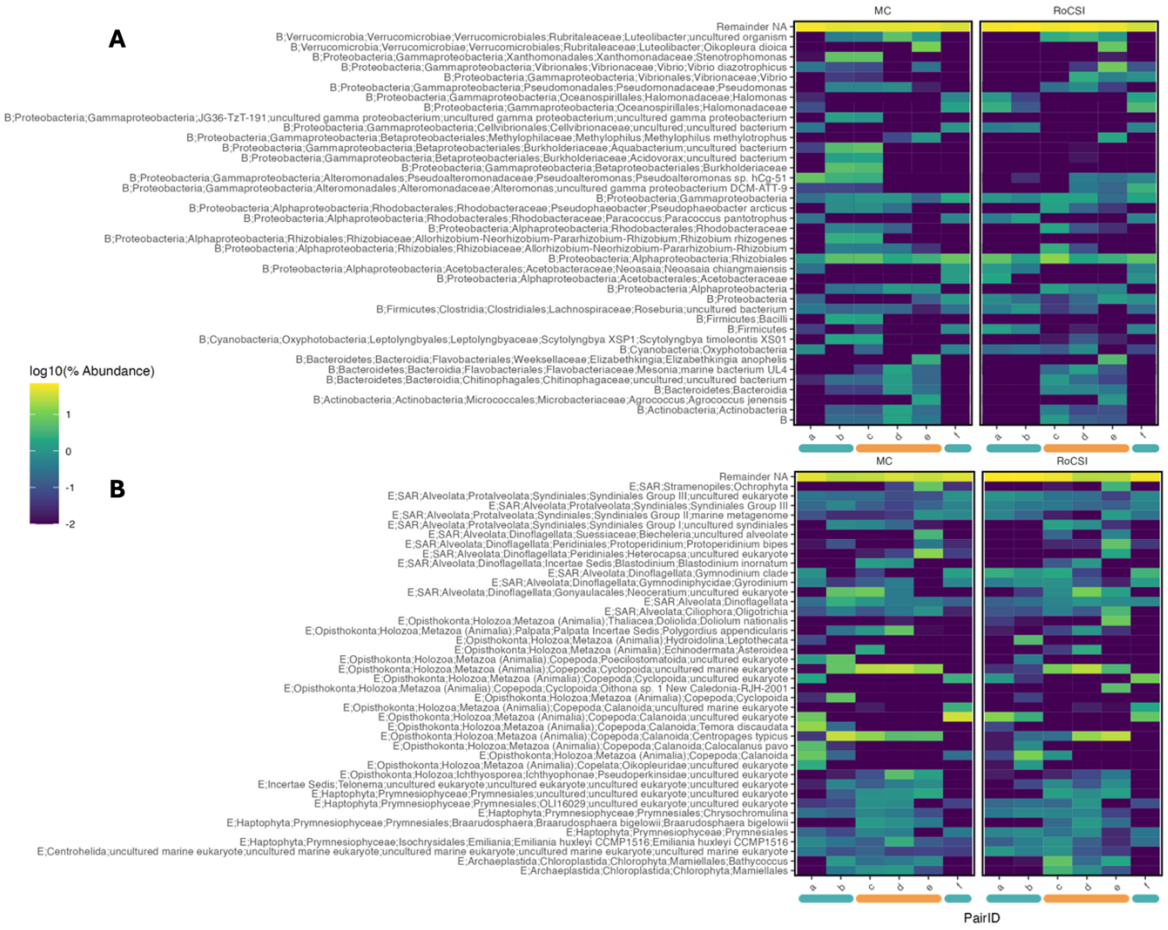


Figure 9 Heat maps of the top 40 most abundant taxa with colour scales representing relative abundance of sequence reads. Taxonomy assignments on the y-axis assigned to the lowest common ancestor. Individual samples along the x-axis grouped by collection method. Samples collected in offshore and coastal regions were illustrated by the blue and orange bars under the sample letters. **A** Shows Bacteria identified through 16S rRNA gene metabarcoding. **B.** Shows Eukaryota identified through 18S rRNA gene metabarcoding.

The 16S rRNA gene heat tree shows Bacteria with significantly greater proportion of reads in coastal samples included *Mesonia*, a flavobacteria associated with coastal blooms; *Pirellula*, a genus with global distribution; *Methylophilus*, methanol-utilizing bacteria; *Coxiella; Chitinophagagales* and *Luteolibacter*. Bacteria with significantly greater proportions of reads offshore included *Bdellovibrionaceae*, a heterotrophic bacterium; *Halomonas*, a saline tolerant genus; *Roseburia* and *Phycisphaerae* (Figure 10).

The 18S rRNA heat tree shows the Eukaryotic taxa that had a significantly greater proportion of reads in coastal regions including; the coastal Chlorophytes *Trebouxiophyceae, Ostreococcus, Bathycoccus*, *Micromonas* (Mamiellales), *Pycnococcus,* and *Pterosperma*; the coastal diatoms *Thalassiosira* and *Guinardia*; the ciliate *Oligotrichia*; and the radiolaria *Arthracanthida*.

Eukaryotic taxa with proportionally greater reads in offshore samples were more diverse. Several of these taxa are associated with coral endosymbiotic dinoflagellates (Symbiodinium genus), harmful algal blooms (Alexandrium, Gonyaulax and Kareniaceae genus); and parasitic lifestyles (Scuticociliatia genus) (Figure 10).

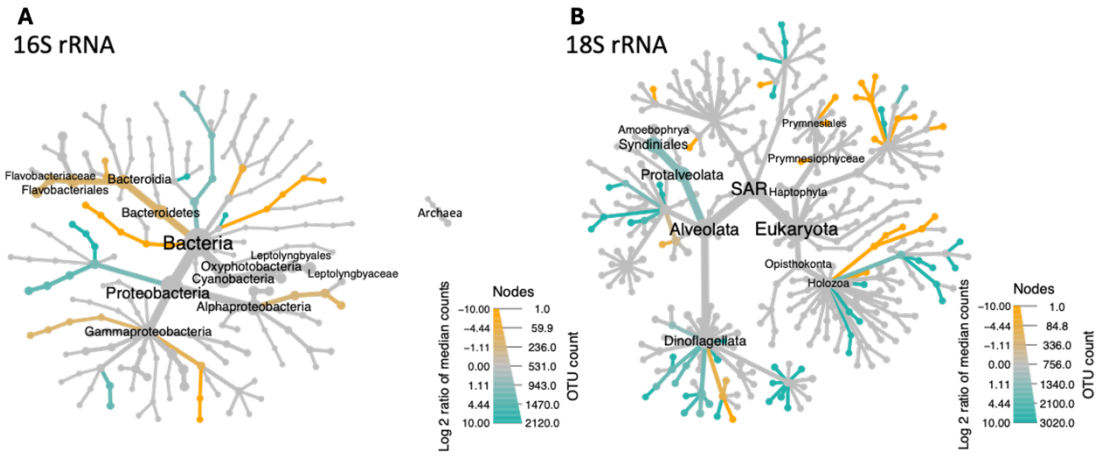


Figure 10 Taxonomic heat tree showing the pairwise comparisons of coastal and offshore communities. The colour of each taxon represents the log-2 ratio of median proportion of reads observed in each region. Only significant differences are coloured. Taxa coloured blue are enriched in offshore communities and taxa in orange are enriched in coastal communities **A** Shows Bacteria and Archaea identified through 16S rRNA gene metabarcoding. **B.** Shows Eukaryota identified through 18S rRNA metabarcoding. (Detailed view in Supplementary Figure 14)

2.5 Discussion

During this cruise, the RoCSI was managed by one crew member who replaced cartridges once daily and manually decontaminated the system. This allowed for more time to be allocated to on-board experimentation and other projects. Furthermore, as long as the pump remained submerged, samples could be safely collected throughout periods of rough sea conditions, increasing overall scientific productivity during research cruises and reducing the risks associated with at-sea research.

The RoCSI successfully filtered and preserved water samples across a range of biomass, from low biomass oligotrophic open ocean to high biomass coastal bloom conditions. However, the automated overpressure protocol, set to stop filtration when inline pressure exceeded 800 mBar for 10 seconds, had to be overridden in high-biomass regions. Instead, the filtration time was limited to 30 mins. Exceeding this overpressure threshold may have cause some cells to lyse, but the impact of cell lysis on metabarcoding results remains unclear. No damage to the tubing or pump components was observed after this safety mechanism was overridden.

Molecular analysis of these samples provided valuable insights into the microbial ecology of the region, and these findings were consistent with results obtained through conventional manual collection methods, as well as with previous studies conducted along the same transect (Tang et al., 2019; Wang et al., 2021).

2.5.1 Nucleic acid yield

DNA and RNA yields from manually collected samples both showed significant positive correlation with RoCSI methods (1:0.4 and 1:1.7, respectively). Although DNA yields were generally higher in manual samples, all samples still had sufficient DNA yield for metabarcoding (>2ng/µl). Conversely, RNA yields were lower in manually collected samples compared to RoCSI samples. RNA is less stable than DNA, therefore obtaining high enough yields for laboratory analysis can be challenging (McGrath et al., 2008). Consistent with previous finding (Ottesen et al., 2011), these results indicate that immediate stabilisation with RNAlater using an autonomous device improves recovery of RNA when compared to manually collected samples that are flash frozen after filtration.

The co-purification kit used in this study, Qiagen AllPrep DNA/RNA Mini Kit, allows both DNA and RNA analyses to be run on a single sample, maximising scientific output. In recent years, studies have been using both eDNA and eRNA to determine the living proportion of DNA reads (Cristescu, 2019; Guardiola et al., 2016; Keeley et al., 2018; Laroche et al., 2017; Pawlowski et al., 2016; Pochon et al., 2017; Wood et al., 2020). To fully utilise the RoCSI collected samples, it is important to consider the best preservation fluid for both DNA and RNA. In this study and others, the preservative RNAlater was highly effective at preserving the RNA from filter samples (Edgcomb et al., 2016). However, DNA yields were typically lower than manually collected flash frozen samples. Further research is needed to identify the most effective preservative for both RNA and DNA, particularly for the long-term preservation of DNA and RNA at ambient temperatures, which would be necessary for long-term deployments of the RoCSI. Preservative for the RoCSI is stored in FlexBoy blood bags (Sartorius) and can be easily exchanged, allowing for future optimization of preservation methods.

2.5.2 Quantitative PCR of UCYN-A1 *nifH*

UCYN-A1 is a unicellular cyanobacterium that fixes nitrogen (Moisander et al., 2010) and forms a symbiotic relationship with the Prymnesiophyte, *Braarudosphaera bigelowii* (Thompson et al., 2012). In the sample survey region during the summer months, there is frequently a strong gradient in UCYN-A1 *nifH* counts, with high abundances nearshore, corresponding with high rates of nitrogen fixation and high chlorophyll-a concentrations, and lower abundance offshore

(Mulholland, 2007; Tang et al., 2020; Figure 4). We saw these expected trends across this gradient, demonstrating that RoCSI can collect samples that deliver comparably quantitative molecular data relative to manually filtered samples, despite wide ranges in biomass. Further, the blank samples (filtered Milli-Q water on RoCSI and on the bench), had undetectable *nifH* abundances, an indication of the suitability of the autosampler to deliver samples with low contamination when measured by targeted techniques.

2.5.3 16S & 18S rRNA gene metabarcoding

The results from metabarcoding show that neither 16S or 18S rRNA gene community composition significantly differs between the RoCSI sampler and best practice manually collected samples. We observed a sharp transition between open ocean and coastal communities, reflecting the transition observed in satellite-derived chlorophyll-a observations. Coastal and offshore sites have distinct community profiles (Clarke, 1940; Stefanni et al., 2018; Zorz et al., 2019) and this is reflected in our results, which revealed significant differences in Bacteria, Archaea and Eukaryotes communities when comparing coastal and offshore samples. Furthermore, taxa recovered from 16S and 18S rRNA gene metabarcoding across the transect samples are similar to those recovered from a separate CTD Niskin based survey from the same cruise (Wang et al., 2021).

Several of the Eukaryotic taxa that had a significantly greater proportion of reads in coastal samples included the diatoms *Thalassiosira* and *Guinardia*, the ciliate *Oligotrichia*, the protozoa *Arthracanthida*, and several genera of green algae, including *Ostreococcus*, *Bathycoccus*, and *Micromonas* (Figure 10). The presence of these coastal taxa is supported by previous studies that have described their association with coastal regions in the northwest Atlantic (Decelle et al., 2013; Doherty et al., 2010; Hernández-Becerril, 1995; Marin & Melkonian, 2010; Prasad et al., 2011).

In addition to the Eukaryotic taxa, the Bacterial taxa *Flavobacteriales*, *Pirellula*, and *Methylophilus* were also found to have a significantly greater proportion of reads in the coastal samples, in line with previous studies that have linked these taxa to coastal blooms (Georges et al., 2014; Morris et al., 2006). The harmful algal species *Aureococcus anophagefferens*, known to form blooms on the US mid-Atlantic shelf, was also present in high abundance in our sample set and confirmed to be blooming during the cruises in a separate study (Wang et al., 2021). The presence of the symbiotic host of UCYN-A1, the coccolithophore *Braarudosphaera* (Hagino et al., 2013), was also high in the coastal region, where UCYN-A1 *nifH* counts reached record abundances (Wang et al., 2021).

In contrast, the heat trees (Figure 10) revealed a higher proportion of taxa associated with harmful algal blooms (HABs) in the offshore samples. This same pattern is not seen in the heatmap, where *Gymnodinium* is among the 40 most abundant taxa. It might therefore be artificially attributed to a bias in HAB taxa representation in the database used, as the filtering process for visualising taxa in a heat tree involves excluding unknown and unnamed taxa which are more abundant offshore. Additionally, offshore communities were found to be more diverse, with a greater evenness between taxa, which could have led to an increase in the proportion of ubiquitous taxa with low read counts in communities with greater evenness.

Results from this study have highlighted the need for additional decontamination steps in subsequent models of the RoCSI. While sequences from three of the Milli-Q blank samples clustered separately from the coastal and offshore communities, 16S rRNA gene sequences from one RoCSI blank clustered with the coastal communities and 18S rRNA gene sequences from two RoCSI blanks clustered with the coastal communities (Figure 7). The most likely source of contamination is the water remaining in the lines from the previous sample (approx. 10ml) and/or biofilm build up on the tubing. Accordingly, subsequent models of the RoCSI prototype have been designed to include bleach rinse and flushing steps to reduce these sources of contamination. Contamination is difficult to avoid even whilst using manual sampling methods and the 18S rRNA sequences from one of the manually-collected blank samples also clustered with coastal communities (Figure 7). This contamination may be residual within the tubing of the peristaltic pump, or a result of carry over during library preparation or sequencing (Sepulveda et al., 2020).

Group average cluster analysis with SIMPROF tests showed that while RoCSI and MC samples clustered into coastal and offshore communities, they did not all link by pairs. This is likely due to the patchiness of phytoplankton communities, similar to patchiness previously identified in high resolution chlorophyll measurements (Gennip et al., 2016). Satellite chlorophyll imaging showed that the outlier RoCSI sample from pair b (Figure 7), was filtered whilst the ship was underway and crossing a chlorophyll front (Figure 4). This highlights an opportunity for autonomous eDNA samplers to explore this smaller scale variability in community composition and functionality and examine the interplay with biogeochemistry on fine scales.

2.5.4 Applications

The RoCSI offers versatile applications for enhancing marine ecological research and conservation efforts. These devices can be deployed in various settings to improve sample collection efficiency, expand the scope of biological observations, and provide comprehensive

data for ecological modelling, as well as contribute to net-zero science goals. Below, we summarise the key applications and benefits of using the RoCSI:

CTD Rosette Samplers: Integration with CTD rosette samplers enables in-situ filtration and preservation at various depths, minimising the time between sample collection and preservation thereby reducing nucleic acid degradation. This approach also captures more accurate gene expression patterns by preserving samples before the pressure and temperature changes associated with sample retrieval can alter gene expression (Edgcomb et al., 2016).

Autonomous Vehicles: The RoCSI can be mounted on Autonomous Underwater Vehicles (AUVs) and Surface Vehicles, facilitating biological observations in remote or difficult-to-access areas like deep seas and underneath ice sheets. This capability can support conservation and sustainable use goals of the Biodiversity Beyond National Jurisdiction treaty through autonomous eDNA surveys.

Long-term Sampling Stations: Deploying the RoCSI at these stations can generate continuous biological datasets, undisrupted by adverse weather events that would otherwise prevent ship-based sampling. Enriching our understanding of marine ecosystems' responses to climate change and anthropogenic pressures. Furthermore, co-deployment with additional biological, chemical, and physical sensors creates metadata-rich ecological observations invaluable for ecological modelling and forecasting.

Ships of Opportunity: Utilising commercial and other non-research vessels for deploying RoCSI presents a cost-effective method for collecting data without the need for dedicated research cruises. This strategy reduces costs and emissions, echoing the Continuous Plankton Recorder's approach (Suter et al., 2021).

The integration of the RoCSI enhances the temporal and spatial resolution of biological observations, allowing for the exploration of global-scale patterns with a detail comparable to chemical and physical data. By pairing with low-power sensors, RoCSI can concurrently record chemical, physical, and biological data, crucial for developing and validating ecological models. These models can provide early warnings for phenomena like harmful algal blooms, enabling proactive management strategies to mitigate their impact on marine ecosystems.

Additionally, the widespread adoption of the RoCSI and similar eDNA samplers would substantially increase the number of environmental samples available for biobanking, supporting future 'futuromics' research. By biobanking eDNA samples, researchers can re-analyse past collections using advanced sequencing and omics techniques, enabling the

application of improved protocols and maintaining continuity in biological observations (Jarman et al., 2018).

In conclusion, this proof-of-concept study demonstrates that the RoCSI is an effective autonomous biomolecular sampling device. Continued changes in oceanic physical, chemical, and biological parameters driven by anthropogenic pressures underscore the importance of innovative monitoring tools like the RoCSI. With its low-effort and high-specificity approach, eDNA analysis offers a transformative method for ecological monitoring, enabling a significant upscaling in the resolution of biological observations across global oceans. By reliable collecting and preserving samples, the RoCSI represents a significant step forward in our ability to observe and manage marine ecosystems effectively.

2.6 Funding

RS was supported by the Natural Environment Research Council (NERC) NEXUSS Studentship (grant number NE/N012070/1). W. T. thank the funding support from Duke Graduate School, Duke Interdisciplinary Studies and Princeton University

2.7 Acknowledgments

We would like to thank the captain and crew onboard cruise AE1714.

2.8 References

Aylagas, E., Borja, A., & Rodríguez-Ezpeleta, N. (2014). Environmental status assessment using DNA metabarcoding: Towards a genetics based marine biotic index (gAMBI). *PLoS ONE*, *9*(3), 1–11. https://doi.org/10.1371/journal.pone.0090529

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology 2019 37:8*, *37*(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

Bombar, D., Taylor, C.D., Wilson, S.T., Robidart, J.C., Rabines, A., Turk-Kubo, K.A., Kemp, J.N., Karl, D.M. & Zehr, J.P. (2015). Measurements of nitrogen fixation in the oligotrophic North Pacific Subtropical Gyre using a free-drifting submersible incubation device. *Journal of Plankton Research*, *37*(4), pp.727-739.

Clarke, G. L. (1940). Comparative Richness of Zooplankton in Coastal and Offshore Areas of the Atlantic. *The Biological Bulletin*, *78*(2), 226–255. https://doi.org/https://doi.org/10.2307/1537775

Cristescu, M. E. (2019). Can Environmental RNA Revolutionize Biodiversity Science? *Trends in Ecology and Evolution*, *34*(8), 694–697. https://doi.org/10.1016/j.tree.2019.05.003

Decelle, J., Martin, P., Paborstava, K., Pond, D. W., Tarling, G., Mahé, F., de Vargas, C., Lampitt, R., & Not, F. (2013). Diversity, Ecology and Biogeochemistry of Cyst-Forming Acantharia (Radiolaria) in the Oceans. *PLoS ONE*, *8*(1). https://doi.org/10.1371/journal.pone.0053598

Doherty, M., Tamura, M., Vriezen, J. A. C., McManus, G. B., & Katz, L. A. (2010). Diversity of oligotrichia and choreotrichia ciliates in coastal marine sediments and in overlying plankton. *Applied and Environmental Microbiology*, *76*(12), 3924–3935. https://doi.org/10.1128/AEM.01604-09

Edgcomb, V. P., Taylor, C., Pachiadaki, M. G., Honjo, S., Engstrom, I., & Yakimov, M. (2016). Comparison of Niskin vs. in situ approaches for analysis of gene expression in deep Mediterranean Sea water samples. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *129*, 213–222. https://doi.org/10.1016/j.dsr2.2014.10.020

Farnelid, H., Turk-Kubo, K., Del Carmen Muñoz-Marín, M., & Zehr, J. P. (2016). New insights into the ecology of the globally significant uncultured nitrogen-fixing symbiont UCYN-A. *Aquatic Microbial Ecology*, *77*(3), 128–138. https://doi.org/10.3354/ame01794

Ficetola, G. F., & Taberlet, P. (2023). Towards exhaustive community ecology via DNA metabarcoding. Molecular Ecology, 32, 6320–6329. https://doi.org/10.1111/mec.16881

Foster, Z. S. L., Sharpton, T. J., & Grünwald, N. J. (2017). Metacoder: An R package for visualization and manipulation of community taxonomic diversity data. *PLoS Computational Biology*, *13*(2), 1–15. https://doi.org/10.1371/journal.pcbi.1005404

Gennip, S. van, Martin, A. P., Srokosz, M. A., Allen, J. T., Pidcock, R., Painter, S. C., Stinchcombe, M. C., & 1National. (2016). Plankton patchiness investigated using simultaneous nitrate and chlorophyll observations. *Journal of Geophysical Research: Oceans*, 4149–4156. https://doi.org/10.1002/2016JC011789.Received

Georges, A. A., El-Swais, H., Craig, S. E., Li, W. K. W., & Walsh, D. A. (2014). Metaproteomic analysis of a winter to spring succession in coastal northwest Atlantic Ocean microbial plankton. *ISME Journal*, *8*(6), 1301–1313. https://doi.org/10.1038/ismej.2013.234

Govindarajan, A.F., McCartin, L., Adams, A., Allan, E., Belani, A., Francolini, R., Fujii, J., Gomez-Ibañez, D., Kukulya, A., Marin, F. & Tradd, K. (2022). Improved biodiversity detection using a large-volume environmental DNA sampler with in situ filtration and implications for marine eDNA sampling strategies. *Deep Sea Research Part I: Oceanographic Research Papers*, *189*, p.103871.

Guardiola, M., Wangensteen, O. S., Taberlet, P., Coissac, E., Uriz, M. J., & Turon, X. (2016). Spatio-temporal monitoring of deep-sea communities using metabarcoding of sediment DNA and RNA. *PeerJ*, *2016*(12), 1–31. https://doi.org/10.7717/peerj.2807

Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E. M., & Troedsson, C. (2014). Characterization of the 18S rRNA Gene for Designing Universal Eukaryote Specific Primers. *PLOS ONE*, *9*(2), e87624. https://doi.org/10.1371/JOURNAL.PONE.0087624

Hagino, K., Onuma, R., Kawachi, M., & Horiguchi, T. (2013). Discovery of an endosymbiotic nitrogen-fixing cyanobacterium UCYN-A in Braarudosphaera bigelowii (Prymnesiophyceae). *PLoS ONE*, *8*(12). https://doi.org/10.1371/journal.pone.0081749

Harper, K. J., Goodwin, K. D., Harper, L. R., Lacasella, E. L., Frey, A., Dutton, P. H., Mayne, B., Simpson, T., & Harper, K. J. (2020). *Finding Crush : Environmental DNA Analysis as a Tool for Tracking the Green Sea Turtle Chelonia mydas in a Marine Estuary*. *6*(January), 1–13. https://doi.org/10.3389/fmars.2019.00810

Hernández-Becerril, D. U. (1995). Planktonic diatoms from the gulf of california and coasts off baja california: The genera rhizosolenia, proboscia, pseudosolenia, and former rhizosolenia species. *Diatom Research*, *10*(2), 251–267. https://doi.org/10.1080/0269249x.1995.9705348

Holman, L. E., de Bruyn, M., Creer, S., Carvalho, G., Robidart, J., & Rius, M. (2019). Detection of introduced and resident marine species using environmental DNA metabarcoding of sediment and water. *Scientific Reports*, *9*(1), 1–10. https://doi.org/10.1038/s41598-019-47899-7

Holman, L. E., de Bruyn, M., Creer, S., Carvalho, G., Robidart, J., & Rius, M. (2021). Animals, protists and bacteria share marine biogeographic patterns. *Nature Ecology and Evolution*. https://doi.org/10.1038/s41559-021-01439-7

Honda, M.C. & Watanabe, S. (2007). Utility of an automatic water sampler to observe seasonal variability in nutrients and DIC in the Northwestern North Pacific. *Journal of Oceanography*, *63*, pp.349-362.

Illumina. (2013). 16S Metagenomic Sequencing Library Preparation. https://emea.support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.html

Keeley, N., Wood, S. A., & Pochon, X. (2018). Development and preliminary validation of a multi-trophic metabarcoding biotic index for monitoring benthic organic enrichment. *Ecological Indicators*, *85*(January 2017), 1044–1057. https://doi.org/10.1016/j.ecolind.2017.11.014

Kelly, R. P., O'Donnell, J. L., Lowell, N. C., Shelton, A. O., Samhouri, J. F., Hennessey, S. M., Feist, B. E., & Williams, G. D. (2016). Genetic signatures of ecological diversity along an urbanization gradient. *PeerJ*, *4*, e2444. https://doi.org/10.7717/peerj.2444

Kitahashi, T., Sugime, S., Inomata, K., Nishijima, M., Kato, S., & Yamamoto, H. (2020). Meiofaunal diversity at a seamount in the pacific ocean: A comprehensive study using environmental DNA and RNA. *Deep-Sea Research Part I: Oceanographic Research Papers*, *160*(February), 103253. https://doi.org/10.1016/j.dsr.2020.103253

Laroche, O., Wood, S. A., Tremblay, L. A., Lear, G., Ellis, J. I., & Pochon, X. (2017). Metabarcoding monitoring analysis: the pros and cons of using co-extracted environmental DNA and RNA data to assess offshore oil production impacts on benthic communities. *PeerJ*, *5*, e3347. https://doi.org/10.7717/peerj.3347

Liu, J., & Zhang, H. (2021). Combining Multiple Markers in Environmental DNA Metabarcoding to Assess Deep-Sea Benthic Biodiversity. *Frontiers in Marine Science*, *8*. https://doi.org/10.3389/fmars.2021.684955

Marin, B., & Melkonian, M. (2010). Molecular Phylogeny and Classification of the Mamiellophyceae class. nov. (Chlorophyta) based on Sequence Comparisons of the Nuclear- and Plastid-encoded rRNA Operons. *Protist*, *161*(2), 304–336. https://doi.org/10.1016/j.protis.2009.10.002

McGrath, K. C., Thomas-Hall, S. R., Cheng, C. T., Leo, L., Alexa, A., Schmidt, S., & Schenk, P. M. (2008). Isolation and analysis of mRNA from environmental microbial communities. *Journal of Microbiological Methods*, *75*(2), 172–176. https://doi.org/10.1016/j.mimet.2008.05.019

McQuillan, J.S. & Robidart, J.C. (2017). Molecular-biological sensing in aquatic environments: recent developments and emerging capabilities. *Current opinion in biotechnology*, *45*, pp.43-50.

Miloslavich, P., Bax, N. J., Simmons, S. E., Klein, E., Appeltans, W., Aburto-Oropeza, O., Andersen Garcia, M., Batten, S. D., Benedetti-Cecchi, L., Checkley, D. M., Chiba, S., Duffy,

J. E., Dunn, D. C., Fischer, A., Gunn, J., Kudela, R., Marsac, F., Muller-Karger, F. E., Obura, D., & Shin, Y. J. (2018). Essential ocean variables for global sustained observations of biodiversity and ecosystem changes. *Global Change Biology*, *24*(6), 2416–2433. https://doi.org/10.1111/gcb.14108

Moisander, P. H., Beinart, R. A., Hewson, I., White, A. E., Johnson, K. S., Carlson, C. A., Montoya, J. P., & Zehr, J. P. (2010). Unicellular cyanobacterial distributions broaden the oceanic N2 fixation domain. *Science (New York, N.Y.)*, *327*(5972), 1512–1514. https://doi.org/10.1126/SCIENCE.1185468

Morris, R. M., Longnecker, K., & Giovannoni, S. J. (2006). Pirellula and OM43 are among the dominant lineages identified in an Oregon coast diatom bloom. *Environmental Microbiology*, *8*(8), 1361–1370. https://doi.org/10.1111/j.1462-2920.2006.01029.x

Mulholland, M. R. (2007). The fate of nitrogen fixed by diazotrophs in the ocean. *Biogeosciences*, *4*, 37–51. www.biogeosciences.net/4/37/2007/

Muller-Karger, F. E., Miloslavich, P., Bax, N. J., Simmons, S., Costello, M. J., Pinto, I. S., Canonico, G., Turner, W., Gill, M., Montes, E., Best, B. D., Pearlman, J., Halpin, P., Dunn, D., Benson, A., Martin, C. S., Weatherdon, L. V., Appeltans, W., Provoost, P., … Geller, G. (2018). Advancing marine biological observations and data requirements of the complementary Essential Ocean Variables (EOVs) and Essential Biodiversity Variables (EBVs) frameworks. *Frontiers in Marine Science*, *5*(JUN), 1–15. https://doi.org/10.3389/fmars.2018.00211

Olejarz, J., Iwasa, Y., Knoll, A. H., & Nowak, M. A. (2021). The Great Oxygenation Event as a consequence of ecological dynamics modulated by planetary change. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-23286-7

Ottesen, E.A., Marin, R., Preston, C.M., Young, C.R., Ryan, J.P., Scholin, C.A. and DeLong, E.F., 2011. Metatranscriptomic analysis of autonomously collected and preserved marine bacterioplankton. The ISME journal, 5(12), pp.1881-1895.

Pawlowski, J., Lejzerowicz, F., Apotheloz-Perret-Gentil, L., Visco, J., & Esling, P. (2016). Protist metabarcoding and environmental biomonitoring: Time for change. *European Journal of Protistology*, *55*, 12–25. https://doi.org/10.1016/j.ejop.2016.02.003

Pecl, G. T., Araújo, M. B., Bell, J. D., Blanchard, J., Bonebrake, T. C., Chen, I. C., Clark, T. D., Colwell, R. K., Danielsen, F., Evengård, B., Falconi, L., Ferrier, S., Frusher, S., Garcia, R. A., Griffis, R. B., Hobday, A. J., Janion-Scheepers, C., Jarzyna, M. A., Jennings, S., … Williams, S. E. (2017). Biodiversity redistribution under climate change: Impacts on ecosystems and

human well-being. In *Science* (Vol. 355, Issue 6332). American Association for the Advancement of Science. https://doi.org/10.1126/science.aai9214

Pochon, X., Zaiko, A., Fletcher, L. M., Laroche, O., & Wood, S. A. (2017). Wanted dead or alive? Using metabarcoding of environmental DNA and RNA to distinguish living assemblages for biosecurity applications. *PLoS ONE*, *12*(11), 1–19. https://doi.org/10.1371/journal.pone.0187636

Prasad, A. K. S. K., Nienow, J. A., & Hargraves, P. (2011). Plicate Species of the Diatom Genus Thalassiosira (Bacillariophyta) from the Atlantic and Gulf Coasts of Southeastern United States, with the Description of T. livingstoniorum sp. nov. . *Proceedings of the Academy of Natural Sciences of Philadelphia*, *161*(1), 1–34. https://doi.org/10.1635/053.161.0101

Rees, H. C., Maddison B. C., Middleditch D. J., Patmore J. R. M., & Gough K. C. (2014). REVIEW: The detection of aquatic animal species using environmental DNA–A review of eDNA as a survey tool in ecology. *J Appl Ecol,* 51(5), 1450-1459. https://doi.org/10.1111/1365-2664.12306

Rey, A., Basurko, O. C., & Rodriguez-Ezpeleta, N. (2020). Considerations for metabarcoding-based port biological baseline surveys aimed at marine nonindigenous species monitoring and risk assessments. *Ecology and Evolution,* 10, 2452–2465.

Robidart, J. C., Church, M. J., Ryan, J. P., Ascani, F., Wilson, S. T., Bombar, D., Marin, R., Richards, K. J., Karl, D. M., Scholin, C. A., & Zehr, J. P. (2014). Ecogenomic sensor reveals controls on N2-fixing microorganisms in the North Pacific Ocean. *ISME Journal*, *8*(6), 1175–1185. https://doi.org/10.1038/ismej.2013.244

Robidart, J. C., Magasin, J. D., Shilova, I. N., Turk-Kubo, K. A., Wilson, S. T., Karl, D. M., Scholin, C. A., & Zehr, J. P. (2019). Effects of nutrient enrichment on surface microbial community gene expression in the oligotrophic North Pacific Subtropical Gyre. *ISME Journal*, *13*(2), 374–387. https://doi.org/10.1038/s41396-018-0280-0

Salter, I., Joensen, M., Kristiansen, R., Steingrund P., & Vestergaard P. (2019). Environmental DNA concentrations are correlated with regional biomass of Atlantic cod in oceanic waters. *Commun Biol 2, 461*. https://doi.org/10.1038/s42003-019-0696-8

Satterthwaite, E. V., Bax, N. J., Miloslavich, P., Ratnarajah, L., Canonico, G., Dunn, D., Simmons, S. E., Carini, R. J., Evans, K., Allain, V., Appeltans, W., Batten, S., Benedetti-Cecchi, L., Bernard, A. T. F., Bristol, S., Benson, A., Buttigieg, P. L., Gerhardinger, L. C., Chiba, S., … Zivian, A. (2021). Establishing the Foundation for the Global Observing System for Marine

Life. *Frontiers in Marine Science*, *8*, 1508. https://doi.org/10.3389/FMARS.2021.737416/XML/NLM

Scholin, C., Doucette, G., Jensen, S., Roman, B., Pargett, D., Marin Iii, R., Preston, C., Jones, W., Feldman, J., Everlove, C. and Harris, A. (2009). Remote detection of marine microbes, small invertebrates, harmful algae, and biotoxins using the Environmental Sample Processor (ESP). *Oceanography*, *22*(2), pp.158-167.

Sepulveda, A. J., Hutchins, P. R., Forstchen, M., Mckeefry, M. N., & Swigris, A. M. (2020). The Elephant in the Lab (and Field): Contamination in Aquatic Environmental DNA Studies. *Frontiers in Ecology and Evolution*, *8*, 440. https://doi.org/10.3389/FEVO.2020.609973/BIBTEX

Siegel, D. A., Behrenfeld, M. J., Maritorena, S., McClain, C. R., Antoine, D., Bailey, S. W., Bontempi, P. S., Boss, E. S., Dierssen, H. M., Doney, S. C., Eplee, R. E., Evans, R. H., Feldman, G. C., Fields, E., Franz, B. A., Kuring, N. A., Mengelt, C., Nelson, N. B., Patt, F. S., … Yoder, J. A. (2013). Regional to global assessments of phytoplankton dynamics from the SeaWiFS mission. *Remote Sensing of Environment*, *135*, 77–91. https://doi.org/10.1016/j.rse.2013.03.025

Stefanni, S., Stanković, D., Borme, D., de Olazabal, A., Juretić, T., Pallavicini, A., & Tirelli, V. (2018). Multi-marker metabarcoding approach to study mesozooplankton at basin scale. *Scientific Reports*, *8*(1), 1–13. https://doi.org/10.1038/s41598-018-30157-7

Suter, L., Polanowski, A. M., Clarke, L. J., Kitchener, J. A., & Deagle, B. E. (2021). Capturing open ocean biodiversity: comparing environmental DNA metabarcoding to the continuous plankton recorder. *Molecular Ecology*, *June*, 1–18. https://doi.org/10.1111/mec.15587

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). Environmental DNA for biodiversity research and monitoring. Oxford University Press.

Takahashi, S., Tomita, J., Nishioka, K., Hisada, T., & Nishijima, M. (2014). Development of a prokaryotic universal primer for simultaneous analysis of Bacteria and Archaea using next-generation sequencing. *PLoS ONE*, *9*(8). https://doi.org/10.1371/journal.pone.0105592

Tang, W., Cerdán, E., Berthelot, H., Polyviou, D., Wang, S., Baylay, H., Whitby, H., Planquette, H., Mowlem, M., & Robidart, J. C. (2020). New insights into the distributions of nitrogen fixation and diazotrophs revealed by high-resolution sensing and sampling methods. *ISME Journal*. https://doi.org/10.1038/s41396-020-0703-6

Tang, W., Wang, S., Fonseca-Batista, D., Dehairs, F., Gifford, S., Gonzalez, A. G., Gallinari, M., Planquette, H., Sarthou, G., & Cassar, N. (2019). Revisiting the distribution of oceanic N2 fixation and estimating diazotrophic contribution to marine production. *Nature Communications 2019 10:1*, *10*(1), 1–10. https://doi.org/10.1038/s41467-019-08640-0

Taylor, C.D. & Doherty, K.W. (1990). Submersible Incubation Device (SID), autonomous instrumentation for the in situ measurement of primary production and other microbial rate processes. *Deep Sea Research Part A. Oceanographic Research Papers*, *37*(2), pp.343-358.

Thompson, A. W., Foster, R. A., Krupke, A., Carter, B. J., Musat, N., Vaulot, D., Kuypers, M. M. M., & Zehr, J. P. (2012). Unicellular cyanobacterium symbiotic with a single-celled eukaryotic alga. *Science*, *337*(6101), 1546–1550. https://doi.org/10.1126/SCIENCE.1222700/SUPPL_FILE/THOMPSON.SM.PDF

Wang, S., Tang, W., Delage, E., Gifford, S., Whitby, H., González, A. G., Eveillard, D., Planquette, H., & Cassar, N. (2021). Investigating the microbial ecology of coastal hotspots of marine nitrogen fixation in the western North Atlantic. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-84969-1

Winslow, L.A., Dugan, H.A., Buelow, H.N., Cronin, K.D., Priscu, J.C., Takacs-Vesbach, C. & Doran, P.T. (2014). Autonomous year-round sampling and sensing to explore the physical and biological habitability of permanently ice-covered Antarctic lakes. *Marine Technology Society Journal*, *48*(5), pp.8-17.

Wood, S. A., Biessy, L., Latchford, J. L., Zaiko, A., von Ammon, U., Audrezet, F., Cristescu, M. E., & Pochon, X. (2020). Release and degradation of environmental DNA and RNA in a marine system. *Science of the Total Environment*, *704*, 135314. https://doi.org/10.1016/j.scitotenv.2019.135314

Yamahara, K. M., Preston, C. M., Birch, J., Walz, K., Marin, R., Jensen, S., Pargett, D., Roman, B., Ussler, W., Zhang, Y., Ryan, J., Hobson, B., Kieft, B., Raanan, B., Goodwin, K. D., Chavez, F. P., & Scholin, C. (2019). In situ Autonomous Acquisition and Preservation of Marine Environmental DNA Using an Autonomous Underwater Vehicle. *Frontiers in Marine Science*, *6*(July), 1–14. https://doi.org/10.3389/fmars.2019.00373

Zaiko, A., Pochon, X., Garcia-Vazquez, E., Olenin, S., & Wood, S. A. (2018). Advantages and Limitations of Environmental DNA/RNA Tools for Marine Biosecurity: Management and Surveillance of Non-indigenous Species. *Frontiers in Marine Science*, *5*(September). https://doi.org/10.3389/fmars.2018.00322

Zehr, J. P. (2011). Nitrogen fixation by marine cyanobacteria. In *Trends in Microbiology* (Vol. 19, Issue 4, pp. 162–173). https://doi.org/10.1016/j.tim.2010.12.004

Zhang, Y., Pavlovska, M., Stoica, E., Prekrasna, I., Yang, J., Slobodnik, J., Zhang, X., & Dykyi, E. (2020). Holistic pelagic biodiversity monitoring of the Black Sea via eDNA metabarcoding approach: From bacteria to marine mammals. *Environment International*, *135*(October 2019), 105307. https://doi.org/10.1016/j.envint.2019.105307

Zorz, J., Willis, C., Comeau, A. M., Langille, M. G. I., Johnson, C. L., Li, W. K. W., & LaRoche, J. (2019). Drivers of regional bacterial community structure and diversity in the northwest atlantic ocean. *Frontiers in Microbiology*, *10*(FEB), 1–24. https://doi.org/10.3389/fmicb.2019.00281

Appendix A Supplementary Material

Table 4 Sample volumes, DNA and RNA yields. Vol, Volume; RoCSI, Robotic Cartridge Sampling Instrument; MC, Manually Collect; NAY, Nucleic Acid Yield.

ID	Start Time	Vol (ml)	RoCSI/MC	MetaB Pair ID	NAY Pair ID	DNA (µg/L)	RNA (µg/L)
2500	29/07/2017 07:24	686	RoCSI	Blank_1	-	13.05	Too Low
2800	29/07/2017 09:05	700	MC	Blank_1	-	9.21	Too Low
2508/9	29/07/2017 20:00	4012	RoCSI	a	1	72.66	29.66
2804/5	29/07/2017 20:41	5500	MC	a	1	251.82	24.91
2539/40	31/07/2017 05:14	1941	RoCSI	b	2	9.58	136.27
2823	31/07/2017 06:20	4000	MC	b	2	372.50	25.00
2570	01/08/2017 00:00	800	RoCSI	Blank_2	-	42.19	83.13
2833	01/08/2017 00:00	1000	MC	Blank_2	-	20.85	Too Low
2579/80	01/08/2017 19:40	1827	RoCSI	c	3	645.87	569.24
2840	01/08/2017 20:36	2000	MC	c	3	1505.00	357.50
2852	02/08/2017 12:42	2000	MC	-	4	770.00	740.00
2594/5	02/08/2017 13:00	1864	RoCSI	-	4	751.07	1174.89

2608/9	03/08/2017 05:42	2310	RoCSI	d	5	350.65	1168.83
2872	03/08/2017 06:35	2000	MC	d	5	492.50	267.50
2888	03/08/2017 23:38	1000	MC	e	6	1575.00	114.88
2626	04/08/2017 00:04	1218	RoCSI	e	6	373.15	194.58
26291	04/08/2017 04:32	200	RoCSI	Blank_3	-	132.25	987.50
2892T	04/08/2017 04:10	200	MC	Blank_3	-	51.25	Too Low
2903	04/08/2017 12:45	1500	MC	-	7	1553.33	-
2639/40	04/08/2017 12:56	1195	RoCSI	-	7	723.85	-
2656	05/08/2017 07:13	1268	RoCSI	-	8	-	78.86
2933/4	05/08/2017 08:00	4000	MC	-	8	-	93.63
2659	05/08/2017 10:13	1234	RoCSI	-	9	-	82.60
2941/2	05/08/2017 12:43	4000	MC	-	9	-	99.50
2945	06/08/2017 06:30	1000	MC	f	10	87.00	137.50
2677	06/08/2017 06:49	1257	RoCSI	f	10	76.98	87.01
2951	06/08/2017 20:25	2000	RoCSI	-	11	20.48	60.50
2692	06/08/2017 21:32	1262	MC	-	11	11.49	163.23

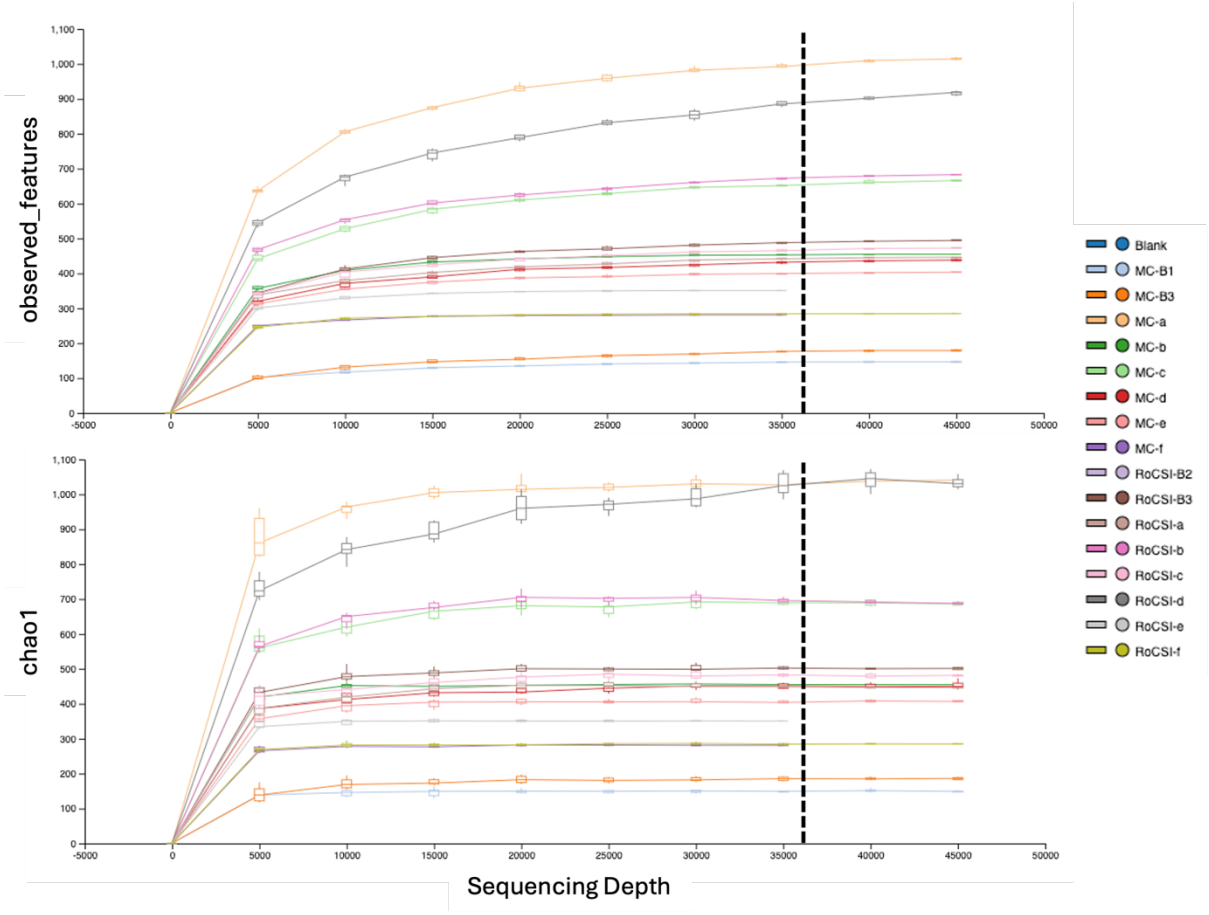


Figure 11 Rarefaction curves from 16S rRNA gene metabarcoding. Samples were rarefied to 36325 sequences as indicated with the dashed line.

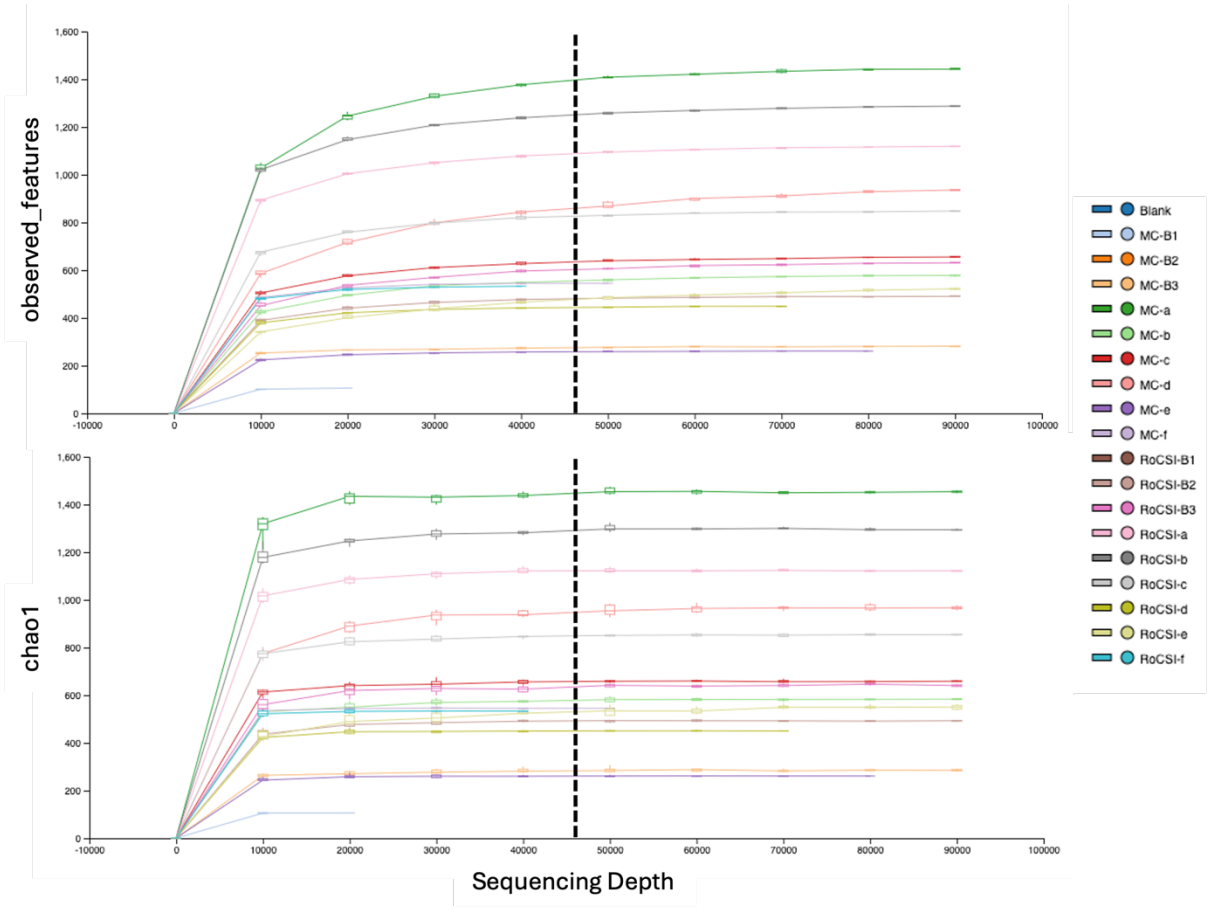


Figure 12 Rarefaction curves from 18S rRNA metabarcoding. Samples were rarefied to 46142 sequences as indicated with the dashed line.

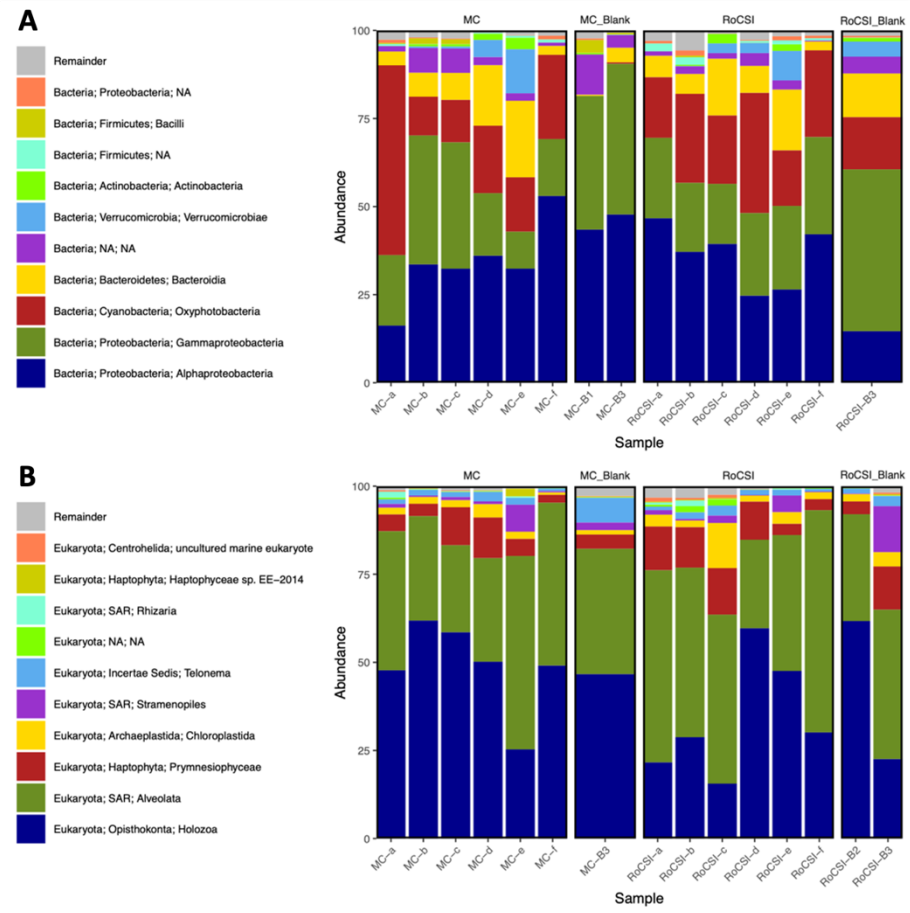


Figure 13 A) Taxa bar plots showing the percentage abundance of taxa at the class level from 16S rRNA gene metabarcoding with data rarefied to 36325 sequences. B) Taxa bar plots showing the percentage abundance of taxa at the class level from 18S rRNA gene metabarcoding with data rarefied to 46142 sequences.

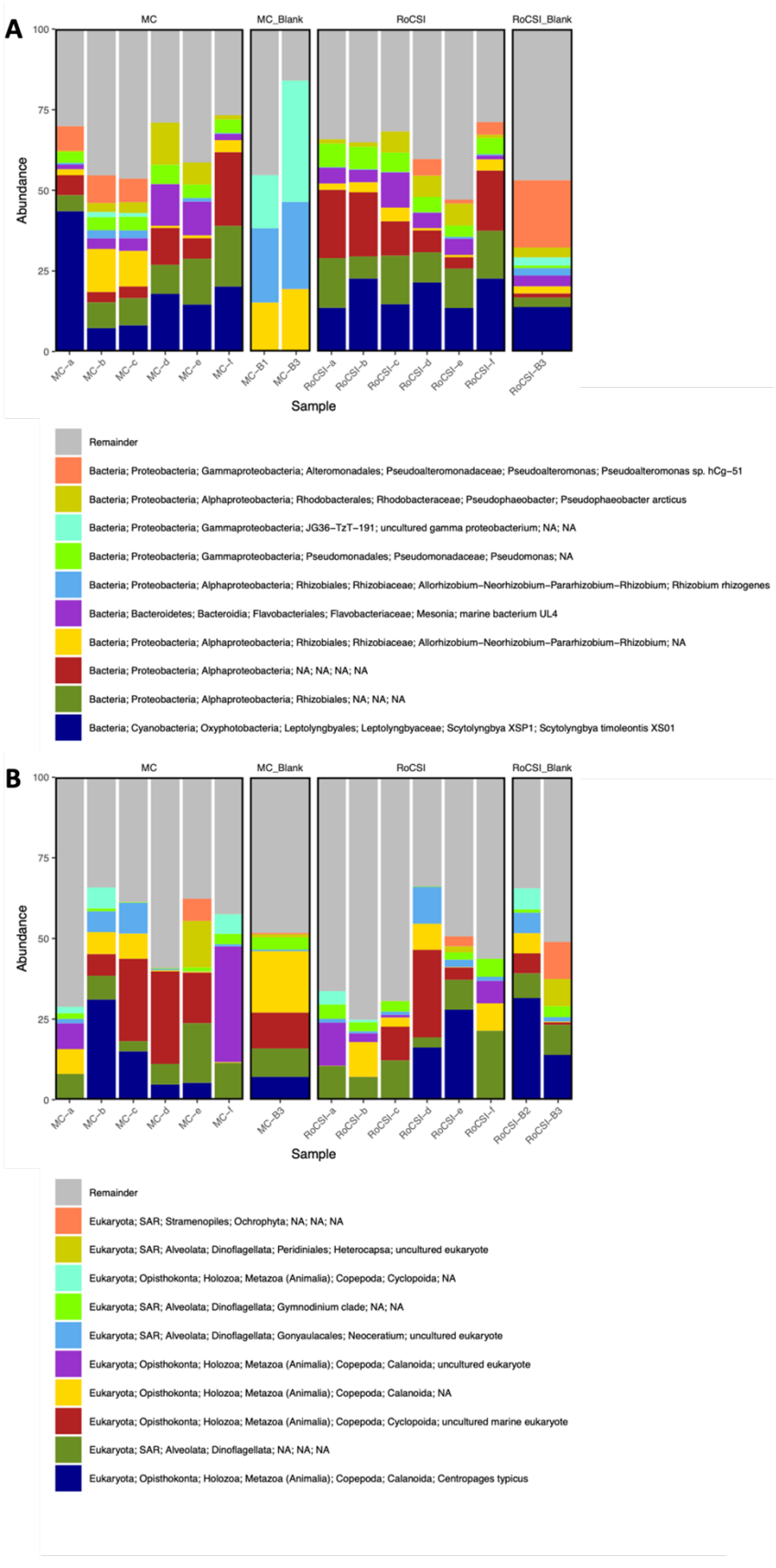


Figure 14 A) Taxa bar plots showing the percentage abundance of taxa at the species level from 16S rRNA gene metabarcoding with data rarefied to 36325 sequences. B) Taxa bar plots showing the percentage abundance of taxa at the species level from 18S rRNA gene metabarcoding with data rarefied to 46142 sequences.

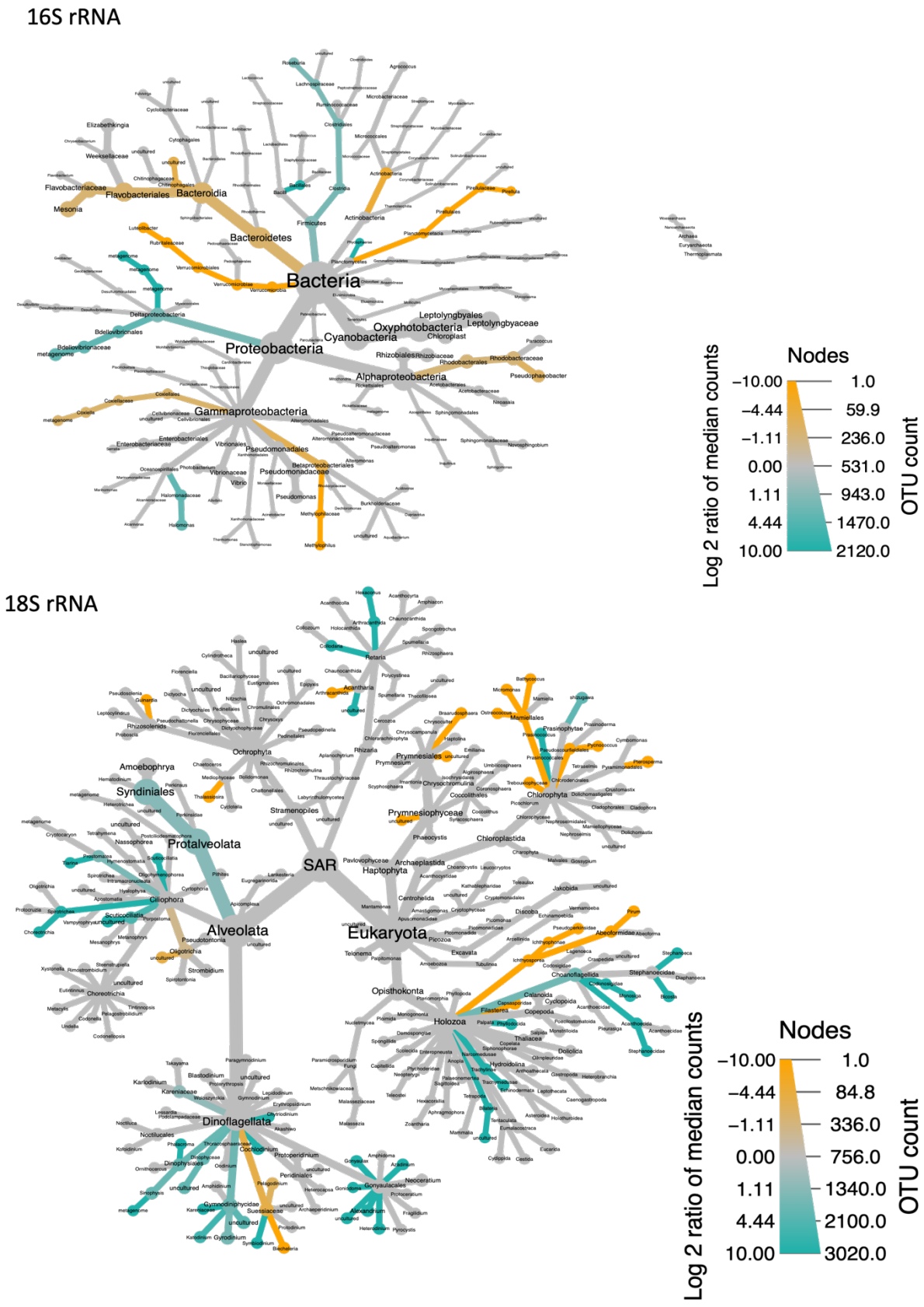


Figure 15 Detailed version of taxonomic heat tree showing the pairwise comparisons of coastal and offshore communities identified through 16S and 18S rRNA gene metabarcoding. The colour of each taxon represents the log-2 ratio of median

proportion of reads observed in each region. Only significant differences are coloured. Taxa coloured blue are enriched in offshore communities and taxa in orange are enriched in coastal communities.

Chapter 3 Diel fluxes in prokaryotic and eukaryotic microbial communities in a highly urbanised tidal estuary

3.1 Abstract

Environmental DNA (eDNA) is increasingly recognised as a key tool for biodiversity monitoring due to its ease of sampling and capacity to detect organisms ranging from microbes to megafauna. Estuaries play a crucial role in life histories of many species, but their tidal and photosynthetic cycles have the potential to influence eDNA results if not properly accounted for in sampling protocols. This study aimed to characterise diel fluctuations in prokaryotic and eukaryotic planktonic communities within a highly urbanised tidal estuary using high-temporal-resolution eDNA time-series data.

Bi-hourly water samples were collected over four days using the Robotic Cartridge Sampling Instrument (RoCSI). Community composition was analysed through 16S and 18S rRNA gene metabarcoding. Both prokaryotic and eukaryotic communities showed significant short-term variability. Prokaryote alpha diversity correlated primarily to barometric pressure, while beta diversity correlated with water temperature and barometric pressure. Eukaryotes alpha and beta diversity correlated with water temperature, but not tidal height. Notably, the hydrocarbonoclastic bacterium *Oleispira* sp., associated with oil pollution, bloomed in the latter part of the study, suggesting its potential as an indicator of oil spills in harbours.

High-temporal-resolution monitoring, such as this, remains costly for long-term studies. However, short-term pilot studies can guide the design of more cost-effective, lower-resolution monitoring. This approach identifies optimal sampling strategies, reducing bias and improving the accuracy of ecological trend assessments. This research demonstrates RoCSI's practicality for high-resolution temporal monitoring and underscores the value in understanding the real-world impacts of anthropogenic activities on marine ecosystems.

3.2 Introduction

Estuaries are diverse habitats providing numerous ecosystem services. They serve as nurseries and feeding grounds for marine life (Gernez et al., 2023), and they play a crucial role in nutrient cycling, improvement of water quality, flood protection, and carbon sequestration (Watson et al. 2020). Despite these benefits, estuarine ecosystems are prime locations for urbanisation and economic development (Barbier, 2017; French 2002). Highly urbanised estuaries, such as the Solent, UK, support a wide range of economic activities such as shipping, fishing, aquaculture, tourism, and renewable energy production as well as bioremediation services

estimated at £1.1 billion (Watson et al., 2020). Economic activities and ecosystem services are intrinsically connected. Therefore, balancing economic and social prosperity with ecological health is vital for resilient and sustainable coastal cities.

Biomonitoring is essential to gauge on-going ecological health, especially in estuarine and marine environments where much of the biodiversity remains hidden beneath the surface. The composition and functionality of microbial planktonic communities are significantly influenced by both urbanisation and fluctuations in physical and chemical conditions (Macé et al., 2024). For example, changes in bacterial community composition have been associated with variations in light intensity (Piwosz et al., 2020), temperature, nutrient availability (Fuhrman et al., 2015), physical mixing, and differing sources of dissolved organic matter (Bruhn et al., 2021; Fuhrman et al., 2015; Piwosz et al., 2020; Zhao et al., 2023). In urbanised estuaries, the planktonic community composition can also be influenced by increased contaminants associated with stormwater drainage or discrete pollution events (Kottuparambil et al., 2023; Macé et al., 2024; Varkey et al., 2018). Temporal monitoring of the ecological impacts from anthropogenic stressors is essential to provide evidence to inform adaptive management strategies. Such evidence is required to balance the ecological value of estuaries with their economic and social value, to ensure their long-term sustainability.

A key challenge in biomonitoring is determining the optimal sampling strategy. High resolution sampling frequency from weekly, to daily, or even hourly intervals enhances our ability to interpret data for long-term trend analysis (Fischer et al., 2021). Higher resolutions offer a more detailed understanding of fine-scale temporal dynamics within an estuary and provide valuable insights into the short-term effects of specific events. In contrast, lower-frequency sampling, such as monthly, seasonal, or annual, is less effective at detecting significant trends because causative factors and their impacts are typically decoupled at these frequencies (Fischer et al., 2021). Understanding temporal trends in ecological health, including declines in key species or the introduction of invasive species, is crucial for informed management decisions. However, obtaining high resolution temporal monitoring, needed to detect these trends, is primarily constrained by the financial cost of fieldwork.

Photosynthesis, providing carbon for food webs in surface oceans, changes on day-to-night cycles. Similarly, tidal change, the dominant physical driver in estuarine environments (Nascimento et al. 2021), has a diel periodicity. Sampling every two hours offers sufficient resolution to observe the influence of diel fluctuations, helping to identify these primary drivers of ecological change. However, sampling using manual collection and visual identification

methods is impractical at these frequencies, requiring out of hours working and morphological taxonomy expertise. Autonomous biomolecular sampling with the Robotic Cartridge Sampling Instrument (RoCSI) can facilitate this scale of high-resolution sampling. With the RoCSI, water samples for biomolecular analysis can be filtered and preserved autonomously. The mission parameters, including sampling frequency, can be set at the start and filter cartridges can be exchanged at convenient daily, weekly or monthly intervals depending on the sampling frequency. Samples can be processed and sequenced in batches allowing for a relatively quick turnaround between sample collection and data output.

This study explores use of the RoCSI for autonomous high-resolution biomolecular sampling. By collecting water samples every two hours over four days, we investigated the fine-scale fluctuations in microbial and eukaryotic planktonic communities and correlated these with environmental variables routinely collected at the port, such as water temperature, tide height, and meteorological variables.

3.3 Methods

Over a four-day period from November 11th to 15th, 2019, time series data were collected outside the National Oceanography Centre (NOC) in the Empress Dock of the Solent Estuary. Water samples were taken to analyse the prokaryotic and eukaryotic community compositions using the RoCSI prototype. Additional environmental data, including air temperature, wind speed, barometric pressure, and tide height, were acquired from the Southampton Dockhead Weather Station (50°53'00.6"N 1°23'39.6"W), located approximately 400 meters due south of the water sampling site. Water temperature data from alongside the sampling station were obtained from a temperature sensor mounted on the hull of the RV Calista, a research vessel operated by the University of Southampton and docked at the NOC.

3.3.1 Sample Collection

The RoCSI prototype was mounted in a waterproof housing above an outside seawater tank at the NOC, Southampton, UK (Figure 16). The RoCSI was configured to filter and preserve 500ml samples of seawater every two hours. Seawater was filtered through a 0.22µm Sterivex™ (Merck Millipore), with RLT+ buffer (Qiagen) added to the cartridge immediately after filtration. Seawater supply to the tank was approximately 200ml/second from 1m depth at the National Oceanography Centre pontoon, a 5mm net pre-filter prevented larger detritus from entering. The tank contained approximate 800L of seawater, with a residence time of approximately 1

hour 7 minutes. Sample collection began on 2019-11-11 at 12:26 and ended on 2019-11-15 at 08:26. The cartridges were removed from the device every 24 hour and stored at -80°C, except for the final sample 56 which remained in the device for 2 weeks prior to freezing at -80°C. A blank sample (500ml of molecular grade water) was taken with the RoCSI prior to sampling.

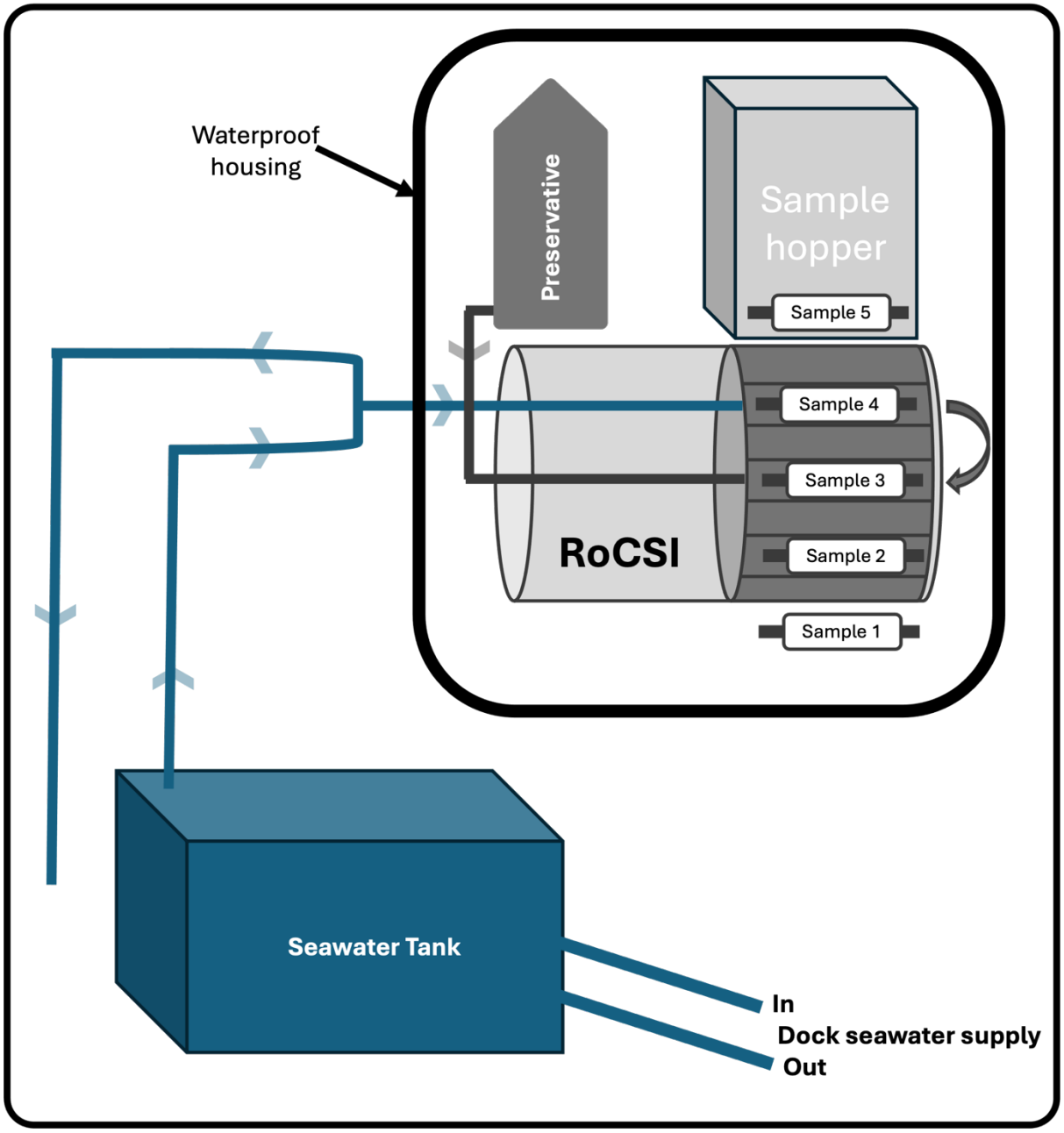


Figure 16 RoCSI dockside configuration, with seawater supply from a dockside seawater tank.

3.3.2 Nucleic acid purification

DNA was purified using a modified protocol for the Qiagen AllPrep DNA/RNA Mini Kit, (For full protocol see Appendix B.1). DNA yields were measured using Qubit High Sensitivity DNA Assay Kits.

3.3.3 Metabarcoding

The 16S rRNA gene barcode region was amplified using 515F-Y (5'-GTGYCAGCMGCCGCGGTAA) forward and 926R (5'-CCGYCAATTYMTTTRAGTTT) reverse primers (Parada et al., 2016). The 18S rRNA gene barcode region was amplified using 1391F (5'-GTACACACCGCCCGTC) forward and EukBr (5'-TGATCCTTCTGCAGGTTCACCTAC) reverse primers (Amaral-Zettler et al. 2009). Primers were modified with overhang sequences and a second, limited cycle PCR was used to add barcodes and sequencing adaptors, as specified in the Illumina sequencing library preparation protocol (Illumina, 2013). PCR indexing, rt-PCR and sequencing was completed at Exeter Sequencing Facility (For full protocol see Appendix B.1).

3.3.4 Bioinformatic Processing

The raw sequencing data was initially processed by removing adaptor sequences and sequences less than 250 bp (raw sequence data available at European Nucleotide Archive (accession number: PRJEB83250; Martin 2011). Paired-end joining, denoising and taxonomic assignment was performed using QIIME2 release 2022.2 (Bolyen et al., 2019). In brief, denoising and pair-end joining were performed using DADA2 to produce amplicon sequence variants (ASVs; Callahan et al., 2016). Phylogenetic trees were built using fasttree and mafft alignment (Bolyen et al. 2019; Katoh & Standley, 2013; Price et al., 2010). Taxonomic classification was performed with Naïve Bayes classifier and classify-sklearn method with the SILVA version 138.1 database, clustered at 99% sequence similarity as a reference (Bokulich et al., 2018; Bolyen et al., 2019; Gurevich et al., 2013; Pedregosa et al., 2011; Pruesse et al., 2007; Rognes et al., 2016). For full scripts see GitHub rms1u18/RoCSI_Soton.

3.3.5 Data Analysis

Data analyses were performed using QIIME2 version 2022.2 and R version 4.3.2 (2023-10-31). Alpha diversity metrics were derived from rarefied sequence data in QIIME2. Sequences were rarefied to 61,400 for prokaryotes and 16,744 for eukaryotes (see Appendix B.2 for rarefaction

curves). Data from QIIME2 was read into R using the qiime2r package version 0.99.6 (Bisanz, 2018).

Alpha diversity and environmental time series bar plots were created using the lubridate v1.9.3, xts v0.12.2, ggplot2 v3.4.4, and ggpubr v0.6.0 packages (Grolemund & Wickham 2011; Ryan & Ulrich 2024; Wickham 2016; Kassambra 2023). Correlations between alpha diversity and environmental variables were calculated using Spearmans rank correlation in QIIME2 (Bolyen et al., 2019). Correlations between beta diversity and environmental variables were calculated in R using Mantel tests with Spearman's rank correlation between Bray-Curtis dissimilarity matrices for ASV abundance and Euclidean distance matrices for environmental parameters with the vegan package version 2.6.4 package (Oksanen et al., 2024).

Non-metric multidimensional scaling (nMDS) analysis was based on Bray-Curtis dissimilarities from Wisconsin double transformed data. Significantly correlated environmental parameters were plotted using vector lines on each nMDS axis. The nMDS plots and stacked line plots for the ten species with highest rarefied sequence counts across the dataset were visualised using the vegan version 2.6.4, ggplot2 version 3.4.4, viridis version 0.6.4 and ggpubr version 0.6.0 packages (Oksanen et al. 2022; Wickham 2016; Garnier et al. 2023; Kassambra 2023).

Correlations between the ten most abundant Prokaryotic/Eukaryotic species and environmental variables were calculated using Spearman's rank correlations and tabulated using sjPlot v2.8.15 package (Lüdecke et al. 2023). For full scripts see GitHub rms1u18/RoCSI_Soton.

3.4 Results

During the mission, the target was for the RoCSI to collect 56 samples. Out of these, 39 samples were successfully processed, each acquiring the intended seawater volume of 500ml and subsequent addition of RLT+ (Qiagen) preservative. However, sample number 38 only reached 486ml before the maximum pressure limit was exceeded; despite this, the preservative was successfully added, and the sample was considered valid for final analysis. An issue with the pressure sensor resulted in the next seven samples (nos. 39-45) only collecting 9ml of filtered seawater, leading to their exclusion from further analysis. To circumvent the sensor issue and continue the project, a manual override was implemented for the overpressure limit for samples 46 through 56. This adjustment ensured each of these samples successfully achieved the intended 500ml volume.

Samples were removed from the RoCSI once a day and stored at -80°C. However, the final sample (no. 56) remained in the device for two weeks before being stored at -80°C, due to a water leak that prevented RoCSI operations. Previous investigation has revealed that the RLT+ (Qiagen) buffer can preserve eDNA for up to 1 month (Chapter 4), therefore sample no. 56 was included in the final analysis.

Metabarcoding of the 16S rRNA gene targeting prokaryotes initially identified 6,049,413 sequences. After bioinformatic processing, 3,712,438 sequences (61.4%) were retained. From these, 132 taxonomic assignments were Archaea, 3,834 were Bacteria, and 26 were Eukaryotes. Sequences attributed to Eukaryotes were excluded from the subsequent analysis of the 16S rRNA gene metabarcoding data. Metabarcoding of the 18S rRNA gene targeting Eukaryotes, initially yielded 1,274,266 sequences, with 1,114,251 (87.4%) remaining post bioinformatic processing. These resulted in 658 taxonomic assignments to Eukaryotes, with an incidental capture of 20 sequences assigned to Bacteria and 2 to Archaea. These non-target sequences were removed from downstream analysis of the 18S rRNA metabarcoding data.

Water temperature in the Solent declined with the rapid transitions over the four days, while air temperature, barometric pressure and winds fluctuated independently (Figure 17 A). Prokaryotic and Eukaryotic diversity also varied across the study (Figure 17 B & C). For prokaryotes, the average number of ASVs was 1,332 (range: 1,018 - 1,738), compared to 467 for eukaryotes (range: 254 - 602). The average Shannon diversity index was 6.9 for prokaryotes (range: 6.4 - 7.1) and 6.5 for eukaryotes, which showed greater variability (range: 2.1 - 7.3). Pielou's evenness averaged 0.66 for prokaryotes (range: 0.64 - 0.69), and 0.73 for eukaryotes, again with greater variability (range: 0.26 - 0.81). Shannon's diversity and Pielou's evenness are plotted on identical axes in Figure 17 B and C to highlight this contrast in variability. Faith's phylogenetic diversity was 260 for prokaryotes (range: 165 - 361) and 66.8 for eukaryotes (range: 43.9 - 99.0).

Spearman's rank correlation revealed significant correlations between environmental factors and alpha diversity measures. For prokaryotes, Pielou's evenness was negatively correlated with wind speed (r_s = -0.352, p = 0.028). Barometric pressure was also negatively correlated with prokaryotic ASVs (r_s = -0.318, p = 0.048), Shannon's diversity (r_s = -0.357, p = 0.026), and Faith's phylogenetic diversity (r_s = -0.459, p = 0.003). For eukaryotes, ASVs increased with the elapsed time since sampling began (r_s = 0.349, p = 0.030). Water temperature was negatively correlated with eukaryotic ASVs (r_s = -0.544, p < 0.001), Shannon's diversity (r_s = -0.327, p =

0.043), and Faith's phylogenetic diversity (r_s = -0.325, p = 0.044). Similarly, air temperature was negatively correlated with eukaryotic ASVs (r_s = -0.361, p = 0.024; see Table 5).

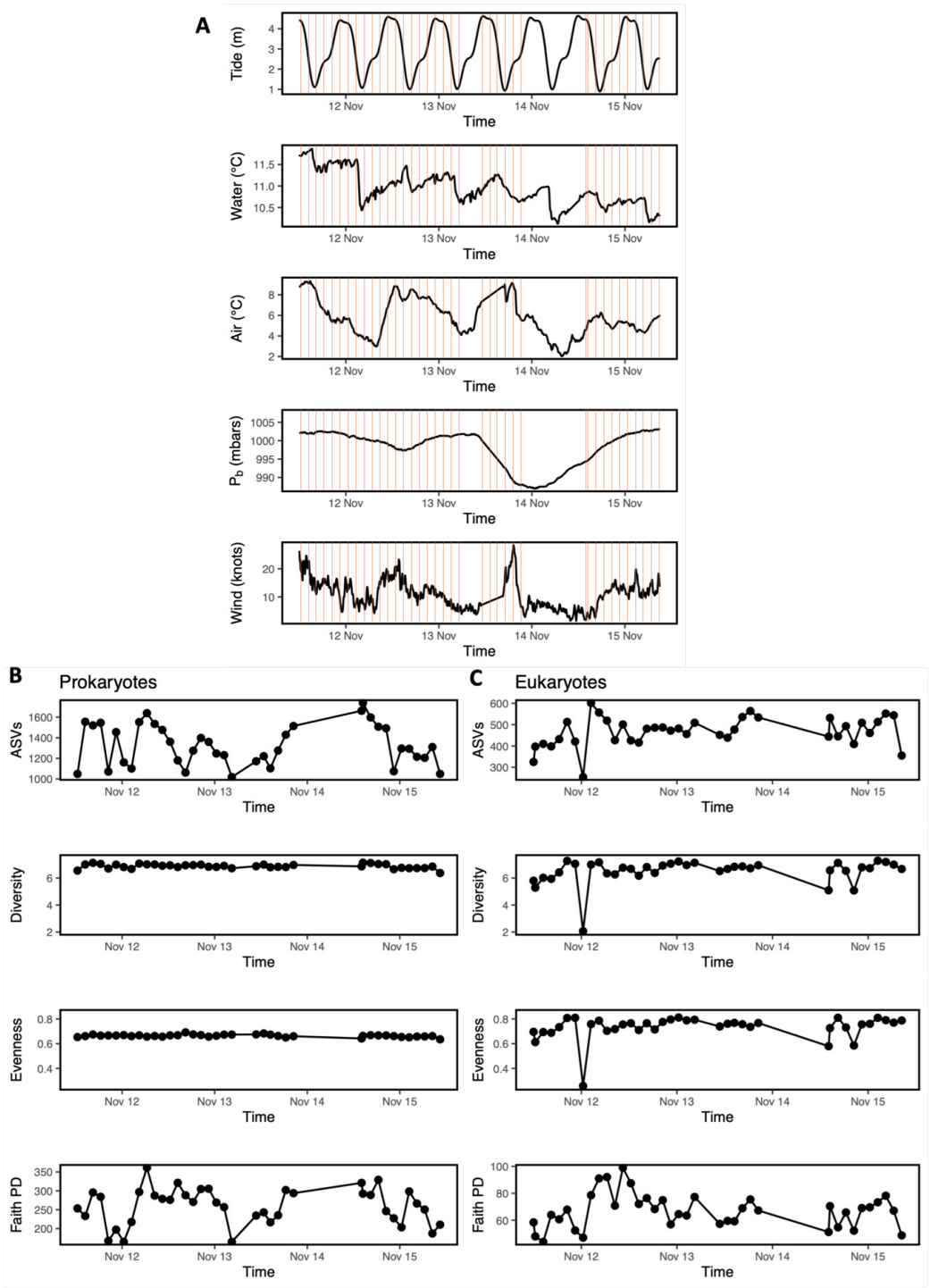


Figure 17 **A** Environmental time series, with all data plotted using 15-minute rolling averages. **B** Time series for prokaryotic alpha diversity measures. **C** Time series of Eukaryotic alpha diversity. Abbreviations: Tide (m) – Tide height (m); Water (°C) - Water

temperature (°C), Air (°C) - Air temperature (°C); P_b (mbars) – Barometric pressure (millibars); Wind (knots) – Wind speed (knots); ASVs – Number of amplicon sequence variants; Diversity – Shannon's diversity; Evenness – Pielou's evenness; and Faith PD – Faith's phylogenetic diversity.

Table 5 Correlation between alpha/beta diversity and environmental variables with correlation coefficient r values above and significant p-values in brackets below. Statistically significant results in bold. Alpha diversity measures (including ASVs, Shannon's Diversity, Pielou's Evenness, and Faith PD) were calculated using Spearman's rank correlation. Correlations between beta diversity and environmental variables were calculated using Mantel tests with Spearman's rank correlation between Bray-Curtis dissimilarity matrices for ASV abundance and Euclidean distance matrices for environmental parameters.

	Time since start	Tide height	Water temp	Air temp	Wind speed	Barometric pressure
<u>Prokaryotes</u>						
ASVs	-0.075 (0.65)	-0.072 (0.66)	-0.132 (0.42)	-0.172 (0.30)	0.017 (0.92)	**-0.318 (0.048)**
Shannon's Diversity	-0.233 (0.15)	-0.219 (0.18)	0.008 (0.96)	0.023 (0.89)	-0.132 (0.42)	**-0.357 (0.026)**
Pielou's Evenness	-0.274 (0.09)	-0.221 (0.18)	0.28 (0.09)	0.262 (0.11)	**-0.352 (0.028)**	-0.182 (0.27)
Faith PD	-0.025 (0.88)	-0.15 (0.36)	-0.155 (0.34)	0.015 (0.93)	0.094 (0.57)	**-0.459 (0.003)**
Beta Diversity	**0.5896 (<.001)**	0.01872 (0.30)	**0.3094 (<.001)**	0.02901 (0.28)	0.06955 (0.16)	**0.2172 (0.005)**
<u>Eukaryotes</u>						
ASVs	**0.349 (0.030)**	-0.079 (0.64)	**-0.544 (<.001)**	**-0.361 (0.024)**	-0.166 (0.31)	-0.22 (0.18)
	0.278	-0.02	**-0.327**	-0.249	-0.22	0.108

Shannon's Diversity	(0.09)	(0.90)	**(0.043)**	(0.13)	(0.18)	(0.51)
Pielou's Evenness	0.305	0.046	-0.246	-0.151	-0.202	0.157
	(0.06)	(0.78)	(0.13)	(0.36)	(0.22)	(0.34)
Faith PD	0.048	-0.001	**-0.325**	-0.151	0.097	-0.232
	(0.77)	(1.00)	**(0.044)**	(0.36)	(0.56)	(0.16)
Beta Diversity	0.151	-0.01736	**0.2309**	0.02534	0.132	0.1275
	(0.01)	(0.61)	**(0.003)**	(0.33)	(0.08)	(0.09)

Prokaryotic beta diversity was positively correlated with the time elapsed since the start of sampling (r_s = 0.59, p < 0.001) and with water temperature (r_s = 0.309, p < 0.001). Additionally, prokaryotic beta diversity was positively correlated with barometric pressure (r_s = 0.217, p = 0.005; see Table 5; Figure 18A). Eukaryotic beta diversity was positively correlated with water temperature (r_s = 0.231, p = 0.003; see Table 5; Figure 18B).

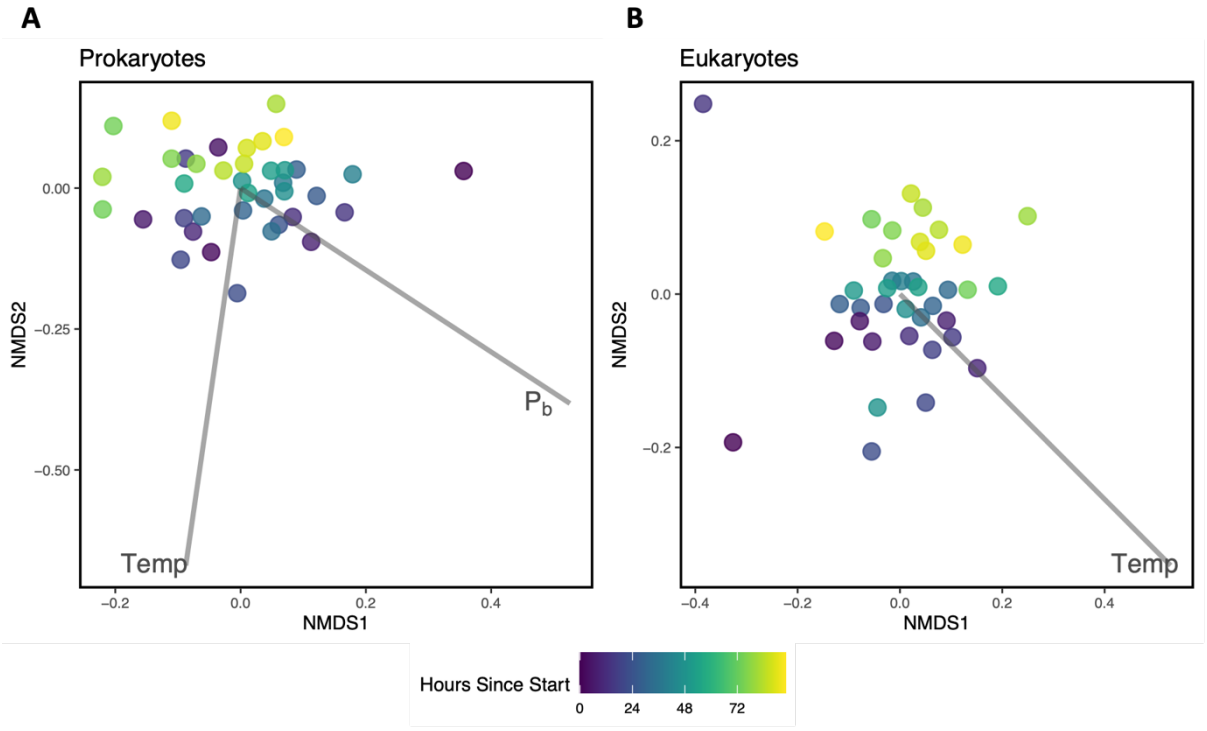


Figure 18 Non-metric multidimensional scaling (nMDS) based on Bray-Curtis dissimilarities of Wisconsin double transformed data. Each circle represents one sample, circle colour scale based on the number of hours since the start of sampling, and vector lines for environmental parameters weighted by their p values with each NMDS

axis. Only significantly correlated environmental parameter were plotted. **A** Prokaryotic community composition with stress of 0.189. Vectors lines for water temperature (Temp) and barometric pressure (Pb). **B** Eukaryotic community composition with stress of 0.187. Vectors line for water temperature (Temp).

Figure 19A highlights the ten prokaryotic taxa with greatest sequence abundance across the dataset. Among these, *Alphaproteobacteria SAR11 Clade Ia* (also known as *Pelagibacteraceae*), *Cryomorphaceae NS3a marine group*, and *Candidatus Puniceispirillum*, all displayed highly significant negative correlations with time since the start of sampling (r_s = -0.512, p = 0.001; r_s = -0.657, p < 0.001; r_s = -0.655, p < 0.001; r_s = -0.532, p < 0.001, respectively). These taxa also showed significant positive correlations with water temperature (r_s = 0.371, p = 0.021; r_s = 0.456, p = 0.004; r_s = 0.472, p = 0.003; r_s = 0.338, p = 0.035, respectively). *Planktomarina* was correlated positively with wind speed (r_s = 0.367, p = 0.021) and barometric pressure (r_s = 0.456, p = 0.003). The *Thioglobaceae SUP05 cluster* was negatively correlated with time since the start of sampling (r_s = -0.411, p = 0.01) and positively correlated with wind speed (r_s = 0.386, p = 0.015) and barometric pressure (r_s = 0.437, p = 0.005). The *Flavobacteriaceae NS5 marine group* did not exhibit significant correlations with any environmental variable. *Oleispira* showed a highly significant positive correlation with time since the start (r_s = 0.695, p < 0.001) and highly significant negative correlation with barometric pressure (r_s = 0.495, p = 0.001), water temperature (r_s = -0.486, p = 0.002), and wind speed (r_s = -0.348, p = 0.03). The *Methylophilaceae OM43 clade* was negatively correlated with time since the start of sampling (r_s = -0.362, p = 0.024). Lastly, *Glaciecola* had a highly significant negative correlation with time since the start (r_s = -0.496, p = 0.001) and a significant positive correlation with barometric pressure (r_s = 0.333, p = 0.038). See Table 6 for the full list of correlations.

Figure 19B illustrates the ten eukaryotic taxa with the highest sequence counts across the dataset. Among these, *Capitellida* was positively correlated with water temperature (r_s = 0.353, p = 0.28). *Cryptophyta sp*. was negatively correlated with water temperature (r_s = -0.455, p = 0.004). *Calanoida* was highly significantly positively correlated with water temperature (r_s = 0.52, p = 0.001), positively correlated with tide height (r_s = 0.372, p = 0.02), and negatively correlated with time since the start of sampling (r_s = -0.339, p = 0.035). The remaining top ten taxonomic assignments did not show any significant correlations with environmental variables (See Table 7 for a full list of correlations).

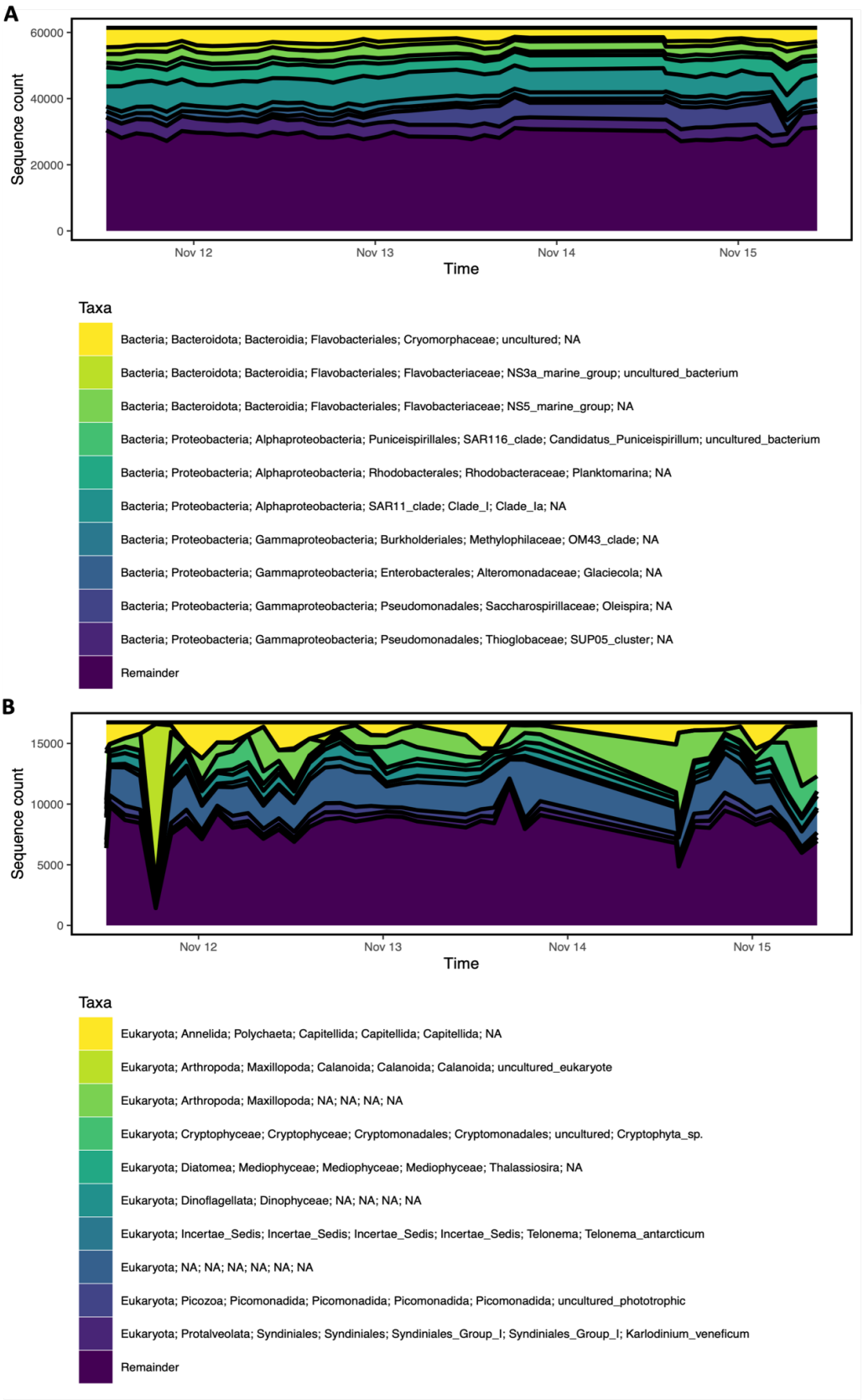


Figure 19 Stacked line plots with rarefied sequences count for the 10 most abundant species across the time series dataset. **A** Prokaryotic species rarefied to 61400. **B** Eukaryotic species rarefied to 16744.

Table 6 Correlations between the 10 most abundant Prokaryotic species and environmental variables with correlation coefficient r values above and significant p-values in brackets below. Statistically significant results in bold.

	Time since start	Tide height	Water temp	Air temp	Wind speed	Barometric pressure
Bacteria; Proteobacteria; Alphaproteobacteria; SAR11_clade;Clade_I; Clade_Ia; NA	**-0.512** **(0.001)**	-0.018 (0.91)	**0.371** **(0.021)**	0.223 (0.17)	-0.069 (0.67)	0.054 (0.74)
Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae; Planktomarina; NA	-0.004 (0.98)	-0.109 (0.51)	-0.013 (0.94)	-0.041 (0.81)	**0.367** **(0.021)**	**0.456** **(0.003)**
Bacteria; Bacteroidota; Bacteroidia; Flavobacteriales; Cryomorphaceae; uncultured; NA	**-0.657** **(<.001)**	-0.29 (0.07)	**0.456** **(0.004)**	0.091 (0.58)	0.253 (0.12)	0.282 (0.08)
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Thioglobaceae; SUP05_cluster; NA	**-0.411** **(0.01)**	0.226 (0.17)	0.244 (0.14)	0.059 (0.72)	**0.386** **(0.015)**	**0.437** **(0.005)**
Bacteria; Bacteroidota; Bacteroidia; Flavobacteriales; Flavobacteriaceae; NS5_marine_group; NA	-0.306 (0.06)	0.062 (0.71)	0.187 (0.25)	-0.129 (0.44)	-0.13 (0.43)	0.201 (0.22)
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Saccharospirillaceae; Oleispira; NA	**0.695** **(<.001)**	0.031 (0.85)	**-0.486** **(0.002)**	-0.125 (0.45)	**-0.348** **(0.030)**	**-0.495** **(0.001)**
Bacteria; Proteobacteria; Gammaproteobacteria; Burkholderiales; Methylophilaceae; OM43_clade; NA	**-0.362** **(0.024)**	0.184 (0.26)	0.177 (0.28)	-0.027 (0.87)	-0.054 (0.74)	0.038 (0.82)

Bacteria; Bacteroidota; Bacteroidia; Flavobacteriales; Flavobacteriaceae; NS3a_marine_group; uncultured_bacterium	**-0.655** (**<.001**)	-0.3 (0.06)	**0.472** (**0.003**)	0.01 (0.95)	0.12 (0.47)	0.15 (0.36)
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacterales; Alteromonadaceae; Glaciecola; NA	**-0.496** (**0.001**)	-0.161 (0.33)	0.272 (0.09)	-0.081 (0.62)	0.279 (0.09)	**0.333** (**0.038**)
Bacteria; Proteobacteria; Alphaproteobacteria; Puniceispirillales; SAR116_clade; Candidatus_Puniceispirillum	**-0.532** (**<.001**)	0.112 (0.50)	**0.338** (**0.035**)	-0.146 (0.38)	0.131 (0.43)	0.274 (0.09)

Table 7 Correlations between the 10 most abundant Eukaryotic species and environmental variables with correlation coefficient r values above and significant p-values in brackets below. Statistically significant results in bold.

	Time since start	Tide height	Water temp	Air temp	Wind speed	Barometric pressure
Eukaryota; NA	-0.05 (0.76)	0.14 (0.39)	0.159 (0.33)	0.118 (0.47)	0.021 (0.90)	-0.247 (0.13)
Eukaryota; Arthropoda; Maxillopoda; NA	0.168 (0.31)	0.185 (0.26)	-0.155 (0.34)	0.046 (0.78)	-0.001 (1.00)	0.033 (0.84)
Eukaryota; Annelida; Polychaeta; Capitellida; NA	-0.314 (0.05)	-0.078 (0.64)	**0.353** (**0.028**)	0.133 (0.42)	0.028 (0.87)	-0.121 (0.46)
Eukaryota; Dinoflagellata; Dinophyceae; NA	-0.195	-0.088	0.115	0.08	0.078	0.294

	(0.23)	(0.59)	(0.48)	(0.63)	(0.64)	(0.07)
Eukaryota; Cryptophyceae; Cryptomonadales; Cryptophyta_sp.	0.309	0.025	**-0.455**	-0.288	-0.043	0.041
	(0.06)	(0.88)	**(0.004)**	(0.08)	(0.80)	(0.80)
Eukaryota; Picozoa; Picomonadida; uncultured_phototrophic	-0.028	0.069	0.111	0.03	0.036	-0.095
	(0.86)	(0.68)	(0.50)	(0.86)	(0.83)	(0.56)
Eukaryota; Diatomea; Mediophyceae; Thalassiosira; NA	0.209	0.215	-0.121	-0.093	-0.107	-0.132
	(0.20)	(0.19)	(0.46)	(0.57)	(0.52)	(0.42)
Eukaryota; Protalveolata; Syndiniales_Group_I; Karlodinium_veneficum	0.225	0.113	-0.008	0.17	-0.22	-0.313
	(0.17)	(0.49)	(0.96)	(0.30)	(0.18)	(0.05)
Eukaryota; Arthropoda; Maxillopoda; Calanoida; uncultured_eukaryote	**-0.339**	**0.372**	**0.52**	0.262	-0.082	0.079
	(0.035)	**(0.020)**	**(0.001)**	(0.11)	(0.62)	(0.63)
Eukaryota; Incertae_Sedis; Telonema_antarcticum	0.031	0.238	0.053	-0.039	-0.144	-0.105
	(0.85)	(0.14)	(0.75)	(0.81)	(0.38)	(0.53)

3.5 Discussion

The RoCSI operated continuously and unattended overnight, enabling the safe sample collection during less favourable working hours. After installation, a single operator could easily program the system and load new cartridges. The RoCSI reliably injected preservative to all samples, ensuring DNA stability for up to two months (Chapter 4). This preservation kept samples viable for analysis even when transfers to the long-term freezer storage were delayed.

Previous high-resolution biomolecular time-series studies with daily sampling have identified cohesive, short-lived coastal plankton communities (Martin-Platero et al., 2018). In the highly dynamic, macrotidal Solent estuary, we hypothesised that the signal from daily eDNA sampling could be influenced by tidal mixing of riverine waters, as has been observed with water quality measures (Nascimento et al. 2021). Such mixing could confound short-lived communities with Lagrangian cohesive communities associated with tidal water movement. However, in this four-day neap-tide time-series, we observed no significant correlation between plankton communities and tidal height.

These results suggest that long-term biomolecular time-series with daily sampling at a fixed time of day in a macrotidal estuary are unlikely to be biased by the stage of the tidal cycle during neap tides. However, further research involving high-resolution (bi-hourly) sampling over a full lunar cycle, encompassing both spring and moderate tides, is needed to determine whether this finding holds true during other tidal phases.

3.5.1 RoCSI Operations

The RoCSI prototype used in this study was designed solely for demonstration purposes and later retrofitted into a waterproof housing for this mission. Consequently, two issues associated with the retrofitted waterproof housing arose during the time-series. The pressure sensor, used to prevent clogging of filter membranes, malfunctioned resulting in six samples being excluded from the analysis. The fault was identified during the daily sample exchange, and the maximum pressure limit was manually overridden to continue the project. Exceeding the 800 mbar pressure threshold may have caused lysis of some microbial cells. However, the impact of cell lysis during metabarcoding sample collection remains unclear. Notably, consistent metabarcoding results have been reported from citizen science eDNA sampling, where filtration pressures will vary between participants (Tøttrup et al. 2021). This suggests that the effect of “overpressure” samples on metabarcoding outcomes is likely to be minimal. However, further

testing is needed to determine the maximum pressure the RoCSI system can withstand without causing damage to its tubing and pump components.

Additionally, water infiltrated the retrofitted waterproof electronic housing at the end of the mission, causing damage to the circuit board and memory card. Consequently, the final sample could not be removed from the preservative engagement position, as the fault prevented operations. In contrast, the commercially available RoCSI model has a submersible electronics housing and has been pressure-tested to 6000 meters to avoid such leaks. This incident highlighted the need for RoCSI to incorporate either a setting that would rotate the Geneva wheel after the preservative has been added to the final sample, or a mechanical fix enabling the user to manipulate the Geneva wheel manually. Implementing this fail-safe would allow for the retrieval of all samples independently of laptop connectivity, a significant advantage in remote or difficult operating environments. However, when using a bandolier to load cartridges with this setting, it may be necessary to insert one or two dummy cartridges after the final sample to ensure that the bandolier remains in place after the final sample has been released.

Due to the RoCSI malfunction at the end of the mission it was not possible to retrieve a final blank sample from the device. A final blank sample is essential for characterising potential contamination between samples, which can occur from residual water left in the lines between samplings (approximately 10 ml). This type of contamination has been observed in previous deployments of the RoCSI prototype (Chapter 2). Consequently, the commercially available RoCSI now includes an optional bleach flushing step specifically designed to mitigate such contamination.

3.5.2 Drivers of biodiversity in high-resolution

An initial exploration of the dataset indicated that barometric pressure, although consistently low throughout the time series, was significantly correlated with both alpha and beta diversity of marine prokaryotes. This finding contrasts with other research in coastal microbial communities such as (Trombetta et al., 2022), which reported no significant correlations between air pressure and alpha or beta diversity in coastal microbial communities. Wind speed was significantly negatively correlated with Pielou’s evenness among prokaryotes, with recorded speeds ranging from light air to near gale on the Beaufort Wind Scale (1 – 31 knots). This is supported by previous observations of decreased Pielou's evenness in coastal microbial communities after hurricane events in the North Atlantic (Garrison et al., 2022). However, associations between meteorological measures and marine ecology are less frequently explored compared to water environmental variables, such as salinity, chlorophyll, and nutrient concentrations.

Eukaryotic alpha and beta diversity measures showed significant negative correlations with water temperature. While temperature is widely recognised as an important driver of estuarine communities (Lo et al., 2023; Nakane et al., 2008; Trombetta et al., 2022), the direction and magnitude of its effect vary by region, time of year, and taxa involved. In macrotidal estuarine environments such as the Solent, temperature covaries with salinity, turbidity and dissolved oxygen due to mixing between freshwater and seawater (Gomez-Castillo et al., 2023; Onabule et al., 2020). Therefore, the fluctuations in alpha and beta diversity may be impacted by a range of independent or autocorrelated variables not measured in this study (Downie et al. 2024).

Differences in how environmental variables influence prokaryotic and eukaryotic alpha and beta diversity may stem from difference in growth rates between the two groups. Prokaryotes typically grow faster than eukaryotes (Lynch et al., 2022; Zubkov, 2014), allowing them to respond more rapidly to short-term environmental fluctuations and making shifts in abundance easier to detect. In contrast, eukaryotic populations adjust more slowly, so physical mixing processes may contribute more to their observed dynamics. Moreover, such broad taxonomic divisions likely encompass diverse and complex responses to environmental factors. Subdividing these groups into functional groups may provide a clearer understanding of the complex interactions at play.

No significant correlations were found between tidal height and diversity indices, and only one of the most abundant taxa, *Calanoida*, showed a significant positive correlation ($r_s = 0.372$, $p = 0.020$). We initially expected high tide to strengthen the marine pelagic eDNA signal and low tide to reflect more freshwater or benthic signals through sediment resuspension. Consistent with this expectation, *Calanoida*, a predominantly pelagic group, exhibited a weak but significant positive correlation with tidal height. Overall, however, tidal height was not a major driver of variation in the eDNA signal, contrary to initial assumptions. In contrast, water temperature correlated significantly with both prokaryotic and eukaryotic community composition, suggesting that water mass properties have a greater influence. Given the complexity of estuarine mixing, salinity observations may provide a more accurate proxy for tracking water mass origin and its effect on the eDNA signal than tidal height alone.

The composition of the most abundant prokaryotes remained relatively consistent throughout the study, except for *Oleispira*, which appeared to bloom in the latter half of the time series. *Oleispira* is an obligate hydrocarbonoclastic bacterium, specializing in the metabolism of hydrocarbons, and is known to bloom following pollution events (Gregson et al., 2020). *Oleispira* populations can expand to comprise 80-90% of the total microbial community post-pollution events (Harayama et al., 1999; Kasai et al., 2002). Ports interested in monitoring their environmental impact may find these short-lived *Oleispira* blooms to be effective biological

indicators of incoming ships with oil leaks or poor environmental practices, though the current study has limited environmental contextual information to evaluate the likelihood of a local pollution event.

The composition of the most abundant eukaryotes demonstrated more variability throughout the time series. The observed spike in calanoid sequences might be attributed to partial or entire calanoids being trapped on the filter. Despite efforts to remove visible organisms from filters before DNA extraction, some filters accumulated a considerable amount of detritus due to water turbidity, consequently some whole or large fragments of organisms may have inadvertently been included. Even though calanoids are small, their multicellularity means that their genetic material can disproportionately dominate sequence data compared to the marine microbiome. To avoid this, the feasibility of using a pre-filter or a series of filters could be explored in future iterations of the RoCSI.

Moreover, the variability observed in the eukaryotic data may be partly due to insufficient sample volumes, as larger volumes are generally required for effective analysis of metazoans (Ohnesorge et al., 2023). To address this, the sampling regime could be adjusted to allow for consecutive filtration of multiple samples until the required volume is reached, with the filters subsequently combined during downstream sample processing. Alternatively, switching from 0.22 µm to 0.45 µm Sterivex™ filters could facilitate the collection of larger volumes, particularly when targeting larger taxa such as fish.

Much of our understanding of anthropogenic impacts on marine ecology has been derived from micro- and mesocosm experiments, which often fail to accurately represent natural communities (Reiber et al., 2022). To understand the real-world consequences of activities such as pollutant release or alterations in water conditions, it is essential to monitor the marine environments where these activities most frequently occur. Long-term, routine monitoring of highly urbanised marine environments can provide the data needed to explore the cumulative effects of anthropogenic activities. This study serves as a preliminary investigation into this high-resolution temporal dataset, but there are many more potential avenues for data exploration. For example, network analysis could be used to explore species co-occurrence networks (Djurhuus et al., 2020), or wavelet analysis could be applied to explore the periodicities and trends with the environmental variables(Martin-Platero et al., 2018).

When analysing long-term time series data from eDNA datasets, it is important to distinguish short-term variability driven by tidal or diurnal cycles from broader longer-term trends. Standardising sampling to the same time each day controls for diurnal effects but not tides, while sampling at the same tidal stage controls for tides but not daily cycles. Therefore, high-

resolution datasets such as this, carried out over a full lunar cycle can provide an opportunity to disentangle these short-term dynamics and better interpret long-term trends.

3.6 Conclusion

Continuous monitoring of microbial communities in highly urbanised estuaries can yield crucial insights into the effects of emerging anthropogenic stressors. Before the development of autonomous samplers like the RoCSI, exploring this type of variability was impractical. Now it is more feasible to include high-resolution temporal datasets to characterise short-term dynamics, during the design phase of longer-term monitoring strategies. These preliminary investigations can help determine the optimal sampling regime for biomolecular observatories, which in turn would provide the evidence base for adaptive management strategies capable of responding dynamically to environmental change.

3.7 References

Barbier, E. B. (2017). Marine ecosystem services. *Current Biology*, *27*(11), R507–R510. https://doi.org/10.1016/j.cub.2017.03.020

Bisanz, J. (2018). *GitHub - jbisanz/qiime2R: Import qiime2 artifacts to R*.

Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., Huttley, G. A., & Gregory Caporaso, J. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome*, *6*(1), 1–17. https://doi.org/10.1186/S40168-018-0470-Z/TABLES/3

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology 2019 37:8*, *37*(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

Bruhn, A. D., Stedmon, C. A., Comte, J., Matsuoka, A., Speetjens, N. J., Tanski, G., Vonk, J. E., & Sjöstedt, J. (2021). Terrestrial Dissolved Organic Matter Mobilized From Eroding Permafrost Controls Microbial Community Composition and Growth in Arctic Coastal Zones. *Frontiers in Earth Science*, *9*, 640580. https://doi.org/10.3389/FEART.2021.640580/BIBTEX

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods 2016 13:7*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Djurhuus, A., Closek, C. J., Kelly, R. P., Pitz, K. J., Michisaki, R. P., Starks, H. A., Walz, K. R., Andruszkiewicz, E. A., Olesin, E., Hubbard, K., Montes, E., Otis, D., Muller-Karger, F. E., Chavez, F. P., Boehm, A. B., & Breitbart, M. (2020). Environmental DNA reveals seasonal shifts and potential interactions in a marine community. *Nature Communications 2020 11:1*, *11*(1), 1–9. https://doi.org/10.1038/s41467-019-14105-1

Downie, A.T., Bennett, W.W., Wilkinson, S., de Bruyn, M. and DiBattista, J.D., 2024. From land to sea: Environmental DNA is correlated with long-term water quality indicators in an urbanized estuary. *Marine Pollution Bulletin*, *207*, p.116887.

Fischer, P., Dietrich, P., Achterberg, E. P., Anselm, N., Brix, H., Bussmann, I., Eickelmann, L., Flöser, G., Friedrich, M., Rust, H., Schütze, C., & Koedel, U. (2021). Effects of Measuring Devices and Sampling Strategies on the Interpretation of Monitoring Data for Long-Term Trend Analysis. *Frontiers in Marine Science*, *8*, 770977. https://doi.org/10.3389/FMARS.2021.770977/BIBTEX

French, P. (2002). *Coastal and estuarine management*. Routledge.

Fuhrman, J. A., Cram, J. A., & Needham, D. M. (2015). Marine microbial community dynamics and their ecological interpretation. *Nature Reviews Microbiology 2015 13:3*, *13*(3), 133–146. https://doi.org/10.1038/nrmicro3417

Garrison, C. E., Roozbehi, S., Mitra, S., Corbett, D. R., & Field, E. K. (2022). Coastal Microbial Communities Disrupted During the 2018 Hurricane Season in Outer Banks, North Carolina. *Frontiers in Microbiology*, *13*, 816573. https://doi.org/10.3389/FMICB.2022.816573/BIBTEX

Gernez, M., Champagnat, J., Rivot, E. and Le Pape, O. (2023). Potential impacts of the restoration of coastal and estuarine nurseries on the stock dynamics of fisheries species. *Estuarine, Coastal and Shelf Science*, *295*, p.108557.

Gomez-Castillo, A. P., Panton, A., & Purdie, D. A. (2023). Temporal variability of phytoplankton biomass and net community production in a macrotidal temperate estuary. *Estuarine, Coastal and Shelf Science*, *280*, 108182. https://doi.org/10.1016/J.ECSS.2022.108182

Gregson, B. H., Metodieva, G., Metodiev, M. V., Golyshin, P. N., & McKew, B. A. (2020). Protein expression in the obligate hydrocarbon-degrading psychrophile Oleispira antarctica RB-8 during alkane degradation and cold tolerance. *Environmental Microbiology*, *22*(5), 1870. https://doi.org/10.1111/1462-2920.14956

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. https://doi.org/10.1093/BIOINFORMATICS/BTT086

Harayama, S., Kishira, H., Kasai, Y., & Shutsubo, K. (1999). Petroleum Biodegradation in Marine Environments. *Petroleum Biodegradation 63 J. Molec. Microbiol. Biotechnol*, *1*(1), 63–70. www.caister.com/bacteria-plant

Illumina. (2013). *16S Metagenomic Sequencing Library Preparation*. https://emea.support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.html

Kasai, Y., Kishira, H., Sasaki, T., Syutsubo, K., Watanabe, K., & Harayama, S. (2002). Predominant growth of Alcanivorax strains in oil-contaminated and nutrient-supplemented sea water. *Environmental Microbiology*, *4*(3), 141–147. https://doi.org/10.1046/J.1462-2920.2002.00275.X

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/MOLBEV/MST010

Kottuparambil, S., Ashok, A., Barozzi, A., Michoud, G., Cai, C., Daffonchio, D., Duarte, C. M., & Agusti, S. (2023). Tracking the early signals of crude oil in seawater and plankton after a major oil spill in the Red Sea. *Environmental Science and Pollution Research International*, *30*(26), 69150. https://doi.org/10.1007/S11356-023-27111-0

Lo, L. S. H., Liu, X., Liu, H., Shao, M., Qian, P. Y., & Cheng, J. (2023). Aquaculture bacterial pathogen database: Pathogen monitoring and screening in coastal waters using environmental DNA. *Water Research X*, *20*, 100194. https://doi.org/10.1016/J.WROA.2023.100194

Lynch, M., Trickovic, B., & Kempes, C. P. (2022). Evolutionary scaling of maximum growth rate with organism size. *Scientific Reports 2022 12:1*, *12*(1), 1–12. https://doi.org/10.1038/s41598-022-23626-7

Macé, B., Mouillot, D., Dalongeville, A., Bruno, M., Deter, J., Varenne, A., Gudefin, A., Boissery, P., & Manel, S. (2024). The Tree of Life eDNA metabarcoding reveals a similar taxonomic richness but dissimilar evolutionary lineages between seaports and marine reserves. *Molecular Ecology*, *33*(12), e17373. https://doi.org/10.1111/MEC.17373

Martin-Platero, A. M., Cleary, B., Kauffman, K., Preheim, S. P., McGillicuddy, D. J., Alm, E. J., & Polz, M. F. (2018). High resolution time series reveals cohesive but short-lived communities in coastal plankton. *Nature Communications 2018 9:1*, *9*(1), 1–11. https://doi.org/10.1038/s41467-017-02571-4

Nakane, T., Nakaka, K., Bouman, H., & Platt, T. (2008). Environmental control of short-term variation in the plankton community of inner Tokyo Bay, Japan. *Estuarine, Coastal and Shelf Science*, *78*(4), 796–810. https://doi.org/10.1016/J.ECSS.2008.02.023

Nascimento, Â., Biguino, B., Borges, C., Cereja, R., Cruz, J.P., Sousa, F., Dias, J., Brotas, V., Palma, C. and Brito, A.C. (2021). Tidal variability of water quality parameters in a mesotidal estuary (Sado Estuary, Portugal). *Scientific Reports*, *11*(1), p.23112.

Ohnesorge, A., John, U., Taudien, S., Neuhaus, S., Kuczynski, L., & Laakmann, S. (2023). Capturing drifting species and molecules—Lessons learned from integrated approaches to assess marine metazoan diversity in highly dynamic waters. *Environmental DNA*, *5*(6), 1541–1556. https://doi.org/10.1002/EDN3.478

Oksanen, J., Gavin L. Simpson, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R.B. O'Hara, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs, Helene Wagner, Matt Barbour, Michael Bedward, Ben Bolker, Daniel Borcard, Gustavo Carvalho, Michael Chirico, Miquel De Caceres, Sebastien Durand, ... James Weedon. (2024). *Package "vegan" Title Community Ecology Package*.

Onabule, O. A., Mitchell, S. B., & Couceiro, F. (2020). The effects of freshwater flow and salinity on turbidity and dissolved oxygen in a shallow Macrotidal estuary: A case study of Portsmouth Harbour. *Ocean & Coastal Management*, *191*, 105179. https://doi.org/10.1016/J.OCECOAMAN.2020.105179

Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2016). Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology*, *18*(5), 1403–1414. https://doi.org/10.1111/1462-2920.13023

Pedregosa FABIANPEDREGOSA, F., Michel, V., Grisel OLIVIERGRISEL, O., Blondel, M., Prettenhofer, P., Weiss, R., Vanderplas, J., Cournapeau, D., Pedregosa, F., Varoquaux, G., Gramfort, A., Thirion, B., Grisel, O., Dubourg, V., Passos, A., Brucher, M., Perrot andÉdouardand, M., Duchesnay, andÉdouard, & Duchesnay EDOUARDDUCHESNAY, Fré. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830. http://jmlr.org/papers/v12/pedregosa11a.html

Piwosz, K., Vrdoljak, A., Frenken, T., González-Olalla, J. M., Šantić, D., McKay, R. M., Spilling, K., Guttman, L., Znachor, P., Mujakić, I., Fecskeová, L. K., Zoccarato, L., Hanusová, M., Pessina, A., Reich, T., Grossart, H.-P., & Koblížek, M. (2020). Light and Primary Production Shape Bacterial Activity and Community Composition of Aerobic Anoxygenic Phototrophic Bacteria in a

Microcosm Experiment. *MSphere*, *5*(4). https://doi.org/10.1128/MSPHERE.00354-20/SUPPL_FILE/MSPHERE.00354-20-S0001.PDF

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE*, *5*(3), e9490. https://doi.org/10.1371/JOURNAL.PONE.0009490

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glöckner, F. O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, *35*(21), 7188–7196. https://doi.org/10.1093/NAR/GKM864

Reiber, L., Foit, K., Liess, M., Karaoglan, B., Wogram, J., & Duquesne, S. (2022). Close to reality? Micro-/mesocosm communities do not represent natural macroinvertebrate communities. *Environmental Sciences Europe*, *34*(1), 1–19. https://doi.org/10.1186/S12302-022-00643-X/TABLES/2

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, *2016*(10), e2584. https://doi.org/10.7717/PEERJ.2584/FIG-7

Shu, L., Ludwig, A., & Peng, Z. (2020). Standards for Methods Utilizing Environmental DNA for Detection of Fish Species. *Genes 2020, Vol. 11, Page 296*, *11*(3), 296. https://doi.org/10.3390/GENES11030296

Tøttrup, A. P., Svenningsen, L., Rytter, M., Lillemark, M. R., Møller, P. D. R., & Knudsen, S. W. (2021). Citizens in the lab: performance and validation of eDNA result

Trombetta, T., Bouget, F. Y., Félix, C., Mostajir, B., & Vidussi, F. (2022). Microbial Diversity in a North Western Mediterranean Sea Shallow Coastal Lagoon Under Contrasting Water Temperature Conditions. *Frontiers in Marine Science*, *9*, 858744. https://doi.org/10.3389/FMARS.2022.858744/BIBTEX

Varkey, D., Mazard, S., Jeffries, T. C., Hughes, D. J., Seymour, J., Paulsen, I. T., & Ostrowski, M. (2018). Stormwater influences phytoplankton assemblages within the diverse, but impacted Sydney Harbour estuary. *PLOS ONE*, *13*(12), e0209857. https://doi.org/10.1371/JOURNAL.PONE.0209857

Watson, S. C. L., Preston, J., Beaumont, N. J., & Watson, G. J. (2020). Assessing the natural capital value of water quality and climate regulation in temperate marine systems using a EUNIS biotope classification approach. *Science of The Total Environment*, *744*, 140688. https://doi.org/10.1016/J.SCITOTENV.2020.140688

Zhao, L., Brugel, S., Ramasamy, K. P., & Andersson, A. (2023). Bacterial community responses to planktonic and terrestrial substrates in coastal northern Baltic Sea. *Frontiers in Marine Science*, *10*, 1130855. https://doi.org/10.3389/FMARS.2023.1130855/BIBTEX

Zubkov, M. V. (2014). Faster growth of the major prokaryotic versus eukaryotic CO2 fixers in the oligotrophic ocean. *Nature Communications 2014 5:1*, *5*(1), 1–6. https://doi.org/10.1038/ncomms4776

Appendix B Supplementary Materials

B.1 Laboratory protocols

Modified AllPrep DNA extraction- QIAGEN w/ bead-beating:

- Extract the **RLT+** preservative from the Sterivex cartridge into labeled bead-beating tubes with beads (2 per Sterivex sample).
- Cut the Sterivex filter in half and put half into each labeled tube.
- Bead beat for 1 min at 30 Hz. Repeat.
- Use clean tweezers to squeeze liquid out of filters and throw them out. Spin tubes if the beads aren't at the bottom.
- Transfer the homogenized lysate to an AllPrep DNA spin column. Centrifuge for 30 s at ≥ 8000 x g.
- Place the AllPrep DNA spin column in a new 2 ml collection tube.
- Add **500 µl Buffer AW1** to the AllPrep DNA spin column. Close the lid gently, and centrifuge for 15 s at ≥ 8000 x g. Discard the flow through.
- Add **500 µl Buffer AW2** to the spin column. Close the lid gently, and centrifuge for 2 min at full speed to wash the spin column membrane. Carefully remove the AllPrep DNA spin column from the collection tube.
- Place the AllPrep DNA spin column in a new 1.5 ml collection tube. Add **100 µl Buffer EB** directly to the spin column membrane. Incubate for 10 min at 65ºC. Centrifuge for 1 min at ≥ 8000 x g to elute the DNA.
- Repeat with the eluate. Incubate for 2 min at room temperature. Centrifuge for 1 min at ≥ 8000 x g to elute the DNA.

Sequencing - Exeter Sequencing Facility completed the PCR indexing for the 16S rRNA and 18S rRNA gene amplicons as follows:

16S rRNA:

515F-Y-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GTGYCAGCMGCCGCGGTAA**

926R -GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**CCGYCAATTYMTTTRAGTTT**

18S rRNA:

1391F-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GTACACACCGCCCGTC**

EukBr-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**TGATCCTTCTGCAGGTTCACCTAC**

Table 8 First round PCR amplification conditions

	16S			18S		
	Temperature	Time	Repeat	Temperature	Time	Repeat
Initial Denaturation	98°C	30s	1	98°C	30s	1
Denaturation	98°C	10s	30	98°C	10s	35
Annealing	54°C	30s	30	69°C	30s	35
Extension	72°C	30s	30	72°C	30s	35
Final Extension	72°C	5 min	1	72°C	5 min	1
Hold	4°C	hold	1	4°C	hold	1

Primer mix:

12.5µl NEB NEXT ULTRA II QS

5µl F primer 1µM

5µl R primer 1µM

2.5µl template DNA/RNA ~4nM

PCR2 indexes

This step attaches dual indices and Illumina sequencing adapters using the Nextera XT Index Kit (Illumina #FC-131-1002, £597.99) Four identical reactions for each sample,

DNA	1 µl
Water	14 µl
Nextera XT Index Primer 1 (N7xx)	5 µl

Nextera XT Index Primer 2 (S5xx)	5 µl
2x NEBNext PCR reaction mix	25 µl
Total	50 µl

Gently pipette up and down 10 times to mix.

Cover the plate with Microseal 'A'

Centrifuge the plate at 1,000 × g at 20°C for 1 minute.

Perform PCR on a thermal cycler using the following program:

95°C	3min	
95°C	30seconds	4 cycles
55°C	30seconds	
72°C	30seconds	
72°C	5min	
4°C	Hold	

PCR clean up 2:

Bring the AMPure XP beads to room temperature for 30 minutes before use.

Centrifuge the plate at 280 × g at 20°C for 1 minute to collect condensation.

Vortex the AMPure XP beads for 30 seconds.

Add 35 µl of AMPure XP beads to each well.

Gently pipette mix up and down 10 times.

Incubate at room temperature without shaking for 5 minutes.

Place the plate on a magnetic stand for 2 minutes or until the supernatant has cleared.

Remove and discard the supernatant.

Wash the beads with freshly prepared 80% ethanol as follows:

Add 200 µl of freshly prepared 80% ethanol to each sample well.

Incubate the plate on the magnetic stand for 30 seconds.

Carefully remove and discard the supernatant

Do a second Ethanol wash.

Allow the beads to air-dry for 3 minutes

Remove the plate from the magnetic stand.

Add 27.5 µl of 10 mM Tris pH 8.5 to each well of the plate.

Pipette mix up and down 10 times until beads are fully resuspended, changing tips after each column.

Incubate at room temperature for 2 minutes.

Place the plate on the magnetic stand for 2 minutes or until the supernatant has cleared.

Transfer 25 µl of the supernatant to a new 96-well PCR plate.

Check on Glowmax

Use 2ul of library

B.2 Rarefaction curves

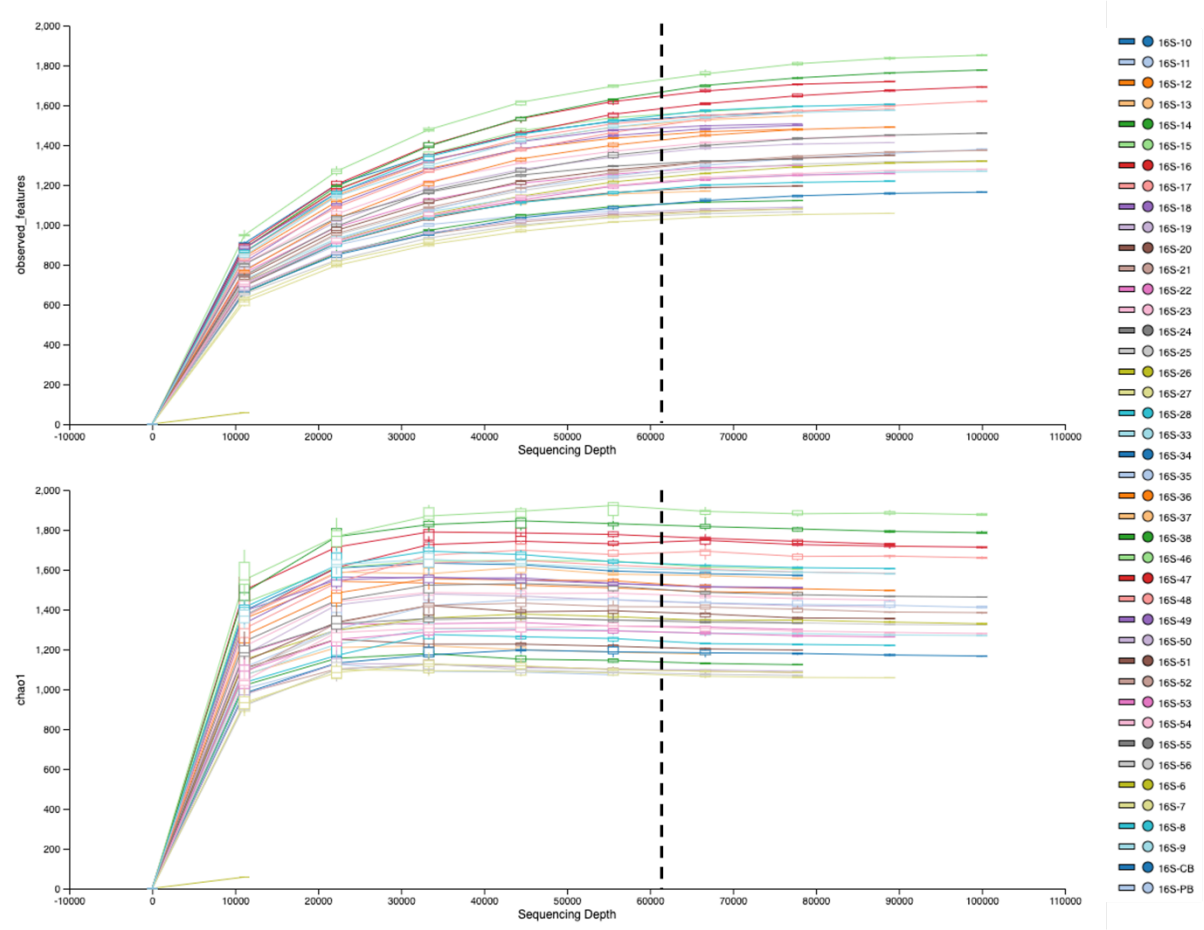


Figure 20 Prokaryotic rarefaction curves for observed features and choa1 index. Black dashed line indicates that samples were rarefied to 61400 sequences.

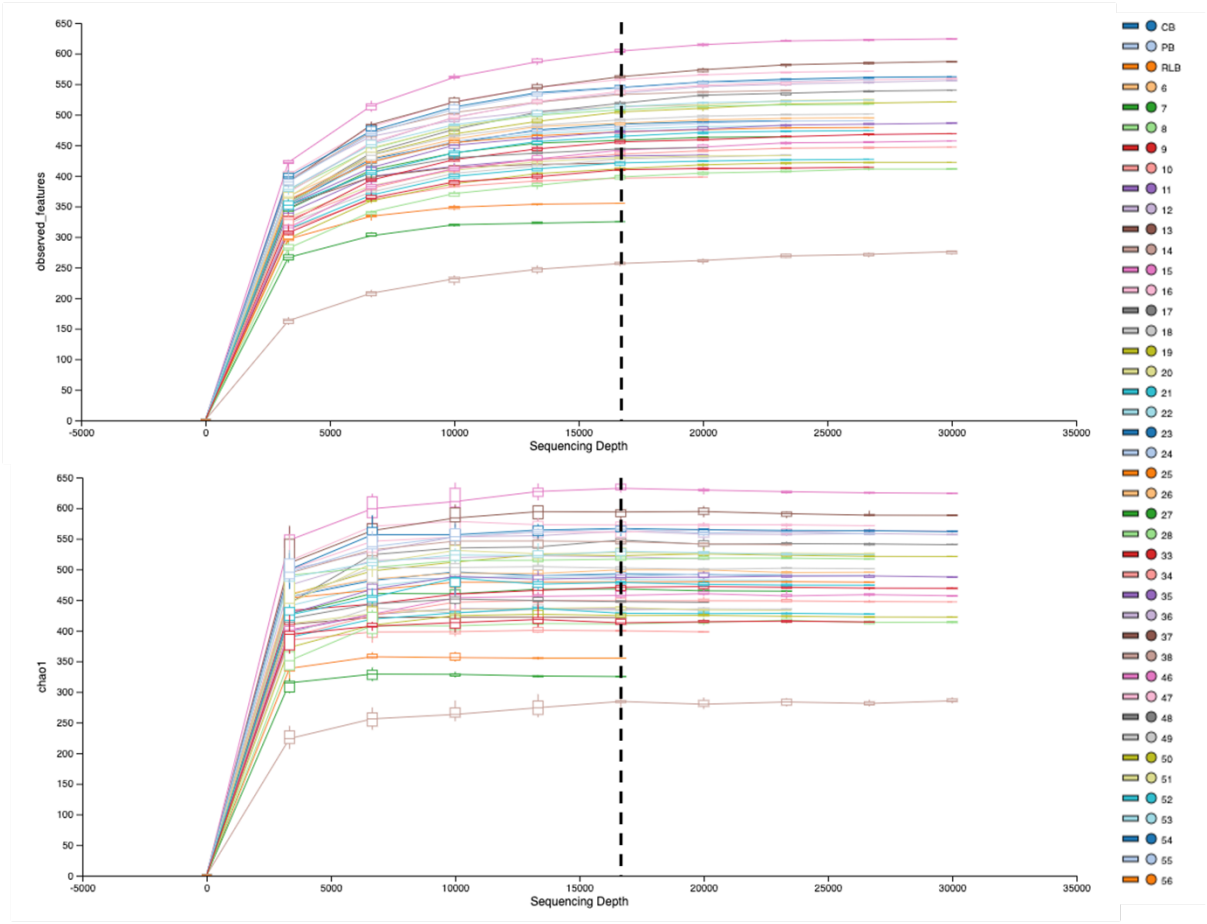


Figure 21 Eukaryotic rarefaction curves for observed features and choa1 index. Black dashed line indicates that samples were rarefied to 16744 sequences.

Chapter 4 Comparison of liquid preservatives for the long-term (2-month) preservation of nucleic acids

4.1 Abstract

Autonomous biomolecular samplers, such as the Robotic Cartridge Sampling Instrument (RoCSI), rely on liquid preservatives to stabilize biomolecules, such as DNA and RNA, for subsequent ex-situ analysis. In this study, we compare four liquid preservatives, RNAlater® (Inivitrogen), RLT+ buffer (Qiagen), DNA/RNA Shield (Zymo Research), and Nucleic Acid Preservative (NAP), against samples that were flash-frozen and subsequently stored at -80°C.

Results showed that both DNA and RNA can be preserved for two months with liquid preservatives, revealing comparable community composition and alpha diversity measures to flash frozen samples. DNA/RNA Shield (Zymo Research) was the only preservative to successfully preserve both DNA and RNA with yields sufficient for metabarcoding analysis. DNA/RNA Shield (Zymo Research) also produced the highest DNA yields and ASV counts from DNA metabarcoding. However, ASV counts from RNA metabarcoding were lower with DNA/RNA Shield (Zymo Research) compared to both RNAlater® (Qiagen) and NAP, with significant differences in beta diversity. This indicates that when RNA is the primary focus of an investigation, RNAlater® (Qiagen) or NAP may be preferable to DNA/RNA Shield (Zymo Research) for long-term (two-month) preservation.

Understanding the temporal limits of liquid preservation methods is essential to maximize the utility of biomolecular samplers. This knowledge allows for extended intervals between sample exchanges, enabling sampling throughout challenging conditions, such as adverse weather events, remote locations, or outside normal working hours. Thereby, greatly increasing our capacity for consistent, high-resolution temporal and spatial biological observations.

4.2 Introduction

DNA and RNA are the building blocks of life and analysing them helps to reveal the structures and functions of environmental microbial and metazoan communities. Environmental DNA (eDNA) is used to identify and quantify the presence of different organisms within a sample (Deiner et al., 2017; Taberlet et al., 2018). While environmental RNA (eRNA), provides insight into gene expression and the metabolic activities of these organisms (Carradec et al., 2018; Tang et al., 2020). It has also been proposed that additional metabarcoding of eRNA can help in distinguishing the most recent or live proportion of a community (Cristescu, 2019; Jo, 2023; Veilleux et al., 2021), with one study finding that eRNA analyses better correlates with morphological indices of diversity when compared with eDNA (Pochon et al., 2017).

Liquid preservation of environmental DNA and RNA is crucial for maintaining the integrity of these biomolecules in fieldwork where access to frozen storage is lacking (Edgcomb et al.,

2016; Gray et al., 2013). Autonomous technologies such as the Robotic Cartridge Sampling Instrument (National Oceanography Centre) and the Environmental Sample Processor (Monterey Bay Aquarium Research Institute) utilise liquid nucleic acid preservatives to minimise degradation, denaturation, and oxidation of eDNA and eRNA (Chapter 2; Truelove et al., 2022; Wang et al., 2021; Yamahara et al., 2019). *In situ* preservation immediately after filtration ensures the most accurate molecular data by stabilising biomolecules before the environmental changes that occur during sample retrieval, such as depressurisation and warming (Edgcomb et al., 2016). There are several commonly used preservative methods that can be used for the preservation of eDNA and eRNA, each with its own set of benefits and limitations.

Flash-freezing in liquid nitrogen or dry ice is commonly used as a best practice for rapidly preserving DNA and RNA, effectively preventing degradation (López-Escardó et al., 2018; Pratte & Kellogg, 2021). However, this technique is not intended for long-term preservation and requires subsequent storage at -80°C in cryogenic facilities. Additionally, flash-freezing is impractical for most *in-situ* environmental research and is incompatible with autonomous technologies. Alternatively, liquid nucleic acid preservatives are often more practical for fixing and preserving nucleic acids *in-situ*.

RNAlater® (Invitrogen) is a widely used preservative, that can stabilize both RNA and DNA in a single solution (Camacho-Sanchez et al., 2013; Gorokhova, 2005; Preston et al., 2024; Truelove et al., 2022). RNA is a more labile molecule than DNA and is considered more difficult to preserve (Littlefair et al., 2022). While RNAlater® (Invitrogen) is specifically designed to preserve RNA from tissue samples and cultured cells (Life Technologies Corporation 2012), many studies have found that RNAlater® (Invitrogen) is also suitable for DNA preservation (Miya et al., 2016; Truelove et al., 2022; Yamahara et al., 2019). The high stability of RNA in RNAlater® (Invitrogen) make it an attractive option for metatranscriptomics studies, which seek to understand gene expression within environmental samples. However, RNAlater® (Invitrogen) is not designed for long-term sample storage and has previously been found to be insufficient for long term storage between 10-50 weeks (Wietz et al., 2022).

Autonomous eDNA sampling devices can significantly enhance biodiversity monitoring by improving both temporal and spatial resolution, particularly when deployed at long-term time-series stations. Given the remoteness of some monitoring stations and limited maintenance access opportunities, liquid preservatives must be suitable for extended periods. However, most commercial liquid nucleic acid preservatives are recommended for use within a short duration of up to two weeks, presenting a challenge for long-term, autonomous eDNA monitoring.

In this study, we evaluated the efficacy of four liquid nucleic acid preservatives for long-term (two-month) preservation of eDNA and eRNA, comparing their performance against flash-frozen samples. The samples were filtered and preserved on Sterivex™ (Merck Millipore) cartridges, then stored submerged in the preservative within custom-designed leur lock cartridges. These cartridges are specifically designed for autonomous deployments of the RoCSI eDNA autosampler. We analysed the samples for total nucleic acid recovery and measures of alpha- and beta-diversity.

4.3 Material and methods

4.3.1 Preservatives

RNAlater® (RL): An Invitrogen reagent designed to stabilize RNA in tissue samples for 1 day at 37°C, 1 week at 25°C, 1 month at 4°C, or indefinitely at -20°C (Sigma-Aldrich, 2016)

Buffer RLT Plus (RLT): A Qiagen lysis buffer containing guanidine-isothiocyanate, which is designed to denature biological samples inactivating DNases and RNases (Qiagen, 2024). The buffer is the initial lysis buffer for use with the RNeasy and AllPrep extraction kits.

DNA/RNA Shield (Shield): A Zymo Research reagent designed to preserve nucleic acids in a range of biological sample types. It is specified to stabilize RNA for at least 1-month in ambient temperatures (4°C- 25°C) and DNA for at least 2 years in ambient temperature (4°C- 25°C), with indefinite DNA and RNA stabilisation when frozen (<-20°C; Zymo Research International, 2024).

Nucleic Acid Preservative (NAP): A buffer consisting of EDTA disodium salt dihydrate, sodium citrate trisodium salt dihydrate, and ammonium sulphate, that has been proven effective at preserving RNA and DNA in tissue samples for 2 months at ambient temperature. The NAP buffer was prepared in an amplicon and culture free molecular lab following Camacho-Sanchez et al. (2013) and autoclaved prior to use. (See full protocol in Appendix C.1)

4.3.2 Sample Collection

On November 14, 2019, at 14:02, a 15-liter sample of seawater was collected from a seawater tank on the dock of the National Oceanography Centre, Southampton, UK. Seawater supply to the tank was approximately 200ml/second from 1m depth at the National Oceanography Centre pontoon, with a 5mm net pre-filter to prevent larger detritus. Samples were filtered in an amplicon and culture free molecular lab on site within 4 hours of collection. Each replicate sample had 500 ml of seawater filtered through 0.22µm Sterivex™ filters, using a peristaltic

pump. Prior to filtration, the seawater underwent thorough mixing by inversion, and throughout the filtration process, the container was agitated every 5 minutes to prevent settling. All equipment was either UV cleaned or soaked with 10% bleach and rinsed with milli-Q water prior to use.

4.3.3 Sample Preservation

Each sample had 4ml of preservative added immediately after filtration. The RL, Shield, RLT and NAP preservatives were added to 5 replicate Sterivex™ filter samples and stored in RoCSI cartridges in a ziplock bag in the laboratory at room temperature (17°C -21°C) for 2 months before being stored at -80°C. The five control group samples were flash frozen immediately after filtration in liquid nitrogen and stored at -80°C.

4.3.4 Nucleic acid purification

DNA and RNA were simultaneously purified from FF, RL, RLT and NAP samples using a modified protocol for the Qiagen AllPrep DNA/RNA Mini Kit, which included an additional Trizol-chloroform RNA extraction prior to the RNeasy on-column extraction (For full protocol see Appendix C.1). Shield-preserved samples were extracted using the ZymoBIOMICS DNA/RNA Miniprep Kit, according to manufacturer recommendations. DNA and RNA yields were measured using Qubit High Sensitivity DNA and RNA Assay Kits.

4.3.5 Metabarcoding

Prior to amplification two of the control group FF samples were lost due to laboratory error. For RNA extracts, cDNA was synthesised using the LunaScript® RT SuperMix Kit (E3010) kit (for full protocol see Appendix C.1). For both DNA and cDNA, the 16S rRNA and the 16S rRNA gene region was amplified using 515F-Y (5'-GTGYCAGCMGCCGCGGTAA) forward and 926R (5'-CCGYCAATTYMTTTRAGTTT) reverse primers (Parada et al., 2016). The 18S rRNA gene region was amplified using 1391F (5'-GTACACACCGCCCGTC) forward and EukBr (5'-TGATCCTTCTGCAGGTTCACCTAC) reverse primers (Amaral-Zettler et al., 2009). Primers were modified with overhang sequences and a second, limited cycle PCR was used to add barcodes and sequencing adaptors, as specified in the Illumina sequencing library preparation protocol (Illumina, 2013). PCR indexing, RT-PCR and sequencing was completed at Exeter Sequencing Facility (For full protocol see Appendix C.1).

4.3.6 Bioinformatic Processing

The raw sequencing data was initially processed by removing adaptor sequences and sequences with less than 250 bp (Martin, 2011), raw sequence data are available at European Nucleotide Archive (accession number: PRJEB97124). Paired-end joining, denoising and taxonomic assignment was performed using QIIME2 release 2022.2 (Bolyen et al., 2019a). In brief, amplicon sequences variants (ASVs) were produced using DADA2 denoising and pair-end joining (Callahan et al., 2016). Phylogenetic trees were built using fasttree and mafft alignment (Bolyen et al. 2019; Price et al. 2010; Katoh et al. 2013). Taxonomic classification was performed with Naïve Bayes classifier and classify-sklearn method with the SILVA version 138.1 database, clustered at 99% sequence similarity as a reference (Bokulich et al., 2018; Bolyen et al., 2019b; Gurevich et al., 2013; Pedregosa et al., 2011; Pruesse et al., 2007; Rognes et al., 2016). For full scripts see GitHub rms1u18/RoCSI_Preservatives.

4.3.7 Data Analysis

All data analysis were conducted in R v4.3.2 (2023-10-31). The ggplot2 v3.5.0 (Wickham, 2016) and ggpubr v0.6 (Kassambara, 2023) packages were used to produce boxplots to visualise difference in nucleic acid yields and alpha diversity measures between preservative groups. The Kruskal-Wallis test for non-parametric data was used to test for significant differences between preservatives for both the nucleic acid yields and alpha diversity measures, with the Wilcoxon rank sum test for pairwise comparison between each preservative group.

To visualise beta diversity, ASV data were rarefied and PCoAs were plotted based on weighted UniFrac distance matrices using the phyloseq v1.46.0 package (McMurdie & Holmes, 2013). PERMANOVA and pairwise comparisons were also based on the weighted UniFrac distances of the rarefied ASV data with Bonferroni corrected p-values to test for significant differences in community composition between preservatives, using the vegan v2.6.4 package (Oksanen et al., 2024). Taxonomic heatmaps and barplots were produced using the qiime2r v0.99.6 package (Bisanz, 2018). For full scripts see GitHub rms1u18/RoCSI_Preservatives.

4.4 Results

4.4.1 Nucleic Acid Yield

After two months preservation at room temperature, the preservative with highest DNA yields was Shield with an average of 6.19 ng/µl (±0.49 SD), followed by RLT (0.94ng/µl ±0.47 SD), then FF (0.59ng/µl ±0.16 SD). RNAlater and NAP had very low yields (0.08 ng/µl ±0.02 SD and 0.07

ng/µl ±0.04 SD, respectively) that were insufficient for further DNA metabarcoding. The preservative with highest RNA yield was NAP with an average of 3.09 ng/µl (±0.14 SD), followed closely by RNAlater (3.04 ng/µl ±0.47 SD), then FF (1.33 ng/µl ±0.23 SD) and Shield (0.92 ng/µl ±0.14 SD). No detectable RNA was preserved in the RLT samples (see Table 9).

Table 9 Average DNA and RNA yields with ± standard deviation (±SD), for each preservative type.

DNA			RNA		
Preservative	Average (ng/µl)	±SD	Preservative	Average (ng/µl)	±SD
Shield	6.19	0.49	NAP	3.09	0.14
RLT	0.94	0.47	RNAlater	3.04	0.47
FF	0.59	0.16	FF	1.13	0.23
RNAlater	0.08	0.02	Shield	0.92	0.14
NAP	0.07	0.04	RLT	0.00	0.00

There was a significant difference for both the DNA and RNA yields between preservative groups. Pairwise comparisons using the Wilcoxon rank sum test revealed significant pair wise difference between the DNA yields for all preservative pairs except the FF - RLT groups (p-value = 0.278) and NAP - RNAlater groups (p-value = 0.917; Figure 22; Table 10). However, the interpretation of these results should be treated with caution as the Kruskal-Wallis is designed to test for significant differences between groups with a minimum of 5 replicates and the FF group only contained 3 replicates for DNA and 2 replicates for RNA. All other preservative groups contained 5 replicates.

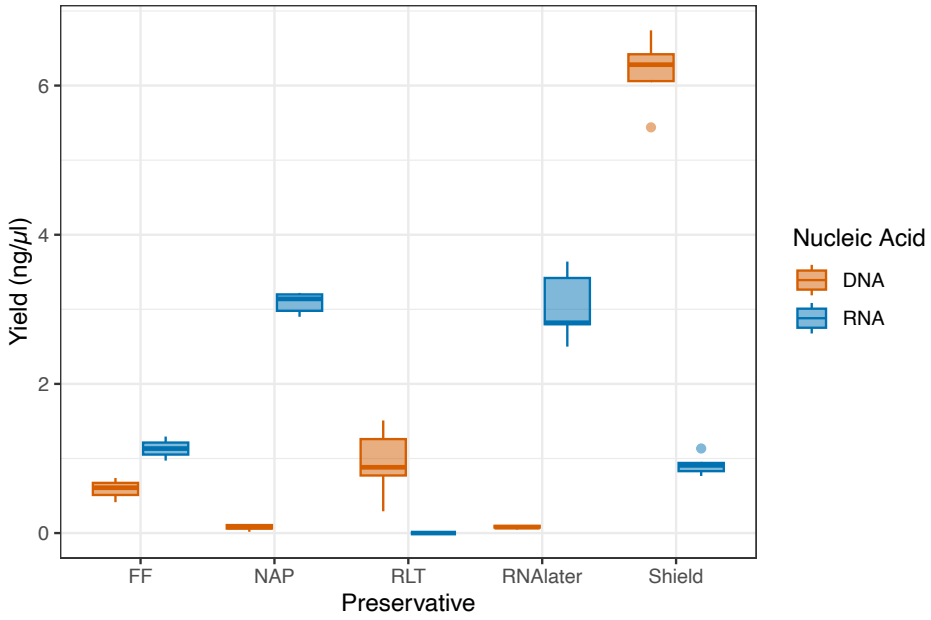


Figure 22 Box plots showing DNA yields in red and RNA yields in blue.

Table 10 Results from Kruskal-Wallis test results showing differences in nucleic acid yields between preservative groups, with pairwise comparisons performed using Dunn's test. Significant differences ($p < 0.05$) are highlighted in bold.

	DNA				RNA			
	K χ^2 = 19.329, df = 4, p-value = **0.0006**				K χ^2= 18.678, df = 4, p-value = **0.0009**			
	FF	NAP	RLT	RNAlater	FF	NAP	RLT	RNAlater
NAP	**0.045**	-	-	-	0.119	-	-	-
RLT	0.278	**0.016**	-	-	0.050	**0.016**	-	-
RNAlater	**0.045**	0.917	**0.016**	-	0.119	0.690	**0.016**	-
Shield	**0.045**	**0.016**	**0.016**	**0.016**	0.212	**0.016**	**0.016**	**0.016**

4.4.2 Alpha Diversity

Observed number of ASVs: Kruskal-Wallis tests showed no significant difference in observed no. of ASVs from 16S or 18S rRNA gene metabarcoding (DNA; Figure 2A) or from 16S or 18S rRNA metabarcoding (RNA; Figure 2B).

Shannon's diversity: Initial Kruskal-Wallis test found a significant difference in Shannon's diversity between preservative groups (K = 9.40, df = 2, p=0.009) from 16S rRNA gene

metabarcoding (DNA). Further pairwise testing revealed that the significant difference was due to a significant difference between the RLT and Shield groups (p=0.048; Figure 2A). No other significant differences in Shannon's diversity were identified from 18S rRNA gene (DNA; Figure 2A) or 16S and 18S rRNA (RNA) metabarcoding (Figure 2B; Table 11).

Simpson's evenness: Initial Kruskal-Wallis test also revealed a significant difference in Simpson's evenness between preservative groups (K = 8.3341, df = 3, p = 0.0155) from 16S rRNA gene (DNA) metabarcoding. Subsequent pairwise testing also revealed a significant difference between RLT and Shield samples (p=0.048; Figure 2A). Akin to Shannon's diversity, no significant differences were identified from 18S rRNA gene (DNA; Figure 2A) or 16S and 18S rRNA (RNA) metabarcoding (Figure 2B; Table 11).

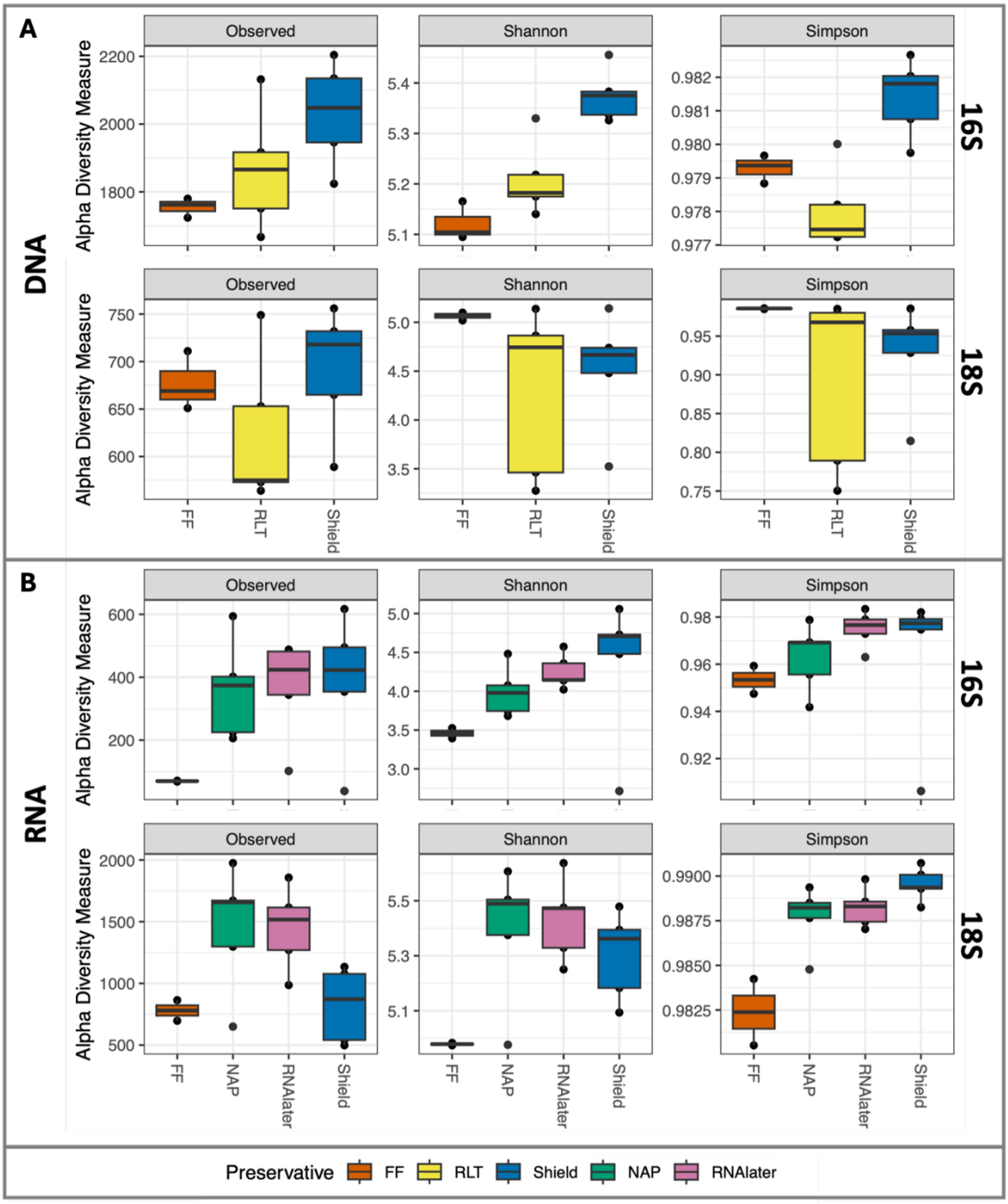


Figure 23 **A** Box plots indicating the alpha diversity measures (observed number of ASVs, Shannon's diversity, and Simpson's evenness) from 16S and 18S rRNA gene metabarcoding (DNA). **B** Box plots indicating the alpha diversity measures (observed number of ASVs, Shannon's diversity, and Simpson's evenness) from 16S and 18S rRNA metabarcoding (RNA).

Table 11 Results from the Kruskal-Wallis test for significant differences in alpha diversity measures between preservative groups, based on metabarcoding of the 16S and

18S rRNA gene (DNA) and 16S and 18S rRNA (RNA), with pairwise comparisons using Dunn's test. Bold text indicates significant differences ($p < 0.05$).

		DNA			RNA	
		16S	18S		16S	18S
Observed	Kruskal-KW	5.4945	2.9363	Kruskal-KW	3.902	7.5098
	df	2	2	df	3	3
	p	0.0641	0.2304	p	0.2722	0.05731
	FF-RLT	0.57	0.57	FF vs NAP	0.29	0.57
	FF-Shield	0.11	0.57	FF vs RNAlater	0.29	0.19
	RLT-Shield	0.23	0.45	FF vs Shield	0.76	0.86
				NAP vs RNAlater	0.84	0.83
				NAP vs Shield	0.84	0.17
				RNAlater vs Shield	0.84	0.17
Shannon	Kruskal-KW	9.3978	2.3209	Kruskal-KW	6.2392	5.4902
	df	2	2	df	3	3
	p	**0.009105**	0.3133	p	0.1005	0.1392
	FF-RLT	0.071	0.38	FF vs NAP	0.27	0.38
	FF-Shield	0.054	0.38	FF vs RNAlater	0.27	0.29
	RLT-Shield	**0.048**	1	FF vs Shield	0.38	0.29
				NAP vs RNAlater	0.27	0.84
				NAP vs Shield	0.27	0.46
				RNAlater vs Shield	0.27	0.5
Simpson	Kruskal-KW	8.3341	4.8352	Kruskal-KW	4.6745	8.0745
	df	2	2	df	3	3

p	**0.0155**	0.08914	p	0.1972	0.0445
FF-RLT	0.25	0.11	FF vs NAP	0.69	0.14
FF-Shield	0.054	0.11	FF vs RNAlater	0.45	0.14
RLT-Shield	**0.048**	1	FF vs Shield	0.57	0.14
			NAP vs RNAlater	0.45	0.84
			NAP vs Shield	0.57	0.14
			RNAlater vs Shield	1	0.18

4.4.3 Beta diversity

DNA: Initial PERMANOVA revealed a significant difference in community composition between 16S and 18S rRNA gene metabarcoding of DNA samples. However, pairwise testing revealed no significant differences between any preservative groups after using Bonferroni adjusted p values (Figure 24A & B; Table 12).

RNA: PERMANOVA test for significant differences in community composition revealed a significant difference between preservative groups from both 16S and 18S rRNA metabarcoding. Further, the pairwise testing of the 16S ASVs showed that the RL and Shield preservative groups had significantly different community composition (p = 0.048). Pairwise testing for the 18S ASVs revealed significant differences in the community composition between both NAP and Shield (p= 0.048) and RNAlater and Shield (p = 0.038; Figure 24C & D; Table 12).

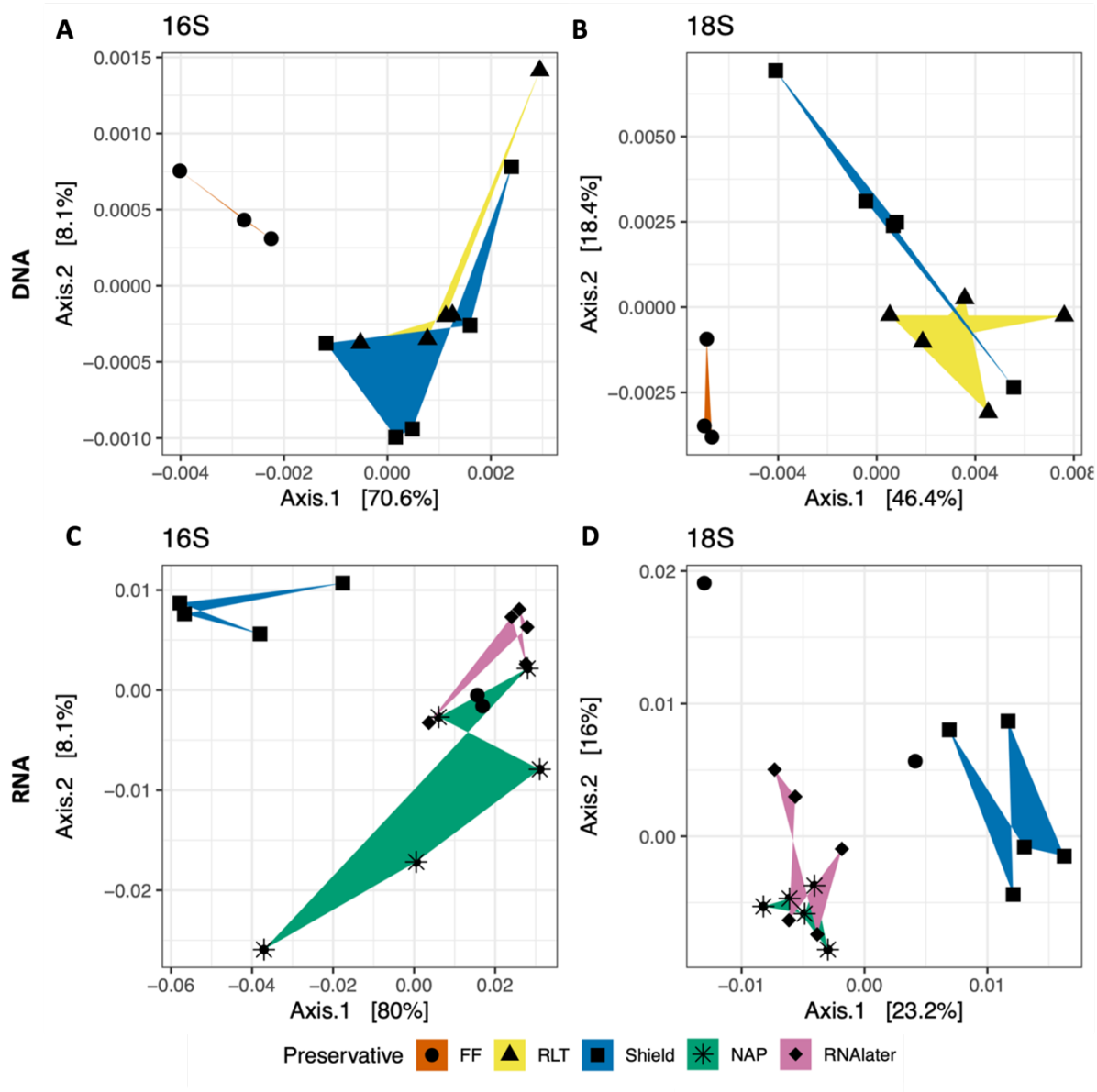


Figure 24 Principal coordinate analysis based on weighted UniFrac distances between samples using rarefied data from metabarcoding of **A** the 16S rRNA gene (DNA), **B** the 18S rRNA gene (DNA), **C** the 16S rRNA (RNA), and **D** the 18S rRNA (RNA).

Table 12 Results from PERMANOVA test for significant difference in beta diversity between preservative groups, with pairwise comparisons and Bonferroni adjusted p values. Bold text indicates significant p-values ($p<0.05$).

			Df	SumOfSqs	R2	F	Pr(>F)	p.adjusted
DNA	16S	Preservative	2	3.9E-05	0.54	5.93	**0.007**	
		Residual	10	3.3E-05	0.46			

		Total	12	7.3E-05	1.00		
		FF-RLT	1	3.3E-05	0.65	11.33	0.019	0.057
		FF-Shield	1	2.8E-05	0.57	8.10	0.027	0.081
		RLT-Shield	1	2.7E-06	0.09	0.77	0.510	1.000
	18S	Preservative	2	2.9E-04	0.47	4.52	**0.003**	
		Residual	10	3.2E-04	0.53		
		Total	12	6.2E-04	1.00		
		FF-RLT	1	2.2E-04	0.60	8.90	0.019	0.057
		FF-Shield	1	1.6E-04	0.43	4.60	0.018	0.055
		RLT-Shield	1	7.6E-05	0.21	2.10	0.063	0.198
RNA	16S	Preservative	3	1.2E-02	0.65	7.31	**0.004**	
		Residual	12	6.6E-03	0.35		
		Total	15	1.9E-02	1.00		
		FF-NAP	1	9.2E-04	0.19	1.14	0.267	1.000
		FF-RNAlater	1	6.1E-04	0.33	2.43	0.094	0.564
		FF-Shield	1	5.1E-03	0.76	12.62	0.067	0.400
		NAP-RNAlater	1	1.2E-03	0.20	2.02	0.100	0.600
		NAP-Shield	1	6.0E-03	0.53	7.90	0.025	0.150
		RNAlater-Shield	1	9.3E-03	0.79	25.87	0.008	**0.048**
	18S	Preservative	3	2.0E-03	0.37	2.54	**0.001**	
		Residual	13	3.4E-03	0.63		
		Total	16	5.4E-03	1.00		
		FF-NAP	1	6.6E-04	0.33	2.46	0.043	0.258
		FF-RNAlater	1	5.3E-04	0.27	1.89	0.057	0.342

FF-Shield	1	7.8E-04	0.33	2.42	0.057	0.270
NAP-RNAlater	1	2.2E-04	0.11	1.01	0.444	1.000
NAP-Shield	1	9.4E-04	0.32	3.80	0.008	**0.048**
RNAlater-Shield	1	8.3E-04	0.29	3.25	0.006	**0.036**

Heatmap from the DNA extracts show very similar abundance across the 30 most abundant features for both 16S and 18S rRNA gene metabarcoding (Figure 25A & B), except for the polychaete *Sabella spallanzanii* which was low abundance in Shield samples and variable abundance in FF and RLT samples (Figure 25B).

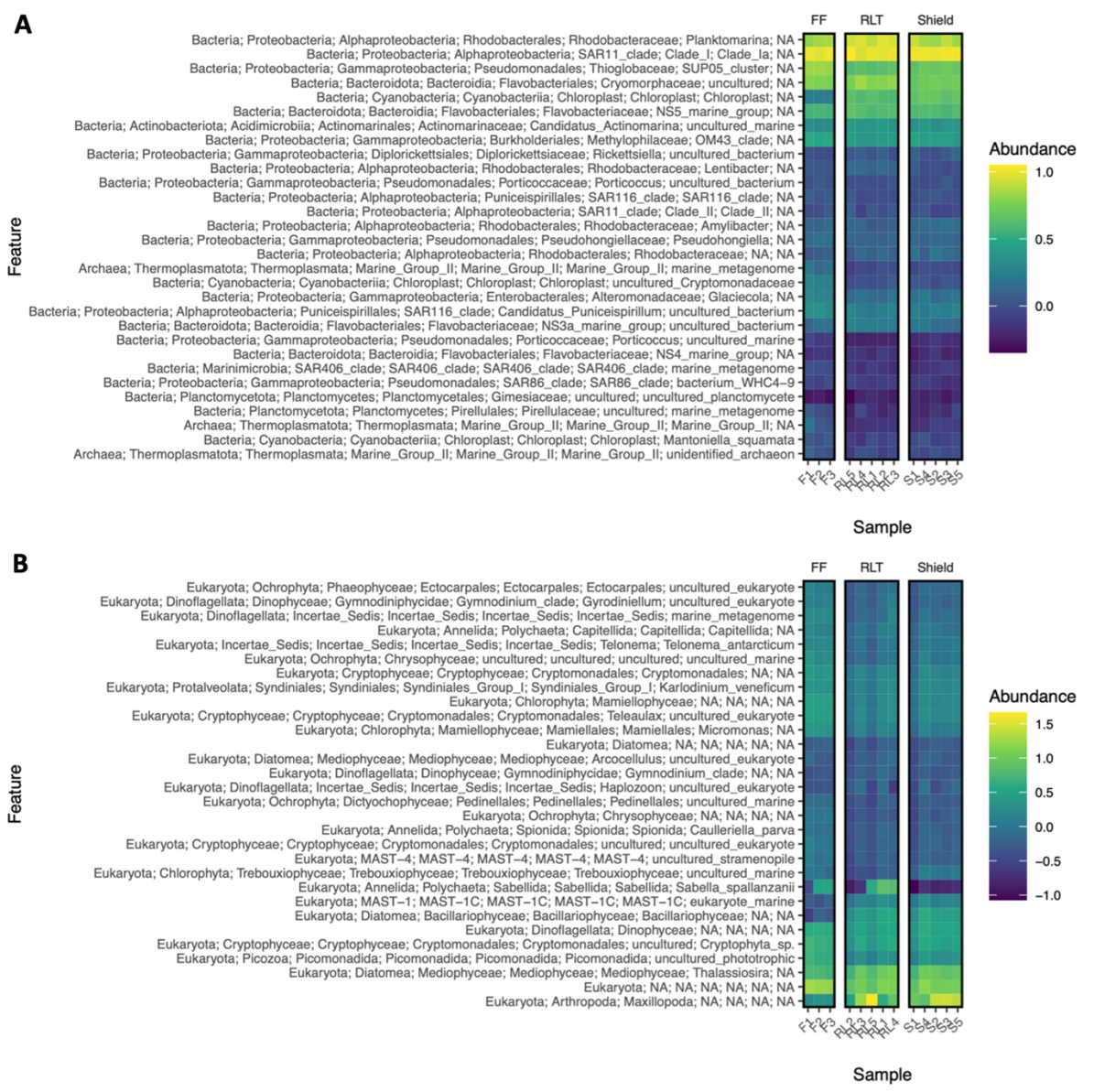


Figure 25 Heatmaps based on rarefied data from **A** 16S rRNA gene metabarcoding and **B** 18S rRNA gene metabarcoding of DNA samples.

Heatmaps from the RNA samples show more variability in the abundance of the top 30 most abundant 16S features, with shield samples appearing the most dissimilar (Figure 26A). In contrast, the 18S features from RNA samples show very similar patterns in abundance across preservative groups, apart from Diatoms which had lower abundances in the FF samples (Figure 26B).

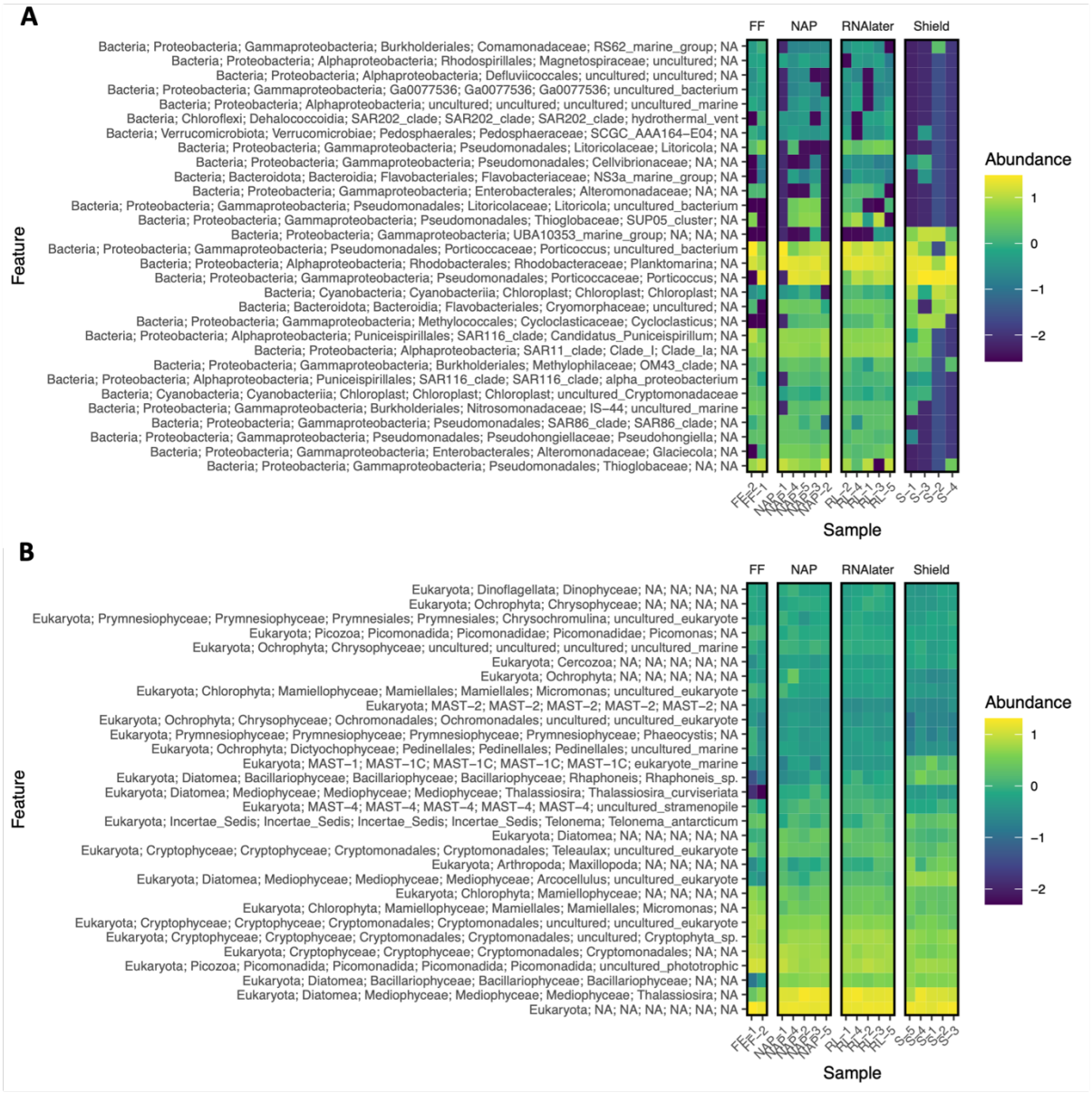


Figure 26 Heatmaps based on rarefied data from **A** 16S rRNA metabarcoding and **B** 18S rRNA metabarcoding of RNA samples.

4.5 Discussion

Liquid preservation of biological samples facilitates rapid stabilisation, thereby limiting the degradation of nucleic acids. This is particularly crucial in warm climates or in environments with high microbial productivity, such as coastal seas, where degradation rates of nucleic acids are known to be elevated (Collins et al., 2018). The RoCSI applies a liquid preservative immediately post-filtration to circumvent sample degradation and the Sterivex™ (Merck Millipore) samples containing preservative, are secured within custom luer lock cartridges, which were found to be suitability for long-term storage. No leakage or reduction in preservative volume was observed after storage at 17–21°C for two months.

Among various liquid preservative methods tested, only the Shield preservative yielded DNA and RNA quantities sufficient for metagenomic sequencing (Figure 22). The ZymoBiomic extraction kits streamlined the nucleic acid purification process, as no additional phenol-chloroform modification were necessary to attain adequate RNA yields for sequencing. However, RNA extracts preserved with Shield preservative exhibited more variability in high-abundance bacterial taxa compared to those preserved by flash-freezing, RNAlater, and NAP methods (Figure 26A). Manufacturer guidelines suggest that the Shield preservative maintains RNA at ambient conditions for up to one month and DNA for at least two years (Zymo Research International, 2024). While there were no significant differences in alpha diversity measures between preservative groups for RNA samples (Figure 23B; Table 11), there were significant differences in beta diversity (Figure 24C & D; Table 12). Our findings suggest caution in exceeding one month for RNA preservation, especially when metabarcoding 16S rRNA, although it remains a viable and easy-to-use option for shorter (≤1month) deployments requiring both DNA and RNA.

Both RLT buffer and Shield preservative were effective in preserving DNA extracts for up to two months. These methods yielded higher DNA concentrations than flash-frozen samples, with Shield averaging 5.6 ng/µl more and RLT averaging 0.35 ng/µl more than flash-frozen samples (Figure 22; Table 9). Flash-frozen samples were preserved less than 30 minutes after preservative addition to the liquid preservative samples, which is unlikely to account for the observed degradation observed between Shield and flash-frozen samples. Consequently, we recommend using the Shield preservative for long term (two-month) preservation of DNA.

The RLT buffer was also suitable for two-month preservation of DNA. However, the yields were not as high as those obtained with the Shield preservative (Figure 22; Table 9). Furthermore, although RLT acts as both a lysis buffer and preservative, any time saved by combining these step was offset by the need to incorporate an additional trizol-chloroform step to the AllPrep extraction protocol to produce sufficient yields for DNA sequencing.

RNAlater and NAP samples had insufficient DNA yields for sequencing (<0.1 ng/µl). These findings corroborate Wu & Minamoto (2023) who also found that using the Qiagen ATL buffer was preferential to using RNAlater for DNA preservation. The Qiagen ATL and RLT+ buffers are both initial lysis buffers for the Qiagen DNeasy and AllPrep extraction kits, respectively. Using these lysis buffers as preservatives offers cost and time efficiencies, eliminating the cost of additional preservative solutions and time required for pre-extraction removal of the preservative (Wu & Minamoto, 2023).

For DNA extracts, high-abundance taxa exhibited consistent trends across various preservation methods in both 16S and 18S rRNA gene metabarcoding. This aligns with (Gray et al., 2013), who also found no clear difference between liquid preservatives for DNA preservation. No significant pairwise differences in community composition were noted with Bonferroni-adjusted p-values, however differences were observed prior to this adjustment (Table 12). PCoA plots suggest minor differences between samples, warranting further investigation with additional replicates to enhance the robustness of these findings (Figure 24A & B). Presently, we can infer that reliable comparisons can be made between high-abundance taxa across datasets preserved using Shield or RLT preservatives. However, caution is advised in interpreting differences in low-abundance taxa between preservative groups due to the need for more replicates to ascertain the authenticity of observed differences.

RNAlater and NAP preservatives resulted in higher RNA yields than flash freezing (Figure 22). Although RNAlater is recommended for use at 25°C for up to one week (Sigma-Aldrich Co. 2016), Ottesen et al. (2011) found that RNAlater yielded high-quality RNA after 30 days at room temperature. Furthering this, our results indicate that RNA of sufficient quality for sequencing remains intact after two months at room temperature (17–21°C). NAP buffer also demonstrated similar performance to RNAlater, with no significant differences in RNA yield (Figure 22; Table 10) or community composition (Figure 24C & D; Table 11). Therefore, NAP may be a feasible alternative for projects with limited resources, requiring large preservative volumes, if the necessary laboratory infrastructure is available.

However, RNAlater, NAP, Shield, and flash-frozen samples displayed variability in high-abundance taxa for 16S rRNA metabarcoding of RNA extracts (Figure 26A), while 18S rRNA metabarcoding results were more consistent (Figure 26B). This variability in 16S rRNA metabarcoding results might be attributable to sample degradation, suggesting that exceeding the 30-day RNA preservation period observed by Ottesen et al. (2011) should be approached with caution, particularly when using longer barcode regions like the 16S marker (~563bp), which are more susceptible to degradation (Wei et al., 2018; Perry et al., 2024). Further research with additional time points is needed to establish the upper limit for RNA preservation with liquid preservatives.

The enhanced RNA yields from liquid preservation when compared to flash freezing align with Edgcomb et al. (2016), concluding that in-situ liquid RNA preservation is preferred over flash freezing. *In situ* liquid preservation via autonomous collection devices like the Microbial Sampler-Submersible Incubation Device and the RoCSI enables immediate nucleic acid stabilisation, thereby mitigating the impact of environmental changes encountered during

sample collection, such as pressure and temperature fluctuations, which can affect RNA expression (Edgcomb et al., 2016; Feike et al., 2012).

Some larger eukaryotes, such as *Sabella spallanzanii* and the class Maxillopoda, showed greater variability across DNA samples (Figure 25B), which may be due to the presence of tissue fragments retained on the filter. This variability appeared less pronounced for Maxillopoda in RNA samples (Figure 26B), potentially due to the faster degradation of RNA within these fragments. However, additional replicates are needed to confirm this trend.

Interestingly, heatmaps of the 30 most abundant taxa suggest that diatoms may be underrepresented in both RNA and DNA samples that are flash frozen without a liquid preservative (Figure 25B & Figure 26B). The cause is unclear but considering that diatoms are a major group frequently detected by visual methods, these results highlight the importance of adding a liquid preservative to ensure their detection.

The findings from this study are limited to inshore Coastal Observatories such as the Western Channel Observatory (WCO), which are generally accessible but may become unreachable during periods of adverse sea conditions. Further research is needed to assess the effectiveness of liquid preservatives beyond two months, particularly for open ocean moorings that are only serviced annually (e.g. Porcupine Abyssal Plain; PAP-SO). Future studies should incorporate more replicates and samples extracted at multiple time points to better define the temporal limits of liquid nucleic acid preservation. Additionally, liquid preservatives are typically tested under stable temperature conditions, but autonomous sampling devices experience fluctuating environmental conditions such as temperature and pressure. Understanding the impact of these variables on eDNA/RNA integrity and recovery will improve the reliability of biomolecular analyses from autonomously collected samples.

4.6 Conclusion

This study demonstrates the effectiveness of liquid preservation methods for stabilising nucleic acids in environmental samples. The Leur-lock cartridges designed for the RoCSI provided secure containment for up to two months. Among the preservatives tested, Shield was the most effective for DNA preservation, offering high DNA yields and ease of use, while RNAlater and NAP were the best suitedfor RNA preservation. However, variability in high-abundance bacterial taxa detected via16S rRNA metabarcoding indicate that two months may be approaching the temporal limit for RNA preservation, particularly for longer barcode regions. Therefore, when both RNA and DNA are required from the same sample, we recommend using the shield preservative and limiting the preservation time to one month prior to freezing, in line with

manufacturer guidelines. Selecting the optimal liquid preservation strategy, based on the nucleic acid type, storage duration before freezing, and, potentially, barcode length, ensures data reliability. Taking these steps to ensure reliable data from autonomously collected samples also supports net-zero science goals by enabling longer deployments and reducing reliance on ship-based sampling for biodiversity observations.

4.7 References

Amaral-Zettler, L. A., McCliment, E. A., Ducklow, H. W., & Huse, S. M. (2009). A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. *PLOS ONE*, *4*(7), e6372. https://doi.org/10.1371/JOURNAL.PONE.0006372

Bisanz, J. (2018). *GitHub - jbisanz/qiime2R: Import qiime2 artifacts to R*.

Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., Huttley, G. A., & Gregory Caporaso, J. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome*, *6*(1), 1–17. https://doi.org/10.1186/S40168-018-0470-Z/TABLES/3

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019a). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology 2019 37:8*, *37*(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019b). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology 2019 37:8*, *37*(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods 2016 13:7*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Camacho-Sanchez, M., Burraco, P., Gomez-Mestre, I., & Leonard, J. A. (2013). Preservation of RNA and DNA from mammal samples under field conditions. *Molecular Ecology Resources*, *13*(4), 663–673. https://doi.org/10.1111/1755-0998.12108

Carradec, Q., Pelletier, E., Da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., Lima-Mendez, G., Rocha, F., Tirichine, L., Labadie, K., Kirilovsky, A., Bertrand, A., Engelen, S., Madoui, M. A., Méheust, R., Poulain, J., Romac, S., Richter, D. J., Yoshikawa, G., … Wincker, P. (2018). A global ocean atlas of eukaryotic genes. *Nature Communications 2018 9:1*, *9*(1), 1–13. https://doi.org/10.1038/s41467-017-02342-1

Collins, R. A., Wangensteen, O. S., O'Gorman, E. J., Mariani, S., Sims, D. W., & Genner, M. J. (2018). Persistence of environmental DNA in marine systems. *Communications Biology 2018 1:1*, *1*(1), 1–11. https://doi.org/10.1038/s42003-018-0192-6

Cristescu, M. E. (2019). Can Environmental RNA Revolutionize Biodiversity Science? *Trends in Ecology and Evolution*, *34*(8), 694–697. https://doi.org/10.1016/j.tree.2019.05.003

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D. M., de Vere, N., Pfrender, M. E., & Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, *26*(21), 5872–5895. https://doi.org/10.1111/MEC.14350

DNA/RNA Shield Stabilization Solution - Free Sample | ZYMO RESEARCH – Zymo Research International. (n.d.). Retrieved May 27, 2024, from https://zymoresearch.eu/products/dna-rna-shield

Edgcomb, V. P., Taylor, C., Pachiadaki, M. G., Honjo, S., Engstrom, I., & Yakimov, M. (2016). Comparison of Niskin vs. in situ approaches for analysis of gene expression in deep Mediterranean Sea water samples. *Deep Sea Research Part II: Topical Studies in Oceanography*, *129*, 213–222. https://doi.org/10.1016/J.DSR2.2014.10.020

Feike, J., Jürgens, K., Hollibaugh, J. T., Krüger, S., Jost, G., & Labrenz, M. (2012). Measuring unbiased metatranscriptomics in suboxic waters of the central Baltic Sea using a new in situ fixation system. *The ISME Journal*, *6*(2), 461. https://doi.org/10.1038/ISMEJ.2011.94

Gorokhova, E. (2005). Effects of preservation and storage of microcrustaceans in RNAlater on RNA and DNA degradation. *Limnology and Oceanography: Methods*, *3*(2), 143–148. https://doi.org/10.4319/LOM.2005.3.143

Gray, M. A., Pratte, Z. A., & Kellogg, C. A. (2013). Comparison of DNA preservation methods for environmental bacterial community samples. *FEMS Microbiology Ecology*, *83*(2), 468–477. https://doi.org/10.1111/1574-6941.12008

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. https://doi.org/10.1093/BIOINFORMATICS/BTT086

Illumina. (2013). *16S Metagenomic Sequencing Library Preparation*. https://emea.support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.html

Jo, T. S. (2023). Utilizing the state of environmental DNA (eDNA) to incorporate time-scale information into eDNA analysis. *Proceedings of the Royal Society B*, *290*(1999). https://doi.org/10.1098/RSPB.2023.0979

Kassambara, A. (2023). ggpubr: "ggplot2" Based Publication Ready Plots. In *DOI:10.32614/CRAN.package.ggpubr*.

Littlefair, J. E., Rennie, M. D., & Cristescu, M. E. (2022). Environmental nucleic acids: A field-based comparison for monitoring freshwater habitats using eDNA and eRNA. *Molecular Ecology Resources*, *22*(8), 2928–2940. https://doi.org/10.1111/1755-0998.13671

López-Escardó, D., Paps, J., De Vargas, C., Massana, R., Ruiz-Trillo, I., & Del Campo, J. (2018). Metabarcoding analysis on European coastal samples reveals new molecular metazoan diversity. *Scientific Reports 2018 8:1*, *8*(1), 1–14. https://doi.org/10.1038/s41598-018-27509-8

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. https://doi.org/10.14806/EJ.17.1.200

McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE*, *8*(4), e61217. https://doi.org/10.1371/JOURNAL.PONE.0061217

Miya, M., Minamoto, T., Yamanaka, H., Oka, S. I., Sato, K., Yamamoto, S., Sado, T., & Doi, H. (2016). Use of a Filter Cartridge for Filtration of Water Samples and Extraction of Environmental DNA. *JoVE (Journal of Visualized Experiments)*, *2016*(117), e54741. https://doi.org/10.3791/54741

Oksanen, J., Gavin L. Simpson, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R.B. O'Hara, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs, Helene Wagner, Matt Barbour, Michael Bedward, Ben Bolker, Daniel Borcard, Gustavo Carvalho, Michael Chirico, Miquel De Caceres, Sebastien Durand, … James Weedon. (2024). *Package "vegan" Title Community Ecology Package*.

Ottesen, E. A., Marin, R., Preston, C. M., Young, C. R., Ryan, J. P., Scholin, C. A., & Delong, E. F. (2011). Metatranscriptomic analysis of autonomously collected and preserved marine bacterioplankton. *The ISME Journal*, *5*(12), 1881–1895. https://doi.org/10.1038/ISMEJ.2011.70

Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2016). Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology*, *18*(5), 1403–1414. https://doi.org/10.1111/1462-2920.13023

Pedregosa FABIANPEDREGOSA, F., Michel, V., Grisel OLIVIERGRISEL, O., Blondel, M., Prettenhofer, P., Weiss, R., Vanderplas, J., Cournapeau, D., Pedregosa, F., Varoquaux, G., Gramfort, A., Thirion, B., Grisel, O., Dubourg, V., Passos, A., Brucher, M., Perrot andÉdouardand, M., Duchesnay, andÉdouard, & Duchesnay EDOUARDDUCHESNAY, Fré. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830. http://jmlr.org/papers/v12/pedregosa11a.html

Perry, W.B., Seymour, M., Orsini, L., Jâms, I.B., Milner, N., Edwards, F., Harvey, R., de Bruyn, M., Bista, I., Walsh, K. & Emmett, B. (2024). An integrated spatio-temporal view of riverine biodiversity using environmental DNA metabarcoding. *Nature Communications*, *15*(1), p.4372.

Pochon, X., Zaiko, A., Fletcher, L. M., Laroche, O., & Wood, S. A. (2017). Wanted dead or alive? Using metabarcoding of environmental DNA and RNA to distinguish living assemblages for biosecurity applications. *PLoS ONE*, *12*(11), 1–19. https://doi.org/10.1371/journal.pone.0187636

Pratte, Z. A., & Kellogg, C. A. (2021). Comparison of Preservation and Extraction Methods on Five Taxonomically Disparate Coral Microbiomes. *Frontiers in Marine Science*, *8*, 684161. https://doi.org/10.3389/FMARS.2021.684161/BIBTEX

Preston, C., Yamahara, K., Pargett, D., Weinstock, C., Birch, J., Roman, B., Jensen, S., Connon, B., Jenkins, R., Ryan, J., & Scholin, C. (2024). Autonomous eDNA collection using an uncrewed surface vessel over a 4200-km transect of the eastern Pacific Ocean. *Environmental DNA*, *6*(1), e468. https://doi.org/10.1002/EDN3.468

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glöckner, F. O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, *35*(21), 7188–7196. https://doi.org/10.1093/NAR/GKM864

Qiagen. (2024). *AllPrep DNA/RNA/Protein Mini Kit*. Https://Www.Qiagen.Com/Us/Products/Discovery-and-Translational-Research/Dna-Rna-Purification/Multianalyte-and-Virus/Allprep-Dnarnaprotein-Mini-Kit.

RNAlater - RNA Stabilization Solution for Tissue - Catalog Number R0901. (2016). *Sigma-Aldrich*. www.sigmaaldrich.com

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, *2016*(10), e2584. https://doi.org/10.7717/PEERJ.2584/FIG-7

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). Environmental DNA: For biodiversity research and monitoring. *Environmental DNA: For Biodiversity Research and Monitoring*, 1–253. https://doi.org/10.1093/oso/9780198767220.001.0001

Tang, W., Cerdán-García, E., Berthelot, H., Polyviou, D., Wang, S., Baylay, A., Whitby, H., Planquette, H., Mowlem, M., Robidart, J., & Cassar, N. (2020). New insights into the distributions of nitrogen fixation and diazotrophs revealed by high-resolution sensing and sampling methods. *The ISME Journal*, *14*(10), 2514–2526. https://doi.org/10.1038/S41396-020-0703-6

Truelove, N. K., Patin, N. V., Min, M., Pitz, K. J., Preston, C. M., Yamahara, K. M., Zhang, Y., Raanan, B. Y., Kieft, B., Hobson, B., Thompson, L. R., Goodwin, K. D., & Chavez, F. P. (2022). Expanding the temporal and spatial scales of environmental DNA research with autonomous sampling. *Environmental DNA*, *4*(4), 972–984. https://doi.org/10.1002/EDN3.299

Veilleux, H. D., Misutka, M. D., & Glover, C. N. (2021). Environmental DNA and environmental RNA: Current and prospective applications for biological monitoring. *Science of The Total Environment*, *782*, 146891. https://doi.org/10.1016/J.SCITOTENV.2021.146891

Wang, S., Tang, W., Delage, E., Gifford, S., Whitby, H., González, A. G., Eveillard, D., Planquette, H., & Cassar, N. (2021). Investigating the microbial ecology of coastal hotspots of marine nitrogen fixation in the western North Atlantic. *Scientific Reports 2021 11:1*, *11*(1), 1–14. https://doi.org/10.1038/s41598-021-84969-1

Wei, N., Nakajima, F. & Tobino, T. (2018). A microcosm study of surface sediment environmental DNA: decay observation, abundance estimation, and fragment length comparison. *Environ. Sci. Technol.* 52, 12428–12435.

Wickham, H. (2016). *ggplot2*. https://doi.org/10.1007/978-3-319-24277-4

Wietz, M., Metfies, K., Bienhold, C., Wolf, C., Janssen, F., Salter, I., & Boetius, A. (2022). Impact of preservation method and storage period on ribosomal metabarcoding of marine microbes: Implications for remote automated samplings. *Frontiers in Microbiology*, *13*, 999925. https://doi.org/10.3389/FMICB.2022.999925/BIBTEX

Wu, Q., & Minamoto, T. (2023). Improvement of recovery yield of macro-organismal environmental DNA from seawater samples. *Analytical Sciences*, *1*, 1–8. https://doi.org/10.1007/S44211-023-00280-1/FIGURES/5

Yamahara, K. M., Preston, C. M., Birch, J., Walz, K., Marin, R., Jensen, S., Pargett, D., Roman, B., Ussler, W., Zhang, Y., Ryan, J., Hobson, B., Kieft, B., Raanan, B., Goodwin, K. D., Chavez, F. P., & Scholin, C. (2019). In situ autonomous acquisition and preservation of marine environmental dna using an autonomous underwater vehicle. *Frontiers in Marine Science*, *6*(JUL), 464097. https://doi.org/10.3389/FMARS.2019.00373/BIBTEX

Appendix C Supplementary Material

C.1 Laboratory protocol

Modified AllPrep DNA/RNA extraction- QIAGEN w/ bead-beating

- Add **700 µl RLT (with β-ME)** and beads to labeled bead-beating tubes (2 per sterivex sample).
- Cut the Sterivex filter in half and put half into each labeled tube.
- Bead beat for 1 min at 30 Hz. Repeat.
- Use clean tweezers to squeeze liquid out of filters and throw them out. Spin tubes if the beads aren't at the bottom.
- Transfer the homogenized lysate to an AllPrep DNA spin column. Centrifuge for 30 s at ≥ 8000 x g.
- Place the AllPrep DNA spin column in a new 2 ml collection tube (supplied), and store at room temperature (15–25ºC) or at 4ºC for later DNA purification. Use the flow-through for RNA purification

RNA

- Add **600 µl Trizol** to the flow through, cap tube tightly and vortex for 30 seconds.
- Centrifuge for 10 min at 12000 x g at 4ºC. RNA remains in the aqueous phase.
- Transfer aqueous phase (~60% of TRI volume used in the lysis) to a new tube.
- Add **1/5 of the volume of chloroform*** (without isoamyl alcohol). Vortex for 15 sec. Incubate at room temperature for 5-10 min.
- BE CAREFUL with chloroform. It drips from the pipette tip so keep tube close to the bottle opening when adding.
- Centrifuge for 10 min at 12000 x g at 4 ºC. RNA remains in the aqueous phase. DNA and proteins are in the interphase and organic phase.
- Transfer aqueous phase (~60% of TRI volume used in the lysis) to a new tube.

- Add **1/2 volume of the aqueous phase 100% ethanol**. Vortex immediately at maximum speed for 5 seconds to avoid RNA precipitation.

On-column extraction

- Transfer **700 µl** of the sample to the RNeasy spin column. Centrifuge for 15 s at ≥ 8000 x g. Discard the flow-through. Repeat for the rest of the sample volume
- Add **700 µl Buffer RW1** to the RNeasy spin column (BE SURE to pipette against the walls of the column!). Centrifuge for 15 s at ≥ 8000 x g to wash the spin column membrane. Carefully discard the flow-through completely.
- Add **500 µl Buffer RPE** to the RNeasy spin column (BE SURE to pipette against the walls of the column!). Close the lid gently, and centrifuge for 15 s at ≥ 8000 x g to wash the spin column membrane. Discard the flow-through.
- Add **500 µl of 80% ethanol** and centrifuge for 2 min at ≥ 8000 x g
- Transfer the Filter Cartridge to a new collection Tube. Centrifuge for 5 min at full speed to dry the membrane
- Place the column in a new 1.5 ml collection tube. Add **50 µl of RNase-free water** to the filter column. Incubate at room temperature for 1 min and centrifuge for 1 min at s at ≥ 8000 x g.
- Repeat using the eluate.

DNA

- Add **500 µl Buffer AW1** to the AllPrep DNA spin column. Close the lid gently, and centrifuge for 15 s at ≥ 8000 x g. Discard the flow through.
- Add **500 µl Buffer AW2** to the spin column. Close the lid gently, and centrifuge for 2 min at full speed to wash the spin column membrane. Carefully remove the AllPrep DNA spin column from the collection tube.
- Place the AllPrep DNA spin column in a new 1.5 ml collection tube. Add **100 µl Buffer EB** directly to the spin column membrane. Incubate for 10 min at 65ºC. Centrifuge for 1 min at ≥ 8000 x g to elute the DNA.
- Repeat with the eluate. Incubate for 2 min at room temperature. Centrifuge for 1 min at ≥ 8000 x g to elute the DNA.

cDNA synthesis

Protocol for LunaScript® RT

SuperMix Kit (E3010)

1. Mix components briefly and spin

down if necessary.

2. Prepare cDNA synthesis reaction as

described below:

COMPONENTS	20 µl REACTION	FINAL CONCENTRATION
LunaScript RT SuperMix (5X)	4 µl	1X
RNA Sample	8 ul	(up to 1 µg)*
Nuclease-free Water	8 µl	
	20 ul	

For no-RT control reaction, mix the following components:

COMPONENTS	20 µl REACTION	FINAL CONCENTRATION
No-RT Control Mix (5X)	4 µl	1X
RNA Sample	0	
Nuclease-free Water	to 20 µl	

*Up to 1 µg total RNA, 1 µg mRNA or 100 ng specific RNA can be used in a 20 µl reaction.

For no template controls, mix the following components:

COMPONENTS	20 µl REACTION	FINAL CONCENTRATION
LunaScript RT SuperMix (5X)	4 µl	1X
Nuclease-free Water	to 20 µl	

Incubate reactions in a thermocycler
with the following steps:

CYCLE STEP	TEMPERATURE	TIME	CYCLES
Primer Annealing	25°C	2 minutes	1
cDNA Synthesis	55°C	10 minutes	
Heat Inactivation	95°C	1 minute	

Sequencing

Exeter Sequencing Facility completed the second stage of PCR indexing for the 16S rRNA and 18S rRNA gene amplicons from the DNA samples and both PCR barcode amplification and PCR indexing for the 16S and 18S rRNA from RNA samples before sequencing.

Nextera Transposase Adapters

515F-Y

TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GTGYCAGCMGCCGCGGTAA**

926R

GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**CCGYCAATTYMTTTRAGTTT**

1391F

TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GTACACACCGCCCGTC**

EukBr

GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**TGATCCTTCTGCAGGTTCACCTAC**

Table 13 First round PCR conditions

		16S		18S		
	Temperature	Time	Repeat	Temperature	Time	Repeat
Initial Denaturation	98°C	30s	1	98°C	30s	1
Denaturation	98°C	10s	30	98°C	10s	35
Annealing	54°C	30s	30	69°C	30s	35
Extension	72°C	30s	30	72°C	30s	35
Final Extension	72°C	5 min	1	72°C	5 min	1
Hold	4°C	hold	1	4°C	hold	1

Primer mix:

12.5µl NEB NEXT ULTRA II QS

5µl F primer 1µM

5µl R primer 1µM

2.5µl template DNA/RNA ~4nM

PCR2 indexes

This step attaches dual indices and Illumina sequencing adapters using the Nextera XT Index Kit (Illumina #FC-131-1002, £597.99) Four identical reactions for each sample,

DNA	1 µl
Water	14 µl
Nextera XT Index Primer 1 (N7xx)	5 µl
Nextera XT Index Primer 2 (S5xx)	5 µl

2x NEBNext PCR reaction mix	25 µl
Total	50 µl

Gently pipette up and down 10 times to mix.

Cover the plate with Microseal 'A'

Centrifuge the plate at 1,000 × g at 20°C for 1 minute.

Perform PCR on a thermal cycler using the following program:

95°C	3min	
95°C	30seconds	4 cycles
55°C	30seconds	
72°C	30seconds	
72°C	5min	
4°C	Hold	

Pcr clean up 2

Bring the AMPure XP beads to room temperature for 30 minutes before use.

Centrifuge the plate at 280 × g at 20°C for 1 minute to collect condensation.

Vortex the AMPure XP beads for 30 seconds.

Add 35 µl of AMPure XP beads to each well.

Gently pipette mix up and down 10 times.

Incubate at room temperature without shaking for 5 minutes.

Place the plate on a magnetic stand for 2 minutes or until the supernatant has cleared.

Remove and discard the supernatant.

Wash the beads with freshly prepared 80% ethanol as follows:

Add 200 µl of freshly prepared 80% ethanol to each sample well.

Incubate the plate on the magnetic stand for 30 seconds.

Carefully remove and discard the supernatant

Do a second Ethanol wash.

Allow the beads to air-dry for 3 minutes

Remove the plate from the magnetic stand.

Add 27.5 µl of 10 mM Tris pH 8.5 to each well of the plate.

Pipette mix up and down 10 times until beads are fully resuspended, changing tips after each column.

Incubate at room temperature for 2 minutes.

Place the plate on the magnetic stand for 2 minutes or until the supernatant has cleared.

Transfer 25 µl of the supernatant to a new 96-well PCR plate.

Check on Glowmax

Use 2ul of library

C.2 Taxa Barplots

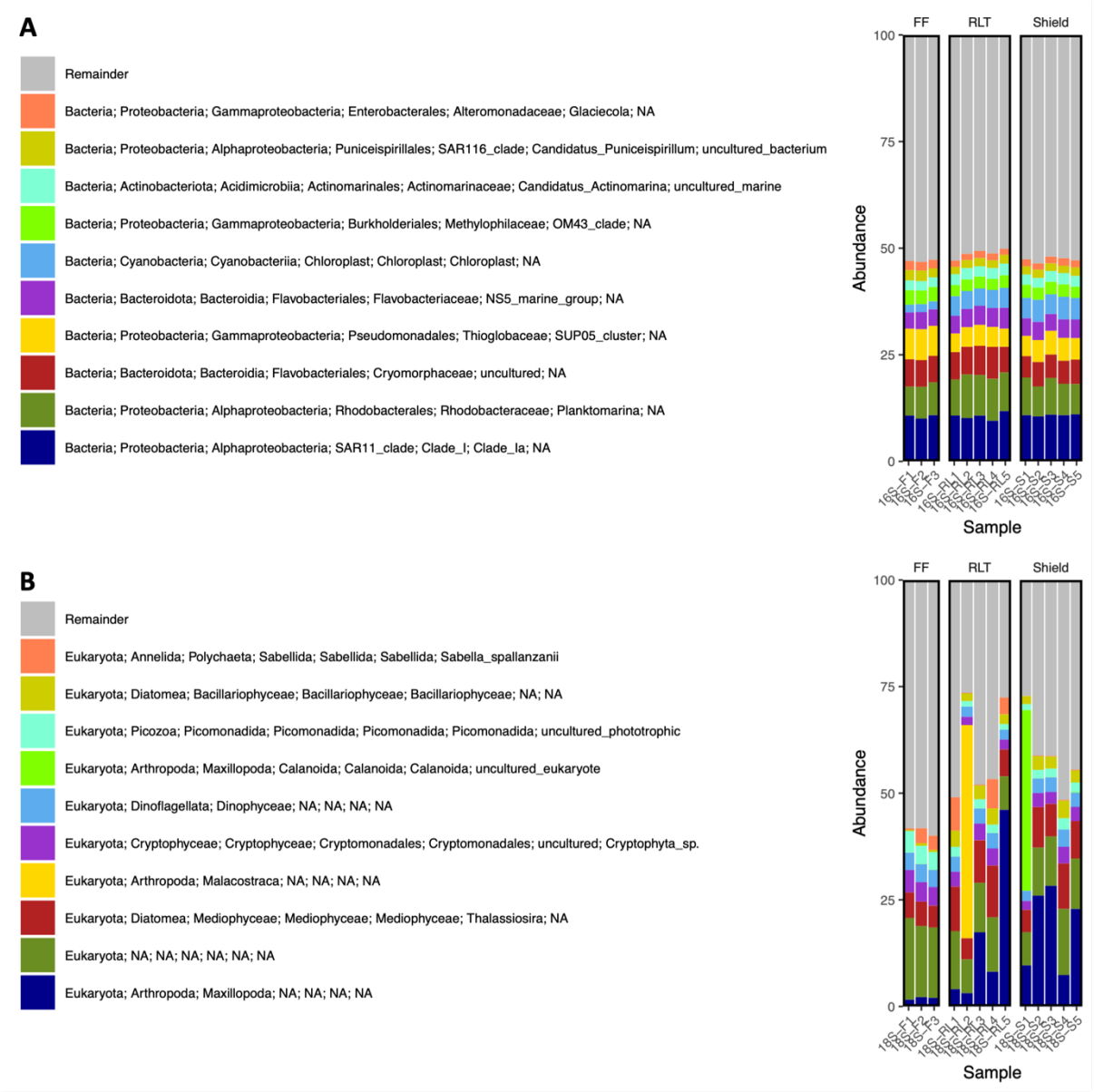


Figure 27 Taxa bar plots based on rarefied data from **A** 16S rRNA gene metabarcoding and **B** 18S rRNA gene metabarcoding of DNA samples

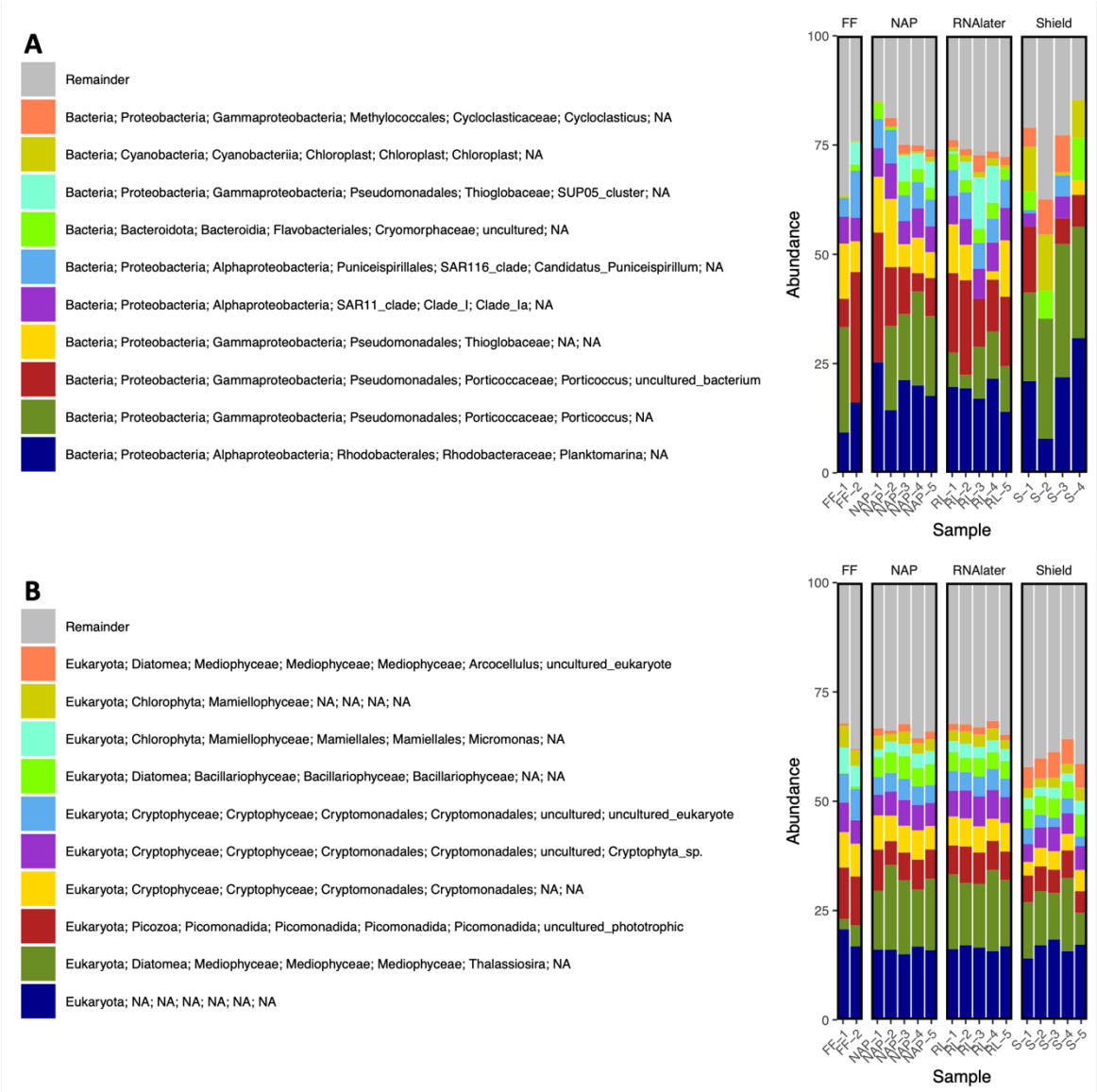


Figure 28 Taxa bar plots based on rarefied data from **A** 16S rRNA metabarcoding and **B** 18S rRNA metabarcoding of RNA samples.

Chapter 5 Towards a Global Public Repository of Community Protocols to Encourage Best Practices in Biomolecular Ocean Observing and Research

Robyn M. Samuel[1,2], Raissa Meyer[3], Pier Luigi Buttigieg[4], Neil Davies[5], Nicholas W. Jeffery[6], Christopher Meyer[7], Christina Pavloudi[8], Kathleen Johnson Pitz[9], Maxime Sweetlove[10], Susanna Theroux[11], Jodie van de Kamp[12], Alison Watts[13]

[1]School of Ocean and Earth Science, University of Southampton, Southampton, United Kingdom

[2]Ocean Technology and Engineering Group, National Oceanography Center, Southampton, United Kingdom

[3]Alfred Wegener Institute Helmholtz Center for Polar and Marine Research, Bremerhaven, Germany

[4]Helmholtz Metadata Collaboration, GEOMAR, Kiel, Germany

[5]Gump South Pacific Research Station, University of California, Berkeley, Moorea, French Polynesia

[6]Fisheries and Oceans Canada, Bedford Institute of Oceanography, Dartmouth, NS, Canada

[7]Department of Invertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC, United States

[8]Hellenic Centre for Marine Research, Institute of Marine Biology, Biotechnology and Aquaculture, Heraklion, Greece

[9]Monterey Bay Aquarium Research Institute, Moss Landing, CA, United States

[10]Royal Belgian Institute of Natural Sciences, Brussels, Belgium

[11]Southern California Coastal Water Research Project, Costa Mesa, CA, United States

[12]Commonwealth Scientific and Industrial Research Organization, Oceans and Atmosphere, Hobart, TAS, Australia

[13]Civil and Environmental Engineering, University of New Hampshire, Durham, NH, United States

5.1 Abstract

Biomolecular ocean observing and research is a rapidly evolving field that uses omics approaches to describe biodiversity at its foundational level, giving insight into the structure and function of marine ecosystems over time and space. It is an especially effective approach for investigating the marine microbiome. To mature marine microbiome research and operations within a global ocean biomolecular observing network (OBON) for the UN Decade of Ocean Science for Sustainable Development and beyond, research groups will need a system to

effectively share, discover, and compare “omic” practices and protocols. While numerous informatic tools and standards exist, there is currently no global, publicly supported platform specifically designed for sharing marine omics [or any omics] protocols across the entire value-chain from initiating a study to the publication and use of its results. Toward that goal, we propose the development of the Minimum Information for an Omic Protocol (MIOP), a community-developed guide of curated, standardized metadata tags and categories that will orient protocols in the value-chain for the facilitated, structured, and user-driven discovery of suitable protocol suites on the Ocean Best Practices System. Users can annotate their protocols with these tags or use them as search criteria to find appropriate protocols. Implementing such a curated repository is an essential step toward establishing best practices. Sharing protocols and encouraging comparisons through this repository will be the first steps toward designing a decision tree to guide users to community endorsed best practices.

5.2 Introduction

The term “omics” generally means studying anything holistically, and here we take a broad view of biomolecular omics that includes, but is not limited to: quantitative target gene amplification (e.g., qPCR, qNASBA etc.), (meta)barcoding, (meta)genomics, (meta)transcriptomics, (meta)proteomics, and metabolomics; and field collection approaches that target organisms or parts thereof, including single-celled organisms (microorganisms), as well as environmental DNA (eDNA). In the marine realm, omic techniques are used to assess and monitor biodiversity, reveal population structure and gene flow, and discover new compounds with applications in medicine and industry. Rapid advances in omic research, and the declining cost of high-throughput sequencing technologies (Wetterstrand, 2020) support the increasing application of omics in marine microbiome research.

The recent expansion in marine omics has led to a proliferation of protocols specific to multiple applications. However, these protocols are rarely shared publicly with sufficient detail to reliably reproduce a study (Dickie et al., 2018). While the omics community has already achieved high standards for sharing sequence data through the International Nucleotide Sequence Database Collaboration, these data often lack sufficient metadata and provenance information on the protocols used (Dickie et al., 2018), undermining efforts to implement the Findable, Accessible, Interoperable and Reusable (FAIR) data principles (Wilkinson et al., 2016). These limitations create challenges for marine microbiome research and operations from individual labs up to global (meta)data analysis efforts such as MGnify (Mitchell et al., 2019), which must identify data collected using comparable methods, in order to integrate and re-use data for meta-analysis (Berry et al., 2020). Moreover, a lack of protocol-sharing impedes the identification of comparable methods needed for global monitoring efforts aiming to understand, and sustainably manage the changing marine ecosystem (Aylagas et al., 2020; Berry et al., 2020; Makiola et al., 2020).

Many projects are looking to develop best practices for omics research: standards organizations, such as the Genomic Standards Consortium's (GSC) Genomic Biodiversity Interest Group, the Biodiversity Information Standards (TDWG) and the Biocode Commons are working collaboratively toward standards specifications for genomic observatories (Davies et al., 2012, 2014). Large campaigns, such as the Earth Microbiome Project (Gilbert et al., 2014; Thompson et al., 2017), TARA Oceans (Sunagawa et al., 2020), and the Australian Microbiome Initiative (AM; Bissett et al., 2016; Brown et al., 2018; doi: 10.4227/71/561c9bc670099), have already developed standardized practices, and innovative software enterprises, such as protocols.io, are providing powerful solutions for sharing protocols. Yet there is currently no global, publicly supported infrastructure developed explicitly for encouraging the exchange and harmonization of omic protocols, so these valuable contributions remain fragmented and underutilized.

For marine ecosystems, the Intergovernmental Oceanographic Commission's Ocean Best Practices System (OBPS) provides a public repository for all ocean research methodological documentation that can interlink protocols, standard specifications, and other guidelines. The OBPS seeks to support continuous convergence of methods as they undergo community refinement to become best practices (Hörstmann et al., 2021). In collaboration with the broader omics community, through the Omic BON initiative (Buttigieg et al., 2019), we propose to develop a best practice system specific to marine omics research, leveraging the framework of the OBPS to curate a global repository for marine omics protocols.

As part of the omics/eDNA session at the 4th OBPS workshop, we discussed recommendations and community needs for an omics/eDNA specific best practices system. Recognizing an urgent need for the ocean omics community to get organized as the UN Decade of Ocean Science for Sustainable Development starts, we identified the demand for publishing protocols into a user-friendly decision tree framework. With such a framework we would aim to support protocol selection, increase protocol findability and improve recognition for protocol developers. In a series of focused follow-up meetings, we identified that an omics decision tree would require a library of constituent parts (the protocols) and framework to: (1) locate where the protocol fits within the entire omics workflow (outlined in section "Ocean Omics Methodology Categories"), and (2) organize protocols using focused descriptive terms (metadata tags), based on what the protocol does and how/why it is used (outlined in section "Essential Metadata for Omics Protocols").

5.3 Ocean Omics Methodology Categories

The typical omics workflow involves a series of protocols, which take a project from ideation, through to publication, and on to societal use. Protocols from each step in the omics workflow hold valuable information for different groups. For example, sample collection protocols may be most relevant to scientists/technicians in the field, whereas local stakeholders and indigenous communities may primarily engage with aspects of how the project and resulting data address and impact important ethical, legal, and societal issues (Nagoya Protocol, 2010; Carroll et al., 2020). Documenting details and provenance for the entire marine omics workflow requires input from multiple parties, as each step of the workflow may be conducted by different individuals or groups. The omics OBPS therefore needs to identify these key methodological categories, to allow protocols and accompanying metadata to be uploaded in modules that link together to form the entire workflow.

We propose twelve protocol categories (Figure 29A) for ocean omics research and operations. Protocols and guidelines are assigned into these categories according to the purpose they serve[1]. Categories 5–12 outline methodological categories for operational activities used in the AM Initiative (van de Kamp et al., 2019). Categories 1–4 were identified to additionally cover cross-cutting documentation in the omics workflow: (1) Society, (2) Sampling/observational design, (3) Ethics and law, and (4) Data management.

1. Society—All workflows should begin and end with society; societal needs inform the question or purpose behind the research, and societal impacts show the value in the research once it has been completed.
2. Design and logistics—This category covers the practical logistics for implementing ocean omics research and operations, including the experimental/observational design formulated to address the societal priorities outlined in 1.
3. Ethics and law—A survey of workshop participants highlighted a need for guidance on sharing data and complying with important ethical and legal requirements (Simpson et al., 2021). This category will include information on permits and permission required to obtain samples and release data. Collating and publishing this information will firstly provide examples for how previous projects have adhered to legal requirements/ethical principles and secondly stimulate discussion on how to facilitate adherence to these requirements and principles, perhaps through checklists, templates, or training materials.
4. Data management—The data management plan (DMP) is designed to support all the downstream steps according to the ethics, legalities and societal needs identified in (1–3), while making sure that the (meta)data flows to the right stakeholders in society that

we need to interface with. DMPs should be drafted prior to data collection and referred to throughout the workflow to ensure that quality assurance and quality checks take place, and that detailed information on (meta)data requirements for both short and long-term (meta)data storage is given. There is a growing body of tools and best practices surrounding DMPs, including principles for making them more machine-actionable, that should be leveraged in omic protocols and associated infrastructure (see Miksa et al., 2019). Publishing documentation on omics specific DMPs will increase transparency for funders by providing direct links to the protocols they refer to. Furthermore, collating examples of omics specific DMPs will provide insight into what the community needs from omics specific data management tools.

A

1. Society	Documenting societal inputs and outputs (e.g. mission statements (linked to broadly accepted societal priorities e.g. SDGs EOVs, EBVs. etc), stakeholder consultation outputs, associated publications/media, links to relevant data repositories, educational materials associated with the workflow).
2. Design and logistics	Documenting plans for the experimental/observational design (e.g. cruise schedules/reports, deployment regimes, statistical replication, logistics, costing/resources/material lists)
3. Ethics and law	Documenting permits, permissions and guidelines for obtaining samples and release data. (e.g., Internationally Recognized Certificate of Compliance (IRCC) code for sampling, or relevant Due Diligence documentation if an IRCC code is not required, for countries that have signed and ratified the Nagoya protocol (Nagoya Protocol, 2010)) checklists showing compliance to ethical principle (e.g. OCAP) for ethical sampling on indigenous lands, etc
4. Data management	Protocols for data and quality management. This will include data management plans with quality assurance and quality check measures to be undertaken throughout the workflow, as well as checklists and guidance documents for the compliance with FAIR/CARE data standards; and links to all data repositories to which the omics study (meta)data should be uploaded.
5. Sampling Collection Procedure	Protocols for the physical collection of samples to be used in omics research. Including both environmental samples (e.g. water, sediments, biofilms) and organismal samples (e.g. gut contents, bulk plankton, sponge).
6. Contextual Metadata Analytical Procedures	Protocols used for the measurement of additional parameters (e.g. temperature, salinity, pH, chlorophyll etc.)
7. Sample Extraction & Purification	Protocols for the extraction and purification of biological molecules (e.g. DNA, RNA, Proteins) from environmental and/or organismal samples.
8. Omics Sequencing Procedures	Protocols for sequencing nucleic acids. This stage may contain sub-stages that occur during the sample extraction and purification stage, if so, links between the two stages should be provided.
9. Bioinformatics	Bioinformatics pipeline for processing sequence data. This can include links to github, docker etc.
10. Quantitative amplification	Protocols for the quantitative amplification of DNA and RNA. Including quantitative PCR (qPCR), reverse transcriptase quantitative PCR (rt-qPCR), nucleic acid sequence-based amplification (NASBA), Recombinase Polymerase Amplification (RPA) etc.
11. Data Analysis	Protocols for analysis of data. This will include scripts for statistical tests, modelling and data visualisations.
12. Sample archiving/biobanking	Protocols for archiving physical samples at all stages of laboratory processing (e.g. whole environmental/organismal samples, DNA/RNA/protein extracts and/or amplicons)

B

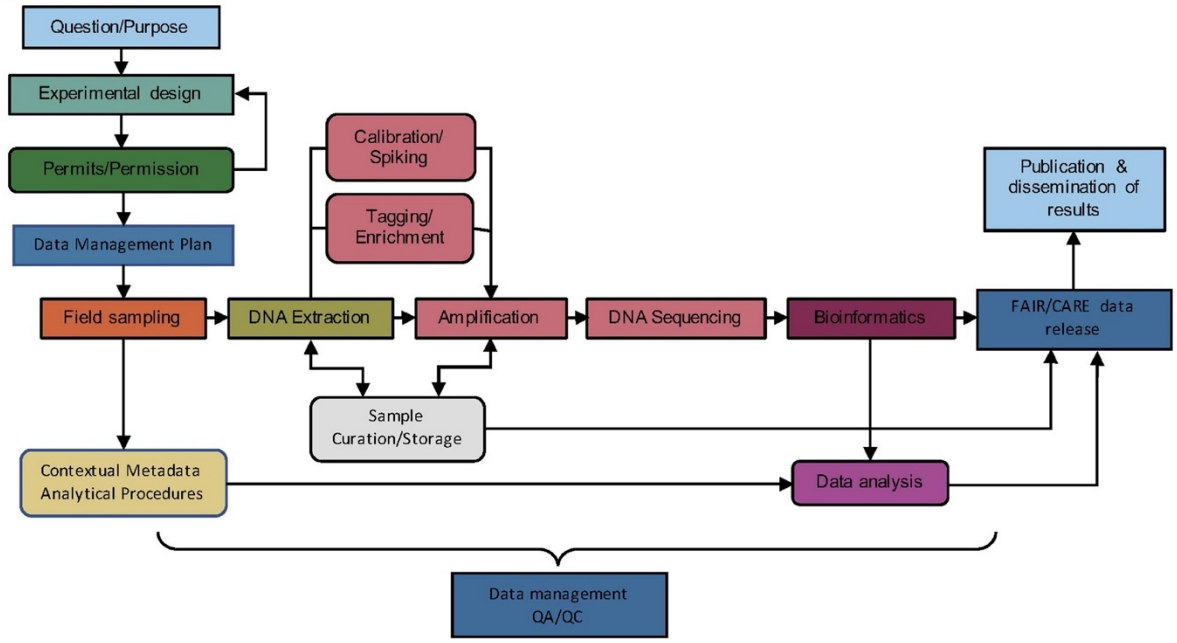


Figure 29**(A)** Proposed methodology categories to enhance exchange of ocean omics analysis knowhow. Protocols, guidelines, and other methodologies in some of these categories (such as Sample archiving/biobanking, Data Management, and Society) are cross-cutting and may apply at multiple points in the workflow. **(B)** Example workflow for a DNA metabarcoding project. Colors correspond to the methodology categories outlined in panel **(A)** and arrows indicate the order of the workflow. Square boxes show essential steps in a metabarcoding workflow, whereas rounded boxes indicate non-essential steps. Data management and QA/QC are required throughout the entire workflow.

In Figure 29B, we give an example of a DNA metabarcoding workflow, where the colour of each step corresponds to a methodology category in Figure 29A. Protocols uploaded to OBPS can be assigned (tagged) to the relevant omics categories. The granularity of protocols uploaded to the OBPS may include individual uploads for sub-stages (i.e., Tagging/Enrichment within 4, Omics sequencing procedures), or single documents spanning multiple methodology categories (i.e., 7, Sample extraction and purification, through to 9, Bioinformatics). To accommodate these levels of granularity, each upload could be tagged with single or multiple methodology category and linked to those protocols pre- and succeeding it. The granular use of methodology categories will increase modularity within the omics workflow and facilitate the mixing and matching of methods from various projects.

The interplay between the activities within and across the steps within a workflow—and how they bring value to the community and society—is complex and beyond the scope of this article; however, we have provided an initial perspective on this using the Porter's value chain approach (Porter, 1985; Supplementary Figure 30).

5.4 Essential Metadata for Omics Protocols

The targeted discovery and reuse of protocols can be improved if protocols are effectively described using standardized metadata terms on upload to OBPS and other platforms. Terms and checklists to standardize metadata about primary sequence or biodiversity data already exist [GSC's Minimum Information about any (x) Sequence checklist (MIxS; Yilmaz et al., 2011) and TDWG's Darwin Core standard (Wieczorek et al., 2012)]; however, no such standards have thus far been published for metadata about omics protocols.

Here we present initial suggestions for the Minimum Information for an Omic Protocol (MIOP), a set of ten metadata categories which could correspond to ten key decision tree questions asked to identify the relevant protocol for any project. The ten MIOP categories (Table 1) consist of five novel categories (methodology category, purpose, resources, analysis, target) and five categories already used in the GSC's MIxS (project, geographic location, broad-scale environmental context, local environmental context, and environmental medium). Each category is linked to a set of predefined keywords (metadata terms) from existing vocabularies or ontologies; except for the "project" category, which contains project names, affiliations, and contact details and the "methodology category" outlined in section "Ocean Omics Methodology Categories" (Figure 29A). Omics users would then select the most appropriate keywords for each category, assigning the terms as metadata for the protocol. This will improve the FAIRness of our protocol data, by allowing consequent users to search the protocol database using the

same set of keywords; thereby, limiting the proliferation of descriptive keywords (e.g., mapping synonyms) and increasing the findability of protocols.

Table 14 Description of keyword categories for protocol metadata and the terminologies (controlled vocabularies, thesauri, and ontologies) containing the relevant keywords.

Categories	Terminology/ ontology	Description
Methodology category	Methodology category (see **Figure 1A**)	Methodology category which the uploaded protocol belongs to. This links to the associated methodology categories which precede and succeed it in the workflow, to facilitate the linking of protocols into entire workflows, while keeping granularity and flexibility. This will enable the mixing and matching of protocol modules from various uploaded workflows.
Project	N/A	Details about the project (e.g., Name, Affiliation, website). May also includes a field for tagging any projects that protocols are compliant with (e.g., Earth Microbiome Project/TARA Oceans). Once submitted the relevant PI may be notified and could choose to endorse or reject the protocol as compliant with their project.
Purpose	EFO, OBI	Terms to describe the purpose of the omics research. [e.g., time series design (OBI:0500020) or taxonomic diversity assessment by targeted gene survey (OBI:0001960)]
Resources	EFO, NCIT	Terms to identify the key resources needed to complete the protocol [e.g., Illumina MiSeq (EFO:0004205), centrifuge (OBI:0400106)]
Analyses	EFO, OBI, and NCIT	Terms to describe the types of analyses used in the protocol [e.g., amplicon sequencing assay (OBI:0002767) or polymerase chain reaction (OBI:0002692)]
Geographic Location	GAZ	Geographic location/s in which the protocol has been used [e.g., Hawaii Ocean Time-series Site (GAZ:00187530), Western English Channel Sampling Stations (GAZ:00187525)]
Broad-scale environmental context (former Biome)	ENVO	Biome/s in which the protocol was successfully used [e.g., oceanic epipelagic zone biome (ENVO:01000033)]
Local environmental context (former Feature)	ENVO, UBERON	Environmental feature/s targeted using the protocol [e.g., seasonal thermocline (ENVO:01000107)]
Environmental medium (former Material)	MIxS environmental packages; ENVO	Identify the environmental or organismal material from which the biological molecule (e.g., DNA/RNA/Protein) was extracted [e.g., ocean water (ENVO:00002151)]
Target	NCIT, NCBITaxon, and EFO	Identify the target taxa, gene and/or molecule for the protocol [e.g., Polaribacter (NCBITaxon:1642819), 16S Mitochondrial Ribosomal RNA (NCIT:C131261)].

Terms would be added at upload and additional metadata would accumulate as the protocols are used in different settings (e.g., Geographic Locations; in the discussion see the section "Learning From Failed Practices").
EFO, Environmental Factor Ontology; OBI, Ontology for Biomedical Investigations; NCIT, NCI Thesaurus; GAZ, Gazetteer; ENVO, Environment Ontology; UBERON, Uber-anatomy ontology; NCBITaxon, NCBITaxon ontology.

5.5 Discussion

Ocean Best Practices System provides a neutral, global public repository for ocean community practices. It is a stable and persistent foundation that can host protocols themselves, or link to other protocol tools and functionalities that can (and should) continue to be developed by other organizations including the private sector. The primary function of Omics OBPS would be to publish and archive omics protocols to enhance their global visibility and discoverability, and provide stable links to the entire workflow of protocols. Expanding and improving the functionality of the OBPS for omics protocols will help the community mature by providing a structured system in which context-based best practices can be discovered and identified. A transparent and structured process for handling our omics protocols will be an essential step toward operationalizing omics observing.

Increasing protocol transparency, through detailed publication on OBPS, also means that simple cited protocol strings can become a core component of methods sections in publications. Those strings can then be harvested by machines to generate a graph of "what came before" and "what came after." When used with the decision tree recommendations this process could point out the most recent protocol development to users and would essentially provide the decision-tree resource we are aiming for. Such an approach enables "practices" (which might be defined as "protocol strings") to emerge from how protocols are actually being used in the community. Assessment of which of these practices represent a "best" practice in a given context is a distinct challenge, but not a unique one in knowledge sectors. Peer endorsement and citation metrics are two commonly employed ranking mechanisms that could also be applied here.

5.5.1 Learning From Community Preferences

Community-use metrics offer a way to capture the community's preference for certain protocols. We suggest that metrics such as times cited, user upvotes, and number of associated data records all be recorded and used to rank lists of relevant protocols. Combined with the MIOP-based grouping into methodology categories, this process will help accelerate the identification of potential best practices within each category. Narrowing down the list of relevant protocols will additionally provide the basis for more targeted and rigorous scientific comparisons between multiple potential best practices for a given scientific endeavor. Outputs of such comparisons may offer further information about the superiority of certain protocols, and could be considered in addition to the more general community-use metrics[2]. Furthermore, focusing on these community driven best practices will help to reveal protocols that are effective and convenient for a broad range of research facilities. This in turn can reduce

literature biases toward novel state of the art practices, which may not be feasible for mainstream use.

5.5.2 Learning From Failed Practices

During the initial workshop, participants outlined a desire for a best practice system to include "failed practices" and flag when a protocol may limit or eliminate a range of downstream applications. While this type of functionality would not be immediately addressed by implementing MIOP metadata, there would be potential for users to provide feedback for protocols using MIOP metadata and Boolean operators. For example, if a protocol, originally designed for seawater, was used with freshwater samples, the user could upload additional MIOP metadata using "AND freshwater" if the protocol was successful or "NOT freshwater" if unsuccessful. Thereby, broadening the findability of successful protocols and documenting potential limitations to be aware of. Documenting these failed attempts has the potential to save both time and resources.

5.5.3 Promoting Collaborative Omic Networks

Minimum Information for an Omic Protocol may additionally promote collaboration between groups. For example, the "Project" category is an administrative metadata field that will describe the project (study or program) for which the protocol was developed, including contact details and affiliated institution. To create links between similar projects and facilitate collaboration, it would be possible to introduce an option to tag a protocol as compliant with pre-existing projects. In such cases, a notification could be sent to the PI of the lead project, allowing them to add or reject the protocol to their list of compliant protocols. Protocols linked this way could form overarching protocol concepts, which may contain a variety of versions and accepted, cross-comparable protocols that include minor adaptations to make them suitable in different circumstances.

An endorsement process for a global observation network has already been developed by Global Ocean Observing System (GOOS) in cooperation with OBPS, to encourage standardized methods for global observations and for reporting on GOOS' Essential Ocean Variables (EOVs) (Miloslavich et al., 2018; Hermes, 2020). To gain this endorsement, protocols will have to undergo a rigorous community review process that will be strengthened if there is a large source of omics protocols to compare with on the OBPS. Standardized practices and official endorsements are likely to become increasingly valuable as countries begin to use legislation to make biodiversity targets legally binding. Any omic method used to measure biodiversity impacts will need to undergo legal scrutiny if it is used as evidence of a country/organization

meeting or failing to meet biodiversity targets. Therefore, protocols officially endorsed through international programmes, such as GOOS, are likely to hold more sway legally. Broad participation from the omics community in open sharing and reviewing of protocols on the OBPS will help to ensure that community endorsed best practices are representative of the wider community needs and not only focused on expensive state of the art methodologies.

5.5.4 Machine Readability

Machine readable tracking of protocol versions presents an opportunity to visually map the progression of protocols by linking all versions to a "concept," as implemented in Zenodo and GitHub. Like software, omic protocols may be updated, corrected, and improved necessitating forms of version control and tracking, such as the use of semantic versioning (Hörstmann et al., 2020; Preston-Werner, 2021). Implementing this would help to increase recognition for the scientists/technicians/students involved in protocol development through citable documentation of their contributions.

Machine-readable and machine-actionable protocols are becoming more important as autonomous technologies evolve. Devices such as the Environmental Sample Processor (ESP) and the Robotic Cartridge Sampling Instrument (RoCSI) are currently being used and developed for autonomous collection, preservation, and *in situ* analysis of omics samples (Yamahara et al., 2019; National Oceanography Centre, 2021). Eventually, smart sensing platforms using these technologies will be able integrate data from various sensors and satellites to implement adaptive sampling regimes or extraction protocols based on real-time environmental observations (Whitt et al., 2020). To reach this goal a variety of protocols will need to be translated into a machine actionable format using common workflow language. A systematic review of protocols will help to devise such machine actionable formats and protocol templates may help to bridge the gap between lab-based protocol development and *in situ* autonomous use.

5.6 Conclusion

Multiple groups within the omics community are actively developing best practices for their field. To ensure that all these efforts are effectively utilized, a concerted and community wide effort will be needed to gather and organize these practices. By harnessing the OBPS infrastructure and further developing the MIOP metadata we can: (1) allow protocols to be searched for within a decision tree framework; (2) establish a system that encourages the systematic review of protocols; and (3) reveal community preferences through the accumulation of community use data. Taking these steps toward a structured and global public

repository of omics protocols will increase transparency and streamline biomolecular ocean observing research to foster the collaborative networks needed to achieve global scale biodiversity observations.

5.7 Author's Note

This manuscript has been released as a preprint at https://zenodo.org/record/5482852#.YVNeo55KjAM (Samuel et al., 2021).

5.8 Data Availability Statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s.

5.9 Author Contributions

AW, CM, and RS constructed the main text figure with input from all authors. RM, PB, and RS developed the supplementary figure. All authors contributed to the discussion and wrote the manuscript.

5.10 Funding

RS was supported by the Natural Environment Research Council (NERC) NEXUSS Studentship (grant number NE/N012070/1). RM's contributions were supported by the European Union's Horizon 2020 Research and Innovation Programmes under grant agreement N° 862923, project AtlantECO (Atlantic ECOsystem assessment, forecasting and sustainability), and grant agreement N° 862626, project EuroSea (Improving and Integrating European Ocean Observing and Forecasting Systems for Sustainable use of the Oceans). NJ was supported by the Fisheries and Oceans Canada Marine Conservation Targets programme. CP was supported by BIOIMAGING-GR (MIS 5002755) implemented under "Action for Strengthening Research and Innovation Infrastructures" funded by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014–2020) and co-financed by Greece and the European Union (European Regional Development Fund), ECCO (project ID 343) funded by the 1st call for the support of Post-doctoral Researchers by the Hellenic Foundation for Research and Innovation (HFRI) and FutureMARES (grant agreement N° 869300) funded by the European Union's Horizon 2020 Research and Innovation Programme. In addition, CP's research is co-financed by Greece and the European Union (European Social Fund- ESF) through the Operational Programme «Human Resources Development, Education and Lifelong Learning» in

the context of the project “Reinforcement of Postdoctoral Researchers—2nd Cycle” (MIS-5033021), implemented by the State Scholarships Foundation (IKY). JK was supported by the CSIRO Environomics Future Science Platform and the Australian Microbiome Initiative. The Australian Microbiome initiative was supported by funding from Parks Australia through the Bush Blitz program funded by the Australian Government and BHP, Bioplatforms Australia and the Integrated Marine Observing System (IMOS) through the Australian Government’s National Collaborative Research Infrastructure Strategy (NCRIS), and the CSIRO. KP was supported by the David and Lucile Packard Foundation and the National Aeronautics and Space Administration (NASA) grants 80NSSC20M0001 and 80NSSX21M003.

5.11 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

5.12 Publisher’s Note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

5.13 Acknowledgments

We would like to thank participants and organisers from the OBPS workshop IV omics/eDNA working group (Simpson et al., 2021), who have contributed to discussions on how to create an ocean best practice system to serve the biomolecular omics community.

5.14 Footnotes

^ Currently, the protocol categories focus on genomics and transcriptomics but we expect this list to expand with further input from the broader omics community, particularly in areas such as proteomics and metabolomics.

^ In certain cases (e.g., for contributing to a standardized global sampling scheme) it may not be about which method is “best,” but about which method delivers reliable results while being applicable throughout all regions of the ocean and inclusive of lower capacity research activities.

5.15 References

Aylagas, E., Borja, A., Pochon, X., Zaiko, A., Keeley, N., Bruce, K., et al. (2020). Translational molecular ecology in practice: linking DNA-based methods to actionable marine environmental management. *Sci. Tot. Environ.* 744:140780. doi: 10.1016/j.scitotenv.2020.140780

Berry, O., Jarman, S., Bissett, A., Hope, M., Paeper, C., Bessey, C., et al. (2020). Making environmental DNA (eDNA) biodiversity records globally accessible. *Environ. DNA* 3, 699–705. doi: 10.1002/edn3.173

Bissett, A., Fitzgerald, A., Meintjes, T., Mele, P., Reith, F., Dennis, P., et al. (2016). Introducing BASE: the Biomes of Australian Soil Environments soil microbial diversity database. *GigaScience* 5:21. doi: 10.1186/s13742-016-0126-5

Brown, M., van de Kamp, J., Ostrowski, M., Seymour, J., Ingleton, T., Messer, L., et al. (2018). Systematic, continental scale temporal monitoring of marine pelagic microbiota by the australian marine microbial biodiversity initiative. *Sci. Data* 5:180130. doi: 10.1038/sdata.2018.130

Buttigieg, P., Appeltans, W., Canonico, G., Chavez, F. P., Deck, J., Glöckner, F. O., et al. (2019). "Building a strategy towards an omic biodiversity observation network (Omic BON)," in *Proceedings of the OceanObs'19 conference*, Honolulu.

Carroll, S., Garba, I., Figueroa-Rodríguez, O., Holbrook, J., Lovett, R., Materechera, S., et al. (2020). The CARE Principles for indigenous data governance. *Data Sci. J.* 19:43. doi: 10.5334/dsj-2020-043

Davies, N., Field, D., Amaral-Zettler, L., Clark, M., Deck, J., Drummond, A., et al. (2014). The founding charter of the genomic observatories network. *GigaScience* 3:2. doi: 10.1186/2047-217x-3-2

Davies, N., Meyer, C., Gilbert, J., Amaral-Zettler, L., Deck, J., Bicak, M., et al. (2012). A call for an international network of genomic observatories (GOs). *GigaScience* 1:5. doi: 10.1186/2047-217x-1-5

Dickie, I., Boyer, S., Buckley, H., Duncan, R., Gardner, P., Hogg, I., et al. (2018). Towards robust and repeatable sampling methods in eDNA -based studies. *Mol. Ecol. Resour.* 18, 940–952. doi: 10.1111/1755-0998.12907

Gilbert, J., Jansson, J., and Knight, R. (2014). The Earth Microbiome project: successes and aspirations. *BMC Biol.* 12:69. doi: 10.1186/s12915-014-0069-1

Hermes, J. (2020). *GOOS Best Practices Endorsement Process. Version 1.* Paris: Global Ocean Observing System.

Hörstmann, C., Buttigieg, P., Simpson, P., Pearlman, J., and Waite, A. (2021). Perspectives on documenting methods to create ocean best practices. *Front. Mar. Sci.* 7:556234. doi: 10.3389/fmars.2020.556234

Hörstmann, C., Buttigieg, P. L., Simpson, P., Pearlman, J., Karstensen, J., and Waite, A. M. (2020). *Towards a Best Practice for Developing Best Practices in Ocean Observation (BP4BP): Supporting Methodological Evolution through Actionable Documentation.* Paris: UNESCO.

Makiola, A., Compson, Z., Baird, D., Barnes, M., Boerlijst, S., Bouchez, A., et al. (2020). Key questions for next-generation biomonitoring. *Front. Environ. Sci.* 7:197. doi: 10.3389/fenvs.2019.00197

Miksa, T., Simms, S., Mietchen, D., and Jones, S. (2019). Ten principles for machine-actionable data management plans. *PLoS Comput. Biol.* 15:e1006750. doi: 10.1371/journal.pcbi.1006750

Miloslavich, P., Bax, N., Simmons, S., Klein, E., Appeltans, W., Aburto-Oropeza, O., et al. (2018). Essential ocean variables for global sustained observations of biodiversity and ecosystem changes. *Glob. Change Biol.* 24, 2416–2433. doi: 10.1111/gcb.14108

Mitchell, A., Almeida, A., Beracochea, M., Boland, M., Burgin, J., Cochrane, G., et al. (2019). MGnify: the microbiome analysis resource in 2020. *Nucleic Acids Res.* 48, D570–D578. doi: 10.1093/nar/gkz1035

Nagoya Protocol (2010). *Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity: (Chapter XXVII: Environment).* Nagoya: United Nations Treaty Collection.

National Oceanography Centre (2021). *Auto-Sampler Device and Method of Autosampling.* Paris: National Oceanography Centre.

Porter, M. (1985). *Competitive Advantage: Creating and Sustaining Superior Performance.* New York, NY: Free Press.

Preston-Werner, T. (2021). *Semantic Versioning 2.0.0. Semantic Versioning.* Available online at: https://semver.org/ (accessed July 28, 2021).

Samuel, R. M., Meyer, R., Buttigieg, P. L., Davies, N., Jeffery, N. W., Meyer, C., et al. (2021). Towards a global public repository of community protocols to encourage best practices in biomolecular ocean observing and research. *Zenodo* [Preprint]. doi: 10.5281/zenodo.5482852

Simpson, P., Pearlman, F., and Pearlman, J. (eds) (2021). *Evolving and Sustaining Ocean Best Practices Workshop IV, 18; 21-25 & 30 Sep 2020 [Online]: Proceedings, Volumes 1 & 2.* IOC Workshop Report No. 294. Paris: UNESCO.

Sunagawa, S., Acinas, S., Bork, P., Bowler, C., Eveillard, D., Gorsky, G., et al. (2020). Tara Oceans: towards global ocean ecosystems biology. *Nat. Rev. Microbiol.* 18, 428–445. doi: 10.1038/s41579-020-0364-5

Thompson, L., Sanders, J., McDonald, D., Amir, A., Ladau, J., Locey, K., et al. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551, 457–463. doi: 10.1038/nature24621

van de Kamp, J., Brown, M., Ostrowski, M., Seymour, J., Smith, M., Raes, E., et al. (2019). The Australian Microbiome Initiative-towards microbial omics at the continental scale. *Geophys. Res. Abstr.* 21:88.

Wetterstrand, K. (2020). *The Cost of Sequencing a Human Genome [WWW Document].* Available online at: www.genome.gov/about -genomics/fact-sheets/Sequencing-Human-Genome-cost (accessed March 28, 2021).

Whitt, C., Pearlman, J., Polagye, B., Caimi, F., Muller-Karger, F., Copping, A., et al. (2020). Future vision for autonomous ocean observations. *Front. Mar. Sci.* 7:697. doi: 10.3389/fmars.2020.00697

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., et al. (2012). Darwin core: an evolving community-developed biodiversity data standard. *PLoS One* 7:e29715. doi: 10.1371/journal.pone.0029715

Wilkinson, M., Dumontier, M., Aalbersberg, I., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018. doi: 10.1038/sdata.2016.18

Yamahara, K., Preston, C., Birch, J., Walz, K., Marin, R., Jensen, S., et al. (2019). In situ autonomous acquisition and preservation of marine environmental DNA using an autonomous underwater vehicle. *Front. Mar. Sci.* 6:373. doi: 10.3389/fmars.2019.00373

Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J., Amaral-Zettler, L., et al. (2011). Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *Nat. Biotechnol.* 29, 415–420. doi: 10.1038/nbt.1823

Keywords: omics, eDNA, ocean best practices, ocean observations, metadata, protocol management, methods

Citation: Samuel RM, Meyer R, Buttigieg PL, Davies N, Jeffery NW, Meyer C, Pavloudi C, Pitz KJ, Sweetlove M, Theroux S, van de Kamp J and Watts A (2021) Toward a Global Public Repository of Community Protocols to Encourage Best Practices in Biomolecular Ocean Observing and Research. *Front. Mar. Sci.* 8:758694. doi: 10.3389/fmars.2021.758694

Received: 14 August 2021; **Accepted:** 21 September 2021;
Published: 11 October 2021.

Edited by:

Stéphane Pesant, European Bioinformatics Institute (EMBL-EBI), United Kingdom

Reviewed by:

Gerhard Josef Herndl, University of Vienna, Austria
Marine Landa, University of California, Santa Cruz, United States

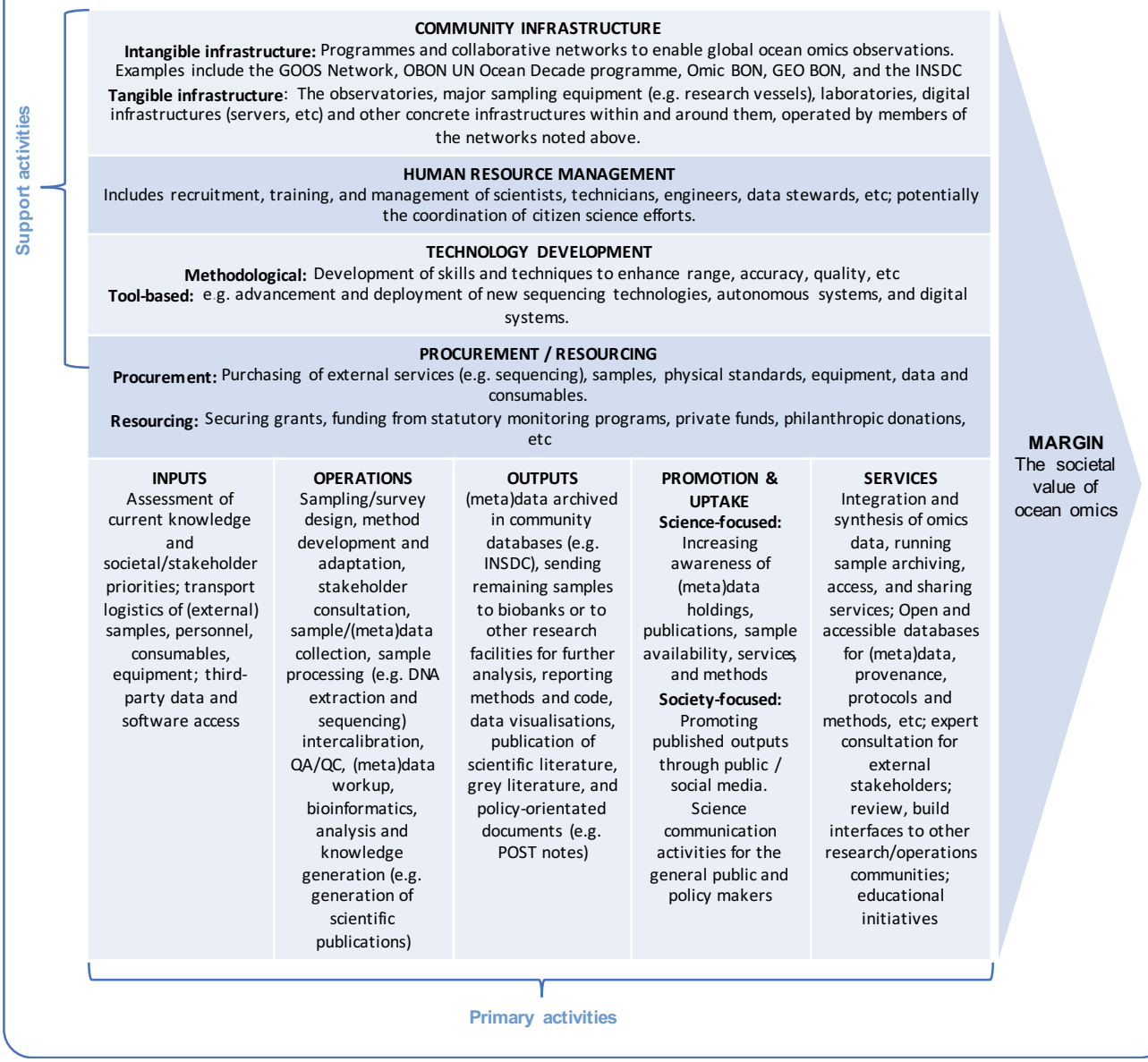
***Correspondence:** Robyn M. Samuel, r.m.samuel@soton.ac.uk

Appendix D Supplementary Material

D.1 Supplementary Figures

Ocean Omics Value Chain

Applying Porter's value chain concept to ocean omics research can help to systematically identify activities that will increase the value of ocean omics research to society at large: A well-functioning ocean omics value chain will generate more downstream value than the costs needed to operate it. This is the basis of justifying increases or decreases in investment in any activity along or across the chain. We believe such mechanisms will help the omics community coherently reflect on the activities needed to advance our methods while ensuring a healthy cost/benefit ratio, especially as we interface with other ocean communities and their own value chains. For example, when considering the uptake of new technologies (e.g. new sequencing technologies), value chain analysis helps to consider whether uptake is worth the disruption to the continuity of long-term ocean omics observatories. The process helps to evaluate what value is compromised (i.e. Are the time series broken? Are the POST notes valid anymore?) and what needs co-investment to mitigate it (e.g. intercalibration across space and time). Essentially answering "is it worth it?" at scale, or should it be a more isolated test by a smaller community (e.g. a scientific team/pilot study).



Supplementary Figure 30. We frame part of our perspective in an adaptation of Porter's (1985) value chain, where a product passes through all portions of the chain, gaining value

from each activity. The language used in Porter's value chain relates to commercial business activities, this adaptation gives examples for the types of omic research activities that could apply to each category. Category headings have been adapted to fit omics operations and research, as follows: Firm infrastructure → Community infrastructure, Procurement → Procurements / Resourcing, Inbound logistics → Inputs, Outbound logistics → Outputs, Marketing & Sales → Promotion & Uptake (note that original headings would be appropriate for omics-focused businesses). It should also be noted that value within omics operations and research does not only refer to monetary transactions (for example, procurement may be facilitated by credit on scientific publications).

Chapter 6 Shipboard shotgun sequencing of eDNA from the Oxygen Minimum Zone at the Porcupine Abyssal Plain Sustained Observatory

Robyn Samuel[1] and CR Young[2]

[1]Ocean and Earth Sciences, University of Southampton Waterfront Campus, Southampton UK

[2]National Oceanography Centre, Southampton, England, United Kingdom

6.1 Abstract

During the DY103 cruise aboard the RRS Discovery, we tested the suitability of the MinION sequencing platform (Oxford Nanopore Technologies) for near real-time, on-board metagenomic sequencing. To enable this, we used a modified DNA extraction protocol tailored for shipboard DNA purification. Here, we present results from shotgun sequencing of an environmental DNA sample collected from the oxygen minimum zone (OMZ) at the Porcupine Abyssal Plain Sustained Observatory (PAP-SO). Sequencing was completed entirely onboard, using the MinION platform. The metagenome consisted of 43,820 sequences with an average length of 1,683 bps and average GC content was 39%. Taxonomic annotations revealed that 93% of taxa assigned to bacteria and archaea and 60% of function annotations assigned to metabolic functions, reflecting the expected characteristics of OMZ microbial communities. We found that MinION sequencing onboard a research vessel is feasible and doing so allows ecological data to be visualised in near-real-time. Having this metagenomic data available at sea will create an opportunity for adaptive sampling strategies to maximise scientific outputs based on the current ecological conditions.

6.2 Introduction

The Porcupine Abyssal Plain sustained observatory (PAP-SO) is located in the Northeast Atlantic (49°N 16.5°W) at a water depth of 4800 m. PAP-SO is an open-ocean multidisciplinary observatory that has produced high-resolution datasets integrating environmental and ecologically relevant variables from the surface to the seabed since mid-1980's (Hartman et al. 2012). The observatory now consists of a full-depth mooring, with autonomous sensors measuring temperature, salinity, chlorophyll-a fluorescence, nitrate, and pCO_2 and a surface buoy, for simultaneous meteorological and ocean variable monitoring (Hartman et al. 2012). PAP-SO is the longest running sustained observatory in the oceans around Europe and has

played an integral role in the long-term monitoring of both essential climatic variables (ECVs) and essential ocean variables (EOVs; Hartman et al. 2012).

Incorporating genomics data into the global sustained observatory network has garnered considerable interest as a means to increase the temporal, spatial, and taxonomic resolution of biodiversity monitoring (Goodwin et al. 2018; Miloslavich et al. 2018; Djurhuus et al. 2020). Historically, biodiversity monitoring has lagged behind the monitoring of physical and biogeochemical properties, primarily due to the time and expense of carrying out ecological surveys based on traditional morphological identification(Muller-Karger et al. 2018). Genomic approaches to assessing marine biodiversity offer a promising solution to reduce this gap, as they can be carried out rapidly using sequencing technologies that are becoming increasingly more affordable and portable (Tyler et al. 2018).

Genomic analysis can provide in depth data on taxonomy and gene functionality, helping to unravel some of the complexities of biogeochemical cycling (McCarren et al. 2010). Consequently, metagenomic data is now recognised as an essential biodiversity variable (EBV) (Muller-Karger et al. 2018). Recent developments in ecogenomic sensors make it possible to collect near real-time ecological data with in-situ genetic assays, such as in-situ qPCR and sandwich hybridization assays (Ussler et al. 2013; Bowers et al. 2017; McQuillan and Robidart 2017). It is likely that portable sequencing devices, such as Oxford Nanopore Technology's MinION, will soon be integrated into in-situ ecogenomic sensors as well. The combination of these technologies has the potential to trigger a step-change in the quantity of metagenomic data collected at ocean observatories.

For the last 5 years eDNA samples have routinely been collected from the water column and sediments at PAP-SO, with sample processing and sequencing completed post cruise. Survey designs are typically planned in advance, although some refinements can be made in response to the depth profiles from physical and biogeochemical sensors deployed alongside the CTD Rossette. Onboard sequencing during the cruise enables near-real-time visualisation of ecological data, supporting adaptive sampling strategies informed by current ecological conditions, as well as physical, and biogeochemical conditions. The MinION sequencing platform (Oxford Nanopore Technologies) makes onboard sequencing feasible due to its low cost, compact size, and portability (Tyler et al. 2018). This dataset was collected as an initial test of onboard sequencing feasibility during PAP-SO research cruises.

Methods

The water sample was collected on 24th June 2019 by CTD rosette cast. Sample DY103-002-N16 (N 49.0001°, W -16.5004°, Figure 31) collected at 850m from the oxygen minimum zone (OMZ;

Figure 32) was selected for sequencing as it had the largest volume (15L) remaining after all other on-board chemical analyses were complete. The sample was filtered through a 0.22µm Sterivex cartridge filter using a Masterflex L/S Digital Precision peristaltic pump (Cole-Parmer, Saint Neots, UK). DNA was extracted on board using a modified version of the Qiagen AllPrep DNA extraction kit. The addition of β-mercaptoethanol (β-ME) was omitted from the initial lysis step; β-ME is used to denature ribonucleases and is therefore only necessary for the extraction of RNA. Centrifuge times were tripled to compensate for the reduced centrifugal force of the low power, portable mini centrifuge available for use onboard. DNA concentration and purity were quantified onboard with a NanoDrop Nanovue Plus.

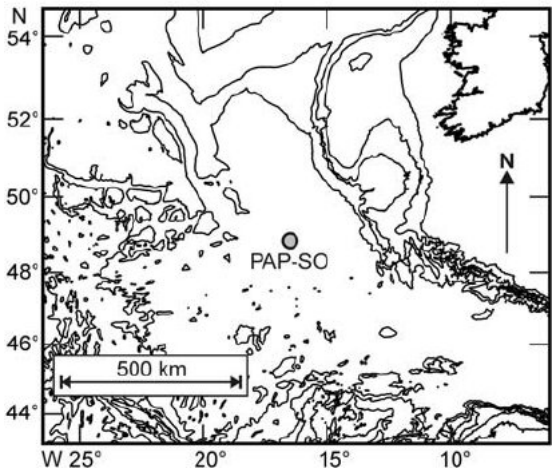


Figure 31 Location map of the Porcupine Abyssal Plain Sustained Observatory (PAP-SO). The depth contours shown are 200, 1000, 2000, 3000, and 4000 m (image from Hartman et al. 2012)

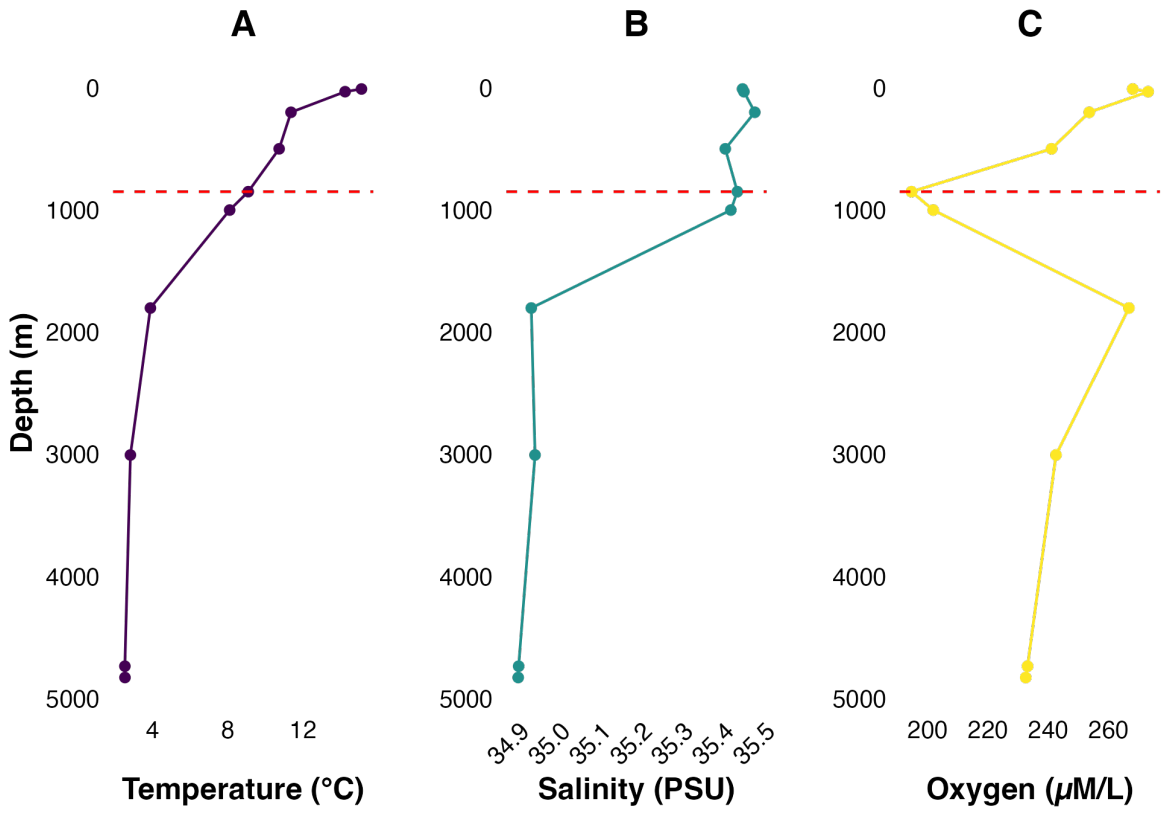


Figure 32 CTD depth profiles from station DY103-002 showing (A) temperature measured by the SBE 11Plus CTD, (B) salinity analysed on board using a Guildline Autosal 8400B, and (C) dissolved oxygen measured with an SBE 43 sensor. Red dashed line: sequencing sample depth (850 m).

The standard Rapid Sequencing Kit protocol (SQ-RAD004) was followed to generate a sequencing library for onboard MinION sequencing. Sequence reads were annotated using the MG-RAST pipeline (Meyer et al. 2008, 2018) with a minimum alignment length of 30, a minimum e-value of 1-e^{-5}, minimum percentage identity at 60% and a minimum abundance of 1, following the recommendations outlined in Randle-Boggis et al. (2016). Taxonomy was assigned to genus level using the RefSeq database (Pruitt et al. 2007) and functional annotations were made using KEGG orthologous groups (KO) database (Kanehisa and Goto 2000). Krona plots were used to interactively visualise hierarchical taxonomy and functional classifications (Ondov et al. 2011).

6.3 Results

DNA yield was approximately 45 ng/μl with a 260/280 and 260/230 ratio of 1.6 and 0.3 respectively. DNA quality was lower than recommended for MinION sequencing; therefore, optimization of a suitable onboard DNA extraction protocol is recommended. None the less, MinION sequencing ran for 15 hours and produced a dataset consisting of 43,820 sequences totalling 73,751,445 base pairs with an average length of 1,683 bps. Average GC content was

39%. A total of 646 sequences (1%) contain ribosomal RNA genes, 8,610 sequences (20%) contain predicted proteins with known functions, and 34,564 sequences (79%) contain predicted proteins with unknown function. Of the sequences that could be assigned to a single taxonomic entity across the entire length of the sequence, 3,540 were assigned a taxonomic classification of Bacteria, 1,346 were assigned to Archaea, 897 to Eukaryota, and 15 to viruses (Figure 33A). Of the 4655 functional annotations, 2801 related to metabolism, 1134 to Genetic Information Processing, 411 to Environmental Information Processing, 181 to cellular processes, 97 to human diseases and 31 to organismal systems (Figure 33B).

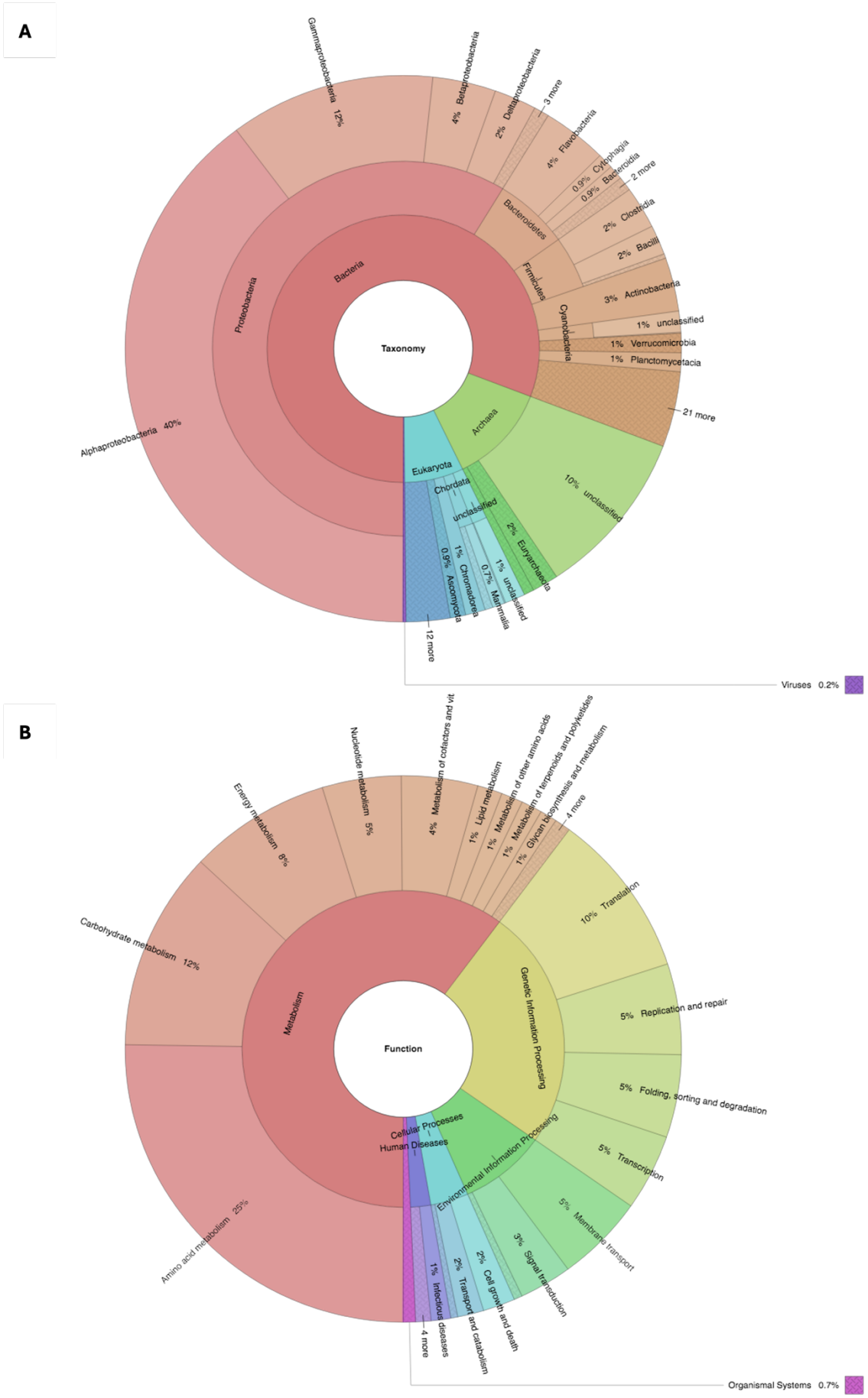

A
Taxonomy
Bacteria
Archaea
Eukaryota
Proteobacteria
Bacteroidetes
Firmicutes
Cyanobacteria
Chordata
unclassified
Alphaproteobacteria 40%
Gammaproteobacteria 12%
Betaproteobacteria 4%
Deltaproteobacteria 2%
3 more
Flavobacteria 4%
Cytophagia 0.9%
Bacteroidia 0.9%
2 more
Clostridia 2%
Bacilli 2%
Actinobacteria 3%
unclassified 1%
Verrucomicrobia 1%
Planctomycetacia 1%
21 more
10% unclassified
Euryarchaeota 2%
unclassified 1%
Mammalia 0.7%
Chromadorea 1%
Ascomycota 0.9%
12 more
Viruses 0.2%
B
Function
Metabolism
Genetic Information Processing
Environmental Information Processing
Cellular Processes
Human Diseases
Carbohydrate metabolism 12%
Energy metabolism 8%
Nucleotide metabolism 5%
Metabolism of cofactors and vit 4%
Lipid metabolism 1%
Metabolism of other amino acids 1%
Metabolism of terpenoids and polyketides 1%
Glycan biosynthesis and metabolism
4 more
Amino acid metabolism 25%
10% Translation
5% Replication and repair
5% Folding, sorting and degradation
5% Transcription
5% Membrane transport
3% Signal transduction
2% Cell growth and death
2% Transport and catabolism
1% Infectious diseases
4 more
Organismal Systems 0.7%

Figure 33 Hierarchical display metagenomic shot-gun sequencing of eDNA from the oxygen minimum zone at the Porcupine Abyssal Plain. **A** Taxonomic sequence classification to class level. **B** Functional gene annotations to two levels.

6.4 Discussion

MinION sequencing revealed a diverse microbial community, with 93% of taxa assigned to bacteria and archaea. Annotated functional genes were predominantly (60%) assigned to metabolism. The diverse microbial communities and metabolic approaches identified here are typical of oxygen minimum zones (Wright et al. 2012). OMZs occur when detritus, formed of predominantly dead organisms, phytodetritus, and faecal matter, sink down from a productive euphotic zone. This downward flux of organic matter is colonised by diverse microbial communities which aerobically metabolise the organic matter and respire oxygen in the process, resulting in lower ambient oxygen concentrations.

Metagenomic sequencing with the MinION sequencer enables onboard characterisation of community composition and gene functionality. Access to this near real-time ecological data, makes it possible to adjust sampling strategies at sea and maximize scientific outputs. Furthermore, the compact design of the MinION sequencer presents a promising opportunity to incorporate the device into existing ecogenomic sensors to enable remote *in situ* sequencing.

Currently, genetic assays used in ecogenomic sensors only target specific taxa or functional genes. Incorporating broad-range sequencing to the suite of genetic analyses available for ecogenomic sensors would expand the scope of ecological monitoring by providing additional non-targeted data. This additional data could enhance our ability to detect rare taxa and understand community dynamics across trophic levels.

However, the use of non-targeted sequencing data at sea presents some technical and ethical challenges. In this study, a small number of spurious taxonomic assignments were observed, including reads classified as Monotremata (n=7) and Anura (n=2). These are likely resulting from sequencing errors or limitations in the reference databases, which are particularly limited for open-ocean species (Yang et al. 2024). Furthermore, long-read sequencing can inadvertently capture human DNA, presenting potential privacy and ethical concerns (Whitmore et al. 2023). This dataset included 17 reads assigned to primates, presumably human contamination. Current best practices encourage the submission of all raw reads to public repositories, such as the INSDC. However, due to the potential for sensitive information to be included unintentionally, additional data screening steps will be necessary before submission. For this reason, this dataset has not been uploaded to the INSDC.

The development of devices capable of remote *in situ* sequencing has the potential to increase the spatial and temporal resolution of EBVs. By integrating metagenomic data collection into routine monitoring of physical and geochemical measurements, we can begin to bridge the data gap between ecological and environmental datasets.

6.5 Acknowledgements

We thank the Captain and crew of the RRS Discovery. C.R.Y was supported by UK Natural Environment Research Council Funding Calibrating eDNA Tools for Biodiversity Monitoring in the Ocean (grant number NE/N006496/1) and National Capability funding to the National Oceanography Centre, as part of the Climate Linked Atlantic Section Science (CLASS) programme (grant number NE/R015953/1).

6.6 References

Bowers HA, Marin R, Birch JM, Scholin CA (2017) Sandwich hybridization probes for the detection of Pseudo-nitzschia (Bacillariophyceae) species: An update to existing probes and a description of new probes. Harmful Algae 70:37–51. doi: 10.1016/j.hal.2017.10.005

Djurhuus A, Closek CJ, Kelly RP, Pitz KJ, Michisaki RP, Starks HA, Walz KR, Andruszkiewicz EA, Olesin E, Hubbard K, Montes E, Otis D, Muller-Karger FE, Chavez FP, Boehm AB, Breitbart M (2020) Environmental DNA reveals seasonal shifts and potential interactions in a marine community. Nat Commun 11:1–9. doi: 10.1038/s41467-019-14105-1

Goodwin KD, Muller-Karger FE, Djurhuus A, Allen LZ, Allen AE, McCrow JP, Canonico Hyde G (2018) Molecular approaches for an operational marine biodiversity observation network, Second Edi. Elsevier Ltd.

Hartman SE, Lampitt RS, Larkin KE, Pagnani M, Campbell J, Gkritzalis T, Jiang Z, Pebody CA, Ruhl HA, Gooday AJ, Bett BJ, Billett DSM, Provost P, Mclachlan R, Turton JD, Lankester S (2012) Marine Science. 69:776–783.

Kanehisa M, Goto S (2000) Comprehensive gene and pathway analysis of cervical cancer progression. Nucleic Acids Res 28:27–30. doi: 10.3892/ol.2020.11439

McCarren J, Becker JW, Repeta DJ, Shi Y, Young CR, Malmstrom RR, Chisholm SW, DeLong EF (2010) Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. Proc Natl Acad Sci U S A 107:16420–16427. doi: 10.1073/pnas.1010732107

McQuillan JS, Robidart JC (2017) Molecular-biological sensing in aquatic environments: recent developments and emerging capabilities. Curr Opin Biotechnol 45:43–50. doi: 10.1016/j.copbio.2016.11.022

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA (2008) The metagenomics RAST server - A public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9:1–8. doi: 10.1186/1471-2105-9-386

Meyer F, Bagchi S, Chaterji S, Gerlach W, Grama A, Harrison T, Paczian T, Trimble WL, Wilke A (2018) MG-RAST version 4-Lessons learned from a decade of low-budget ultra-high-throughput metagenome analysis. Brief Bioinform 20:1151–1159. doi: 10.1093/bib/bbx105

Miloslavich P, Bax NJ, Simmons SE, Klein E, Appeltans W, Aburto-Oropeza O, Andersen Garcia M, Batten SD, Benedetti-Cecchi L, Checkley DM, Chiba S, Duffy JE, Dunn DC, Fischer A, Gunn J, Kudela R, Marsac F, Muller-Karger FE, Obura D, Shin YJ (2018) Essential ocean variables for global sustained observations of biodiversity and ecosystem changes. Glob Chang Biol 24:2416–2433. doi: 10.1111/gcb.14108

Muller-Karger FE, Miloslavich P, Bax NJ, Simmons S, Costello MJ, Pinto IS, Canonico G, Turner W, Gill M, Montes E, Best BD, Pearlman J, Halpin P, Dunn D, Benson A, Martin CS, Weatherdon L V., Appeltans W, Provoost P, Klein E, Kelble CR, Miller RJ, Chavez FP, Iken K, Chiba S, Obura D, Navarro LM, Pereira HM, Allain V, Batten S, Benedetti-Checchi L, Emmett Duffy J, Kudela RM, Rebelo LM, Shin Y, Geller G (2018) Advancing marine biological observations and data requirements of the complementary Essential Ocean Variables (EOVs) and Essential Biodiversity Variables (EBVs) frameworks. Front Mar Sci 5:1–15. doi: 10.3389/fmars.2018.00211

Ondov BD, Bergman NH, Phillippy AM (2011) Interactive metagenomic visualization in a Web browser.

Pruitt KD, Tatusova T, Maglott DR (2007) NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 35:61–65. doi: 10.1093/nar/gkl842

Randle-Boggis RJ, Helgason T, Sapp M, Ashton PD (2016) Evaluating techniques for metagenome annotation using simulated sequence data. FEMS Microbiol Ecol 92:1–15. doi: 10.1093/femsec/fiw095

Tyler AD, Mataseje L, Urfano CJ, Schmidt L, Antonation KS, Mulvey MR, Corbett CR (2018) Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome

Sequencing Applications. Sci Rep 8:1–12. doi: 10.1038/s41598-018-29334-5

Ussler W, Preston C, Tavormina P, Pargett D, Jensen S, Roman B, Marin R, Shah SR, Girguis PR, Birch JM, Orphan V, Scholin C (2013) Autonomous application of quantitative PCR in the deep sea: In Situ surveys of aerobic methanotrophs using the deep-sea environmental sample processor. Environ Sci Technol 47:9339–9346. doi: 10.1021/es4023199

Whitmore, L., McCauley, M., Farrell, J. A., Stammnitz, M. R., Koda, S. A., Mashkour, N., ... & Duffy, D. J. (2023). Inadvertent human genomic bycatch and intentional capture raise beneficial applications and ethical concerns with environmental DNA. *Nature ecology & evolution*, *7*(6), 873-888.

Wright JJ, Konwar KM, Hallam SJ (2012) Microbial ecology of expanding oxygen minimum zones. Nat Rev Microbiol 10:381–394. doi: 10.1038/nrmicro2778

Yang, N., Jin, D., & Govindarajan, A. F. (2024). Applying environmental DNA approaches to inform marine biodiversity conservation: the case of the ocean twilight zone. *Marine Policy*, *165*, 106151.

Chapter 7 Conclusion: Autonomous eDNA sampling for UK Marine Biomolecular Monitoring

7.1 Introduction

On land, ecological changes are often visible or audible: we might notice a new bird species frequenting our garden or notice fewer insects on our car compared to a decade ago. These sensory observations spark curiosity and drive the scientific investigation to explore their underlying causes. In the marine realm, however, such direct perception is far more limited, and many ecological shifts may go unnoticed. Environmental DNA (eDNA) monitoring provides a means to detect these hidden ecological changes, providing insights into marine biodiversity and ecological trends that would otherwise remain invisible.

Traditionally, our understanding of anthropogenic impacts on marine ecology has come from micro- and mesocosm experiments, which do not accurately represent natural communities or their rates of change (Reiber et al., 2022). To understand the real-word consequences of combined effects of anthropogenic impacts, place-based monitoring is required to observe their cumulative effects. Empirical observations are needed to evidence the effects of adaptive

management strategies. Achieving this requires high-resolution spatial and temporal ecological datasets, obtained through long-term monitoring in a variety of representative marine locations (Wisz et al., 2013, 2020).

This scale of marine biodiversity monitoring is currently prohibitively expensive and often hazardous using direct human observation. Since 2008/09, UK biodiversity monitoring has seen a 42% real-term decrease in public funding (JNCC, 2023). Most biodiversity monitoring is also sporadic, dependant on short-term funding. While this approach can yield high-quality data, the inconsistency limits our ability to derive local-scale trends, which are essential for an evidenced-based approach to marine management (Wilding et al., 2017; IPBES, 2019).

Autonomous eDNA filtration and preservation technologies, such as the Robotic Cartridge Sampling Instrument (RoCSI), offer a safe solution to achieve consistent and long-term biomolecular monitoring necessary to identify ecological changes in the marine environment. Establishing a network of autonomous eDNA samplers requires considerable initial investments but once in place, it can generate consistent, low-cost biodiversity data. Additionally, storing samples in biobanks allows analysis to occur as and when funding is available, mitigating the effects of fluctuating funding. This approach circumvents the issues involved with inconsistent collection of biodiversity data.

Throughout this NEXUSS doctoral training programme, I have explored how biomolecular monitoring and autonomous technologies, such as the RoCSI, can enhance the spatial and temporal resolution of marine biodiversity monitoring.

- Chapter 2 describes how the RoCSI can autonomously filter and preserve samples using the ship's underway system, significantly enhancing the spatial resolution of biomolecular sampling (Objective 2). A comparison between the RoCSI-collected and manually collected samples demonstrated that the RoCSI can produce comparable results (Objective 1). This chapter also highlights the need for high-resolution sampling in open ocean environments, where patchiness in planktonic communities may obscure ecological patterns at coarser spatial scales.
- Chapter 3 demonstrates how the RoCSI can capture high temporal resolution datasets from a highly urbanised estuary (Objective 3). The findings reveal that short-lived microbial community shifts may serve as indicators of small-scale pollution events typical in urbanised environments. This highlights the potential of using the RoCSI for near real-time environmental monitoring in dynamic urban coastal systems.
- Chapter 4 compares a range of liquid nucleic acid preservatives to optimise the RoCSI for long-term (2-month) deployments at ocean observatories (Objective 4). The results support the use of ZymoBIOMICS DNA/RNA Shield, for 2-month preservation. However, for RNA preservation of the same sample, the manufacturers limit of 1 month should only be exceed with caution. If only RNA is required from the sample, Qiagen's RNAlater or lab-prepared NAP buffer were better suited, though the results indicate that two

months may be approaching the upper limit for RNA stability with these preservatives as well.

- Chapter 5 examines strategies to improve data management in biomolecular research, enabling RoCSI-produced datasets to contribute to a global network of biomolecular observations. This chapter arose from challenges encountered while reviewing methods sections to identify extraction protocols for different preservatives. The frequent lack of methodological detail prompted co-leading the Ocean Best Practice System (OBPS) Workshop on protocol sharing for Omics and eDNA. Discussions from this workshop, and subsequent OBPS Task Team, informed the data management approach outlined in this chapter.
- Chapter 6 tests the use of the portable MinION sequencing platform onboard the RRS Discovery, employing a modified DNA extraction protocol for use at sea. The MinION successfully generated 43,820 sequences with an average length of 1,683 bp. These results demonstrate the feasibility of integrating such a device with the RoCSI system for near real-time in-situ metagenomic analysis. However, the findings also highlight several ethical and technical challenges, such as data privacy and sequencing accuracy, that must be addressed as this technology develops.
- Additionally, during an internship with Natural England's DNA team, I explored the implementation of eDNA monitoring for national biodiversity assessments. I produced internal guidance for using DNA methods to monitor inshore fish and created a template for Natural England's staff and contractors to record taxonomic data and metadata from metabarcoding studies. This work aimed to improve methodological transparency and ensure compatibility with global data repositories such as the National Biodiversity Network Atlas (NBN Atlas), Global Biodiversity Information Facility (GBIF), and the International Nucleotide Sequence Data Centre (INSDC).

7.2 Increasing the temporal and spatial resolution of marine biomolecular observations with RoCSI

Ecological datasets often contain missing data (Lopucki et al., 2022), particularly in marine environments when adverse sea conditions can hinder data collection. Collecting high spatial resolution data throughout a cruise often involves out-of-hours work, and at least two personnel for safety. Chapter 2 demonstrated that the RoCSI prototype can provide high resolution spatial data with only one operator reloading samples and decontaminating the system at convenient daily intervals. Autonomous devices reduce personnel time, increase safety, and generate consistent high-resolution datasets. Such data also enable comparisons with satellite observations, facilitating exploration of taxonomy within short-lived or moving oceanographic features like eddies, gyres, and coastal fronts.

High-resolution spatial datasets can also be captured in three dimensions by deploying the RoCSI on autonomous underwater vehicles (AUVs). Since the completion of Chapter 2, the commercially available RoCSI has also been successfully deployed on remotely operated vehicles (ROVs) and Autosub-AUVs, enabling high-resolution sampling across various regions,

with a maximum depth reached of 4,719 meters. Extending the capacity for deep-sea sampling in difficult-to-reach environments.

Chapter 3, demonstrated the suitability of the RoCSI device for high-resolution time-series monitoring. Sampling manually every two hours is prohibitively labour-expensive. However, this type of high-resolution dataset can be useful to understand the frequency of sampling needed to address specific science questions in highly dynamic environments, such as estuaries. For a long-term monitoring programme, it is important to interpret eDNA-based ecology within the context of diel variation. Understanding these nuances can help when developing long-term sampling regimes, to ensure that the sampling is unbiased by the tidal or diurnal cycles.

To optimise the RoCSI for long-term biomolecular monitoring Chapter 4 compares different liquid preservatives for their effectiveness in preserving DNA and RNA over a two-month period. The preservative DNA/RNA shield (Zymo Research) could preserve both DNA and RNA at room temperature for the entire two months. However, the homemade Nucleic Acid Preservative (NAP) and RNAlater (Qiagen) proved to be more effective at preserving the RNA within a sample. Understanding the limits of various nucleic acid preservatives is crucial for planning RoCSI missions, enabling the RoCSI to operate autonomously for extended periods. This capability is particularly valuable for long-term deployments at offshore monitoring stations when harsh storm conditions might prevent the safe retrieval of cartridges for extended periods throughout the winter months. Moreover, minimising visits to offshore sites required to exchange samples can contribute to achieving net-zero science goals through reduced fuel consumption (NZOC, 2022)

7.3 Integrating autonomous biomolecular monitoring with other environmental observations

7.3.1 Physical and chemical observations

Integrating biomolecular observations with satellite data and physio-chemical information gathered from sensor-rich oceanographic and meteorological monitoring stations can reveal the underlying mechanisms driving ecological changes. The results presented in both Chapter 2 and Chapter 3 provide examples of this integrated monitoring approach through the co-deployment of various sensors and comparisons with satellite data. This combination of parameters can significantly enhance the assessment of marine ecosystem conditions (Smit et al., 2021). Moreover, considering the ecological interplay between physical and chemical parameters is likely to improve long-term meteorological forecasting (Tagliabue, 2023).

Additionally, the potential consistency of data obtained through autonomous eDNA sampling helps mitigate issues related to missing data in ecological models (Łopucki et al., 2022).

7.3.2 Ecological observations

One of the key advantages of autonomous DNA monitoring is that samples can be collected consistently, ensuring continuous data streams that facilitate the analysis of trends within datasets. These trends, along with target species identifications can then guide confirmatory and targeted research using traditional visual monitoring methods. For example, Holman et al. (2019), detected the invasive *Arcuatula senhousia* (Asian date mussel) through eDNA analysis, which was subsequently confirmed through targeted visual surveys.

However, as highlighted in Chapter 3, eDNA from larger, more mobile eukaryotes, such as Arthropoda, interacting with sampling equipment can disrupt the temporal signal by shedding disproportionately more DNA to an individual sample. Therefore, autonomous biomolecular monitoring should not be viewed as a replacement for manual sampling or visual/acoustic biomonitoring surveys but rather as a complimentary tool that can direct high-confidence and more labour-intensive methods, like visual surveys, towards sites with the highest-likelihood of relevant findings.

7.3.3 Citizen science observations

Citizen science observations complement autonomous biomolecular monitoring by offering excellent spatial resolution at a single time point, while autonomy ensures consistent data collection over time. For instance, in the national-scale Danish BioBlitz, 100 sites were sampled simultaneously across two seasons, providing a valuable snapshot of coastal biodiversity across multiple locations (Agersnap et al., 2022). Autonomous sampling, on the other hand, allows for the collection of higher-resolution data on temporal fluctuations, including sampling through the night or adverse sea conditions, which would be challenging for citizen scientists.

Beyond the advantage of higher resolution data, combining autonomous sampling with citizen science offers a valuable opportunity for knowledge exchange between coastal communities and scientists. The Danish BioBlitz engaged 360 citizen scientists over two surveys, providing an opportunity to explain the science behind eDNA monitoring and potentially increasing trust in the data produced (Agersnap et al., 2022). This engagement also allows scientists to gain insights into local trends that may warrant further investigation within the eDNA datasets. For example, if citizens report sightings of new species or declines in certain local species, these observations can guide targeted analysis in the eDNA data. This collaborative approach to

science has a proven track record of delivering real-world benefits to society (Baker et al., 2023; Nichols et al., 2019).

7.4 National Biomolecular Observatory

To fully harness biomolecular data from diverse sources, whether from autonomous samplers like RoCSI or citizen scientists, a national-scale infrastructure is essential. This requires the development of standardised practices for sample processing, bioinformatics, and data and metadata management, as well as the establishment of central repositories for long-term data and sample storage.

At present, biomolecular data is underutilised in ocean biogeochemical models, which typically rely on biological indicators such as nutrient or chlorophyll concentrations (Tagliabue, 2023). This underutilisation is largely due to the complexity of genomic data and the challenges associated with integrating data from various sources that employ different methodologies. A national data repository for ocean biomolecular data, collected through standardised practices, would reduce uncertainties arising from methodological variability and simplify the often-time-consuming data wrangling required to merge datasets. This approach would promote the reuse of biomolecular data, enabling the exploration of broader ecological trends across multiple projects.

Maintaining this consistent biomolecular data and associated metadata will facilitate the extension of biogeochemical models to include genomic diversity. This is a huge task to tackle over the next decade, but incorporating this complexity into future models has the potential to greatly improve both climate and species distribution forecasting (Keck et al., 2023).

7.4.1 Standard Practices

A national biomolecular data repository that uses standard practices for key universal primer sets such as 12S rRNA for vertebrates, 16S rRNA for prokaryotes, 18S rRNA for eukaryotes, CO1 for invertebrates, and ITS for fungi, would enable monitoring across the tree of life. Many organisations see the benefit of standard practices and are beginning to establish their own best or standard practices (De Brauwer et al., 2023; Nature Metrics 2022; Minamoto et al. 2021). However, in general, eDNA methods are becoming more diverse as the field expands (Hakimzadeh et al., 2024).

The process of selecting standard practices for a national repository could be contentious as the selection of a standard practice could be misconstrued as an endorsement of the practice as the best practice. While a standard practice must be high quality and thoroughly tested, it

should also meet practical requirements for use on a national scale. Therefore, practices that have lower costs, less specialist equipment, and utilise fully open-source software may be preferential to more state-of-the-art practices. As stated by Stepien et al. (2024) in a commentary on the US National Workshop on Environmental DNA,

> "It is time to stop letting perfect be the enemy of good and to focus future efforts on method harmonization and a national strategy towards method adoption".
>
> **Stepien et al., 2024**

As research continuously builds upon and improves methodologies, the best practice at any point in time will inevitably be improved upon. Consequently, standard practices should not be perceived as completely static and resistant to change or updates. Rather, they serve as a starting point that requires regular review and adaptation to evolving research. Assessing the efficacy of new methods becomes challenging when numerous diverse 'old' methods are in use. In contrast, evaluating a single standardised method against various new approaches is more manageable.

Standard practices can also improve biomolecular literacy by simplifying the learning process. Learning resources focused on a single set of standardised methods can reduce the complexity and time investment required to understand the full breadth of biomolecular processes involved in acquiring a dataset, from sample collection through to bioinformatics. This streamlined approach makes it feasible for a greater number of people to understand the underlying techniques, ultimately facilitating a deeper and more widespread understanding of biomolecular data collection and analysis.

7.4.2 Data Management

Developing standard practices includes standard practices for data management, to increase the Findability Accessibility Interoperability and Reusability (FAIR) of data (Wilkinson et al., 2016). Initiating projects with a clear understanding of how and where the data and metadata will be shared and stored throughout the project lifetime, streamlines data sharing and decreases the risk of data loss that arises when data is handled across multiple parties.

Effective data management requires knowledge of the appropriate data repository for storing and sharing data at the conclusion of a project. For instance, repositories like NBN Atlas and GBIF are suitable for DNA-derived species data (NBN Trust, 2024; Abarenkov et al., 2023), while the INSDC is designed for storing raw sequence data (Arita et al., 2021). Methods documentation can be housed in the Ocean Best Practices Systems repository or on platforms like protocols.io (Pearlman et al., 2019; Teytelman et al., 2016). Each of these repositories

requires data in certain formats with key metadata requirements. Chapter 5 highlights the benefits of sharing biomolecular observation methods through the OBPS platform (Samuel et al., 2021).

Developing standardised data management practices is essential to ensure that all necessary metadata and data are collected throughout the project's lifespan in the correct format for repository submission (Thompson & Thielen, 2023). Without such practices, sharing data via these platforms can become an arduous task, which is often overlooked at the end of a project when funding and resources are limited. With the RoCSI, this process could be automated through software that syncs RoCSI data outputs with ship-derived geolocation data, ensuring compatibility with repository upload formats.

7.4.3 Biobanks

Another key piece of infrastructure that will greatly facilitate ocean biomolecular monitoring is the establishment of national biobanking facilities for the storage of physical samples (Jarman et al., 2018). With autonomous monitoring devices such as the RoCSI, surplus samples can be easily collected and stored and used later for targeted research of historical reference samples. Replicate samples can also be re-analysed with future technologies to facilitate the transition between old and new standard practices. Biobanks can also ensure consistency in long term monitoring projects throughout uncertain financial circumstances, enabling samples to be stored until funding becomes available to analyse them.

7.5 Future technology development

Since the start of this PhD and the development of the RoCSI prototype, new compact samplers similar in size and shape to the RoCSI have been developed, such as the DOT eDNA sampler. The DOT sampler, like RoCSI, includes features such as self-cleaning and in-situ preservation but is limited to nine discrete samples per deployment (Hendricks et al., 2023). In contrast, the commercially available RoCSI can collect up to 48 samples per deployment, with the potential for more through custom sample storage configurations. This allows greater flexibility for deployments at offshore long-term monitoring stations where access, and therefore opportunities, for sample exchange is limited. This advantage is particularly significant given that several liquid preservatives tested in Chapter 4 were found to effectively preserve nucleic acids for up to two months.

Future developments in RoCSI technology may include additional modular components to support in-situ sample analysis. For example, incorporating devices for in-situ amplification,

such as the Amplitron (Wilson, 2020), would enable in-situ detection of key species like *Oleispira sp* as potential early warning of pollution events (Chapter 3), or for hazardous species, like *Chiropsoides buitendijki* (box jellyfish), that pose public health risks in regions reliant on coastal recreation activities (Osathanunkul, 2024).

Additionally, integrating technology to prepare samples for analysis with portable sequencers like the MinION would enable near real-time sequencing in the field. This capability would allow for adaptive sampling regimes, where areas of interest are identified *in situ*, enabling the tracking of endangered species and guiding decisions on when and where to conduct more intensive sampling efforts. As demonstrated in Chapter 6, MinION sequencing has already been incorporated into offshore observatory maintenance cruises. If samples are collected and processed on first arrival at the observatory, before maintenance operations begin, sequencing and preliminary interpretation can be completed within 24 hours. This could enable the adoption of adaptive sampling strategies based on the taxa or functional annotations identified in the initial samples.

At the time of writing Chapter 6, MinION technology was error-prone and unreliable for eDNA monitoring. However, recent advancements in the Oxford Nanopore Technologies' R10.4 flow cell have achieved a modal read accuracy of over 99%, making the MinION a much more viable tool for near-real time biomolecular monitoring (Ni et al., 2023; Sereika et al., 2022).

7.6 Conclusion

Autonomous biomolecular sampling is an emerging technology that offers a comprehensive approach to monitoring marine ecosystems on a large scale. It serves as a valuable complement to existing biodiversity monitoring methods and is particularly well-suited for sentinel monitoring of diverse taxa, providing rapid and consistent data that can guide when and where to deploy more resource-intensive observational methods. As long-term biomolecular time series expand, the value of broadscale autonomous monitoring will continue to grow. To fully realise the potential of this technology, standard practices and robust data management is required to support a national scale ocean biomolecular observatory. Developing this infrastructure would improve our capacity to predict and respond to ecological changes. Ultimately, the detailed ecological insights provided by biomolecular monitoring are essential for evidence-based decision-making to ensure the sustainability of marine environments.

7.7 References

Abarenkov K, Andersson A F, Bissett A, Finstad A G, Fossøy F, Grosjean M, Hope M, Jeppesen T S, Kõljalg U, Lundin D, Nilsson R H, Prager M, Provoost P, Schigel D, Suominen S, Svenningsen C, Frøslev T G. 2023. Publishing DNA-derived data through biodiversity data platforms. v1.3.0 2023. https://doi.org/10.35035/doc-vf1a-nr22

Agersnap, S., Sigsgaard, E. E., Jensen, M. R., Avila, M. D. P., Carl, H., Møller, P. R., Krøs, S. L., Knudsen, S. W., Wisz, M. S., & Thomsen, P. F. (2022). A National Scale "BioBlitz" Using Citizen Science and eDNA Metabarcoding for Monitoring Coastal Marine Fish. Frontiers in Marine Science, 9, 824100. https://doi.org/10.3389/FMARS.2022.824100/BIBTEX

Arita, M., Collaboration, on behalf of the I. N. S. D., Karsch-Mizrachi, I., Collaboration, on behalf of the I. N. S. D., Cochrane, G., & Collaboration, on behalf of the I. N. S. D. (2021). The international nucleotide sequence database collaboration. Nucleic Acids Research, 49(D1), D121–D124. https://doi.org/10.1093/NAR/GKAA967

Baker, M. R., Alverson, R., Christiansen, R., Criddle, K., Eilertsen, D., Foy, R. J., Gauvin, J., Goodman, S. E., Habegger, L., Harris, B. P., Kimball, N., Mercer, A. M., Poulsen, E., Robinson, M., Rusin, J., Skinner, R., Sparrevohn, C. R., Stokesbury, K. D. E., & Sykes, D. R. (2023). Mechanisms and models for industry engagement in collaborative research in commercial fisheries. Frontiers in Marine Science, 10, 1077944. https://doi.org/10.3389/FMARS.2023.1077944/BIBTEX

De Brauwer, M., Clarke, L. J., Chariton, A., Cooper, M. K., de Bruyn, M., Furlan, E., MacDonald, A. J., Rourke, M. L., Sherman, C. D. H., Suter, L., Villacorta-Rath, C., Zaiko, A., & Trujillo-González, A. (2023). Best practice guidelines for environmental DNA biomonitoring in Australia and New Zealand. Environmental DNA, 5(3), 417–423. https://doi.org/10.1002/EDN3.395

Hakimzadeh, A., Abdala Asbun, A., Albanese, D., Bernard, M., Buchner, D., Callahan, B., Caporaso, J. G., Curd, E., Djemiel, C., Brandström Durling, M., Elbrecht, V., Gold, Z., Gweon, H. S., Hajibabaei, M., Hildebrand, F., Mikryukov, V., Normandeau, E., Özkurt, E., M. Palmer, J., … Anslan, S. (2024). A pile of pipelines: An overview of the bioinformatics software for metabarcoding data analyses. Molecular Ecology Resources, 24(5), e13847. https://doi.org/10.1111/1755-0998.13847

Hendricks, A., Mackie, C. M., Luy, E., Sonnichsen, C., Smith, J., Grundke, I., Tavasoli, M., Furlong, A., Beiko, R. G., LaRoche, J., & Sieben, V. (2023). Compact and automated eDNA sampler for in situ monitoring of marine environments. Scientific Reports 2023 13:1, 13(1), 1–13. https://doi.org/10.1038/s41598-023-32310-3

Holman, L. E., de Bruyn, M., Creer, S., Carvalho, G., Robidart, J., & Rius, M. (2019). Detection of introduced and resident marine species using environmental DNA metabarcoding of sediment and water. Scientific Reports 2019 9:1, 9(1), 1–10. https://doi.org/10.1038/s41598-019-47899-7

IPBES (2019). Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. S. Díaz, J. Settele, E. S. Brondízio, H. T. Ngo, M. Guèze, J. Agard, A. Arneth, P. Balvanera, K. A. Brauman, S. H. M. Butchart, K. M. A. Chan, L. A. Garibaldi, K. Ichii, J. Liu, S. M. Subramanian, G. F. Midgley, P. Miloslavich, Z. Molnár, D. Obura, A. Pfaff, S. Polasky, A. Purvis, J. Razzaque, B. Reyers, R. Roy Chowdhury, Y. J. Shin, I. J. Visseren-Hamakers, K. J. Willis, and C. N. Zayas (eds.). IPBES secretariat, Bonn, Germany. 56 pages. https://doi.org/10.5281/zenodo.3553579

Jarman, S. N., Berry, O., & Bunce, M. (2018). The value of environmental DNA biobanking for long-term biomonitoring. Nature Ecology and Evolution, 2(8), 1192–1193. https://doi.org/10.1038/s41559-018-0614-3

Keck, F., Brantschen, J., & Altermatt, F. (2023). A combination of machine-learning and eDNA reveals the genetic signature of environmental change at the landscape levels. Molecular Ecology, 32(17), 4791–4800. https://doi.org/10.1111/MEC.17073

Łopucki, R., Kiersztyn, A., Pitucha, G., & Kitowski, I. (2022). Handling missing data in ecological studies: Ignoring gaps in the dataset can distort the inference. Ecological Modelling, 468, 109964. https://doi.org/10.1016/J.ECOLMODEL.2022.109964

Minamoto T, Miya M, Sado T, Seino S, Doi H, Kondoh M, Nakamura K, Takahara T, Yamamoto S, Yamanaka H, Araki H, Iwasaki W, Kasai A, Masuda R, Uchii K (2021) An illustrated manual for environmental DNA research: Water sampling guidelines and experimental protocols. Environmental DNA 3:8-13 https://doi.org/10.1002/edn3.121

NatureMetrics. 2022. Protocol for Sampling Marine Benthic Sediment for DNA-based Analysis. JNCC Report No. 706, JNCC, Peterborough, ISSN 0963-8091

NBN Trust. 2024. The National Biodiversity Network (NBN) Atlas. https://ror.org/00mcxye41.

Net Zero Oceanographic Capability | National Oceanography Centre. (n.d.). Retrieved August 22, 2024, from https://noc.ac.uk/facilities/ships/net-zero-oceanographic-capability

Ni, Y., Liu, X., Simeneh, Z. M., Yang, M., & Li, R. (2023). Benchmarking of Nanopore R10.4 and R9.4.1 flow cells in single-cell whole-genome amplification and whole-genome shotgun sequencing.

Computational and Structural Biotechnology Journal, 21, 2352–2364. https://doi.org/10.1016/j.csbj.2023.03.038

Nichols, C. R., Wright, L. D., Bainbridge, S. J., Cosby, A., Hénaff, A., Loftis, J. D., Cocquempot, L., Katragadda, S., Mendez, G. R., Letortu, P., Le Dantec, N. L., Resio, D., & Zarillo, G. (2019). Collaborative science to enhance coastal resilience and adaptation. Frontiers in Marine Science, 6(JUL), 433620. https://doi.org/10.3389/FMARS.2019.00404/BIBTEX

Osathanunkul, M. (2024). Species-specific eDNA assay development for enhanced box jellyfish risk management in coastal environments. Science of The Total Environment, 931, 172900. https://doi.org/10.1016/J.SCITOTENV.2024.172900

Pearlman, J., Bushnell, M., Coppola, L., Karstensen, J., Buttigieg, P. L., Pearlman, F., Simpson, P., Barbier, M., Muller-Karger, F. E., Munoz-Mas, C., Pissierssens, P., Chandler, C., Hermes, J., Heslop, E., Jenkyns, R., Achterberg, E. P., Bensi, M., Bittig, H. C., Blandin, J., … Whoriskey, F. (2019). Evolving and sustaining ocean best practices and standards for the next decade. Frontiers in Marine Science, 6(JUN), 433901. https://doi.org/10.3389/FMARS.2019.00277/BIBTEX

Reiber, L., Foit, K., Liess, M., Karaoglan, B., Wogram, J., & Duquesne, S. (2022). Close to reality? Micro-/mesocosm communities do not represent natural macroinvertebrate communities. Environmental Sciences Europe, 34(1), 1–19. https://doi.org/10.1186/S12302-022-00643-X/TABLES/2

Samuel, R. M., Meyer, R., Buttigieg, P. L., Davies, N., Jeffery, N. W., Meyer, C., Pavloudi, C., Pitz, K. J., Sweetlove, M., Theroux, S., van de Kamp, J., & Watts, A. (2021). Toward a Global Public Repository of Community Protocols to Encourage Best Practices in Biomolecular Ocean Observing and Research. Frontiers in Marine Science, 8, 758694. https://doi.org/10.3389/FMARS.2021.758694/BIBTEX

Sereika, M., Kirkegaard, R. H., Karst, S. M., Michaelsen, T. Y., Sørensen, E. A., Wollenberg, R. D., & Albertsen, M. (2022). Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. Nature Methods 2022 19:7, 19(7), 823–826. https://doi.org/10.1038/s41592-022-01539-7

Smit, K. P., Bernard, A. T. F., Lombard, A. T., & Sink, K. J. (2021). Assessing marine ecosystem condition: A review to support indicator choice and framework development. Ecological Indicators, 121, 107148. https://doi.org/10.1016/J.ECOLIND.2020.107148

Stepien, C. A., Theroux, S., Weisberg, S. B., Carol Stepien, C. A., & California, S. (2024). The Second National Workshop on Marine eDNA: A workshop to accelerate the incorporation of eDNA science into environmental management applications. Environmental DNA, 6(1), e379. https://doi.org/10.1002/EDN3.379

Tagliabue, A. (2023). 'Oceans are hugely complex': modelling marine microbes is key to climate forecasts. Nature 2023 623:7986, 623(7986), 250–252. https://doi.org/10.1038/d41586-023-03425-4

Teytelman, L., Stoliartchouk, A., Kindler, L., & Hurwitz, B. L. (2016). Protocols.io: Virtual Communities for Protocol Development and Discussion. PLOS Biology, 14(8), e1002538. https://doi.org/10.1371/JOURNAL.PBIO.1002538

Thompson, L.R. & Thielen, P. (2023). Decoding dissolved information: environmental DNA sequencing at global scale to monitor a changing ocean. *Current Opinion in Biotechnology*, *81*, p.102936.

UKBI - E2. Biodiversity expenditure | JNCC - Adviser to Government on Nature Conservation. (n.d.). Retrieved August 22, 2024, from https://jncc.gov.uk/our-work/ukbi-e2-biodiversity-expenditure/

Wilding, T. A., Gill, A. B., Boon, A., Sheehan, E., Dauvin, J. C., Pezy, J. P., O'Beirn, F., Janas, U., Rostin, L., & De Mesel, I. (2017). Turning off the DRIP ('Data-rich, information-poor') – rationalising monitoring with a focus on marine renewable energy developments and the benthos. Renewable and Sustainable Energy Reviews, 74, 848–859. https://doi.org/10.1016/J.RSER.2017.03.013

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 2016 3:1, 3(1), 1–9. https://doi.org/10.1038/sdata.2016.18

Wilson M W. 2020 Development of Automated Nucleic Acid Technologies for Marine Point of Sample Diagnostics. PhD Thesis. University of Southampton. https://eprints.soton.ac.uk/447354/1/Wilson_Matthew_Thesis_final_Jan21.pdf

Wisz, M. S., Pottier, J., Kissling, W. D., Pellissier, L., Lenoir, J., Damgaard, C. F., Dormann, C. F., Forchhammer, M. C., Grytnes, J. A., Guisan, A., Heikkinen, R. K., Høye, T. T., Kühn, I., Luoto, M., Maiorano, L., Nilsson, M. C., Normand, S., Öckinger, E., Schmidt, N. M., … Svenning, J. C. (2013). The role of biotic interactions in shaping distributions and realised assemblages of

species: implications for species distribution modelling. Biological Reviews, 88(1), 15–30. https://doi.org/10.1111/J.1469-185X.2012.00235.X

Wisz, M. S., Satterthwaite, E. V., Fudge, M., Fischer, M., Polejack, A., St. John, M., Fletcher, S., & Rudd, M. A. (2020). 100 Opportunities for More Inclusive Ocean Research: Cross-Disciplinary Research Questions for Sustainable Ocean Governance and Management. Frontiers in Marine Science, 7, 541934. https://doi.org/10.3389/FMARS.2020.00576/BIBTEX