frontiers | Research Topics

Marine microbiomes: Towards standard methods and best practices

Edited by
Stéphane Pesant, Kelly D. Goodwin, Catarina Magalhães and Anaïs Lacoursière-Roussel

Published in
Frontiers in Microbiology
Frontiers in Marine Science

ISSN 1664-8714
ISBN 978-2-8325-4131-9
DOI 10.3389/978-2-8325-4131-9

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Marine microbiomes: Towards standard methods and best practices

Topic editors

Stéphane Pesant — European Bioinformatics Institute (EMBL-EBI), United Kingdom
Kelly D. Goodwin — National Oceanic and Atmospheric Administration (NOAA), United States
Catarina Magalhães — University of Porto, Portugal
Anaïs Lacoursière-Roussel — St. Andrews Biological Station, Fisheries and Oceans Canada (DFO), Canada

Citation

Pesant, S., Goodwin, K. D., Magalhães, C., Lacoursière-Roussel, A., eds. (2023). *Marine microbiomes: Towards standard methods and best practices*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4131-9

Table of contents

frontiers | Frontiers in Microbiology

TYPE Editorial
PUBLISHED 01 June 2023
DOI 10.3389/fmicb.2023.1219958

OPEN ACCESS

EDITED AND REVIEWED BY
Marcelino T. Suzuki,
Sorbonne Universite, France

*CORRESPONDENCE
Kelly D. Goodwin
kelly.goodwin@noaa.gov

RECEIVED 09 May 2023
ACCEPTED 22 May 2023
PUBLISHED 01 June 2023

CITATION
Goodwin KD and Lacoursière-Roussel A (2023) Editorial: Marine microbiomes: towards standard methods and Best Practices. *Front. Microbiol.* 14:1219958. doi: 10.3389/fmicb.2023.1219958

Editorial: Marine microbiomes: towards standard methods and Best Practices

Kelly D. Goodwin [1]* and Anaïs Lacoursière-Roussel[2,3]

[1]National Oceanic and Atmospheric Administration, NOAA Ocean Exploration, Stationed at SWFSC, La Jolla, CA, United States, [2]St. Andrews Biological Station, Fisheries and Oceans Canada, St. Andrews, NB, Canada, [3]Biological Sciences, University of New Brunswick, Saint John, NB, Canada

KEYWORDS

microbiome, eDNA, standards and codes of practice, Best Practices, marine microbial biodiversity

Editorial on the Research Topic
Marine microbiomes: towards standard methods and Best Practices

The microbiome is key to understanding and sustaining the services that ocean ecosystems provide (Bolhuis et al., 2020). The marine microbiome—an ensemble of microscopic organisms that inhabit water columns, sediments, and aquatic organisms—contains members spanning in size from viruses of a few tens of nanometers to metazoans of several centimeters. Together, the microbiome forms the base of the food web, maintains animal health, and regulates most fluxes of energy and matter. Marine microbiome discovery is part of a great campaign to explore the earth's oceans, and rapid advances in high throughput sequencing are allowing a glimpse into this hidden world (Figure 1). Furthermore, these techniques have been adapted to detect DNA in the environment (eDNA) from organisms of all trophic levels.

Biomolecular observations can provide important insights into ecosystem structure and function, development of new indicators of ecosystem health, and warnings of potential hazards to living resources and humans. Endorsements by the United Nations Ocean Decade[1] reflect the growing demand for affordable, large-scale biological observations provided by biomolecule detection. Examples include the Ocean Biomolecular Observing Network (OBON) Program (Leinen et al., 2022), which aims to transform how we sense, harvest, protect and manage ocean life. OBON actions supporting these aims include the Observing and Promoting Atlantic Microbiomes[2] project hosted by the Atlantic Ocean Research Alliance (AORA) Marine Microbiome Working Group[3] (Bolhuis et al., 2020) that called for this Research Topic.

This Frontiers Research Topic was motivated by the recognition that a number of cross-cutting challenges need to be addressed to fully unlock the marine microbiome for environmental and societal benefit. Such challenges include the development and adoption of standards, common methods, Best Practices, and FAIR (Findable, Accessible, Interoperable, and Re-usable) data principles (Bolhuis et al., 2020). For some authors, the

1 https://oceandecade.org/
2 https://oceandecade.org/actions/ocean-biomolecular-observing-network-obon/
3 https://www.marinemicrobiome.org/

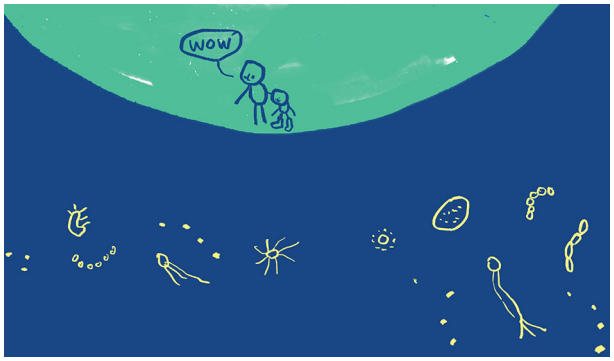

FIGURE 1
The marine microbiome is a largely unexplored treasure for society. Illustration credit: Rán Flygenring.

emphasis was on cyberinfrastructure to ensure that both sequence and environmental data are FAIR (Blumberg et al.). Others focused on developing a Minimum Information for an Omic Protocol (MIOP) and a public repository of protocols that can be both searched and prioritized for use (Samuel et al.). Both of these manuscripts highlighted the importance of machine readable data and products to the achievement of FAIR principles. The need to implement and sustain a global and publicly supported platform to share, discover, and compare practices and protocols was emphasized.

Papers in this Research Topic highlighted that harmonization across the full workflow—from methods through data reporting—is needed to achieve global scale biodiversity observations that can be integrated over space and time. Some manuscripts offered general overviews and "tricks of the trade" to guide microbiome sample collection and processing for coral tissues (Silva et al.) or pelagic waters for a variety of molecular targets and size fractions (Patin and Goodwin). These papers reviewed methods for multiple sections of the overall workflow with detailed guidance provided for sample collection, preservation, and processing. Other manuscripts focused on specific details, such as DNA isolation. For example, Wietz et al. described extraction of DNA from samples preserved in formalin or $HgCl_2$, preservatives commonly used in sediment trap studies. Korlević et al. described a procedure to specifically isolate DNA and protein from macrophyte epiphytic communities to avoid overwhelming microbiome samples with host DNA. Gu et al. described a new analytical protocol to determine Protoporphyrin IX (PPIX) in microbial cells and provided results with coastal aquatic samples to demonstrate the potential to use PPIX as an indicator of microbial productivity. This diversity of topics underscores the large range of microbiome applications.

The growth of publicly available sequence data has increased the ability to perform meta-analysis to investigate broad scale environmental change. However, the rapid expansion of molecular techniques has created disparate protocols and workflows. A number of authors thus addressed the question of whether datasets can be combined across studies by exploring the sensitivity of taxonomic annotation to variations in sequencing methods. For example, taxonomic assignments were compared for the 16S rRNA gene V3-V4 and V4-V5 primer sets as applied to a variety of sample types collected from Arctic Ocean marine systems. In this case, V4-V5 was recommended due to superior inclusion of archaeal taxa (Fadeev et al.). In another case, a single primer set was applied to coral tissues that were processed separately (DNA extraction through library preparation) and then sequenced on different platforms (MiSeq and HiSeq). Despite past studies suggesting that MiSeq and HiSeq data could be combined to provide microbiome taxonomic analysis, the study here cautioned that significant differences in compositional assignments could arise from protocol variations (Epstein et al.). This work suggested that projects that seek to understand and overcome sources of technical variation remain needed. Multiple studies also highlighted the continued need to build out reference databases to improve annotation of sequence data.

This Research Topic fostered cross-community exchange of standards and Best Practices. It provided an opportunity for different communities working on marine microbiomes to communicate the advantages and limitations of various sampling, laboratory, and data processing methods and to open community discussion on how to move toward large scale operationalization. Although the need for harmonization was recognized, workflows must be fit for purpose; meaning that they must meet the objectives and logistical constraints of a study, application, management objective, or time series (including legacy sampling). Even with Best Practices in hand, pilot studies must be conducted to validate and optimize workflows against different sample types, geographies, or molecular targets. To support international efforts to develop guidance on Best Practices, a centralized platform could compile protocols, metadata, and data produced by specific methods. Such could include successful, unsuccessful, anecdotal, or unpublished information to provide real-world feedback. Machine Learning approaches could potentially help define optimal workflows from the growing observations. Overall, the pursuit of cross-community standards and Best Practices will foster data integration across heterogeneous methods, improve future ocean observations, and expand the trusted use of microbiome and eDNA science.

Author contributions

All authors contributed to the writing and editing of the manuscript.

Acknowledgments

We thank Stéphane Pesant and Catarina Magalhães for their service as guest editors for this Research Topic.

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.

Publisher's note

References

Bolhuis, H., Buttigieg, P. L., Goodwin, K., Groben, R., Ludicone, D., Lacoursière-Rousselm, A., et al. (2020). The Atlantic Ocean Research Alliance-marine microbiome roadmap. doi: 10.5281/zenodo.3632526

Leinen, M., Chavez, F., Meyer, R., Meyer, R., Buttigieg, P. L., Davies, N. et al. (2022). The ocean biomolecular observing network (OBON). *Mar. Technol. Soc. J.* 56, 106. doi: 10.4031/MTSJ.56.3.20

frontiers in Microbiology

ORIGINAL RESEARCH
published: 16 February 2021
doi: 10.3389/fmicb.2021.637526

Comparison of Two 16S rRNA Primers (V3–V4 and V4–V5) for Studies of Arctic Microbial Communities

Eduard Fadeev[1,2†], Magda G. Cardozo-Mino[1,2], Josephine Z. Rapp[3], Christina Bienhold[1,2], Ian Salter[1,4], Verena Salman-Carvalho[5], Massimiliano Molari[2], Halina E. Tegetmeyer[2,6], Pier Luigi Buttigieg[1,2] and Antje Boetius[1,2,7]*

[1] Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany, [2] Max Planck Institute for Marine Microbiology, Bremen, Germany, [3] School of Oceanography, University of Washington, Seattle, WA, United States, [4] Faroe Marine Research Institute, Tórshavn, Faroe Islands, [5] Department of Microbiology, Morrill Science Center IVN, University of Massachusetts, Amherst, MA, United States, [6] Center for Biotechnology, Bielefeld University, Bielefeld, Germany, [7] MARUM, University of Bremen, Bremen, Germany

OPEN ACCESS

Edited by:
Catarina Magalhães,
University of Porto, Portugal

Reviewed by:
Miguel Semedo,
University of Porto, Portugal
Alessandro Vezzi,
University of Padua, Italy
Lise Øvreås,
University of Bergen, Norway

***Correspondence:**
Eduard Fadeev
dr.eduard.fadeev@gmail.com

†Present address:
Eduard Fadeev,
Department of Functional and Evolutionary Ecology, University of Vienna, Vienna, Austria

Specialty section:
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Microbiology

Received: 03 December 2020
Accepted: 28 January 2021
Published: 16 February 2021

Citation:
Fadeev E, Cardozo-Mino MG, Rapp JZ, Bienhold C, Salter I, Salman-Carvalho V, Molari M, Tegetmeyer HE, Buttigieg PL and Boetius A (2021) Comparison of Two 16S rRNA Primers (V3–V4 and V4–V5) for Studies of Arctic Microbial Communities. Front. Microbiol. 12:637526. doi: 10.3389/fmicb.2021.637526

Microbial communities of the Arctic Ocean are poorly characterized in comparison to other aquatic environments as to their horizontal, vertical, and temporal turnover. Yet, recent studies showed that the Arctic marine ecosystem harbors unique microbial community members that are adapted to harsh environmental conditions, such as near-freezing temperatures and extreme seasonality. The gene for the small ribosomal subunit (16S rRNA) is commonly used to study the taxonomic composition of microbial communities in their natural environment. Several primer sets for this marker gene have been extensively tested across various sample sets, but these typically originated from low-latitude environments. An explicit evaluation of primer-set performances in representing the microbial communities of the Arctic Ocean is currently lacking. To select a suitable primer set for studying microbiomes of various Arctic marine habitats (sea ice, surface water, marine snow, deep ocean basin, and deep-sea sediment), we have conducted a performance comparison between two widely used primer sets, targeting different hypervariable regions of the 16S rRNA gene (V3–V4 and V4–V5). We observed that both primer sets were highly similar in representing the total microbial community composition down to genus rank, which was also confirmed independently by subgroup-specific catalyzed reporter deposition-fluorescence *in situ* hybridization (CARD-FISH) counts. Each primer set revealed higher internal diversity within certain bacterial taxonomic groups (e.g., the class *Bacteroidia* by V3–V4, and the phylum *Planctomycetes* by V4–V5). However, the V4–V5 primer set provides concurrent coverage of the archaeal domain, a relevant component comprising 10–20% of the community in Arctic deep waters and the sediment. Although both primer sets perform similarly, we suggest the use of the V4–V5 primer set for the integration of both bacterial and archaeal community dynamics in the Arctic marine environment.

Keywords: microbial communities, amplicon sequencing, method comparison, universal primers, Arctic Ocean, molecular observatory

INTRODUCTION

The Arctic Ocean is the most rapidly changing marine region on the planet due to its fast warming causing substantial sea-ice loss (Peng and Meier, 2018; Dai et al., 2019), as well as increasing pollution (Peeken et al., 2018). To assess the impact of global climate change on marine food web dynamics and elemental cycles, it is important to monitor variations in microbial community structure with time (Karl and Church, 2014; Fuhrman et al., 2015; Buttigieg et al., 2018). However, the Arctic Ocean is generally under-sampled in ice-covered regions and in winter (Wassmann et al., 2011), particularly with regard to assessments of its microbial communities and their biogeochemical functions (Boetius et al., 2015). Until recently, microbial monitoring efforts in the deep Arctic Ocean consisted of 1 year-round long-term time series at the HAUSGARTEN observatory in the Fram Strait (Soltwedel et al., 2005, 2015), as well as a few other process studies (e.g., Kirchman et al., 2007; Alonso-Sáez et al., 2008; Nikrad et al., 2012; Wilson et al., 2017; Müller et al., 2018).

The Arctic Ocean features substantial vertical structure that may select for specific microbial types in the sea ice (Boetius et al., 2015; Rapp et al., 2018), in the ice-free and the ice-covered highly stratified surface waters (Wilson et al., 2017; Fadeev et al., 2018), in the sinking particles (further addressed as "marine snow"; Fadeev et al., 2020), as well as in the water and sediments of the deep-sea where temperatures are year-round close to freezing point temperatures (Bienhold et al., 2012; Hoffmann et al., 2017; Wilson et al., 2017; Rapp et al., 2018; Fadeev et al., 2020). Throughout the annual cycle, Arctic surface waters bacterial and archaeal communities exhibit pronounced fluctuations of the dominant taxonomic groups (Alonso-Sáez et al., 2008; Wilson et al., 2017; Müller et al., 2018), which are strongly associated with presence of sea ice and the seasonal phytoplankton blooms (Kirchman et al., 2007; Nikrad et al., 2012; Fadeev et al., 2018; Cardozo-Mino et al., 2020). In winter, as well as under ice-covered conditions, the communities are dominated by the bacterial classes *Alphaproteobacteria* (mainly the SAR11 clade), *Dehalococcoidia* (mainly SAR202 clade), and the archaeal class *Nitrososphaeria* (Alonso-Sáez et al., 2008; Wilson et al., 2017; Müller et al., 2018). In the summer, and under ice-free conditions, the communities are dominated by the bacterial classes *Bacteroidia* (mainly the order *Flavobacteriales*) and *Gammaproteobacteria* (mainly the orders *Alteromonadales* and *Oceanospirillales*; Wilson et al., 2017; Fadeev et al., 2018). During the summer, differences between ice-covered and ice-free communities also affect the microbial diversity of the deep ocean and the seafloor *via* alterations of microbial communities on marine snow (Fadeev et al., 2020).

In the framework of the FRAM Microbial Observatory (FRontiers in Arctic marine Monitoring), we are aiming to develop a standardized methodology for long-term observations of microbial communities in these highly diverse Arctic Ocean environments, which will be also comparable to other long-term microbial time series locations (e.g., HOT and BATS). Unlike other time series sites of the world, the ice-cover and the harsh conditions of the Arctic Ocean are limiting the accessibility of the sampling sites to the summer months. Sampling campaigns during the winter (when microbial biomass is low; Kirchman et al., 2007; Alonso-Sáez et al., 2008) are rare and have only recently been achieved using autonomous samplers with limited sampling capacities (Liu et al., 2020). Therefore, the unique conditions and the currently available technologies constrain year-round microbial observations to PCR-based approaches (i.e., 16S rRNA gene amplicon sequencing), which can be realized with low concentrations of DNA (Thomas et al., 2012). Metagenomics approaches suggest that the functional capacity of marine microbial communities is strongly linked to their taxonomic composition (Galand et al., 2018; McNichol et al., 2020). Thus, when supported by curated taxonomic databases (e.g., SILVA 16S rRNA gene reference; Quast et al., 2013), 16S rRNA gene amplicon sequencing provides an affordable high-throughput tool for addressing traditional community ecology questions, especially under the constrained sampling conditions of the Arctic marine environment.

A critical step in 16S rRNA gene sequencing studies is the selection of PCR primers for DNA amplification (Armougom, 2009; Wang and Qian, 2009). Throughout the years, many primer sets were designed for diversity studies of specific taxonomic groups (e.g., SAR11 clade; Apprill et al., 2015), and attempts have been made to develop a more universal 16S rRNA gene primer sets that could cover close to the entire diversity of a natural microbial community (e.g., Earth Microbiome Project; Caporaso et al., 2012; Gilbert et al., 2014). The development of primer sets for the amplification of 16S rRNA genes is conducted *in silico* using reference databases (e.g., Klindworth et al., 2013). The Arctic Ocean is the smallest and shallowest of all five oceans, representing 4% of the area and 1% of the volume of the global ocean. Nevertheless, it plays an important role in global processes that are strongly affected by the ongoing climatic changes and is considered relevant for several Earth System tipping points (Wassmann and Reigstad, 2011; Lenton et al., 2019). Furthermore, being the coldest among the oceans, with strong stratification and only limited deep-water exchange, the Arctic Ocean is likely to contain unique endemic microbial diversity that drives its biogeochemical cycles (Kirchman et al., 2009; Ghiglione et al., 2012; Pedrós-Alió et al., 2015). An example for such locally adapted Arctic diversity was recently found with Arctic specific members of the ubiquitous SAR11 clade (Kraemer et al., 2020). Furthermore, despite its global importance, sampling effort in the Arctic Ocean is low, especially in and under the sea ice and in the deep basin, as well as generally during the wintertime (Wassmann et al., 2011; Royo-Llonch et al., 2020). Thus, the reference databases are likely lacking proper coverage of the complexity and dynamics of the Arctic Ocean microbiomes that may result in biased representations of them by currently available 16S rRNA gene primers.

One of the most extensively used primer set for the investigation of bacterial diversity in various environments is the 341F/785R (targeting the V3–V4 hypervariable regions of the 16S rRNA gene) that was developed by Klindworth et al. (2013). For the investigation of marine microbiomes, an alternative primer set 515F-Y/926R (targeting the V4–V5 hypervariable regions of the 16S rRNA gene), which is also able to capture

the diversity of the archaeal communities, has been developed by Parada et al. (2016). Currently, both V3–V4 and V4–V5 primer sets are widely used in studies of marine microbial communities and were extensively tested using mock and natural communities of temperate waters (e.g., Wear et al., 2018; Willis et al., 2019; McNichol et al., 2020). However, no study has systematically tested the performance of these primer sets on microbial communities of the Arctic Ocean.

In an attempt to select the most suitable primer set for the long-term monitoring of Arctic microbial communities as part of the FRAM Molecular Observatory, we present here a performance comparison of the 16S rRNA gene primer sets V3–V4 (341F/785R) and V4–V5 (515F-Y/926R). Our hypothesis was that due to relatively low representation of Arctic microbial communities in public databases (due to low number of existing studies), the 16S rRNA gene primer sets may capture different parts of microbial diversity in these unique environments. To test this hypothesis, we have conducted a direct comparison of the taxonomic coverage and potential biases of the two primer sets in 37 field samples collected from various environments of the Arctic Ocean, including sea-ice, surface and deep water column, marine snow, and deep-sea sediment. As an independent line of validation, we performed cell counting of five key taxonomic subgroups in a subset of the field samples *via* CARD-FISH (catalyzed reporter deposition-fluorescence *in situ* hybridization).

MATERIALS AND METHODS

Sample Collection

The samples included in this study were collected at the long-term ecological research (LTER) site HAUSGARTEN in Fram Strait and the central Arctic Ocean (**Supplementary Figure 1** and **Supplementary Table 1**). The samples were collected as follows:

- The sea-ice cores were collected using an ice corer (9 cm diameter; Kovacs Enterprise, Roseburg, OR, United States) and broken into subsections to facilitate quicker melting. The lower 30–50 cm of the sea ice (depending on total core length) was melted in plastic containers (rinsed with ethanol and ultrapure water) at 4°C in the dark. The melting of the sea ice took ∼24 h and the samples were immediately filtered on 0.22 μm Sterivex™ membranes as soon as the last piece of sea ice melted. Additional samples for microscopy counts were filtered onto 0.22 μm polycarbonate membranes (Whatman Nucleopore, Buckinghamshire, United Kingdom), with sterile filtered formalin at a final concentration of 2% and stored at −20°C.
- The water sampling was carried out using 12 L Niskin bottles mounted on a CTD rosette (Sea-Bird Electronics Inc., SBE 911 plus probe, Bellevue, WA, United States) and filtered on 0.22 μm Sterivex™ membranes. The Sterivex™ membranes were then stored at −20°C until further processing. Additional samples for microscopy counts were filtered onto 0.22 μm polycarbonate membranes (Whatman Nucleopore, Buckinghamshire, United Kingdom), with sterile filtered formalin at a final concentration of 2% and stored at −20°C.
- The deep-sea sediment cores were retrieved by a TV-guided multicorer, and subsamples of the uppermost centimeter of the cores were collected with syringes and immediately stored at −20°C until further processing.
- The marine snow samples were collected using sediment traps of the long-term moorings at the LTER site HAUSGARTEN (Bauerfeind et al., 2009; Lalande et al., 2013). Collection cups (400 ml) were filled with filtered seawater, adjusted to a salinity of 40 and poisoned with $HgCl_2$ (0.14% final solution) to preserve samples during deployment and after recovery (Metfies et al., 2017). After recovery, samples were stored at +4°C, swimmers were removed and samples were split by a wet splitting procedure (Bodungen et al., 2013). In this study, we used 1/32 splits of the original trap sample. Sinking particles from the sediment trap samples were collected on 0.22 μm Sterivex filters and stored at −20°C.

All metadata of the samples are accessible *via* the Data Publisher for Earth and Environmental Science PANGAEA[1], the PANGAEA event IDs are listed in **Supplementary Table 1**. Sampling map was produced using Ocean Data View v5.2.1 (Schlitzer, 2018).

DNA Isolation and 16S rRNA Gene Amplicon Sequencing

Genomic DNA was isolated in a combined chemical and mechanical procedure using the PowerWater DNA Isolation Kit for sea ice, water, and sediment traps and using the PowerSoil DNA Isolation Kit for sediment samples (MO BIO Laboratories, Inc., Carlsbad, CA, United States). Prior to DNA isolation, the 0.22 μm Sterivex™ membrane cartridges of the seawater and sea ice samples were cracked open in order to place the filters into the kit-supplied bead beating tubes. The isolation was continued according to the manufacturer's instructions, and DNA was stored at −20°C. Library preparation was performed according to the standard instructions of the 16S Metagenomic Sequencing Library Preparation protocol (Illumina™, Inc., San Diego, CA, United States). Two different hypervariable regions of the bacterial 16S rRNA gene were amplified using aliquots of the isolated DNA from each sample. The V3–V4 region was amplified using the S-D-Bact-0341-b-S-17 (5′-CCTACGGGNGGCWGCAG-3′) and the S-D-Bact-0785-a-A-21 (5′-GACTACHVGGGTATCTAATCC-3′) primers (Klindworth et al., 2013). The V4–V5 regions was amplified using the 515F-Y (5′-GTGYCAGCMGCCGCGGTAA-3′) and the 926R (5′-CCGYCAATTYMTTTRAGTTT-3′) primers (Parada et al., 2016). Sequences were obtained on the Illumina MiSeq™ platform in a 2 × 300 bp paired-end run and for surface water samples on the Illumina HiSeq™ platform in a 2 × 250 bp paired-end run (CeBiTec, Bielefeld, Germany), following the

[1]www.pangaea.de

standard instructions of the 16S Metagenomic Sequencing Library Preparation protocol.

Raw paired-end, primer-trimmed reads were deposited in the European Nucleotide Archive (ENA; Harrison et al., 2019) under accession number PRJEB31938. The data were archived using the brokerage service of the German Federation for Biological Data (GFBio; Diepenbroek et al., 2014).

Bioinformatics and Statistical Analyses

The raw paired-end reads were primer-trimmed using cutadapt (Martin, 2011). Further analyses were conducted using R v4.0.0[2] in RStudio v1.2.5042[3]. The libraries were processed using DADA2 v1.16 (Callahan et al., 2016a), following the suggested workflow (Callahan et al., 2016b). The reads in MiSeq libraries were truncated at 255 bp length for forward reads and at 200 bp length for reverse reads, to facilitate the technical quality drop at the end of the reads. Reads in both MiSeq and HiSeq were then trimmed for low-quality bases and merged based on a minimum overlap of 10 bp. Chimeras and amplicon sequence variants (ASVs) that were observed in only one sample were filtered out. The representative sequences were taxonomically classified against SILVA 16S rRNA gene reference database release 138 (Quast et al., 2013; Yilmaz et al., 2014). The ASVs that were taxonomically unclassified at phylum rank or were not assigned to bacterial or archaeal lineages were excluded from further analysis. Furthermore, all ASVs that were taxonomically assigned to mitochondria and chloroplast were removed from the dataset.

Sample data matrices were managed using the R package "phyloseq" v1.32 (McMurdie and Holmes, 2013), and plots were generated using the R package "ggplot2" v3.3.0 (Gómez-Rubio, 2017). The sample rarefaction analyses were conducted using the R package "iNEXT " v2.0.20 (Hsieh et al., 2016). To test the effect of the different primer sets on the taxonomic composition of the microbial communities, as well as to test for differences between microbial communities of different types of samples, a two-way permutation multivariate analysis of variance ("Two-way PERMANOVA") of Jensen–Shannon Divergence distance matrix was conducted (using the function "adonis2" in the R package "vegan" v.2.5.6; Oksanen et al., 2007).

Scripts for processing data can be accessed at https://github.com/edfadeev/Arctic-16S-Primers-comparison/.

Catalyzed Reporter Deposition Fluorescence *in situ* Hybridization

Both sea ice and seawater samples were directly fixed in 4% formalin for 4 h at 4°C, filtered onto 0.22 μm polycarbonate Track-Etched™ membranes (Whatman Nucleopore, Buckinghamshire, United Kingdom), and stored at −20°C. The CARD-FISH was applied based on the protocol established by Pernthaler et al. (2002), using horseradish-peroxidase (HRP)–labeled oligonucleotide probes ([4] Ulm, Germany; **Supplementary Table 4**). All probes were checked for specificity and coverage of their target groups against the SILVA 16S rRNA gene reference. All filters were embedded in 0.2% low-gelling-point agarose and treated with 10 mg mL^{-1} lysozyme solution (Sigma-Aldrich Chemie GmbH, Hamburg, Germany) for 1 h at 37°C. Subsequently, endogenous peroxidases were inactivated by submerging the filter pieces in 0.15% H_2O_2 in methanol for 30 min, before rinsing in Milli-Q water and dehydration in 96% ethanol. Then, the filters were covered in a hybridization buffer and a probe concentration of 0.2 ng μL^{-1}. Hybridization was performed at 46°C for 2.5 h, followed by washing in a pre-warmed washing buffer at 48°C for 10 min, and 15 min in 1x PBS. Signal amplification was carried out for 45 min at 46°C with an amplification buffer containing either tyramide-bound Alexa 488 (1 $\mu g/mL^{-1}$) or Alexa 594 (0.33 μg mL^{-1}). Afterward, the cells were counterstained in 1 $\mu g/mL^{-1}$ DAPI (4′,6-diamidino-2-phenylindole; Thermo Fisher Scientific GmbH, Bremen, Germany) for 10 min at 46°C. After rinsing with Milli-Q water and 96% ethanol, the filter pieces were embedded in a 4:1 mix of Citifluor (Citifluor Ltd., London, United Kingdom) and Vectashield (Vector Laboratories, Inc., Burlingame, United States) and stored overnight at −20°C for later microscopy evaluation.

Automated Image Acquisition and Cell Counting

The filters were evaluated microscopically under a Zeiss Axio Imager.Z2 stand (Carl Zeiss MicroImaging GmbH, Jena, Germany), equipped with a multipurpose fully automated microscope imaging system (MPISYS), a Colibri LED light source illumination system, and a multi-filter set 62HE (Carl Zeiss MicroImaging GmbH, Jena, Germany). Pictures were taken *via* a cooled charged-coupled-device (CCD) camera (AxioCam MRm; Carl Zeiss AG, Oberkochen, Germany) with a 63x oil objective, a numerical aperture of 1.4, and a pixel size of 0.1016 μm/pixel, coupled to the AxioVision SE64 Rel.4.9.1 software (Carl Zeiss AG, Oberkochen, Germany) as described by Bennke et al. (2016). Exposure times were adjusted after manual inspection with the AxioVision Rel.4.8 software coupled to the SamLoc 1.7 software (Zeder et al., 2011), which was also used to define the coordinates of the filters on the slides. For image acquisition, channels were defined with the MPISYS software, and a minimum of 55 fields of view with a minimum distance of 0.25 mm were acquired of each filter piece by recording a *z*-stack of seven images in autofocus.

Cell enumeration was performed with the software Automated Cell Measuring and Enumeration Tool (ACMETool3, 2018-11-09; M. Zeder, Technobiology GmbH, Buchrain, Switzerland). Cells were counted as objects according to manually defined parameters separately for the DAPI and FISH channels.

RESULTS AND DISCUSSION

In this study, aliquots of 37 DNA samples from different environments in the Arctic Ocean (sea ice, surface and deep ocean water, marine snow, and seafloor sediment; **Supplementary Table 1**) were sequenced using two common primers sets that amplify either the V3–V4 or the V4–V5 hypervariable regions in the 16S rRNA gene and were subjected

[2]http://www.Rproject.org/

[3]http://www.rstudio.com/

[4]Biomers.net

to the same bioinformatic workflow. Both primer sets showed a similar decrease in the number of sequences throughout the workflow, with 62 ± 13% and 68 ± 9% of sequences retained per sample, respectively. The final datasets consisted of 3,318,649 sequences in the V3–V4 dataset that were assigned to 12,045 ASVs and 3,340,628 sequences in the V4–V5 dataset that were assigned to 14,505 ASVs (**Supplementary Table 2**). In addition, the ASVs which were taxonomically assigned to eukaryotic, mitochondrial or chloroplast sequences, as well as ASVs unclassified at phylum rank, were also removed from further analysis (ca. 9% and ca. 17% of sequences in V3–V4 and V4–V5 datasets, respectively). In both datasets, an asymptotic extrapolation of the rarefaction curves did not further increase the number of observed ASVs (**Supplementary Figure 2**). Although, most likely further microbial diversity remains to be uncovered in all sampled environments, the rarefaction curves suggest that our samples contained most of the potential community richness covered by both primer sets. In sea ice, surface water (<30 m depth) and marine snow, both primer sets showed similar community richness (**Figure 1**). However, in the deep-water communities (>600 m depth), richness was significantly different between the primer sets (Wilcoxon Signed-Rank Test; $p < 0.01$), with ca. 40% more bacterial ASVs in the V3–V4. In contrast, the sediment community richness was significantly higher in the V4–V5 dataset (Wilcoxon signed-rank test; $p < 0.01$), with up to double the amount of bacterial ASVs compared to the V3–V4 dataset. The main taxonomic groups, typically observed in the Arctic marine environment, such as the classes *Alphaproteobacteria*, *Bacteroidia*, and *Gammaproteobacteria*, dominated both datasets (each comprising 10–30% of sequences in V3–V4 and V4–V4 datasets, respectively). However, within these groups significant differences between datasets in the number of observed ASVs were detected.

In the V3–V4 dataset, the *Bacteroidia* and *Gammaproteobacteria* showed the highest differences in number of observed ASVs within each class (i.e., type richness) compared to the V4–V5 dataset (**Supplementary Table 2**). The family *Flavobacteriaceae* (class *Bacteroidia*) comprised 18% of all sequences in both datasets; however, in the V3–V4 dataset, it consisted of one third more ASVs compared to the V4–V5 dataset (total of 278 and 196 ASVs, respectively; **Figure 2**). This difference in the number of observed ASVs was mainly associated with ASVs of the genus *Polaribacter* (total of 28 and 14 ASVs, respectively), a key heterotrophic bacterium that responds to phytoplankton blooms in mid- and high-latitudes (Gómez-Pereira et al., 2010; Fadeev et al., 2018; Avcı et al., 2020). The orders *Alteromonadales*, *Cellvibrionales*, and *Oceanospirillales* (all within the class *Gammaproteobacteria*), which comprised 4–6% of all sequences in the V3–V4 dataset and 3% of all sequences in the V4–V5 dataset, also showed differences between datasets in the number of observed ASVs (**Supplementary Table 3**). Each of these *Gammaproteobacteria* orders contained two times more ASVs in the V3–V4 dataset, compared to the V4–V5 dataset (the largest difference was in the order *Alteromonadales*, with total of 113 and 49 ASVs, respectively). These taxonomic groups are typically associated with organic matter degradation (Buchan et al., 2014), and were previously shown to dominate sea ice microbial communities associated with algal aggregates (Rapp et al., 2018), as well as surface waters during phytoplankton blooms (Fadeev et al., 2018). Furthermore, the family *Woeseiaceae* (class *Gammaproteobacteria*) also consisted of ca. 30% more ASVs in the V3–V4 dataset, compared to the V4–V5 dataset (total of 127 and 98 ASVs, respectively; **Figure 2**). This bacterial family is abundant in deep-sea sediments around the globe, including the Arctic Ocean (Bienhold et al., 2016; Hoffmann et al., 2020).

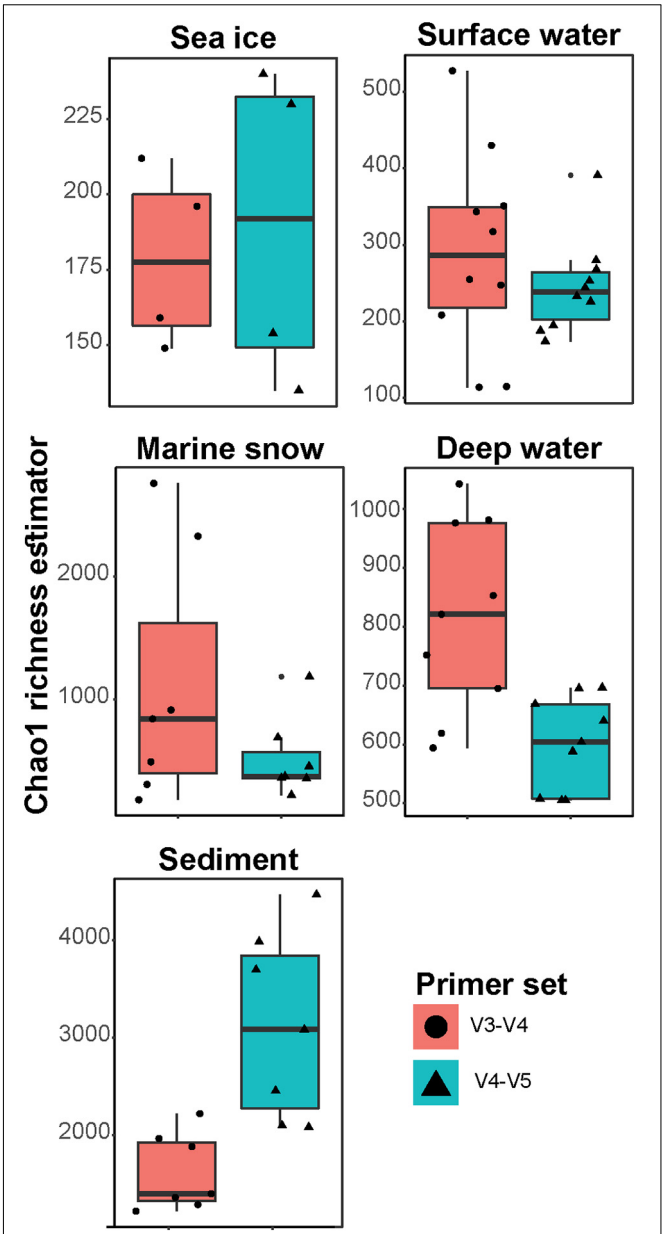

FIGURE 1 | Chao1 richness estimates in the different sample types. Different primer sets represented by colors and shapes. Please note the differences of *y*-axis between the panels.

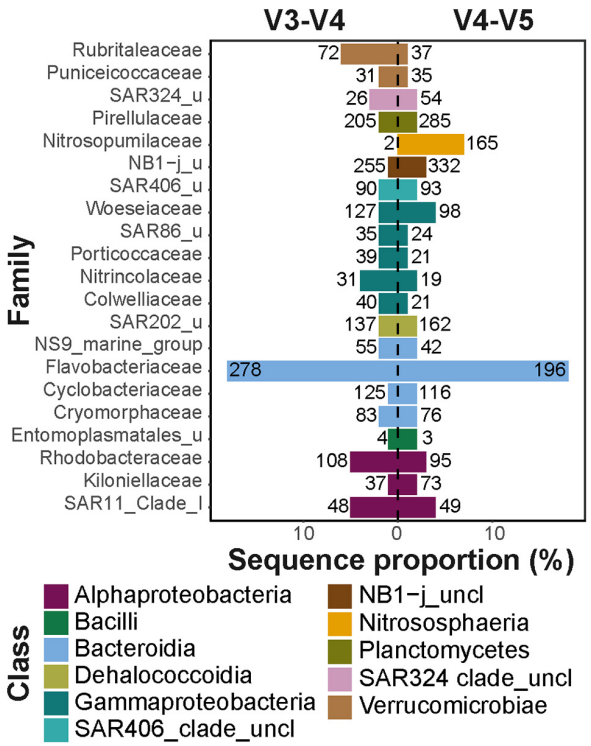

FIGURE 2 | Major taxonomic families in V3–V4 and V4–V5 datasets. The *x*-axis represents the total sequence proportion of each family in V3–V4 (left panel) and V4–V5 (right panel) datasets. The numbers at each column represent the number of observed ASVs affiliated with each taxonomic family. Different taxonomic classes are represented by color code. Only families that comprised at least 1% of sequences in at least one of the datasets were included in the visualization.

Compared to the V3–V4 dataset, the V4–V5 dataset consisted of at least one third more ASVs in the classes *Phycisphaerae* (total of 206 and 117 ASVs, respectively) and *Planctomycetes* (total of 299 and 244 ASVs, respectively). This difference in the number of observed ASVs was mainly associated with the families *Pirellulaceae* that comprised ca. 2% of all sequences in both datasets (**Figure 2**), as well as *Phycisphaeraceae* that comprised less than 1% of all sequences (**Supplementary Table 3**) in both datasets. These taxonomic groups have been previously shown to be associated with sinking particles in the deep ocean and are also abundant in Arctic deep-sea sediments (Fadeev et al., 2020). Furthermore, the archaeal class *Nitrososphaeria* was almost absent from the V3–V4 dataset, with only a few sequences associated with four ASVs, compared to 168 ASVs in the V4–V5 dataset that comprised 7% of the total sequences (**Figure 2**). Marine members of the *Archaea* in general, and the class *Nitrososphaeria* in particular, are abundant in the Arctic marine environment and can reach up to one fifth of the cells in Arctic microbial communities (Müller et al., 2018; Cardozo-Mino et al., 2020). Taken together, these observations suggest that on ASV level the diversity of different taxonomic groups are captured differently by the two primer sets. This is potentially a result of differences in the regional hypervariability of the 16S rRNA gene within different taxonomic groups (Yang et al., 2016; Kerrigan et al., 2019). In addition, as was previously shown for various taxonomic groups, such as the SAR11 clade, differences in captured diversity may occur also due to specificity differences of the primer sets to the targeted 16S rRNA gene region (Parada et al., 2016).

Despite the observed differences on an ASV level, the overall taxonomic composition was consistent between the datasets (**Figure 3**). Sampled sea ice, surface water, and marine snow communities were dominated by heterotrophic bacteria of the classes *Bacteroidia* (mainly the genus *Polaribacter*) and *Gammaproteobacteria* (mainly the genera in the order *Alteromonadales*), with equivalent relative sequence abundances to those described in previous reports (Bowman et al., 2012; Eronen-Rasimus et al., 2016; Hatam et al., 2016; Wilson et al., 2017; Fadeev et al., 2018, 2020; Rapp et al., 2018). At depth, pelagic communities were dominated by sequences of the class *Alphaproteobacteria*, SAR324 clade, and the archaeal class *Nitrososphaeria*, all of which were previously observed to dominate deep Arctic waters, as well as surface communities during the Arctic winter (Wilson et al., 2017; Fadeev et al., 2020). The sediment communities, which have previously been shown to harbor the highest taxonomic diversity among the described Arctic environments by far (Bienhold et al., 2012; Hoffmann et al., 2017; Rapp et al., 2018), were dominated in sequence abundance of *Gammaproteobacteria*.

In order to compare the differences in representation of taxonomic groups between the primer sets, we combined sequence abundances of all ASVs according to their taxonomic affiliation at genus rank (i.e., the highest possible shared between the datasets taxonomic resolution). In the V3–V4 dataset, the ASVs were merged into 306 different genera and 279 lineages that were affiliated to higher taxonomic ranks (i.e., were unclassified on a genus rank). In the V4–V5 dataset, the ASVs were merged into 280 different genera and 299 lineages that were affiliated to higher taxonomic ranks. Overall, 489 (72% of the total) lineages were observed in both datasets at this level of taxonomic resolution. In the V3–V4 dataset there were 96 (14% of the total) lineages that were absent from the V4–V5 dataset, but together they comprised less than 1% of the sequences in the V3–V4 dataset. On the other hand, in the V4–V5 dataset there were 90 (13% of the total) lineages that were absent from the V3–V4 dataset, and together they comprised 5% of the sequences in the V4–V5 dataset. In addition, the dissimilarity of community compositions in merged V3–V4 and V4–V5 datasets revealed consistent and significant difference between the microbiomes captured by both primer sets (Two-way PERMANOVA test; $F_{4,64} = 86.29$, $R^2 = 0.83$, p value < 0.001; **Figure 4**). Only a small fraction of the total variance was associated with the difference between the primer sets (Two-way PERMANOVA test; $F_{1,64} = 7.59$, $R^2 = 0.02$, p value < 0.001). No significant combined effect of different primer sets on different sample types was observed (Two-way PERMANOVA test; p value > 0.05). Taken together, these results confirm that, even though the primer sets showed different sensitivity to diversity at the ASV level, both of them reflect similar taxonomic composition down to the genus rank.

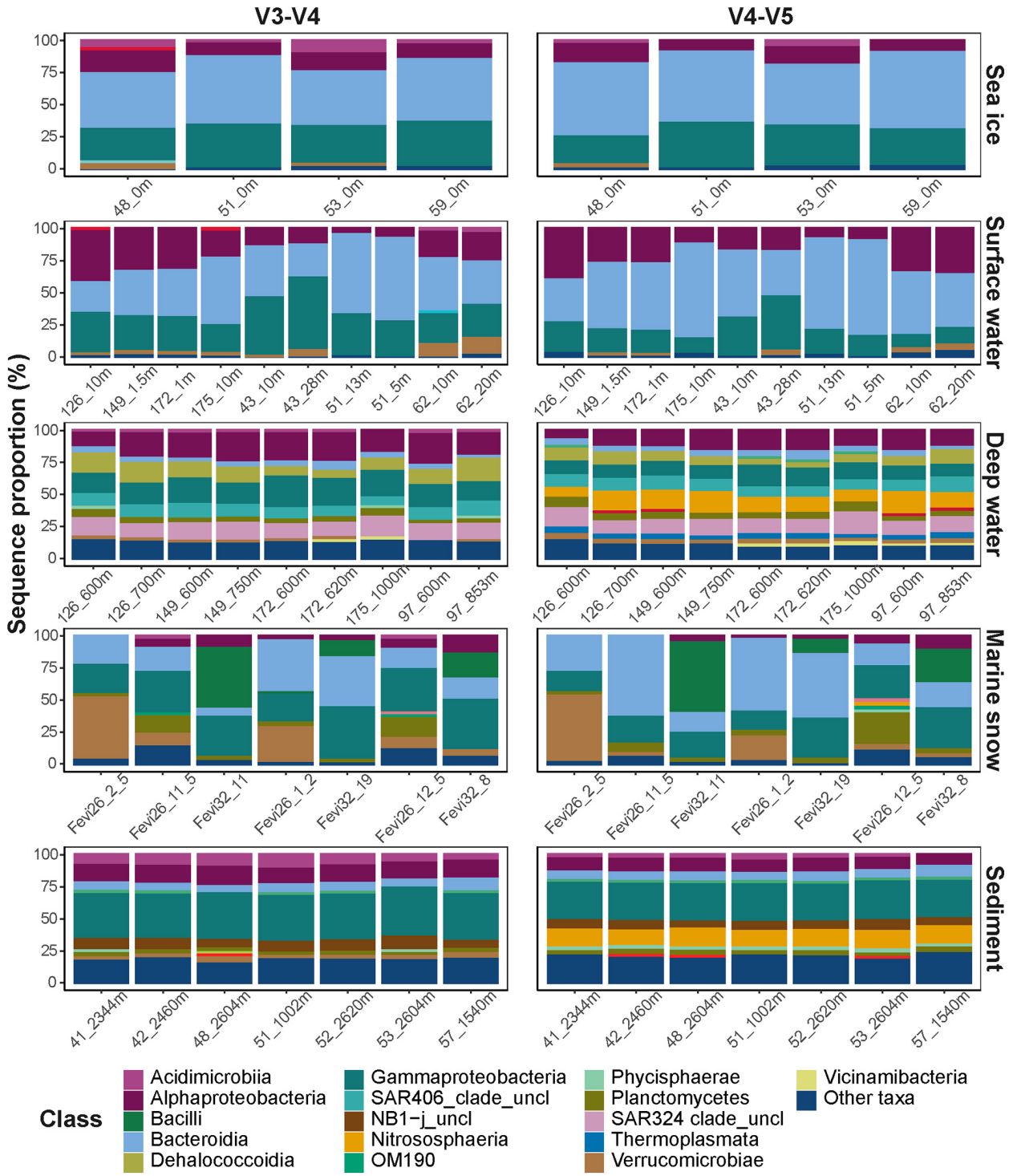

FIGURE 3 | Taxonomic compositions of the microbial communities. Different taxonomic classes are represented by color code. Classes with sequence proportion below 2% were classified as "Other taxa".

The research at the FRAM Microbial Observatory is focused on the seasonal and interannual dynamics of the Arctic Ocean associated with changes in sea ice extent and primary production in the surface ocean (e.g., Metfies et al., 2017; Fadeev et al., 2018, 2020). To further evaluate the performance of the two primer sets in these long-term monitored environments, we compared the sequence representation of selected taxonomic groups, which are associated with distinct stages of seasonal

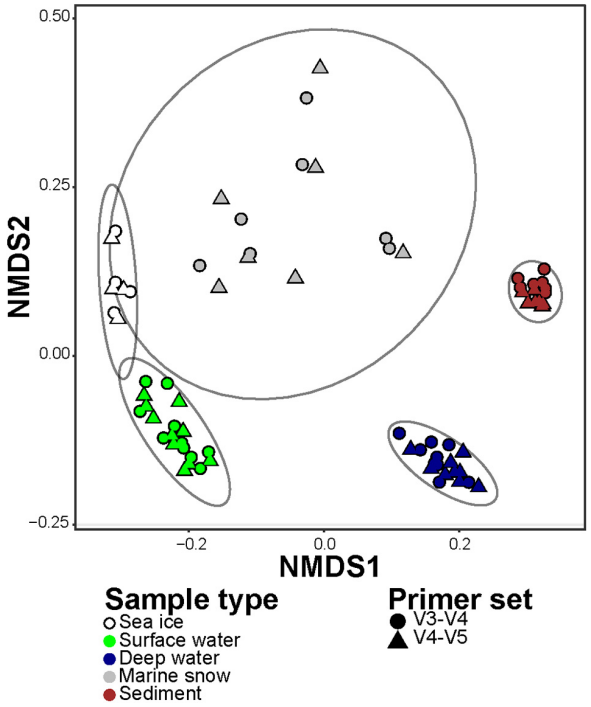

FIGURE 4 | Non-metric Multi-dimensional Scaling (NMDS) of the microbial communities in merged V3–V4 and V4–V5 datasets, based on Jensen-Shannon Divergence. The different types of samples are represented by colors, and the primer set are represented by shapes. Ellipses encompass clustering of each microbiome type with normal confidence of 0.95.

dynamics (Fadeev et al., 2018), to microscopically counted cells using CARD-FISH combined with an automated image acquisition (Cardozo-Mino et al., 2020). Fluorescence *in situ* hybridization techniques have the advantage of providing absolute abundances of (viable) cells that can be directly compared between samples. In microscopy counts of both sea ice and surface water communities the highest observed cell abundance was of the class *Bacteroidia* (up to 35 and 26% of the total microbial community, respectively), which was consistent with the representation of this taxonomic group by both primer sets. In surface water communities, high levels of consistency between the microscopy counts and both primer sets were observed also in the representation of *Alteromonadales* and *Polaribacter* (**Table 1**). On the other hand, the representation, by both primer sets, of the class *Gammaproteobacteria* in sea ice and surface water communities was 2–4 times higher in comparison to the proportion observed in microscopy counts (up to 9 and 18%, respectively). In contrast, the proportional abundance of the SAR11 clade was 5–10 times higher in the microscopy counts, compared to its representation by both primer sets (**Table 1**). Our results suggest that at least for some taxonomic groups (i.e., *Polaribacter*), both primer sets may provide a consistent semi-quantitative representation. However, microscopy results must be interpreted under several methodological caveats, knowing that low cellular ribosome content or low efficiency of the probe may alter the representation of individual taxa in our cell counts (Amann and Fuchs, 2008). Therefore, the observed inconsistency in the representation of some taxonomic groups (i.e., SAR11 clade) may also result from these limitations. In order to further investigate the quantitative performance of the primer sets, further investigation, using techniques such as mock communities (Yeh et al., 2019) or metagenomics (McNichol et al., 2020), is required.

CONCLUDING REMARKS

To understand the links between the rapid environmental changes in the Arctic region and the dynamics of microbial communities in the Arctic Ocean, there is a need for robust methods addressing changes in diversity and relative abundance. In order to conduct such observations using a 16S rRNA gene tag-sequencing approach, optimally a similar extraction method and a single PCR primer set should be selected, which can be applied to all environments of the Arctic Ocean (sea ice, water column, and deep-sea sediment). The most suitable primer set for 16S rRNA amplification and sequencing from environmental samples should produce high-quality amplicon libraries and cover with minimum biases the variety of present

TABLE 1 | Overview of cell abundances and sequence proportions range in selected taxonomic groups.

Sample	Taxonomic group	Abundance (10^5 cells mL^{-1})	% of DAPI counts	% of total community (V3–V4 dataset)	% of total community (V4–V5 dataset)
Sea ice	*Gammaproteobacteria* (c)	1.0–2.2	8–9%	34–35%	28–35%
	Alteromonadales (o)	0.2–0.7	1–6%	23–25%	19–26%
	Bacteroidia (c)	2.1–3.9	9–35%	48–52%	55–60%
	Polaribacter (g)	1.0–1.5	4–13%	16–17%	14–20%
	SAR11 clade (o)	0.2–0.7	1–6%	0.2–0.4%	0.4–0.8%
Surface water	*Gammaproteobacteria* (c)	4.9–9.8	13–18%	32–55%	19–42%
	Alteromonadales (o)	0.8–2.6	2–7%	1–8%	1–4%
	Bacteroidia (c)	10.0–12.0	23–26%	25–61%	35–70%
	Polaribacter (g)	5.0–7.7	9–20%	11–35%	13–36%
	SAR11 clade (o)	11.5–15.4	29–30%	1–6%	3–10%

The selected probes and their coverage are described in ***Supplementary Table 3****. c, class; o, order; g, genus.*

organisms, as well as their relative abundances. We have found that at all taxonomic ranks down to genus, both primer sets represent the overall richness of the major bacterial taxonomic groups at comparable levels across the different Arctic Ocean biomes. The relative sequence abundance of some dominant taxonomic groups, such as the *Polaribacter*, corresponds with their proportional representation *via* microscopic cell counts. Other taxonomic groups such as the SAR11 clade strongly differ between the molecular and the microscopical representations. However, this discrepancy may be due to limitations of the microscopical quantification. On an ASV level, both primer sets capture the diversity within the most abundant taxonomic groups differently, and thus, the use of each primer set may depend on the target groups. However, the main advantage of the V4–V5 primer set is its additional coverage of the archaeal domain, without compromising the detection of other taxonomic groups. Members of the *Archaea* comprise a substantial fraction of Arctic marine microbial communities, particularly during the dark season and in deep waters. Thus, given the demonstrated similarities and differences, we endorse the use of the V4–V5 primer set for capturing comprehensive insights into microbial community dynamics of the Arctic marine environment.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

EF and AB designed the study. CB, IS, MM, and JR provided the environmental samples for the study. HT conducted the sequencing of the samples. MC-M and VS-C conducted the CARD-FISH counts. EF analyzed the data and wrote the manuscript. All authors contributed to the final version of the manuscript.

FUNDING

This work was conducted in the framework of the HGF Infrastructure Program FRAM of the Alfred-Wegener-Institute Helmholtz Center for Polar and Marine Research. The project has received funding from the European Research Council (ERC) under the European Union's Seventh Framework Program (FP7/2007-2013) research project ABYSS (grant agreement no. 294757) awarded to AB. Additional funding came from the Austrian Science Fund (FWF) M-2797 to EF. This publication is Eprint ID 53010 of the Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany.

ACKNOWLEDGMENTS

We thank Daniel Sher for a critical review of the manuscript. We would also like to thank Jana Bäger, Jakob Barz, and Theresa Hargesheimer for DNA extractions and library preparations.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.637526/full#supplementary-material

REFERENCES

Alonso-Sáez, L., Sánchez, O., Gasol, J. M., Balagué, V., and Pedrós-Alio, C. (2008). Winter-to-summer changes in the composition and single-cell activity of near-surface Arctic prokaryotes. *Environ. Microbiol.* 10, 2444–2454. doi: 10.1111/j.1462-2920.2008.01674.x

Amann, R., and Fuchs, B. M. (2008). Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. *Nat. Rev. Microbiol.* 6, 339–348. doi: 10.1038/nrmicro1888

Apprill, A., McNally, S., Parsons, R., and Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat. Microb. Ecol.* 75, 129–137. doi: 10.3354/ame01753

Armougom, F. (2009). Exploring microbial diversity using 16S rRNA high-throughput methods. *J. Comput. Sci. Syst. Biol.* 2, 74–92. doi: 10.4172/jcsb.1000019

Avcı, B., Krüger, K., Fuchs, B. M., Teeling, H., and Amann, R. I. (2020). Polysaccharide niche partitioning of distinct Polaribacter clades during North Sea spring algal blooms. *ISME J.* 14, 1369–1383. doi: 10.1038/s41396-020-0601-y

Bauerfeind, E., Nöthig, E.-M., Beszczynska, A., Fahl, K., Kaleschke, L., Kreker, K., et al. (2009). Particle sedimentation patterns in the eastern Fram Strait during 2000–2005: results from the Arctic long-term observatory HAUSGARTEN. *Deep Sea Res. Part I Oceanogr. Res. Pap.* 56, 1471–1487. doi: 10.1016/j.dsr.2009.04.011

Bennke, C. M., Reintjes, G., Schattenhofer, M., Ellrott, A., Wulf, J., Zeder, M., et al. (2016). Modification of a high-throughput automatic microbial cell enumeration system for shipboard analyses. *Appl. Environ. Microbiol.* 82, 3289–3296. doi: 10.1128/AEM.03931-15

Bienhold, C., Boetius, A., and Ramette, A. (2012). The energy–diversity relationship of complex bacterial communities in Arctic deep-sea sediments. *ISME J.* 6, 724–732. doi: 10.1038/ismej.2011.140

Bienhold, C., Zinger, L., Boetius, A., and Ramette, A. (2016). Diversity and biogeography of bathyal and abyssal seafloor bacteria. *PLoS One* 11:e0148016. doi: 10.1371/journal.pone.0148016

Bodungen, B. V., Wunsch, M., and Fürderer, H. (2013). "Sampling and analysis of suspended and sinking particles in the Northern North Atlantic," in *Marine Particles: Analysis and Characterization*, eds D. C. Hurd, and D. W. Spencer, (Washington, DC: American Geophysical Union), 47–56. doi: 10.1029/GM063p0047

Boetius, A., Anesio, A. M., Deming, J. W., Mikucki, J. A., and Rapp, J. Z. (2015). Microbial ecology of the cryosphere: sea ice and glacial habitats. *Nat. Rev. Microbiol.* 13, 677–690. doi: 10.1038/nrmicro3522

Bowman, J. S., Rasmussen, S., Blom, N., Deming, J. W., Rysgaard, S., and Sicheritz-Ponten, T. (2012). Microbial community structure of Arctic multiyear sea ice and surface seawater by 454 sequencing of the 16S RNA gene. *ISME J.* 6, 11–20. doi: 10.1038/ismej.2011.76

Buchan, A., LeCleir, G. R., Gulvik, C. A., and Gonzalez, J. M. (2014). Master recyclers: features and functions of bacteria associated with

phytoplankton blooms. *Nat. Rev. Microbiol.* 12, 686–698. doi: 10.1038/nrmicro3326

Buttigieg, P. L., Fadeev, E., Bienhold, C., Hehemann, L., Offre, P., and Boetius, A. (2018). Marine microbes in 4D — using time series observation to assess the dynamics of the ocean microbiome and its links to ocean health. *Curr. Opin. Microbiol.* 43, 169–185. doi: 10.1016/j.mib.2018.01.015

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016a). DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869

Callahan, B. J., Sankaran, K., Fukuyama, J. A., McMurdie, P. J., and Holmes, S. P. (2016b). Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses. *F1000 Res.* 5:1492. doi: 10.12688/f1000research.8986.2

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624. doi: 10.1038/ismej.2012.8

Cardozo-Mino, M. G., Fadeev, E., Salman-Carvalho, V., and Boetius, A. (2020). Spatial dynamics in Arctic bacterioplankton community densities are strongly linked to distinct physical and biological processes (Fram Strait, 79° N). *bioRxiv[Preprint]* doi: 10.1101/2020.09.02.277574

Dai, A., Luo, D., Song, M., and Liu, J. (2019). Arctic amplification is caused by sea-ice loss under increasing CO_2. *Nat. Commun.* 10:121. doi: 10.1038/s41467-018-07954-9

Diepenbroek, M., Glöckner, F. O., Grobe, P., Güntsch, A., Huber, R., König-Ries, B., et al. (2014). "Towards an integrated biodiversity and ecological research data management and archiving platform: the German federation for the curation of biological data (GFBio)," in *Informatik 2014*, eds E. Plödereder, L. Grunske, E. Schneider, and D. Ull, (Bonn: Gesellschaft für Informatik e.V), 1711–1721.

Eronen-Rasimus, E., Piiparinen, J., Karkman, A., Lyra, C., Gerland, S., and Kaartokallio, H. (2016). Bacterial communities in Arctic first-year drift ice during the winter/spring transition. *Environ. Microbiol. Rep.* 8, 527–535. doi: 10.1111/1758-2229.12428

Fadeev, E., Rogge, A., Ramondenc, S., Nöthig, E.-M., Wekerle, C., Bienhold, C., et al. (2020). Sea-ice retreat may decrease carbon export and vertical microbial connectivity in the Eurasian Arctic basins. *Nat. Res.* [Preprint]. doi: 10.21203/rs.3.rs-101878/v1

Fadeev, E., Salter, I., Schourup-Kristensen, V., Nöthig, E.-M., Metfies, K., Engel, A., et al. (2018). Microbial communities in the east and west fram strait during sea ice melting season. *Front. Mar. Sci.* 5:429. doi: 10.3389/fmars.2018.00429

Fuhrman, J. A., Cram, J. A., and Needham, D. M. (2015). Marine microbial community dynamics and their ecological interpretation. *Nat. Rev. Microbiol.* 13, 133–146. doi: 10.1038/nrmicro3417

Galand, P. E., Pereira, O., Hochart, C., Auguet, J. C., and Debroas, D. (2018). A strong link between marine microbial community composition and function challenges the idea of functional redundancy. *ISME J.* 12, 2470–2478. doi: 10.1038/s41396-018-0158-1

Ghiglione, J.-F., Galand, P. E., Pommier, T., Pedrós-Alió, C., Maas, E. W., Bakker, K., et al. (2012). Pole-to-pole biogeography of surface and deep marine bacterial communities. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17633–17638. doi: 10.1073/pnas.1208160109

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

Gómez-Pereira, P. R., Fuchs, B. M., Alonso, C., Oliver, M. J., van Beusekom, J. E. E., and Amann, R. (2010). Distinct flavobacterial communities in contrasting water masses of the North Atlantic Ocean. *ISME J.* 4, 472–487. doi: 10.1038/ismej.2009.142

Gómez-Rubio, V. (2017). *ggplot2 - Elegant Graphics for Data Analysis*, 2nd Edn. doi: 10.18637/jss.v077.b02

Harrison, P. W., Alako, B., Amid, C., Cerdeño-Tárraga, A., Cleland, I., Holt, S., et al. (2019). The European nucleotide archive in 2018. *Nucleic Acids Res.* 47, D84–D88. doi: 10.1093/nar/gky1078

Hatam, I., Lange, B., Beckers, J., Haas, C., and Lanoil, B. (2016). Bacterial communities from Arctic seasonal sea ice are more compositionally variable than those from multi-year sea ice. *ISME J.* 10, 2543–2552. doi: 10.1038/ismej.2016.4

Hoffmann, K., Bienhold, C., Buttigieg, P. L., Knittel, K., Laso-Pérez, R., Rapp, J. Z., et al. (2020). Diversity and metabolism of Woeseiales bacteria, global members of marine sediment communities. *ISME J.* 14, 1042–1056. doi: 10.1038/s41396-020-0588-4

Hoffmann, K., Hassenrück, C., Salman-Carvalho, V., Holtappels, M., and Bienhold, C. (2017). Response of bacterial communities to different detritus compositions in arctic deep-sea sediments. *Front. Microbiol.* 8:266. doi: 10.3389/fmicb.2017.00266

Hsieh, T. C., Ma, K. H., and Chao, A. (2016). iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods Ecol. Evol.* 7, 1451–1456. doi: 10.1111/2041-210X.12613

Karl, D. M., and Church, M. J. (2014). Microbial oceanography and the Hawaii Ocean Time-series programme. *Nat. Rev. Microbiol.* 12, 699–713. doi: 10.1038/nrmicro3333

Kerrigan, Z., Kirkpatrick, J. B., and D'Hondt, S. (2019). Influence of 16S rRNA hypervariable region on estimates of bacterial diversity and community composition in seawater and marine sediment. *Front. Microbiol.* 10:1640. doi: 10.3389/fmicb.2019.01640

Kirchman, D. L., Elifantz, H., Dittel, A. I., Malmstrom, R. R., and Cottrell, M. T. (2007). Standing stocks and activity of archaea and bacteria in the western Arctic Ocean. *Limnol. Oceanogr.* 52, 495–507. doi: 10.4319/lo.2007.52.2.0495

Kirchman, D. L., Morán, X. A. G., and Ducklow, H. (2009). Microbial growth in the polar oceans - Role of temperature and potential impact of climate change. *Nat. Rev. Microbiol.* 7, 451–459. doi: 10.1038/nrmicro2115

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41:e1. doi: 10.1093/nar/gks808

Kraemer, S., Ramachandran, A., Colatriano, D., Lovejoy, C., and Walsh, D. A. (2020). Diversity and biogeography of SAR11 bacteria from the Arctic Ocean. *ISME J.* 14, 79–90. doi: 10.1038/s41396-019-0499-4

Lalande, C., Bauerfeind, E., Nöthig, E., and Beszczynska-Möller, A. (2013). Impact of a warm anomaly on export fluxes of biogenic matter in the eastern Fram Strait. *Prog. Oceanogr.* 109, 70–77. doi: 10.1016/j.pocean.2012.09.006

Lenton, T. M., Rockström, J., Gaffney, O., Rahmstorf, S., Richardson, K., Steffen, W., et al. (2019). Climate tipping points — too risky to bet against. *Nature* 575, 592–595. doi: 10.1038/d41586-019-03595-0

Liu, Y., Blain, S., Crispi, O., Rembauville, M., and Obernosterer, I. (2020). Seasonal dynamics of prokaryotes and their associations with diatoms in the Southern Ocean as revealed by an autonomous sampler. *Environ. Microbiol.* 22, 3968–3984. doi: 10.1111/1462-2920.15184

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17:10. doi: 10.14806/ej.17.1.200

McMurdie, P. J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. doi: 10.1371/journal.pone.0061217

McNichol, J. C., Berube, P. M., Biller, S. J., and Fuhrman, J. A. (2020). Evaluating and improving SSU rRNA PCR primer coverage via metagenomes from global ocean surveys. *bioRxiv [Preprint]* doi: 10.1101/2020.11.09.375543

Metfies, K., Bauerfeind, E., Wolf, C., Sprong, P., Frickenhaus, S., Kaleschke, L., et al. (2017). Protist communities in moored long-term sediment traps (Fram Strait, Arctic)–preservation with mercury chloride allows for PCR-based molecular genetic analyses. *Front. Mar. Sci.* 4:301. doi: 10.3389/fmars.2017.00301

Müller, O., Wilson, B., Paulsen, M. L., Ruminska, A., Armo, H. R., Bratbak, G., et al. (2018). Spatiotemporal dynamics of ammonia-oxidizing Thaumarchaeota in Distinct Arctic water masses. *Front. Microbiol.* 9:24. doi: 10.3389/fmicb.2018.00024

Nikrad, M. P., Cottrell, M. T., and Kirchman, D. L. (2012). Abundance and single-cell activity of heterotrophic bacterial groups in the Western Arctic Ocean in summer and winter. *Appl. Environ. Microbiol.* 78, 2402–2409. doi: 10.1128/AEM.07130-11

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M. H. H., Oksanen, M. J., et al. (2007). The vegan package. *Commun. Ecol. Packag.* 10:719.

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

Pedrós-Alió, C., Potvin, M., and Lovejoy, C. (2015). Diversity of planktonic microorganisms in the Arctic Ocean. *Prog. Oceanogr.* 139, 233–243. doi: 10.1016/j.pocean.2015.07.009

Peeken, I., Primpke, S., Beyer, B., Gütermann, J., Katlein, C., Krumpen, T., et al. (2018). Arctic sea ice is an important temporal sink and means of transport for microplastic. *Nat. Commun.* 9:1505. doi: 10.1038/s41467-018-03825-5

Peng, G., and Meier, W. N. (2018). Temporal and regional variability of Arctic sea-ice coverage from satellite data. *Ann. Glaciol.* 59, 191–200. doi: 10.1017/aog.2017.32

Pernthaler, A., Pernthaler, J., and Amann, R. (2002). Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl. Environ. Microbiol.* 68, 3094–3101. doi: 10.1128/AEM.68.6.3094-3101.2002

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Glo, F. O., et al. (2013). The SILVA ribosomal RNA gene database project : improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219

Rapp, J. Z., Fernández-Méndez, M., Bienhold, C., and Boetius, A. (2018). Effects of ice-algal aggregate export on the connectivity of bacterial communities in the central Arctic Ocean. *Front. Microbiol.* 9:1035. doi: 10.3389/fmicb.2018.01035

Royo-Llonch, M., Sánchez, P., Ruiz-González, C., Salazar, G., Pedrós-Alió, C., Labadie, K., et al. (2020). Ecogenomics of key prokaryotes in the arctic ocean. *bioRxiv [Preprint]* doi: 10.1101/2020.06.19.156794

Schlitzer, R. (2018). *Ocean Data View*. Available online at: https://odv.awi.de (accessed July 28, 2019).

Soltwedel, T., Bauerfeind, E., Bergmann, M., Bracher, A., Budaeva, N., Busch, K., et al. (2015). Natural variability or anthropogenically-induced variation? Insights from 15 years of multidisciplinary observations at the arctic marine LTER site HAUSGARTEN. *Ecol. Indic.* 65, 89–102. doi: 10.1016/j.ecolind.2015.10.001

Soltwedel, T., Bauerfeind, E., Bergmann, M., Budaeva, N., Hoste, E., Jaeckisch, N., et al. (2005). HAUSGARTEN: multidisciplinary investigations at a deep-sea, long-term observatory in the Arctic Ocean. *Oceanography* 18, 46–61. doi: 10.5670/oceanog.2005.24

Thomas, T., Gilbert, J., and Meyer, F. (2012). Metagenomics - a guide from sampling to data analysis. *Microb. Inform. Exp.* 2:3. doi: 10.1186/2042-5783-2-3

Wang, Y., and Qian, P.-Y. (2009). Conservative fragments in bacterial 16S rRNA genes and primer design for 16S Ribosomal DNA amplicons in metagenomic studies. *PLoS One* 4:e7401. doi: 10.1371/journal.pone.0007401

Wassmann, P., Duarte, C. M., Agustí, S., and Sejr, M. K. (2011). Footprints of climate change in the Arctic marine ecosystem. *Glob. Chang. Biol.* 17, 1235–1249. doi: 10.1111/j.1365-2486.2010.02311.x

Wassmann, P., and Reigstad, M. (2011). Future Arctic Ocean seasonal ice zones and implications for pelagic-benthic coupling. *Oceanography* 24, 220–231. doi: 10.5670/oceanog.2011.74

Wear, E. K., Wilbanks, E. G., Nelson, C. E., and Carlson, C. A. (2018). Primer selection impacts specific population abundances but not community dynamics in a monthly time-series 16S rRNA gene amplicon analysis of coastal marine bacterioplankton. *Environ. Microbiol.* 20, 2709–2726. doi: 10.1111/1462-2920.14091

Willis, C., Desai, D., and Laroche, J. (2019). Influence of 16S rRNA variable region on perceived diversity of marine microbial communities of the Northern North Atlantic. *FEMS Microbiol. Lett.* 366:fnz152. doi: 10.1093/femsle/fnz152

Wilson, B., Müller, O., Nordmann, E. L., Seuthe, L., Bratbak, G., and Øvreås, L. (2017). Changes in marine prokaryote composition with season and depth over an Arctic polar year. *Front. Mar. Sci.* 4:95. doi: 10.3389/fmars.2017.00095

Yang, B., Wang, Y., and Qian, P.-Y. (2016). Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinform.* 17:135. doi: 10.1186/s12859-016-0992-y

Yeh, Y.-C., McNichol, J., Needham, D. M., Fichot, E. B., and Fuhrman, J. A. (2019). Comprehensive single-PCR 16S and 18S rRNA community analysis validated with mock communities and denoising algorithms. *bioRxiv [Preprint]* doi: 10.1101/866731

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., et al. (2014). The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res.* 42, D643–D648. doi: 10.1093/nar/gkt1209

Zeder, M., Ellrott, A., and Amann, R. (2011). Automated sample area definition for high−throughput microscopy. *Cytom. Part A* 79, 306–310.

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.

frontiers in Microbiology

METHODS
published: 24 May 2021
doi: 10.3389/fmicb.2021.665999

Selective DNA and Protein Isolation From Marine Macrophyte Surfaces

Marino Korlević [1], Marsej Markovski [1], Zihao Zhao [2], Gerhard J. Herndl [2,3,4] and Mirjana Najdek [1]*

[1] Center for Marine Research, Ruđer Bošković Institute, Rovinj, Croatia, [2] Department of Functional and Evolutionary Ecology, University of Vienna, Vienna, Austria, [3] Department of Marine Microbiology and Biogeochemistry, Royal Netherlands Institute for Sea Research (NIOZ), Utrecht University, Den Burg, Netherlands, [4] Vienna Metabolomics Center, University of Vienna, Vienna, Austria

Studies of unculturable microbes often combine methods, such as 16S rRNA sequencing, metagenomics, and metaproteomics. To apply these techniques to the microbial community inhabiting the surfaces of marine macrophytes, it is advisable to perform a selective DNA and protein isolation prior to the analysis to avoid biases due to the host material being present in high quantities. Two protocols for DNA and protein isolation were adapted for selective extractions of DNA and proteins from epiphytic communities inhabiting the surfaces of two marine macrophytes, the seagrass *Cymodocea nodosa* and the macroalga *Caulerpa cylindracea*. Protocols showed an almost complete removal of the epiphytic community regardless of the sampling season, station, settlement, or host species. The obtained DNA was suitable for metagenomic and 16S rRNA sequencing, while isolated proteins could be identified by mass spectrometry. Low presence of host DNA and proteins in the samples indicated a high specificity of the protocols. The procedures are based on universally available laboratory chemicals making the protocols widely applicable. Taken together, the adapted protocols ensure an almost complete removal of the macrophyte epiphytic community. The procedures are selective for microbes inhabiting macrophyte surfaces and provide DNA and proteins applicable in 16S rRNA sequencing, metagenomics, and metaproteomics.

Keywords: selective isolation, DNA, proteins, marine macrophytes, *Cymodocea nodosa*, *Caulerpa cylindracea*

OPEN ACCESS

Edited by:
Catarina Magalhães, University of Porto, Portugal

Reviewed by:
Frank O'Neill Aylward, Virginia Tech, United States
Olga Lage, University of Porto, Portugal

****Correspondence:***
Marino Korlević
marino.korlevic@irb.hr

Specialty section:
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Microbiology

Received: *09 February 2021*
Accepted: *30 March 2021*
Published: *24 May 2021*

Citation:
Korlević M, Markovski M, Zhao Z, Herndl GJ and Najdek M (2021) Selective DNA and Protein Isolation From Marine Macrophyte Surfaces. Front. Microbiol. 12:665999. doi: 10.3389/fmicb.2021.665999

INTRODUCTION

Surfaces of marine macrophytes are colonized by a diverse microbial community whose structure and function are poorly understood (Egan et al., 2013). As <1% of all prokaryotic species are culturable, molecular methods, such as 16S rRNA sequencing, metagenomics, and metaproteomics are indispensable to study these organisms (Amann et al., 1995; Su et al., 2012). Applying these techniques requires an initial isolation step with the purpose of obtaining high-quality DNA and proteins.

Biological material (i.e., proteins and DNA) from pelagic microbial communities is usually isolated by collecting cells onto filters and subsequently isolating the target organisms or communities (Gilbert et al., 2009). If a specific microbial size fraction is aimed, sequential filtration is applied (Massana et al., 1997; Andersson et al., 2010). In contrast, obtaining microorganisms associated to surfaces require either a cell detachment procedure prior to isolation or the host

material is co-extracted with the target material. Methods for separating microbial cells from the host include shaking of host tissue (Gross et al., 2003; Nõges et al., 2010), scraping of macrophyte surfaces (Uku et al., 2007), or applying ultrasonication (Weidner et al., 1996; Cai et al., 2014). It was shown that shaking alone is not sufficient to remove microbial cells from surfaces, at least not from plant root surfaces (Richter-Heitmann et al., 2016). Manual separation methods, such as scraping and brushing are time consuming and subjective, as the detachment efficiency depends on host tissue and the person performing the procedure (Cai et al., 2014). Ultrasonication was proposed as an alternative method as it is providing better results in terms of detachment efficiency (Cai et al., 2014; Richter-Heitmann et al., 2016). The downside of this procedure is that complete cell removal is still not obtained and tissue disruption was observed especially after the application of probe ultrasonication (Richter-Heitmann et al., 2016). An alternative to these cell detachment procedures is the isolation of target epiphytic compounds together with host material (Staufenberger et al., 2008; Jiang et al., 2015). This procedure can lead to problems in the following processing steps, such as mitochondrial and chloroplast 16S rRNA sequence contaminations from the host (Longford et al., 2007; Staufenberger et al., 2008). In addition, when performing metagenomics and metaproteomics host material can cause biased results toward more abundant host DNA and proteins.

An alternative to these procedures is a direct isolation of the target material by incubating macrophyte tissues in an extraction buffer. After the incubation, the undisrupted host tissue is removed followed by the isolation procedure, omitting host material contaminations. To our knowledge, the only procedure describing a direct and selective epiphytic DNA isolation from the surfaces of marine macrophytes was described by Burke et al. (2009). In contrast to previously described methods, this protocol enables an almost complete removal of the surface community. It was used for 16S rRNA gene clone library construction (Burke et al., 2011b) and metagenome sequencing (Burke et al., 2011a). This method, although providing a selective isolation procedure, uses a rapid multi-enzyme cleaner (3M) that is not available worldwide and the chemical constituents are unknown (Burke et al., 2009). Also to our knowledge, no selective isolation protocol to perform (meta)proteomics of epiphytic communities associated with marine macrophytes has been developed yet.

In the present study, we adapted a protocol widely used for DNA isolation from filters (Massana et al., 1997) and a protocol used for protein isolation from soils (Chourey et al., 2010; Hultman et al., 2015). These two adapted methods allowed for a selective extraction of DNA and proteins from epiphytic communities inhabiting the surfaces of two marine macrophytes, the seagrass *Cymodocea nodosa*, and the macroalga *Caulerpa cylindracea*. In addition, we tested the removal efficiency of the protocol and the suitability of obtained DNA and proteins for 16S rRNA sequencing, metagenomics, and metaproteomics.

MATERIALS AND METHODS

Sampling

Leaves of *C. nodosa* were sampled in a *C. nodosa* meadow in the Bay of Saline, northern Adriatic Sea (45°7′5″ N, 13°37′20″ E) and in a *C. nodosa* meadow invaded by *C. cylindracea* in the Bay of Funtana, northern Adriatic Sea (45°10′39″ N, 13°35′42″ E). Thalli of *C. cylindracea* were sampled in the same *C. nodosa* invaded meadow in the Bay of Funtana and at a locality of only *C. cylindracea* located in the proximity of the invaded meadow. Leaves and thalli for 16S rRNA analysis, metagenomics, and metaproteomics were collected in two contrasting seasons, on December 4, 2017 (16S rRNA analysis and metaproteomics), December 14, 2017 (metagenomics), and June 18, 2018 (16S rRNA analysis, metagenomics, and metaproteomics). During spring 2018, the *C. nodosa* meadow in the Bay of Saline decayed to an extent that no leaves could be retrieved (Najdek et al., 2020). In addition, as not enough DNA for both metagenomic and 16S RNA analysis were obtained during the sampling on December 4, 2017, an additional sampling on December 14, 2017 was carried out in the Bay of Funtana. Leaves and thalli were collected by diving and transported to the laboratory in containers placed on ice and filled with seawater from the site. Upon arrival to the laboratory, *C. nodosa* leaves were cut into sections of 1–2 cm, while *C. cylindracea* thalli were cut into 5–8 cm long sections. Leaves and thalli were washed three times with sterile artificial seawater (ASW) to remove loosely attached microbial cells.

DNA Isolation

The DNA was isolated according to the protocol for isolation from filters described in Massana et al. (1997). This protocol was modified and adapted for microbial DNA isolation from macrophyte surfaces as described below. Five milliliter of lysis buffer (40 mM EDTA, 50 mM Tris-HCl, 0.75 M sucrose; pH 8.3) was added to 1 g wet weight of leaves or 2 g wet weight of thalli. For every sample, duplicate isolations were performed. Lysozyme was added (final concentration 1 mg ml^{-1}) and the mixture was incubated at 37°C for 30 min. Subsequently, proteinase K (final concentration 0.5 mg ml^{-1}) and SDS (final concentration 1%) were added and the samples were incubated at 55°C for 2 h. Following the incubation, tubes were vortexed for 10 min and the mixture containing lysed epiphytic cells was separated from host leaves or thalli by transferring the solution into a clean tube. The lysate was extracted twice with a mixture of phenol:chloroform:isoamyl alcohol (25:24:1; pH 8) and once with chloroform:isoamyl alcohol (24:1). After each addition of an organic solvent mixture, tubes were slightly vortexed and centrifuged at 4,500 × g for 10 min. Following each centrifugation, the aqueous phases were retrieved. After the final extraction, 1/10 of chilled 3 M sodium acetate (pH 5.2) was added. DNA was precipitated by adding 1 volume of chilled isopropanol, incubating the mixtures overnight at −20°C and centrifuging at 16,000 × g and 4°C for 20 min. The pellet was washed twice with 1 ml of chilled 70% ethanol and centrifuged after each washing step at 20,000 × g and 4°C for 10 min. After the first washing step, duplicate pellets from the same sample were pooled

and transferred to a clean 1.5 ml tube. The dried pellet was re-suspended in 100 µl of deionized water.

Illumina 16S rRNA Sequencing

An aliquot of isolated DNA was treated with RNase A (final concentration 200 µg ml^{-1}) for 2 h at 37°C. The DNA concentration was determined using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen, USA) according to the manufacturer's instructions and diluted to 1 ng µl^{-1}. The V4 region of the 16S rRNA gene was amplified using a two-step PCR procedure. In the first PCR, the 515F (5′-GTGYCAGCMGCCGCGGTAA-3′) and 806R (5′-GGACTACNVGGGTWTCTAAT-3′) primers from the Earth Microbiome Project (https://earthmicrobiome.org/protocols-and-standards/16s/) were used to amplify the target region (Caporaso et al., 2012; Apprill et al., 2015; Parada et al., 2016). These primers contained on their 5′ end a tagged sequence. Each sample was amplified in four parallel 25 µl reactions, in which each contained 1 × Q5 reaction buffer, 0.2 mM dNTPmix, 0.7 mg ml^{-1} BSA (bovine serum albumin), 0.2 µM forward and reverse primers, 0.5 U of Q5 High-Fidelity DNA Polymerase (New England Biolabs, USA), and 5 ng of DNA template. Cycling conditions were as follows: initial denaturation at 94°C for 3 min, 20 cycles of denaturation at 94°C for 45 s, annealing at 50°C for 60 s and elongation at 72°C for 90 s, finalized by an elongation step at 72°C for 10 min. The four parallel reactions volumes were pooled and PCR products were purified using the GeneJET PCR Purification Kit (Thermo Fisher Scientific, USA) according to the manufacturer's instructions and following the protocol that included isopropanol addition for better small DNA fragment yield. The column was eluted in 30 µl of deionized water. Purified PCR products were sent for Illumina MiSeq sequencing (2 × 250 bp) at IMGM Laboratories, Martinsried, Germany. Before sequencing at IMGM, the second PCR amplification of the two-step PCR procedure was performed using primers targeting the tagged region incorporated in the first PCR. In addition, these primers contained adapter and sample-specific index sequences. The second PCR was carried out for eight cycles. Beside samples, a positive and negative control were sequenced. A negative control was composed of four parallel PCR reactions without DNA template, while for a positive control a mock community composed of evenly mixed DNA material originating from 20 bacterial strains (ATCC MSA-1002, ATCC, USA) was used. Partial 16S rRNA sequences obtained in this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession numbers SAMEA6786270, SAMEA6648792–SAMEA6648794, SAMEA6648809–SAMEA6648811, and SAMEA6648824.

Obtained sequences were analyzed on the computer cluster Isabella (University Computing Center, University of Zagreb) using mothur (version 1.43.0) (Schloss et al., 2009) according to the MiSeq Standard Operating Procedure (MiSeq SOP; https://mothur.org/wiki/MiSeq_SOP) (Kozich et al., 2013) and recommendations given from the Riffomonas project to enhance data reproducibility (http://www.riffomonas.org/). For alignment and classification of sequences, the SILVA SSU Ref NR 99 database (release 138; http://www.arb-silva.de) was used (Quast et al., 2013; Yilmaz et al., 2014). Sequences classified as chloroplasts by SILVA were exported and reclassified using mothur and the RDP (Ribosomal Database Project; http://rdp.cme.msu.edu/) training set (version 16) reference files adapted for mothur (Cole et al., 2014). In comparison to SILVA, RDP allows a more detailed classification of chloroplast sequences. Based on the ATCC MSA-1002 mock community included in the analysis, a sequencing error rate of 0.009% was determined, which is in line with previously reported values for next-generation sequencing data (Kozich et al., 2013; Schloss et al., 2016). In addition, the negative control processed together with the samples yielded only 2 sequences after sequence quality curation.

Metagenomics

Four DNA samples were selected and sent on dry ice to IMGM Laboratories, Martinsried, Germany for metagenomic sequencing. DNA was purified using AMPure XP Beads (Beckman Coulter, USA) applying a bead:DNA ratio of 1:1 (v/v), quantified with a Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, USA) and check for integrity on a 1% agarose gel. Metagenomic sequencing libraries were generated from 300 ng of input DNA using a NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England Biolabs, USA) according to the manufacturer's instructions. Fragments were selected (500–700 bp) using AMPure XP Beads, PCR enriched for 3–5 cycles and quality controlled. Libraries generated from different DNA samples were pooled and sequenced on an Illumina NovaSeq 6000 sequencing system (2 × 250 bp).

Obtained sequences were analyzed on the Life Science Compute Cluster (LiSC) (CUBE–Computational Systems Biology, University of Vienna). Individual sequences were assembled using MEGAHIT (version 1.1.2) (Li et al., 2015) under default settings. Putative genes were predicted from contigs longer than 200 bp using Prodigal (version 2.6.3) (Hyatt et al., 2010) in metagenome mode (-p meta). Abundances of predicted genes were expressed as Reads Per Kilobase Million (RPKM) and calculated using the BWA algorithm (version 0.7.16a) (Li and Durbin, 2009). All predicted genes were functionally annotated using the eggNOG-mapper (Huerta-Cepas et al., 2017) and eggNOG database (version 5.0) (Huerta-Cepas et al., 2019). Sequence taxonomy classification was determined using the lowest common ancestor algorithm adapted from DIAMOND (version 0.8.36) (Buchfink et al., 2015) and by searching against the NCBI non-redundant database (NR). To determine the phylogeny, the top 10% hits with an e-value < 1×10^{-5} were used (--top 10). Sequence renaming, coverage information computing, and metagenomic statistics calculations were performed using software tools from BBTools (https://jgi.doe.gov/data-and-tools/bbtools). Metagenomic sequences obtained in this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession numbers SAMEA6648795, SAMEA6648797, SAMEA6648809, and SAMEA6648811.

Protein Isolation

Proteins were isolated according to the protocol for protein isolation from soil described in Chourey et al. (2010) and modified by Hultman et al. (2015). This protocol was further

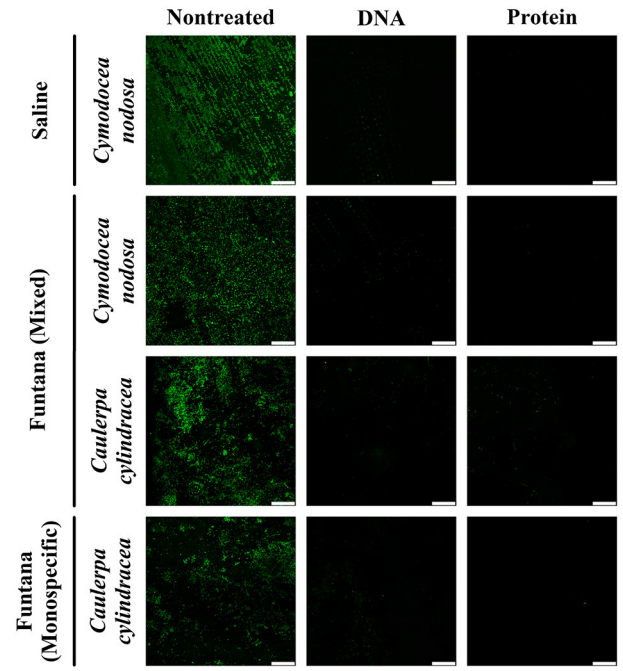

FIGURE 1 | Confocal microscope images of *C. nodosa* and *C. cylindracea* surfaces from the Bay of Saline and the Bay of Funtana (mixed and monospecific settlements) sampled on December 4, 2017 and stained with SYBR Green I. Scale bar at all images is 60 μm.

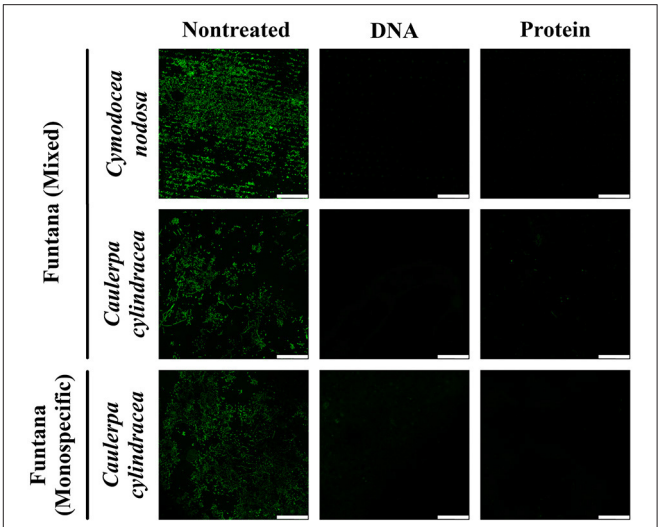

FIGURE 2 | Confocal microscope images of *C. nodosa* and *C. cylindracea* surfaces from the Bay of Funtana (mixed and monospecific settlements) sampled on June 19, 2018 and stained with SYBR Green I. Scale bar at all images is 60 μm.

modified and adapted for microbial protein isolation from macrophyte surfaces as described below. Twenty milliliters of protein extraction buffer (4% SDS, 100 mM Tris-HCl [pH 8.0]) was added to 5 g wet weight of leaves or 10 g wet weight of thalli. The mixture was incubated in boiling water for 5 min, vortexed for 10 min, and incubated again in boiling water for 5 min. After a brief vortex, the lysate was transferred to a clean tube separating the host leaves or thalli from the mixture containing lysed epiphytic cells. Dithiothreitol (DTT; final concentration 24 mM) was added and proteins were precipitated with chilled 100% trichloroacetic acid (TCA; final concentration 20%) overnight at −20°C. Precipitated proteins were centrifuged at 10,000 × g and 4°C for 40 min. The obtained protein pellet was washed three times with chilled acetone. During the first washing step, the pellet was transferred to a clean 1.5-ml tube. After each washing step, samples were centrifuged at 20,000 × g and 4°C for 5 min. Dried pellets were stored at −80°C until further analysis.

Metaproteomics

Isolated proteins were whole trypsin digested using the FASP (filter-aided sample preparation) Protein Digestion Kit (Expedeon, UK) according to the manufacturer's instructions with small modifications (Wiśniewski et al., 2009). Prior to loading the solution onto the column, protein pellets were solubilized in urea sample buffer included in the kit amended with DTT (final concentration 100 mM) for 45 min at room temperature and centrifuged at 20,000 × g for 2–5 min at room temperature to remove larger particles. The first washing step after protein solution loading was repeated twice. In addition, the centrifugation steps were prolonged if the column was clogged. Trypsin digestion was performed on column filters at 37°C overnight for 18 h. The final filtrate containing peptides was acidified with 1% (final concentration) trifluoroacetic acid (TFA), freezed at −80°C, lyophilized, and sent to VIME–Vienna Metabolomics Center (University of Vienna) for metaproteomic analysis. Peptides were re-suspended in 1% (final concentration) TFA, desalted using the Pierce C18 Tips (Thermo Fisher Scientific, USA) according to the manufacturer's instructions, and sequenced on a Q Exactive Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Fisher Scientific). Obtained MS/MS spectra were searched against a protein database composed of combined sequenced metagenomes obtained in this study using SEQUEST-HT engines and validated with Percolator in Proteome Discoverer 2.1 (Thermo Fisher Scientific). The target-decoy approach was used to reduce the probability of false peptide identification. Results whose false discovery rate at the peptide level was <1 % were kept. For protein identification, a minimum of two peptides and one unique peptide were required. For protein quantification, a chromatographic peak area-based free quantitative method was applied.

Data Processing and Visualization

Processing and visualization of 16S rRNA, metagenomic, and metaproteomic data were done using R (version 3.6.0) (R Core Team, 2019), package tidyverse (version 1.3.0) (Wickham et al., 2019), and multiple other packages (Neuwirth, 2014; Xie, 2014, 2015, 2019a,b; Wilke, 2018; Xie et al., 2018; Allaire et al., 2019; Zhu, 2019; Bengtsson, 2020). The detailed analysis procedure including the R Markdown file for this paper are available as a GitHub repository (https://github.com/MicrobesRovinj/Korlevic_SelectiveRemoval_FrontMicrobiol_2021).

Confocal Microscopy

Host leaves and thalli from DNA and protein isolation steps were washed seven times in deionized water and fixed with formaldehyde (final concentration ~3%). In addition, non-treated leaves and thalli, washed three times in ASW to remove loosely attached microbial cells, were fixed in the same concentration of formaldehyde and used as a positive control. For long-term storage, fixed leaves and thalli were immersed in a mixture of phosphate-buffered saline (PBS) and ethanol (1:1) and stored at −20°C. Treated and untreated segments of leaves and thalli were stained in a 2 × solution of SYBR Green I and examined under a Leica TCS SP8 X FLIM confocal microscope (Leica Microsystems, Germany).

RESULTS

To assess the removal efficiency of the DNA and protein isolation procedures, leaves and thalli were examined under a confocal microscope before and after treatments were performed. The modified procedures resulted in an almost complete removal of the surface community of both, *C. nodosa* and *C. cylindracea*. In addition, a similar removal efficiency was observed for communities sampled in contrasting months, December 2017 (**Figure 1**) and June 2018 (**Figure 2**). Also, no effect of station, settlement, or isolation procedure (DNA or protein) on the removal efficiency was observed (**Figures 1**, **2**).

To evaluate whether the obtained DNA is suitable to determine the composition of the microbial community, Illumina sequencing of the V4 region of the 16S rRNA was performed. Sequencing yielded a total of 292,888 sequences after quality curation and exclusion of eukaryotic, mitochondrial, and no relative sequences. The number of sequences classified as chloroplasts was 97,331. After excluding these sequences, the total number of retrieved reads was 195,557, ranging from 13,667 to 41,842 sequences per sample (**Supplementary Table 1**). Even when the highest sequencing effort was applied, the rarefaction curves did not level off which is commonly observed in high-throughput 16S rRNA amplicon sequencing (**Supplementary Figure 1**). Sequences clustering at a similarity level of 97% yielded a total of 8,355 different OTUs. Taxonomic classification of reads revealed a macrophyte-associated epiphytic community mainly composed of *Alphaproteobacteria* (14.9 ± 3.5%), *Bacteroidota* (12.5 ± 2.4%), *Gammaproteobacteria* (11.6 ± 4.3%), *Desulfobacterota* (7.8 ± 7.5%), *Cyanobacteria* (6.5 ± 4.7%), and *Planctomycetota* (2.9 ± 1.7%) (**Figure 3**).

Primers used in this study, as in many other 16S rRNA amplicon procedures, also amplified chloroplast 16S rRNA genes. We observed a high proportion of chloroplast sequences in all analyzed samples (33.4 ± 9.4%) (**Figure 3**). To determine whether chloroplast sequences originate from the host or eukaryotic epiphytic organisms, we exported SILVA-classified chloroplast sequences and reclassified them using the RDP training set that allows for a more detailed chloroplast classification. The largest proportion of sequences was classified as Bacillariophyta (89.7 ± 5.7%) indicating that the DNA removal procedure resulted in only minor co-extracted quantities of host DNA (**Figure 4**). Chloroplast sequences classified as Streptophyta constituted 3.3 ± 2.8% of all chloroplast sequences originating from *C. nodosa* samples, while sequences classified as Chlorophyta comprised only 0.02 ± 0.01% of all chloroplast sequences associated with *C. cylindracea* samples.

To determine whether the extracted DNA can be used for metagenomic sequencing, four samples containing epiphytic DNA were selected and shotgun sequenced using an Illumina platform. Metagenomic sequencing yielded between 207,149,524 and 624,029,930 sequence pairs (**Supplementary Table 2**). Obtained sequences were successfully assembled into contigs whose L50 ranged from 654 to 1,011 bp. In addition, predicted coding sequences were functionally annotated (9,066,667–20,256,215 annotated sequences; **Figure 5A**) and taxonomically classified. Functional annotation allowed for an assessment of the relative contribution of each COG (Clusters of Orthologous Groups) functional category to the total number of annotated coding sequences (**Figure 5A**). Functional categories containing the highest number of sequences were C (energy production and conversion), E (amino acid transport and metabolism), M (cell wall/membrane/envelope biogenesis), L (replication, recombination and repair), and P (inorganic ion transport and metabolism). If host DNA is co-extracted with epiphytes, it should be detected in large proportions in sequenced metagenomes. However, no large proportions of coding sequences classified as Streptophyta and Chlorophyta were detected (**Supplementary Table 3**). Sequenced metagenomic DNA originating from the surface of *C. nodosa* contained 1.3% of coding sequences classified as Streptophyta in December 2017 and 0.7% in June 2018. Furthermore, the summed RPKM (reads per kilobase million) of these sequences constituted 1.7% of total RPKM of all successfully classified sequences in December 2017 and 1.1% in June 2018. Similar low proportions of host coding sequences were detected in metagenomic samples originating from the surfaces of *C. cylindracea*. Of all successfully classified coding sequences, 0.2% sequences were classified as Chlorophyta in December 2017 and 0.1% in June 2018. A relatively higher proportion of RPKM of these sequences than in the case of *C. nodosa* was observed, indicating a higher co-extraction of host DNA in *C. cylindracea*. In December, the proportion of RPKM of sequences classified as Chlorophyta was 8.2%, while in June 2018 it reached 13.6%.

To evaluate whether the procedure for protein extraction is suitable for metaproteomic analysis, obtained proteins were trypsin digested and sequenced using a mass spectrometer. Obtained MS/MS spectra were searched against a protein database from sequenced metagenomes. From 14,219 to 16,449 proteins were identified in isolated protein samples (**Figure 5B**). In addition, successful identification of proteins allowed for an assessment of the relative contribution of each COG functional category to the total number of identified proteins (**Figure 5B**). Functional categories containing the highest number of identified proteins were C (energy production and conversion), G (carbohydrate transport and metabolism), P (inorganic ion

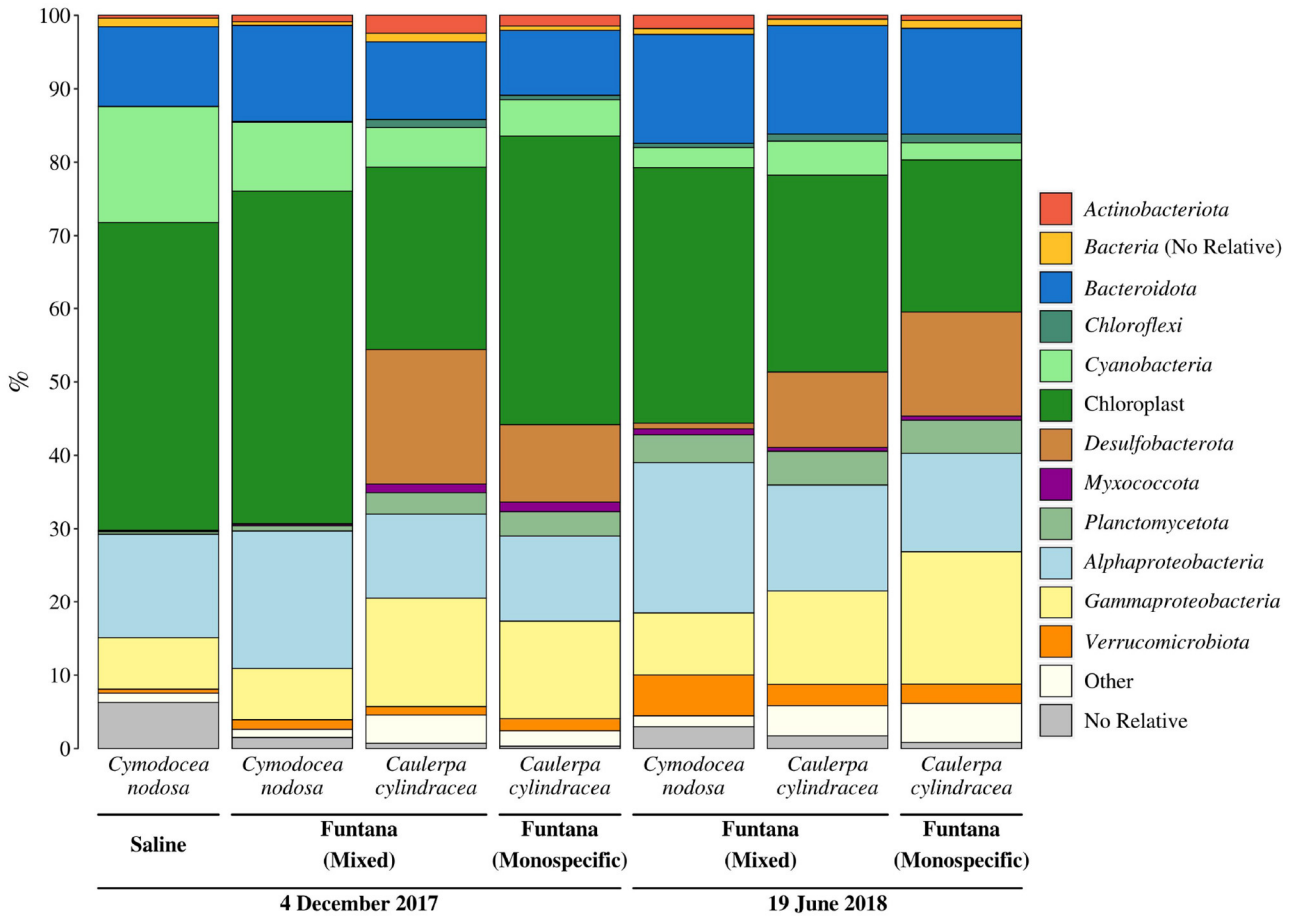

FIGURE 3 | Taxonomic classification and relative contribution of the most abundant (≥1%) bacterial and archaeal sequences from surfaces of two marine macrophytes (*C. nodosa* and *C. cylindracea*) sampled in the Bay of Saline and the Bay of Funtana (mixed and monospecific settlements) and in two contrasting seasons (December 4, 2017 and June 19, 2018).

transport and metabolism), O (post-translational modification, protein turnover, chaperones), and E (amino acid transport and metabolism). Isolated proteins could originate from epiphytic organisms inhabiting the macrophyte surface and/or from macrophyte tissue underlying them. The contribution of proteins originating from host tissues was evaluated by identifying all the proteins predicted to belong to any taxonomic group within the phyla Streptophyta and Chlorophyta and by calculating their contribution to the number and abundance (NAAF [normalized abundance area factor]) of all identified proteins. On average, proteins isolated from the surface of *C. nodosa* contained 1.8 ± 0.06% of proteins associated with Streptophyta, contributing to 2.2 ± 0.8% of total proteins. Similar to metagenomes, proteins associated with Chlorophyta contributed more to *C. cylindracea* than proteins associated with Streptophyta to *C. nodosa*. Chlorophyta-associated proteins composed 5.2 ± 0.06% of all identified proteins in *C. cylindracea*, contributing 19.2 ± 1.5% to the total protein abundance.

DISCUSSION

To test whether the developed DNA and protein isolation protocols efficiently detach microbes from the macrophyte surface, we selected *C. nodosa* and *C. cylindracea* as representatives of seagrass and macroalgal species. These species differ morphologically. While *C. nodosa* leaves are flat, *C. cylindracea* thalli are characterized by an uneven surface (Kuo and den Hartog, 2001; Verlaque et al., 2003). The developed protocol led to an almost complete removal of epiphytic cells from the surfaces of both species comparable to the result of Burke et al. (2009), indicating that structural differences do not impact the removal efficiency. In addition, isolation protocols were tested in two contrasting seasons, as it is known that macrophytes are harboring more algal epiphytes during autumn and winter (Reyes and Sansón, 2001). No differences in the removal efficiency was observed between seasons, suggesting that these protocols can be used on macrophyte samples retrieved throughout the year. Also, no removal differences

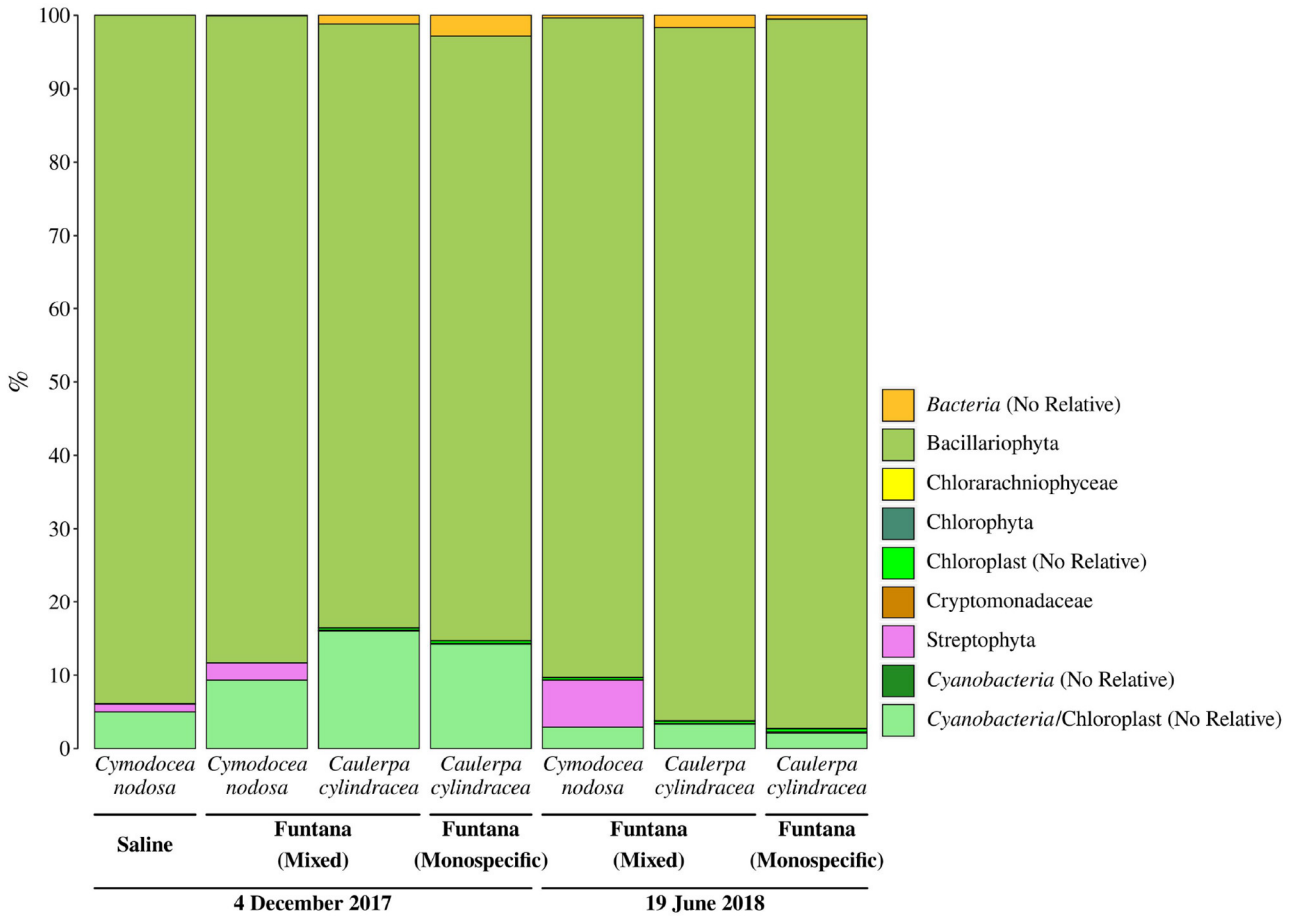

FIGURE 4 | Taxonomic classification and relative contribution of chloroplast sequences from surfaces of two marine macrophytes (*C. nodosa* and *C. cylindracea*) sampled in the Bay of Saline and the Bay of Funtana (mixed and monospecific settlements) and in two contrasting seasons (December 4, 2017 and June 19, 2018).

were observed on samples derived from the same host but from different locations.

Successful amplification and sequencing of the V4 region of the 16S rRNA gene proved that the isolated DNA can be used to estimate the microbial epiphytic diversity. Taxonomic groups detected in this step can also be often found in epiphytic communities associated with other macrophytes (Burke et al., 2011b; Morrissey et al., 2019). A problem often encountered in studies focusing on epiphytic communities is the presence of large proportions of chloroplast 16S rRNA sequences in the pool of amplified molecules, especially if the epiphytic DNA was isolated without prior selection (Staufenberger et al., 2008). These sequences can derive from host chloroplasts or from eukaryotic epiphytic chloroplast DNA. Although the proportion of obtained chloroplast 16S rRNA sequences in our samples was substantial, they derived almost exclusively from eukaryotic epiphytes. High proportion of chloroplast 16S rRNA sequences in studies applying selective procedures that include direct cellular lysis on host surfaces were observed before (Michelou et al., 2013). It is possible that chloroplast-specific sequences even in these studies originated from eukaryotic epiphytic cells and not from host chloroplasts. Indeed, it is common during 16S rRNA profiling of pelagic microbial communities to observe high proportions of chloroplast sequences (Gilbert et al., 2009; Korlević et al., 2016). One of the solutions to further reduce chloroplast 16S rRNA sequence contamination is to use primers that minimize the amplification of these reads if the sequencing and study design allow it (Hanshew et al., 2013). In addition, a very low proportion of chloroplast 16S rRNA sequences in samples originating from *C. cylindracea* in comparison to *C. nodosa* could be explained by the presence of three introns in the gene for 16S rRNA in some members of the genus *Caulerpa* that could hamper the amplification process (Lam and Lopez-Bautista, 2016).

High-quality DNA is also needed for metagenomics. The obtained number of metagenomic sequences and assembly statistics were comparable to metagenomes and metatranscriptomes derived from similar surface-associated communities (Crump et al., 2018; Cúcio et al., 2018). In addition, functional annotation of predicted coding sequences to COG functional categories showed that the obtained metagenomes can be used to determine the metabolic capacity

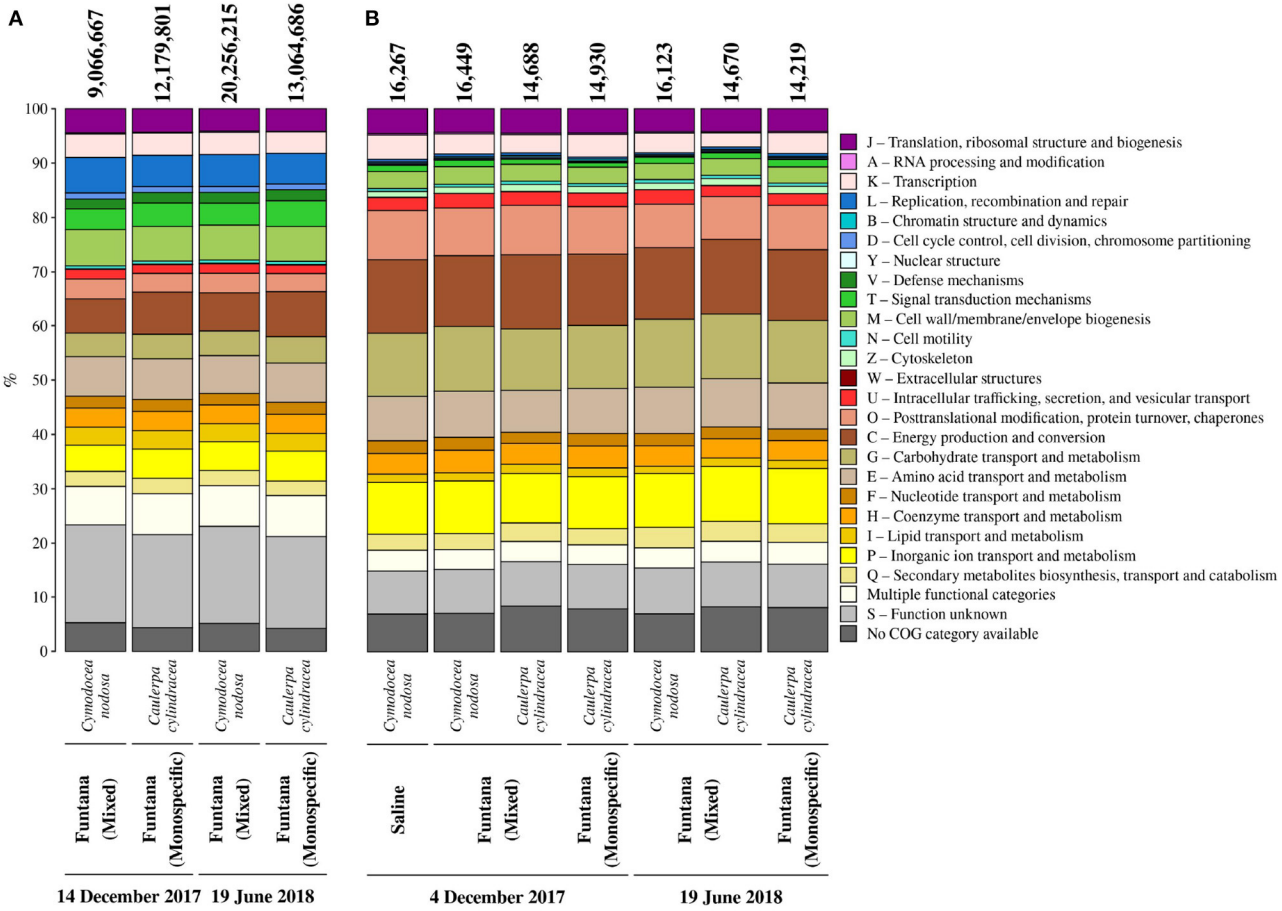

FIGURE 5 | Relative contribution of each COG (Clusters of Orthologous Groups) category to the total number of annotated coding sequences **(A)** from metagenomes and identified proteins **(B)** from metaproteomes associated with surfaces of two marine macrophytes (*C. nodosa* and *C. cylindracea*) sampled in the Bay of Saline and the Bay of Funtana (mixed and monospecific settlements) and in two contrasting seasons (December 4/14, 2017 and June 19, 2018). Total number of annotated coding sequences and identified proteins is given above the corresponding bar.

of surface-associated communities (Leary et al., 2014; Cúcio et al., 2018). The proportion of coding sequences, including their RPKM, originating from *C. nodosa* metagenomes and classified as Streptophyta was low indicating that the isolation procedure was specific for epiphytic cells. DNA samples isolated from the surface of *C. cylindracea* exhibited a low proportion of Chlorophyta coding sequences; however, their RPKM was higher than in the samples originating from *C. nodosa*. One of the reasons for this elevated RPKM of Chlorophyta sequences in *C. cylindracea* could be the differences in the tissue structure between these two host species. While *C. nodosa* leaves are composed of individual cells, the thallus of *C. cylindracea* is, like in other siphonous algal species, composed of a single large multinucleate cell (Coneva and Chitwood, 2015). The absence of individual cells in *C. cylindracea* could cause a leakage of genetic material into the extraction buffer causing an elevated presence of host sequences in the samples for metagenome analyses.

To obtain insight into the metabolic status of uncultivated prokaryotes, a metaproteomic approach is required (Saito et al., 2019). The applied protocol for epiphytic protein isolation followed by a metaproteomic analysis identified between 14,219 and 16,449 proteins, which is higher than previously reported for soils (Chourey et al., 2010; Hultman et al., 2015), seawater (Williams et al., 2012), and biofilms (Leary et al., 2014). The functional annotation of identified proteins into COG functional categories showed that the protein isolation protocol can be used to assess the metabolic status of the epiphytic community (Leary et al., 2014). Similar to the results of the metagenomic analysis, the number and abundance of identified proteins affiliated to Streptophyta in *C. nodosa* samples were low, indicating that the procedure is selective for epiphytic cell proteins. In addition, a higher number and abundance of identified proteins associated with Chlorophyta were observed in *C. cylindracea* samples. The cause of this elevated presence of Chlorophyta-associated proteins can be, similar to the DNA isolation protocol, explained by the absence of individual cells in this siphonous alga (Coneva and Chitwood, 2015).

In conclusion, the developed protocols for DNA and protein isolation from macrophyte surfaces almost completely remove the epiphytic community from both, *C. nodosa* and *C. cylindracea*, in different seasons. Also, the obtained DNA and proteins are suitable for 16S rRNA sequencing, metagenomics and metaproteomics analyses while the obtained material contains low quantities of host DNA and proteins making the protocols specific for epiphytes. Furthermore, the protocols are based on universally available laboratory chemicals hence, making them widely applicable.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repositories and accession numbers can be found in the article.

AUTHOR CONTRIBUTIONS

MK designed the work with the intellectual contribution from GH and MN, and prepared the manuscript with editorial help from MM, ZZ, GH, and MN. MK, MM, ZZ, and MN performed the sampling and laboratory analysis. MK and ZZ analyzed the data. All authors contributed to the article and approved the final submitted version.

FUNDING

This work was funded by the Croatian Science Foundation through the MICRO-SEAGRASS project (project number IP-2016-06-7118). ZZ and GH were supported by the Austrian Science Fund (FWF) project ARTEMIS (project number P28781-B21).

ACKNOWLEDGMENTS

We would like to thank Margareta Buterer for technical support, Paolo Paliaga for help during sampling, Lucija Horvat for technical support in confocal microscopy, Dušica Vujaklija for help in accessing the confocal microscope, and Sonja Tischler for technical support in mass spectrometry. In addition, we would like to thank the University Computing Center of the University of Zagreb for access to the computer cluster Isabella and CUBE–Computational Systems Biology of the University of Vienna for access to the Life Science Compute Cluster (LiSC).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.665999/full#supplementary-material

REFERENCES

Allaire, J. J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins,A., et al. (2019). *rmarkdown: Dynamic documents for R*.

Amann, R. I., Ludwig, W., and Schleifer, K. H. (1995). Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* 59, 143–169.

Andersson, A. F., Riemann, L., and Bertilsson, S. (2010). Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. *ISME J.* 4, 171–181. doi: 10.1038/ismej.2009.108

Apprill, A., McNally, S., Parsons, R., and Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat. Microb. Ecol.* 75, 129–137. doi: 10.3354/ame01753

Bengtsson, H. (2020). *matrixStats: Functions That Apply to Rows and Columns of Matrices (and to Vectors)*.

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176

Burke, C., Kjelleberg, S., and Thomas, T. (2009). Selective extraction of bacterial DNA from the surfaces of macroalgae. *Appl. Environ. Microbiol.* 75, 252–256. doi: 10.1128/AEM.01630-08

Burke, C., Steinberg, P., Rusch, D., Kjelleberg, S., and Thomas, T. (2011a). Bacterial community assembly based on functional genes rather than species. *Proc. Natl. Acad. Sci. U.S.A.* 108, 14288–14293. doi: 10.1073/pnas.1101591108

Burke, C., Thomas, T., Lewis, M., Steinberg, P., and Kjelleberg, S. (2011b). Composition, uniqueness and variability of the epiphytic bacterial community of the green alga *Ulva australis*. *ISME J.* 5, 590–600. doi: 10.1038/ismej.2010.164

Cai, X., Gao, G., Yang, J., Tang, X., Dai, J., Chen, D., et al. (2014). An ultrasonic method for separation of epiphytic microbes from freshwater submerged macrophytes. *J. Basic Microbiol.* 54, 758–761. doi: 10.1002/jobm.201300041

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624. doi: 10.1038/ismej.2012.8

Chourey, K., Jansson, J., VerBerkmoes, N., Shah, M., Chavarria, K. L., Tom, L. M., et al. (2010). Direct cellular lysis/protein extraction protocol for soil metaproteomics. *J. Proteome Res.* 9, 6615–6622. doi: 10.1021/pr100787q

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., et al. (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* 42, D633–D642. doi: 10.1093/nar/gkt1244

Coneva, V., and Chitwood, D. H. (2015). Plant architecture without multicellularity: Quandaries over patterning and the soma-germline divide in siphonous algae. *Front. Plant Sci.* 6:287. doi: 10.3389/fpls.2015.00287

Crump, B. C., Wojahn, J. M., Tomas, F., and Mueller, R. S. (2018). Metatranscriptomics and amplicon sequencing reveal mutualisms in seagrass microbiomes. *Front. Microbiol.* 9:388. doi: 10.3389/fmicb.2018.00388

Cúcio, C., Overmars, L., Engelen, A. H., and Muyzer, G. (2018). Metagenomic analysis shows the presence of bacteria related to free-living forms of sulfur-oxidizing chemolithoautotrophic symbionts in the rhizosphere of the seagrass *Zostera marina*. *Front. Mar. Sci.* 5:171. doi: 10.3389/fmars.2018.00171

Egan, S., Harder, T., Burke, C., Steinberg, P., Kjelleberg, S., and Thomas, T. (2013). The seaweed holobiont: Understanding seaweed-bacteria interactions. *FEMS Microbiol. Rev.* 37, 462–476. doi: 10.1111/1574-6976.12011

Gilbert, J. A., Field, D., Swift, P., Newbold, L., Oliver, A., Smyth, T., et al. (2009). The seasonal structure of microbial communities in the Western English Channel. *Environ. Microbiol.* 11, 3132–3139. doi: 10.1111/j.1462-2920.2009.02017.x

Gross, E. M., Feldbaum, C., and Graf, A. (2003). Epiphyte biomass and elemental composition on submersed macrophytes in shallow eutrophic lakes. *Hydrobiologia* 506-509, 559–565. doi: 10.1023/B:HYDR.0000008538.68268.82

Hanshew, A. S., Mason, C. J., Raffa, K. F., and Currie, C. R. (2013). Minimization of chloroplast contamination in 16S rRNA gene pyrosequencing of insect herbivore bacterial communities. *J. Microbiol. Methods* 95, 149–155. doi: 10.1016/j.mimet.2013.08.007

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., et al. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* 34, 2115–2122. doi: 10.1093/molbev/msx148

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47,D309–D314. doi: 10.1093/nar/gky1085

Hultman, J., Waldrop, M. P., Mackelprang, R., David, M. M., McFarland, J., Blazewicz, S. J., et al. (2015). Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* 521, 208–212. doi: 10.1038/nature14238

Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119

Jiang, Y. F., Ling, J., Dong, J. D., Chen, B., Zhang, Y. Y., Zhang, Y. Z., et al. (2015). Illumina-based analysis the microbial diversity associated with *Thalassia hemprichii* in Xincun Bay, South China Sea. *Ecotoxicology* 24, 1548–1556. doi: 10.1007/s10646-015-1511-z

Korlević, M., Šupraha, L., Ljubešić, Z., Henderiks, J., Ciglenečki, I., Dautović, J., et al. (2016). Bacterial diversity across a highly stratified ecosystem: a salt-wedge Mediterranean estuary. *Syst. Appl. Microbiol.* 39, 398–408. doi: 10.1016/j.syapm.2016.06.006

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., and Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* 79, 5112–5120. doi: 10.1128/AEM.01043-13

Kuo, J., and den Hartog, C. (2001). "Seagrass taxonomy and identification key," in *Global Seagrass Research Methods*, eds F. T. Short and R. G. Coles (Amsterdam: Elsevier Science B.V.), 31–58.

Lam, D. W., and Lopez-Bautista, J. M. (2016). Complete chloroplast genome for *Caulerpa racemosa* (Bryopsidales, Chlorophyta) and comparative analyses of siphonous green seaweed plastomes. *Cymbella* 2, 23–32.

Leary, D. H., Li, R. W., Hamdan, L. J., Judson Hervey, W. IV, Lebedev, N., Wang, Z., et al. (2014). Integrated metagenomic and metaproteomic analyses of marine biofilm communities. *Biofouling* 30, 1211–1223. doi: 10.1080/08927014.2014.977267

Li, D., Liu, C. M., Luo, R., Sadakane, K., and Lam, T. W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. doi: 10.1093/bioinformatics/btv033

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Longford, S., Tujula, N., Crocetti, G., Holmes, A., Holmström, C., Kjelleberg, S., et al. (2007). Comparisons of diversity of bacterial communities associated with three sessile marine eukaryotes. *Aquat. Microb. Ecol.* 48, 217–229. doi: 10.3354/ame048217

Massana, R., Murray, A. E., Preston, C. M., and DeLong, E. F. (1997). Vertical distribution and phylogenetic characterization of marine planktonic *Archaea* in the Santa Barbara Channel. *Appl. Environ. Microbiol.* 63, 50–56.

Michelou, V. K., Caporaso, J. G., Knight, R., and Palumbi, S. R. (2013). The ecology of microbial communities associated with *Macrocystis pyrifera*. *PLoS ONE* 8:e67480. doi: 10.1371/journal.pone.0067480

Morrissey, K. L., Çavas, L., Willems, A., and De Clerck, O. (2019). Disentangling the influence of environment, host specificity and thallus differentiation on bacterial communities in siphonous green seaweeds. *Front. Microbiol.* 10:717. doi: 10.3389/fmicb.2019.00717

Najdek, M., Korlević, M., Paliaga, P., Markovski, M., Ivančić, I., Iveša, L., et al. (2020). Dynamics of environmental conditions during the decline of a *Cymodocea nodosa* meadow. *Biogeosciences* 17, 3299–3315. doi: 10.5194/bg-17-3299-2020

Neuwirth, E. (2014). *RColorBrewer: ColorBrewer Palettes*.

Nõges, T., Luup, H., and Feldmann, T. (2010). Primary production of aquatic macrophytes and their epiphytes in two shallow lakes (Peipsi and Võrtsjärv) in Estonia. *Aquat. Ecol.* 44, 83–92. doi: 10.1007/s10452-009-9249-4

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

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Reyes, J., and Sansón, M. (2001). Biomass and production of the epiphytes on the leaves of *Cymodocea nodosa* in the Canary Islands. *Bot. Mar.* 44, 307–313. doi: 10.1515/BOT.2001.039

Richter-Heitmann, T., Eickhorst, T., Knauth, S., Friedrich, M. W., and Schmidt, H. (2016). Evaluation of strategies to separate root-associated microbial communities: a crucial choice in rhizobiome research. *Front. Microbiol.* 7:773. doi: 10.3389/fmicb.2016.00773

Saito, M. A., Bertrand, E. M., Duffy, M. E., Gaylord, D. A., Held, N. A., Hervey, W. J., et al. (2019). Progress and challenges in ocean metaproteomics and proposed best practices for data sharing. *J. Proteome Res.* 18, 1461–1476. doi: 10.1021/acs.jproteome.8b00761

Schloss, P. D., Jenior, M. L., Koumpouras, C. C., Westcott, S. L., and Highlander, S. K. (2016). Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system. *PeerJ* 4:e1869. doi: 10.7717/peerj.1869

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09

Staufenberger, T., Thiel, V., Wiese, J., and Imhoff, J. F. (2008). Phylogenetic analysis of bacteria associated with *Laminaria saccharina*. *FEMS Microbiol. Ecol.* 64, 65–77. doi: 10.1111/j.1574-6941.2008.00445.x

Su, C., Lei, L., Duan, Y., Zhang, K.-Q., and Yang, J. (2012). Culture-independent methods for studying environmental microorganisms: Methods, application, and perspective. *Appl. Microbiol. Biotechnol.* 93, 993–1003. doi: 10.1007/s00253-011-3800-7

Uku, J., Björk, M., Bergman, B., and Díez, B. (2007). Characterization and comparison of prokaryotic epiphytes associated with three East African seagrasses. *J. Phycol.* 43, 768–779. doi: 10.1111/j.1529-8817.2007.00371.x

Verlaque, M., Durand, C., Huisman, J. M., Boudouresque, C. F., and Le Parco, Y. (2003). On the identity and origin of the Mediterranean invasive *Caulerpa racemosa* (Caulerpales, Chlorophyta). *Eur. J. Phycol.* 38, 325–339. doi: 10.1080/0967026031000161259 2

Weidner, S., Arnold, W., and Puhler, A. (1996). Diversity of uncultured microorganisms associated with the seagrass *Halophila stipulacea*estimated by restriction fragment length polymorphism analysis of PCR-amplified 16S rRNA genes. *Appl. Environ. Microbiol.* 62, 766–771.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., et al. (2019). Welcome to the tidyverse. *J. Open Source Softw.* 4:1686. doi: 10.21105/joss.01686

Wilke, C. O. (2018). *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*.

Williams, T. J., Long, E., Evans, F., DeMaere, M. Z., Lauro, F. M., Raftery, M. J., et al. (2012). A metaproteomic assessment of winter and summer bacterioplankton from Antarctic Peninsula coastal surface waters. *ISME J.* 6, 1883–1900. doi: 10.1038/ismej.2012.28

Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359–362. doi: 10.1038/nmeth.1322

Xie, Y. (2014). "knitr: A comprehensive tool for reproducible research in R," in *Implementing Reproducible Computational Research*, eds V. Stodden, F. Leisch, and R. D. Peng (New York, NY: Chapman and Hall; CRC), 3–32.

Xie, Y. (2015). *Dynamic Documents With R and knitr, 2nd Edn.* Boca Raton, FL: Chapman and Hall; CRC.

Xie, Y. (2019a). TinyTeX: a lightweight, cross-platform, and easy-to-maintain LaTeX distribution based on TeX Live. *TUGboat* 40, 30–32.

Xie, Y. (2019b). *tinytex: Helper Functions to Install and Maintain 'TeX Live', and Compile 'LaTeX' Documents*.

Xie, Y., Allaire, J. J., and Grolemund, G. (2018). *R Markdown: The Definitive Guide, 1st Edn.* Boca Raton, FL: Chapman and Hall; CRC.

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., et al. (2014). The SILVA and "All-Species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res.* 42, D643–D648. doi: 10.1093/nar/gkt1209

Zhu, H. (2019). *kableExtra: Construct Complex Table With 'kable' and Pipe Syntax*.

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.

frontiers in Marine Science

PERSPECTIVE
published: 11 October 2021
doi: 10.3389/fmars.2021.758694

Toward 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] and 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

OPEN ACCESS

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

Specialty section:
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Marine Science

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

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

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.

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

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").

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 1A**) 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

[1]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.

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

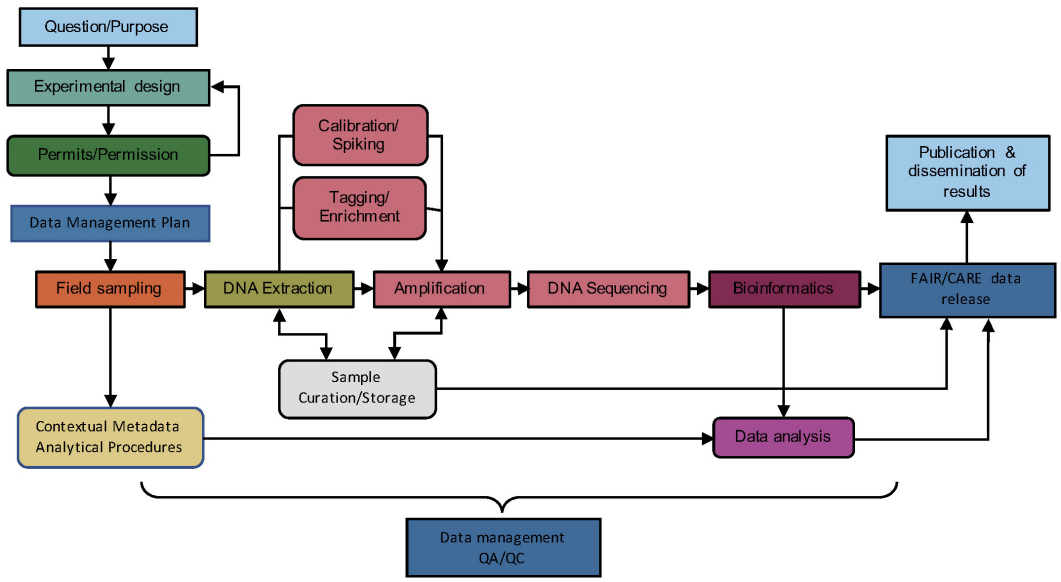

FIGURE 1 | (A) Proposed methodology categories to enhance exchange of ocean omics analysis knowhow. Protocols, guidelines, and other methodologies in some of these categories (such as Sample archiving/biobanking, Data Management, and Society) are cross-cutting and may apply at multiple points in the 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.

(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.

In **Figure 1B**, we give an example of a DNA metabarcoding workflow, where the colour of each step corresponds to a methodology category in **Figure 1A**. 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 1**).

TABLE 1 | 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.

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 1A**). 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.

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.

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.

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.

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

[2]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.

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.

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.

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.

AUTHOR'S NOTE

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

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.

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.

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.

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.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2021.758694/full#supplementary-material

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

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 25 November 2021
doi: 10.3389/fmicb.2021.740932

Inconsistent Patterns of Microbial Diversity and Composition Between Highly Similar Sequencing Protocols: A Case Study With Reef-Building Corals

Hannah E. Epstein[1], Alejandra Hernandez-Agreda[2], Samuel Starko[3], Julia K. Baum[3] and Rebecca Vega Thurber[1]*

[1] Department of Microbiology, Oregon State University, Corvallis, OR, United States, [2] California Academy of Sciences, San Francisco, CA, United States, [3] Department of Biology, University of Victoria, Victoria, BC, Canada

OPEN ACCESS

Edited by:
Catarina Magalhães, University of Porto, Portugal

Reviewed by:
Peng Xing, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences (CAS), China
Rodrigo Costa, University of Lisbon, Portugal
Francisco Pascoal, University of Porto Matosinhos, Portugal, in collaboration with reviewer RC

***Correspondence:**
Hannah E. Epstein
epsteinh@oregonstate.edu

Specialty section:
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Microbiology

Received: *13 July 2021*
Accepted: *01 November 2021*
Published: *25 November 2021*

Citation:
Epstein HE, Hernandez-Agreda A, Starko S, Baum JK and Vega Thurber R (2021) Inconsistent Patterns of Microbial Diversity and Composition Between Highly Similar Sequencing Protocols: A Case Study With Reef-Building Corals. Front. Microbiol. 12:740932. doi: 10.3389/fmicb.2021.740932

16S rRNA gene profiling (amplicon sequencing) is a popular technique for understanding host-associated and environmental microbial communities. Most protocols for sequencing amplicon libraries follow a standardized pipeline that can differ slightly depending on laboratory facility and user. Given that the same variable region of the 16S gene is targeted, it is generally accepted that sequencing output from differing protocols are comparable and this assumption underlies our ability to identify universal patterns in microbial dynamics through meta-analyses. However, discrepant results from a combined 16S rRNA gene dataset prepared by two labs whose protocols differed only in DNA polymerase and sequencing platform led us to scrutinize the outputs and challenge the idea of confidently combining them for standard microbiome analysis. Using technical replicates of reef-building coral samples from two species, *Montipora aequituberculata* and *Porites lobata*, we evaluated the consistency of alpha and beta diversity metrics between data resulting from these highly similar protocols. While we found minimal variation in alpha diversity between platform, significant differences were revealed with most beta diversity metrics, dependent on host species. These inconsistencies persisted following removal of low abundance taxa and when comparing across higher taxonomic levels, suggesting that bacterial community differences associated with sequencing protocol are likely to be context dependent and difficult to correct without extensive validation work. The results of this study encourage caution in the statistical comparison and interpretation of studies that combine rRNA gene sequence data from distinct protocols and point to a need for further work identifying mechanistic causes of these observed differences.

Keywords: amplicon sequencing, 16S rRNA, protocol comparison, coral microbiome, microbial diversity

INTRODUCTION

Microbial ecology has benefited tremendously from recent technological advances in areas such as high throughput sequencing (Schuster, 2008). The generation of large volumes of genomic data (e.g., 16S rRNA gene sequencing data) has encouraged large-scale collaborative efforts, including the Human Microbiome Project [HMP[1]; (Human Microbiome Project Consortium, 2012)], the Earth Microbiome Project [EMP[2]; (Thompson et al., 2017)] and TARA Oceans (Sunagawa et al., 2015), which aim to catalog all microbial life associated with humans, other animal hosts and across ecosystems. Publicly available sequencing data resulting from these initiatives provide opportunities for the production of meta-analyses and for researchers with smaller scale projects to make comparisons and/or combine their dataset with a much broader set of samples, allowing increased impact of their finer sequencing efforts.

Large-scale collaborations also provide standardized protocols for replication and adequate comparison. For example, the EMP standardized protocols for 16S rRNA gene sequencing are optimized for repeatedly processing large numbers of samples and benefit from automation and high throughput sequencing on an Illumina HiSeq platform. Smaller, individual research laboratories are, in many cases, processing fewer samples less frequently, likely without access to automation, but with the capacity to shift reagents and polymerase chain reaction (PCR) conditions to achieve optimized results. Smaller numbers of samples are also more often sequenced on the Illumina MiSeq platform due to cost effectiveness and increased read length. It has been previously accepted that HiSeq and MiSeq platforms produce comparable results (see Caporaso et al., 2012). In fact, there are few differences in the two sequencing platforms: apart from the discrepancy in read length (HiSeq: 150bp; MiSeq: up to 300bp) and sequencing depth (HiSeq: 150M reads/lane; MiSeq: 20–25M reads/lane), the chemistry between the two methods is almost identical, except for the slightly different concentrations of sodium hydroxide (NaOH) used to denature the libraries for sequencing [HiSeq: 0.1N NaOH; MiSeq: 0.2N NaOH; outlined in (Wu et al., 2018)]. As a result, meta-analyses of 16S rRNA gene data across microbial study systems already utilize cross-protocol and platform data that are stored in public repositories (see Duvallet et al., 2017; Pammi et al., 2017; Mo et al., 2020).

However, when attempting to combine 16S rRNA gene data for a large, longitudinal coral microbiome dataset, we found that the data derived from our in-house preparation and MiSeq sequencing runs clustered separately from those prepared and sequenced by EMP, despite following a highly similar preparation protocol. This led us to re-evaluate if the two protocols utilizing different sequencing platforms provide comparable results. Using 24 coral samples that were sequenced in parallel both in-house (MiSeq) and by EMP (HiSeq), we examined if methodological biases lie within these complex microbial communities, and how (or whether) results obtained from the two protocols are comparable when running standard microbial ecology analyses on alpha diversity, beta diversity, dispersion and differential abundances. Large collaborative sequencing efforts and public sharing of these data are central to understanding general, cosmopolitan patterns in the coral microbiome, which makes effective comparison of sequencing data originating from multiple laboratories vital.

[1]https://www.hmpdacc.org/hmp

[2]https://earthmicrobiome.org/

MATERIALS AND METHODS

Sample Collection, DNA Extraction, Library Preparation and Sequencing

Coral samples were originally collected from Kiritimati (Christmas) Island in May 2015 from two species: *Porites lobata* (n = 13) and *Montipora aequituberculata* (n = 11). Frozen tissue for each individual sample was split in two: one portion was sent directly to EMP (University of California, San Diego) for DNA extraction, PCR, library preparation and sequencing on an Illumina HiSeq 2 × 150bp run (Ul-Hasan et al., 2019) and the other processed in-house at Oregon State University (see previously published methods in McDevitt-Irwin et al., 2019) using a highly similar protocol as EMP but sequenced on an Illumina MiSeq 2 × 300bp run. Both protocols targeted the V4 region of the 16S rRNA gene with the following primers: 515F (Parada et al., 2016) 5′–TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGYCAGCMGCCGCGGTAA–3′ and 806R (Apprill et al., 2015) 5′–GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGACTACNVGGGTWTCTAAT– 3′, with the Illumina adapter overhangs underlined. The only difference between the two protocols was the Taq used for PCR: EMP used Platinum Hot Start PCR MasterMix (Thermofisher) and in-house used Accustart™ II PCR ToughMix (QuantaBio). Hereafter, the two protocols will be referred to by their most significant difference: the "HiSeq protocol" run by EMP and the "MiSeq protocol" run in-house.

Bioinformatics

Sequences from both HiSeq and MiSeq protocol outputs were processed using the QIIME2 pipeline to undergo trimming, quality control, identification of amplicon sequence variants (ASVs), and taxonomic assignment. To ensure comparability between the two protocols and accuracy in ASV-picking, we chose to follow similar treatment of sequencing data by EMP (Thompson et al., 2017); only forward reads were used and trimmed to 120bp. Primers were removed using the plug-in cutadapt (Martin, 2011), and denoising and ASV picking was performed using the DADA2 plug-in (Callahan et al., 2016) on sequences from HiSeq and MiSeq protocols separately, after which were combined into a single dataset for downstream processing. For comparison, ASVs were simultaneously clustered using the plug-in vsearch (Rognes et al., 2016) by 97% similarity, resulting in two output biom tables: one for ASVs and one for 97% clustered operational taxonomic units (hereafter, "OTUs"). Taxonomic assignment for both tables was performed using a naïve Bayes classifier with the SILVA v. 132 database (Quast et al., 2013), trained on each set of representative sequences from the two pipelines.

Data Import Into R

All statistical analyses were performed in R v. 4.0.2 (R Core Team, 2020); graphics were conducted in R using the package ggplot2 (Wickham, 2016). QIIME feature tables, taxonomic assignments, and tree files for the ASV and OTU datasets were imported into phyloseq (McMurdie and Holmes, 2013) via qiime2R (Bisanz, 2020) for downstream analyses. The SILVA annotations characterized some reads as Phylum: Alphaproteobacteria, Family: Mitochondria. This annotated family contained a mix of bacterial and mitochondrial (eukaryotic) reads: thus eukaryotic mitochondria were further identified using BLASTn (Altschul et al., 1990) and subsequently removed from the two datasets. In the absence of blank controls from the EMP dataset, contaminants were identified using the 4 blank control samples from the MiSeq Data (McDevitt-Irwin et al., 2019). Contaminants were identified and removed (n = 102) by prevalence using the decontam package (Callahan, 2020) with a threshold value of 0.5 to ensure all sequences that were more prevalent in negative controls than positive samples were removed. Samples with less than 1000 and 998 reads, respectively, were removed from all analyses for ASV and OTU data. These two numbers differ slightly due to differences in contaminant and mitochondrial read removals as a result of ASV identification versus OTU picking.

Diversity Metrics and Differential Abundance

We conducted all diversity tests on the two coral species separately due to well-established differences in both alpha and beta diversity measures across host species (Hernandez-Agreda et al., 2017; Epstein et al., 2019; Ziegler et al., 2019) that could have obscured significant differences between protocols. Three alpha diversity metrics were calculated to account for richness, evenness and phylogenetic diversity: observed species richness, Shannon diversity index and Faith's Phylogenetic Diversity (PD) were calculated on rarefied data (1000 and 998 reads/sample for ASV and OTU data, respectively; these depths were chosen for each dataset to maintain comparability using the highest sample sizes without severely compromising rarefied alpha diversity). For each host species and data type (ASV and OTU), the three alpha diversity metrics were checked for normality using standardized residual plots, Q-Q plots and Shapiro-Wilk tests. If required, log and square-root transformations were performed to meet normality assumptions when data were non-normal (see **Supplementary Table 2**). Differences in alpha diversity indices between protocols were tested using paired t-tests. We also quantified four metrics of beta diversity to examine differences in microbial communities accounting for microbial abundance (Bray-Curtis), presence/absence (binary Jaccard), phylogeny coupled with abundance (weighted UniFrac) and phylogeny coupled with presence/absence (unweighted UniFrac). For each host species, we constructed Bray-Curtis and weighted Unifrac dissimilarity matrices using the relative abundances of taxa to account for differences in sequencing depth between data derived from HiSeq and MiSeq protocols and constructed binary Jaccard and unweighted Unifrac dissimilarity using unrarefied counts. Dissimilarity matrices for all metrics were also built with unrarefied data after removing rare taxa (abundance below 0.5% and 1% threshold per sample). Differences in beta diversity [i.e., both multivariate location ("turnover" and variation)] were tested using permutational analyses of variance (PERMANOVAs) with 999 permutations blocked by coral colony ID (strata = sample label) using the adonis function from the package vegan (Oksanen et al., 2019) implemented in phyloseq (McMurdie and Holmes, 2013). Homogeneity of variances were further tested between protocols using betadisper (PERMDISP) with 999 permutations and communities were visualized using non-metric multidimensional scaling (NMDS) plots. To identify specific significant differences in taxon abundance in the two protocols, differential abundance analyses were performed using DESeq2 (Love et al., 2020) on unrarefied count data with an alpha cut-off of 0.05. All analyses were performed on both ASV and OTU datasets unless otherwise specified. To assess any differences in secondary structure between specific ASVs differentially abundant in HiSeq vs. MiSeq protocols resulting from library denaturation, GC content (%) and melting temperatures were verified through the TmCalculator package in R (Li, 2020). Differences in mean GC content and melting temperature were tested among ASVs present in MiSeq, HiSeq and both protocols using analyses of variance (ANOVAs). We chose to present results from unrarefied data unless otherwise specified in the above methods; however, all analyses were run on both rarefied and unrarefied data and showed no major differences in significance (see **Supplementary Tables 3**, **5–7**).

RESULTS

Sequencing Results

To test whether coral microbiome sequence data generated from the two protocols were comparable, we analyzed paired sequence libraries, combined for comparative downstream analyses, using four standard variables for assessing microbiome variations at both the ASV and OTU levels: alpha diversity, beta diversity, beta dispersion, and differential abundance measures. The final ASV dataset included all 24 samples (total n = 48 to account for 2 technical replicates per sample, one from each protocol) with a combined total of 1,444,493 reads consisting of 5,512 distinct ASVs for analysis. In the OTU dataset, two *P. lobata* HiSeq protocol samples contained less than 998 reads and were removed along with their MiSeq protocol counterparts leaving 22 samples (total n = 44 to account for 2 technical replicates per sample, one from each protocol) for comparison consisting of 953,396 reads and 2,174 OTUs. All comparisons were done using identical read values for both protocols (for sequence read count variation by host species and protocol, see **Supplementary Table 1**).

Sensitivity of Alpha Diversity to Protocol

When using ASVs for analysis, there was a slight tendency for alpha diversity to be lower when calculated from MiSeq protocol data, but diversity did not differ significantly between protocols for any of the four alpha diversity metrics measured ($p > 0.05$: **Supplementary Table 2**), in either species (**Figure 1**).

However, when using OTUs there was a host species-specific effect on some measures of diversity. Specifically, alpha diversity was not significantly different between protocols for *M. aequituberculata*, but there was protocol sensitivity for *P. lobata* when using Shannon diversity and Faith's PD measures (but not observed richness), both of which were significantly greater in data resulting from the HiSeq protocol (**Figure 1** and **Supplementary Table 2**).

Protocol Explains Large Amount of Variation in Community Beta-Diversity

Significant differences in microbial community composition were found between protocols for both *M. aequituberculata* and *P. lobata* in all beta diversity metrics except for weighted UniFrac distances for *M. aequituberculata* in both ASV and OTU datasets (**Figure 2**, **Supplementary Figure 1**, and **Supplementary Table 3**). All beta diversity metrics maintained similar dispersions (homogeneity of variances as calculated by the function "betadisper" and referred to as "PERMDISP"; **Supplementary Table 3**), aside from Bray-Curtis for *P. lobata* in both ASV and OTU datasets (**Figure 2G**, **Supplementary Figure 1G**, and **Supplementary Table 3**), as well as Unweighted UniFrac distances for *M. aequituburculata* in the ASV dataset (**Figure 1B** and **Supplementary Table 3**) and *P. lobata* in the OTU dataset (**Supplementary Figure 1F** and **Supplementary Table 3**). While the communities did not show consistent, distinct visual segregation of nMDS data clouds according to protocol (**Figure 2** and **Supplementary Figure 1**), some individual samples had highly different relative abundances of bacterial taxa (**Figure 3**) and community structure (**Supplementary Figures 2**, **3**). While the top 10 most abundant taxa were similar between protocols and across datasets, differences in the relative abundances and detection of some phyla were present (**Figure 3** and **Supplementary Table 4**). In the ASV dataset, seven out of ten phyla were detected in both MiSeq and HiSeq protocols, and phylum-level bacterial community compositions across samples were dominated by Proteobacteria, followed by Firmicutes. However, when clustered as ASVs, these two phyla account for 74.71% versus 54.67% of the composition in MiSeq and HiSeq protocols, respectively, and three different phyla were alternatively detected between the platforms. One of them was phyla Euryarchaeota, which was present in the ASV dataset for MiSeq protocol samples with a mean relative abundance of 4.45% (**Supplementary Table 4**), but absent in the top 10 most abundant taxa for HiSeq protocol samples, in which the mean relative abundance was less than 0.002%. Although differences in the relative abundances were persistent when clustering at the 97% OTU level, fewer discrepancies were observed (**Figure 3** and **Supplementary Table 4**). For example, nine out of ten phyla were detected in both protocols, and the two dominant phyla (Proteobacteria and Firmicutes) account for 79.1% and 75.15% for MiSeq and HiSeq protocols, respectively. Interestingly, in both ASV and OTU datasets, the most abundant phyla were more evenly represented across samples from the HiSeq protocol as opposed to the MiSeq protocol (see "n" in **Supplementary Table 4**). However, ASV libraries derived from the HiSeq and MiSeq protocols also contained several unclassified bacterial ASVs that were resolved when clustering at the 97% OTU level (see "Unclassified Bacteria" in **Figure 3**). Further investigation into the make-up of these unclassified reads using NCBI blast (Altschul et al., 1990) found a few close hits to eukaryotes, yet the majority remained unidentified; comparing reads against the more prokaryote-focused RDP database (Cole et al., 2014) did not better resolve unclassified bacteria ($< 0.1\%$ of hits passed a 97% identity threshold). To ensure these putative eukaryotes that passed our automated quality control measures did not affect the results of this study, we also manually removed them and re-ran statistical tests (**Supplementary Table 5**) and re-plotted relative abundance (**Supplementary Figure 4**). This additional quality control measure did not change the results of this study. Regardless, this suggests there may be challenges in the taxonomic assignment of ASVs from short read data and requires further attention.

Standard Normalizations Do Not Overcome Protocol Induced Variability in Microbiome Diversity

To examine whether standard normalization methods used in the field could overcome the differences between protocols, the datasets were manipulated by either removing rare taxa or grouping at higher taxonomic classifications, including Family and Phylum level. Truncating the microbial communities by removing rare taxa did not eliminate the beta diversity differences between the two protocols (**Supplementary Table 6**). Removing rare taxa reduced the communities to less than 20 taxa (at ASV or OTU level), representing less than 1% of the total and suggesting that these bacterial communities are predominantly composed of low abundance taxa. Datasets of both species and taxonomic assignments (ASVs and OTUs) maintained the previously seen significant differences between protocols for all dissimilarity metrics aside from Weighted UniFrac for *M. aequituberculata* for both 0.5% and 1% rare taxa cut-offs (**Supplementary Table 6**). To reduce the effects of minor differences in closely related bacterial taxa, we also ran PERMANOVAs and homogeneity of variance tests on communities at both the Family and Phylum classification level. Significant differences were found again between protocols, however, this varied according to both host species and taxonomic level (**Supplementary Table 7**). *Porites lobata* showed significant differences between protocols even at the Phylum level, whereas *M. aequituberculata* communities were significantly different between protocols at the Family level, but only the two dissimilarity metrics utilizing presence/absence data (binary Jaccard and Unweighted UniFrac) showed significant differences at the Phylum level.

Differential Abundance Analysis Is Not Protocol Agnostic

Differential abundance analyses showed that only a few specific ASVs were significantly enriched in one protocol or the other (**Figure 4** and **Supplementary Table 8**). The most enriched taxa belong to the dominant phyla, Proteobacteria and Firmicutes, with the magnitude of enrichments ranging between an

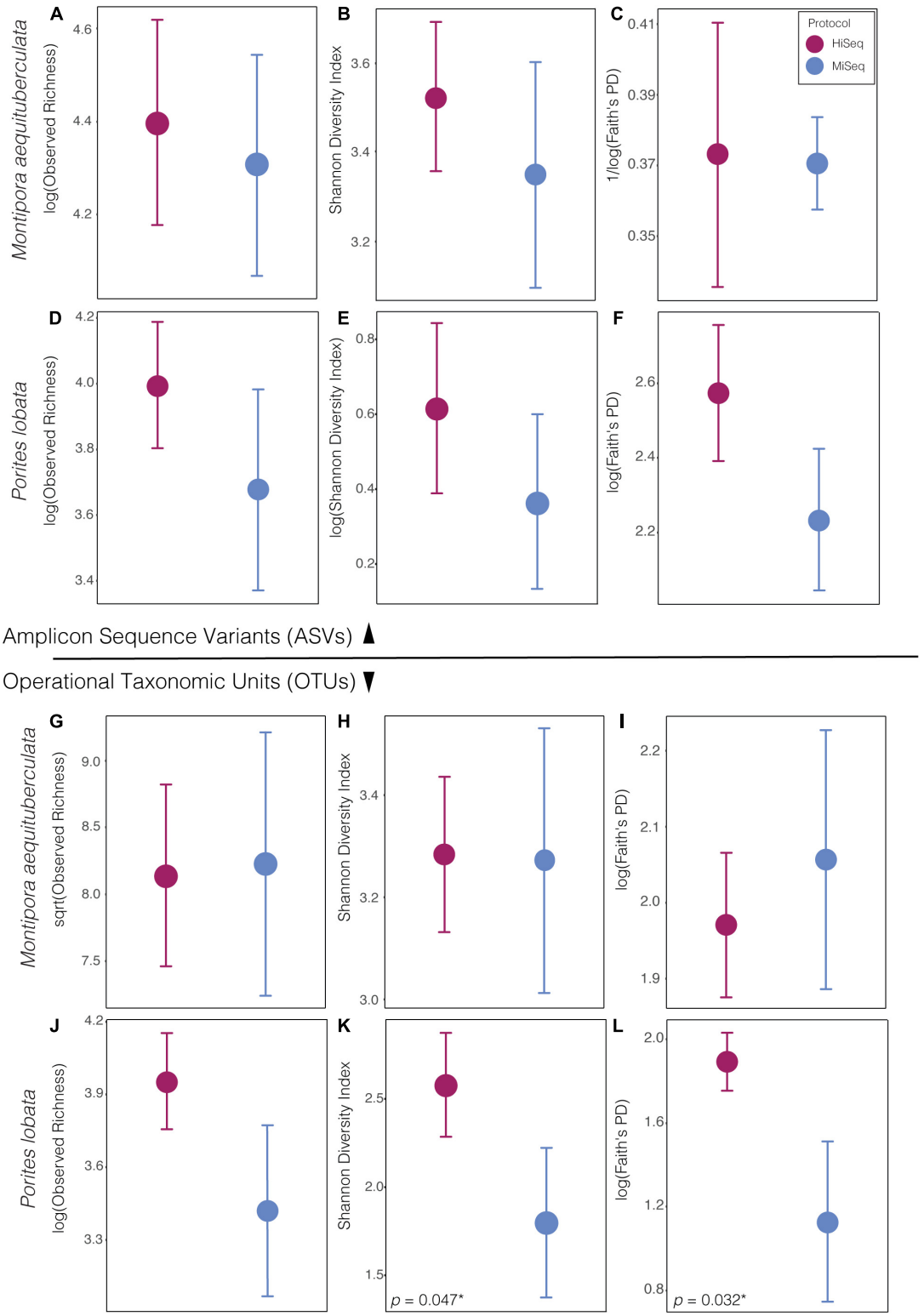

FIGURE 1 | Alpha diversity metrics of bacterial ASVs (top) and OTUs (bottom) between protocols in both species, *Montipora aequituberculata* **(A–C** and **G–I)** and *Porites lobata* **(D–F** and **J–L)**, for observed species richness **(A,D** and **G,F)**, Shannon diversity **(B,E** and **H,K)**, and Faith's phylogenetic distance **(C,F** and **I,L)** [significant p-values are reported in lower left-hand corner of panel with an asterisk (*). For all other p-values, see **Supplementary Table 2**].

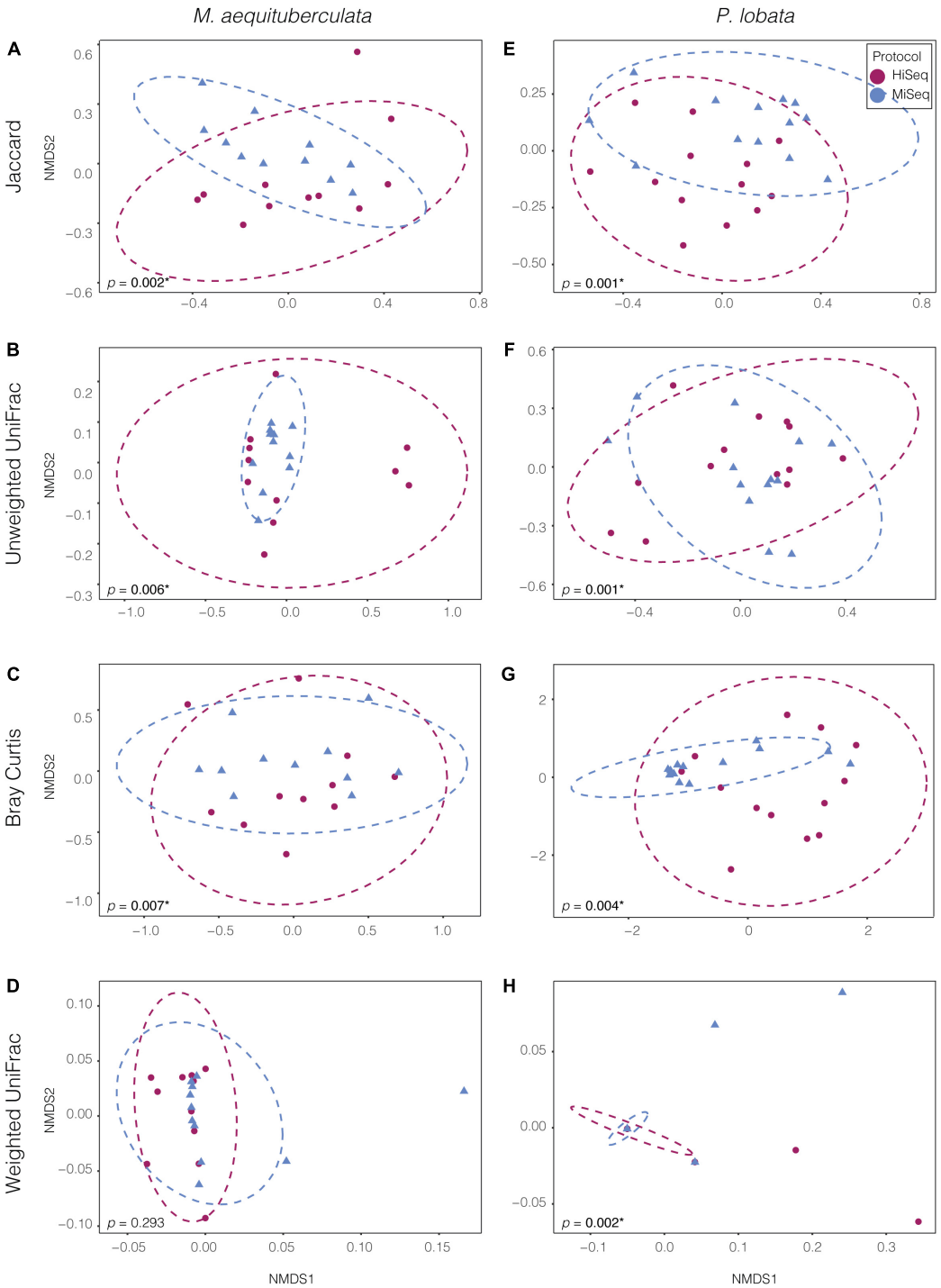

FIGURE 2 | Non-metric multidimensional scaling (NMDS) ordinations of ASV bacterial communities between the two species, *M. aequituberculata* **(A–D)** and *P. lobata* **(E–H)** for each of the four tested dissimilarity metrics: Jaccard, Unweighted UniFrac, Bray-Curtis and Weighted UniFrac. P-values with asterisks (*) refer to significant PERMANOVA results (see **Supplementary Table 3**).

approximately 7- and 29-fold change. While there was variation in differentially abundant taxa between protocols according to species and clustering method, some taxa were consistently different. For instance, HiSeq protocol libraries for both species had consistently higher abundances of *Geobacillus* sp., and lower abundances of Xenococcus PCC-7305 using the OTU dataset (**Figures 4B,D**). The magnitude of these enrichments was also consistent between coral species (**Supplementary Table 8**).

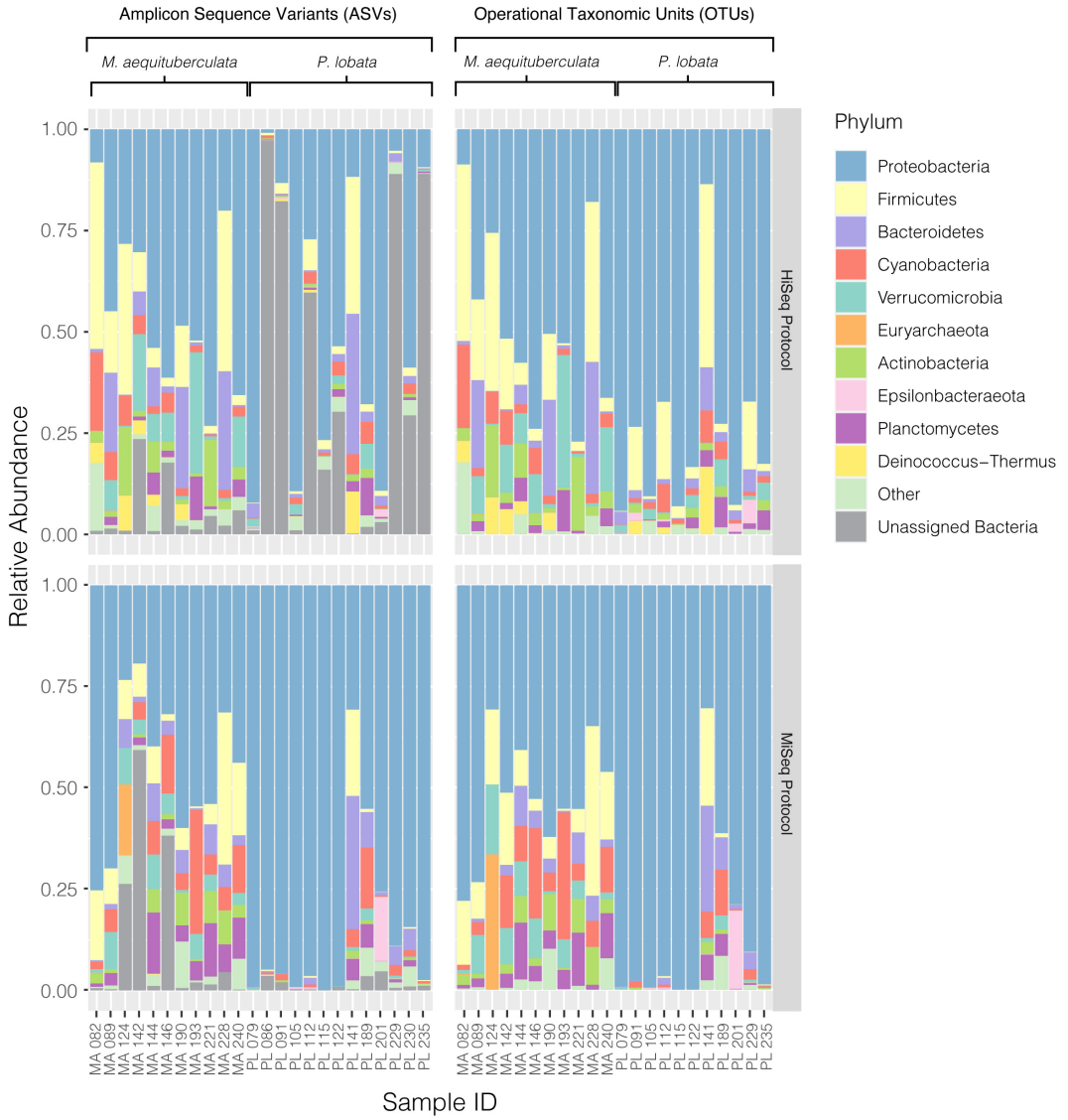

FIGURE 3 | Relative abundances of the top ten most abundant bacterial phyla present in each coral sample from both *M. aequituberculata* and *P. lobata* prepared and sequenced using the HiSeq protocol (top) and the MiSeq protocol (bottom), using both the ASV (left) and OTU (right) datasets. "Other" groups phyla that are not in the top ten most abundant, and "Unassigned Bacteria" refers to unassigned bacterial reads.

Importantly, differential enrichment between the protocols was observed in two taxa identified as crucial players of coral health and resilience, *Endozoicomonas* and *Vibrio* spp. *Endozoicomonas* exhibited significantly higher abundances in data derived from the HiSeq protocol in both species and datasets (ASV vs. OTU), except in *P. lobata* using the OTU dataset (**Figures 4A–C**). *Porites lobata* showed significantly higher abundances of a *Vibrio* ASV when sequences were prepared with the MiSeq protocol (**Figure 4C**), but this difference was not maintained in the OTU dataset (**Figure 4D**). A closer look at all *Vibrio* and *Endozoicomonas* ASVs found in the sequencing output from HiSeq, MiSeq, or both protocols revealed no significant differences in mean GC content or melting temperatures (**Supplementary Table 9**).

DISCUSSION

In Scleractinian corals, 16S rRNA gene profiling remains a common and cost-effective tool for quantifying diversity of bacteria and some archaea in holobionts (Hernandez-Agreda et al., 2019). With the increase in sequencing of coral hosts by large collaborative groups such as EMP, and subsequent public sharing of sequencing data, it has become a common goal to examine widespread patterns through meta-analyses that combine datasets from multiple laboratories, making direct comparability a necessity.

In this small-scale comparative analysis of technical replicates, we found the greatest differences between the HiSeq and MiSeq protocols in the beta diversity and dispersion measures of

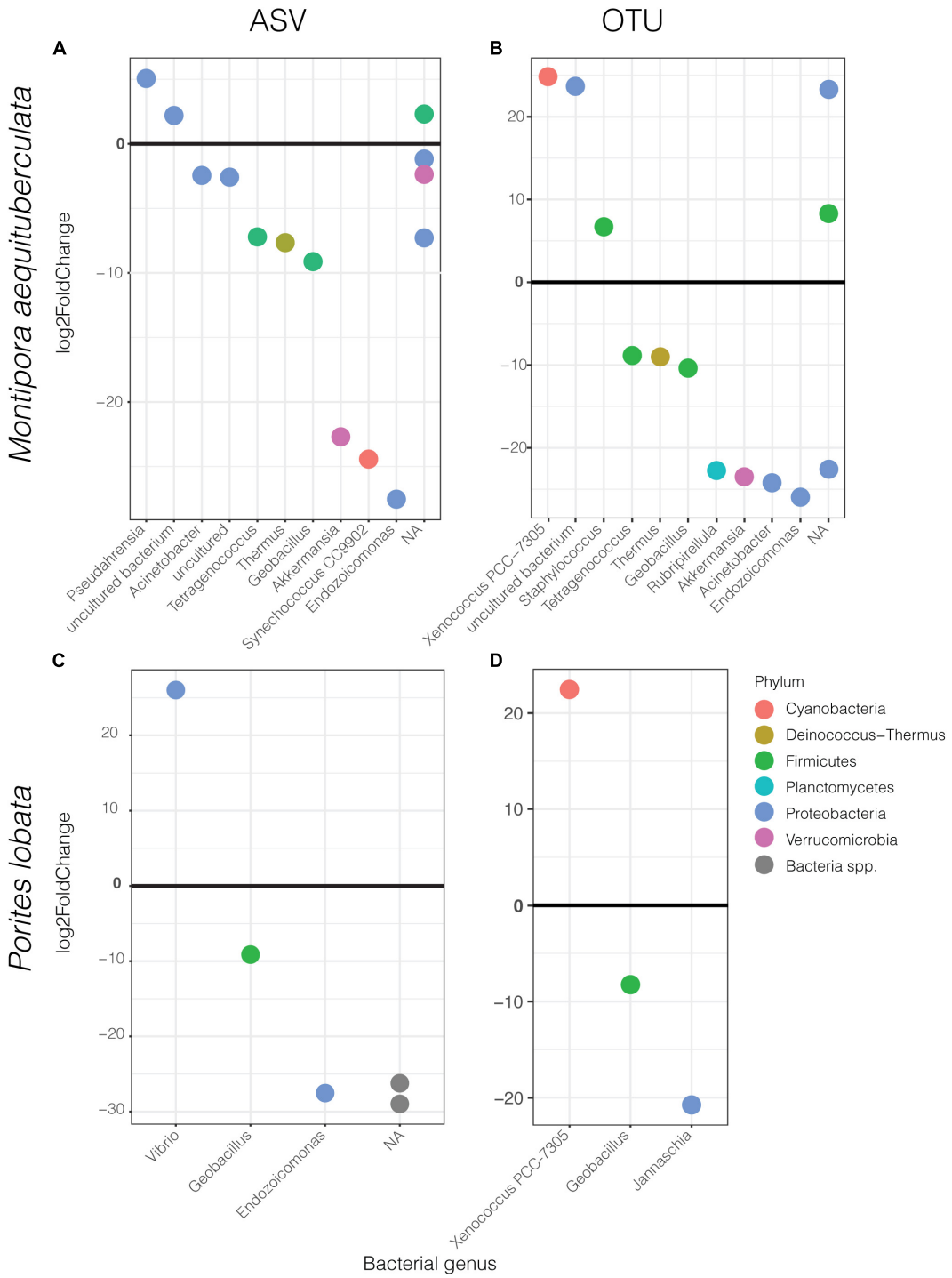

FIGURE 4 | Significantly different ASVs (left) and OTUs (right) between protocols labeled by bacterial genus and colored by phylum for *M. aequituberculata* **(A,B)** and *P. lobata* **(C,D)**. Positive log2fold change refers to those significantly enriched in MiSeq protocol samples and negative log2fold change are those significantly enriched in HiSeq protocol samples. NA refers to bacteria unclassified at genus level.

coral microbiomes. Specifically, beta diversity and dispersion metrics were inconsistent between protocols, host species, and dissimilarity metrics, with differences in protocol explaining between 4 and 28% of the microbiome variability. Certain taxa were also significantly enriched in only one of the protocols, including those with known ecological importance. For example, *Vibrio* spp. and *Endozoicomonas* spp. ASVs were significantly enriched in MiSeq and HiSeq protocols, respectively. These two taxa have been identified as important in the health and maintenance of coral homeostasis and are often used to make

statements about the health of the coral host (Bourne et al., 2008, 2016): *Vibrio* spp. have been implicated in disease (Ben-Haim et al., 2003) but remain common partners in healthy corals, while *Endozoicomonas* spp. are hypothesized to benefit to coral health via synthesis of dimethylsulfoniopropionate (DMSP) (Tandon et al., 2020), carbohydrate cycling and protein provisioning (Neave et al., 2017a), and may be considered a potential symbiont (Neave et al., 2017b). The differential abundances of these two taxa between protocols are particularly troubling for coral-specific studies and further indicate that care must be taken when comparing coral microbiome datasets resulting from even highly similar protocols.

Alpha diversity metrics on ASV data (both abundance-based and phylogenetic) were consistent between protocols in both coral species. Alpha diversity using OTU data were comparable between protocols with all metrics for *M. aequituberculata*, but significant differences were present with Shannon Diversity and Faith's PD indices for *P. lobata*, suggesting some variability in alpha diversity in regard to both relative abundances and phylogenetic makeup when sequences are grouped with 97% similarity only. Nonetheless, our results suggest that comparisons of some alpha diversity metrics between protocols may be more reliable than comparisons of community composition. The results found here should be benchmarked in other systems and tested more broadly across species to determine the extent to which small differences in protocol might bias the perceived composition of host-associated or environmental microbiome sequencing.

Regardless of rarefaction, removal of low abundance reads, or comparisons of the resulting data at higher taxonomic levels, the bacterial community composition and relative abundances of taxa maintained differences between the two protocols but in ways that were inconsistent across host species and analytical metric. This can result, for instance, from one set of samples containing taxa that may never be present if they are prepared with a different protocol or sequenced on a different platform, likely due to differences in sequencing depth, where the deeper sequencing of the HiSeq platform can provide a greater opportunity to identify rare taxa (Caporaso et al., 2012), thus shifting the overall community composition. We note that while we used a relatively high threshold for removal of rare taxa in the present study, it may be useful to use a lower threshold (e.g., removing taxa with below 0.1 or 0.01% relative abundance) depending on dataset. However, differences were also apparent between OTU and ASV datasets, suggesting that how we characterize bacterial species and/or strains, and at what taxonomic level we choose to analyze these data, may result in unintended biases. We found no evidence of differences in secondary structure of two differentially abundant taxa (*Vibrio* and *Endozoicomonas*) that could have resulted from differential denaturation of sequences in the two platforms due to differences in platform chemistry (Nakamura et al., 2011). Specifically, there was no indication of high GC content in these sequences, which has previously been found to affect read numbers from Illumina sequencing runs due to intermittent halting of polymerase during amplification (Lyubetsky et al., 2006; Price et al., 2017). While this was not an exhaustive dive into the effects of platform chemistry on sequencing outcome, it suggests that differential abundances of specific taxa are unlikely to be caused by the presence of differential secondary structures. However, further research is necessary to rule this out completely.

The samples used in this study were not initially intended to test differences between protocols or sequencing platforms, but rather provided an opportunity to examine an overlapping set of technical replicates that arose from a larger study comprised of similarly prepared and differentially sequenced samples. Thus, we cannot clearly identify the specific mechanism(s) involved in driving the found community differences. Biases in these complex microbial communities could be a result of (1) differences in sequencing depth that are not overcome by rarefaction or other *in silico* normalizations, (2) library denaturation and/or sequencing platform chemistry, (3) differences in reagents and/or batches of reagents, such as the type of Hi Fidelity Taq used in PCR or other extraction, PCR or library preparation reagents, and potentially even (4) user and/or facility bias (Rausch et al., 2016; Parker et al., 2018). Regardless, the results shown here reveal not only the necessity to design a targeted study to examine procedural and mechanistic differences in sequencing protocols, but also the responsibility of researchers to proceed with extreme caution when combining and interpreting datasets that are generated from subtly and seemingly innocuously different methodologies.

CONCLUSION

The present study found limitations in our ability to compare coral microbiome 'technical' replicates that were generated in almost identical fashions but then sequenced on different platforms. Despite attempts to rectify these issues with some commonly used normalization methods, we still found significant differences in some alpha diversity metrics and in most beta diversity metrics between the two protocols. These inconsistencies make it difficult to identify a "cure-all" adjustment for comparability between even highly similar protocols and, instead, differences among protocols and sequencing platforms are more likely to be specific to the microbiome host and specific set of microbiomes found in each dataset. Studies that aim to compare beta diversity may find more confidence in their results if overlapping technical replicates for each dataset and host species are run to ensure correct adjustments are used for these specific datasets. Based on these results, we urge caution in the statistical comparison and interpretation of 16S rRNA gene datasets that combine data resulting from different protocols and sequencing platforms. While we continue to encourage meta-analyses to discover of cosmopolitan patterns in microbiome dynamics, we advise researchers to be cognizant that even minor variations in the protocol can significantly affect microbiome composition, and those running longitudinal studies be rigorous in the consistency of their methods through time.

DATA AVAILABILITY STATEMENT

The raw data from the Illumina MiSeq samples are publicly available from McDevitt-Irwin et al. (2019) at Harvard

DataVerse: https://doi.org/10.7910/DVN/3QZTT1. The raw data from the Illumina HiSeq samples are available on the NCBI Sequence Read Archive (SRA) under the BioProject accession PRJNA687031. All bioinformatics and statistical codes are freely accessible via github at github.com/hannaheps/comparison. The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

JB was responsible for the original sample collection and access to sequencing data used in this manuscript. HE and RVT developed the research question. HE performed bioinformatic and statistical analyses with technical support from SS and AH-A and wrote the initial draft of the manuscript and received editorial support from all co-authors. All co-authors contributed to the interpretation of results.

ACKNOWLEDGMENTS

We would like to thank the Kiritimati Project Manager, K. Tietjen, for her contribution to the organization and conduction of field work that provided samples for this comparative study, and to J. McDevitt-Irwin for personal communications about handling of the sequencing data. Thanks also go to the Earth Microbiome Project (EMP) and the Global Coral Microbiome Project (GCMP) who provided the funding and resources for the HiSeq sequencing, and to G. Ackerman for her unyielding assistance in data access. HE acknowledges support from the National Science Foundation (NSF) through a Postdoctoral Research Fellowship in Biology (PRFB grant #2006244). SS acknowledges support from a National Sciences and Engineering Research Council Postdoctoral Research Fellowship (NSERC PDF). JB acknowledges support from a National Sciences Foundation RAPID Grant (OCE-1446402), an NSERC Discovery Grant, the Canadian Foundation for Innovation, the Packard Foundation, a Pew Fellowship, and the University of Victoria. RVT acknowledges support from NSF: grant #2023424.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.740932/full#supplementary-material

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Apprill, A., McNally, S., Parsons, R., and Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat. Microb. Ecol.* 75, 129–137. doi: 10.3354/ame01753

Ben-Haim, Y., Thompson, F. L., Thompson, C. C., Cnockaert, M. C., Hoste, B., Swings, J., et al. (2003). Vibrio coralliilyticus sp. nov., a temperature-dependent pathogen of the coral *Pocillopora damicornis*. *Int. J. Syst. Evol. Microbiol.* 53, 309–315. doi: 10.1099/ijs.0.02402-0

Bisanz, J. (2020). *jbisanz/qiime2R*. Available Online at: https://github.com/jbisanz/qiime2R (accessed July 29, 2020).

Bourne, D., Iida, Y., Uthicke, S., and Smith-Keune, C. (2008). Changes in coral-associated microbial communities during a bleaching event. *ISME J.* 2, 350–363. doi: 10.1038/ismej.2007.112

Bourne, D. G., Morrow, K. M., and Webster, N. S. (2016). Insights into the coral microbiome: underpinning the health and resilience of reef ecosystems. *Annu. Rev. Microbiol.* 70, 317–340. doi: 10.1146/annurev-micro-102215-095440

Callahan, B. (2020). *benjjneb/decontam*. Available Online at: https://github.com/benjjneb/decontam (accessed July 29, 2020).

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

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624. doi: 10.1038/ismej.2012.8

Cole, J. R., Want, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., et al. (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* 42, D633–D642. doi: 10.1093/nar/gkt1244

Duvallet, C., Gibbons, S. M., Gurry, T., Irizarry, R. A., and Alm, E. J. (2017). Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat. Commun.* 8:1784. doi: 10.1038/s41467-017-01973-8

Epstein, H. E., Smith, H. A., Cantin, N., Mocellin, V. J. L., Torda, G., and van Oppen, M. J. H. (2019). Temporal variation in the microbiome of Acropora coral species does not reflect seasonality. *Front. Microbiol.* 10:1775. doi: 10.3389/fmicb.2019.01775

Hernandez-Agreda, A., Gates, R. D., and Ainsworth, T. D. (2017). Defining the Core Microbiome in Corals' Microbial Soup. *Trends Microbiol.* 25, 125–140. doi: 10.1016/j.tim.2016.11.003

Hernandez-Agreda, A., Leggat, W., and Ainsworth, T. D. (2019). A place for taxonomic profiling in the study of the coral prokaryotic microbiome. *FEMS Microbiol. Lett.* 366:fnz063. doi: 10.1093/femsle/fnz063

Human Microbiome Project Consortium (2012). A framework for human microbiome research. *Nature* 486, 215–221. doi: 10.1038/nature11209

Li, J. (2020). *TmCalculator:Melting Temperature of Nucleic Acid Sequences*. Available Online at: https://cran.r-project.org/web/packages/TmCalculator/TmCalculator.pdf (accessed July 21, 2020).

Love, M., Ahlmann-Eltze, C., Anders, S., and Huber, W. (2020). *DESeq2: Differential Gene Expression Analysis Based On The Negative Binomial Distribution*. Bioconductor version: Release (3.11). Available Online at: https://github.com/mikelove/DESeq2 (accessed December 5, 2020).

Lyubetsky, V. A., Rubanov, L. I., Seliverstov, A. V., and Pirogov, S. A. (2006). Model of gene expression regulation in bacteria via formation of RNA secondary structures. *Mol. Biol.* 40, 440–453. doi: 10.1134/S0026893306030113

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

McDevitt-Irwin, J. M., Garren, M., McMinds, R., Vega Thurber, R., and Baum, J. K. (2019). Variable interaction outcomes of local disturbance and El Niño-induced heat stress on coral microbiome alpha and beta diversity. *Coral Reefs* 38, 331–345. doi: 10.1007/s00338-019-01779-8

McMurdie, P. J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. doi: 10.1371/journal.pone.0061217

Mo, Z., Huang, P., Yang, C., Xiao, S., Zhang, G., Ling, F., et al. (2020). Meta-analysis of 16S rRNA microbial data identified distinctive and predictive microbiota dysbiosis in colorectal carcinoma adjacent tissue. *mSystems* 5, e00138–20. doi: 10.1128/mSystems.00138-20

Nakamura, K., Oshima, T., Morimoto, T., Ikeda, S., Yoshikawa, H., Shiwa, Y., et al. (2011). Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res.* 39:e90. doi: 10.1093/nar/gkr344

Neave, M. J., Michell, C. T., Apprill, A., and Voolstra, C. R. (2017a). Endozoicomonas genomes reveal functional adaptation and plasticity in bacterial strains symbiotically associated with diverse marine hosts. *Sci. Rep.* 7:40579. doi: 10.1038/srep40579

Neave, M. J., Rachmawati, R., Xun, L., Michell, C. T., Bourne, D. G., Apprill, A., et al. (2017b). Differential specificity between closely related corals and abundant Endozoicomonas endosymbionts across global scales. *ISME J.* 11, 186–200. doi: 10.1038/ismej.2016.95

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., et al. (2019). *vegan*. Available Online at: https://cran.r-project.org/web/packages/vegan/vegan.pdf (accessed November 28, 2020).

Pammi, M., Cope, J., Tarr, P. I., Warner, B. B., Morrow, A. L., Mai, V., et al. (2017). Intestinal dysbiosis in preterm infants preceding necrotizing enterocolitis: a systematic review and meta-analysis. *Microbiome* 5:31. doi: 10.1186/s40168-017-0248-8

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

Parker, K. D., Albeke, S. E., Gigley, J. P., Goldstein, A. M., and Ward, N. L. (2018). Microbiome composition in both wild-type and disease model mice is heavily influenced by mouse facility. *Front. Microbiol.* 9:1598. doi: 10.3389/fmicb.2018.01598

Price, A., Garhyan, J., and Gibas, C. (2017). The impact of RNA secondary structure on read start locations on the Illumina sequencing platform. *PLoS One* 12:e0173023. doi: 10.1371/journal.pone.0173023

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219

R Core Team (2020). *R: A Language and Environment For Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rausch, P., Basic, M., Batra, A., Bischoff, S. C., Blaut, M., Clavel, T., et al. (2016). Analysis of factors contributing to variation in the C57BL/6J fecal microbiota across German animal facilities. *Int. J. Med. Microbiol.* 306, 343–355. doi: 10.1016/j.ijmm.2016.03.004

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584. doi: 10.7717/peerj.2584

Schuster, S. C. (2008). Next-generation sequencing transforms today's biology. *Nat. Methods* 5, 16–18. doi: 10.1038/nmeth1156

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., et al. (2015). Structure and function of the global ocean microbiome. *Science* 348:1261359. doi: 10.1126/science.1261359

Tandon, K., Lu, C.-Y., Chiang, P.-W., Wada, N., Yang, S.-H., Chan, Y.-F., et al. (2020). Comparative genomics: dominant coral-bacterium Endozoicomonas acroporae metabolizes dimethylsulfoniopropionate (DMSP). *ISME J.* 14, 1290–1303. doi: 10.1038/s41396-020-0610-x

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

Ul-Hasan, S., Bowers, R. M., Figueroa-Montiel, A., Licea-Navarro, A. F., Beman, J. M., Woyke, T., et al. (2019). Community ecology across bacteria, archaea and microbial eukaryotes in the sediment and seawater of coastal Puerto Nuevo, Baja California. *PLoS One* 14:e0212355. doi: 10.1371/journal.pone.0212355

Wickham, H. (2016). *tidyverse/ggplot2. tidyverse*. Available Online at: https://github.com/tidyverse/ggplot2 (accessed July 29, 2020).

Wu, W. W., Phue, J.-N., Lee, C.-T., Lin, C., Xu, L., Wang, R., et al. (2018). Robust sub-nanomolar library preparation for high throughput next generation sequencing. *BMC Genomics* 19:326. doi: 10.1186/s12864-018-4677-y

Ziegler, M., Grupstra, C. G. B., Barreto, M. M., Eaton, M., BaOmar, J., Zubier, K., et al. (2019). Coral bacterial community structure responds to environmental change in a host-specific manner. *Nat. Commun.* 10:3092. doi: 10.1038/s41467-019-10969-5

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.

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 08 December 2021
doi: 10.3389/fmicb.2021.765268

Ontology-Enriched Specifications Enabling Findable, Accessible, Interoperable, and Reusable Marine Metagenomic Datasets in Cyberinfrastructure Systems

*Kai L. Blumberg[1], Alise J. Ponsero[1], Matthew Bomhoff[1], Elisha M. Wood-Charlson[2], Edward F. DeLong[3] and Bonnie L. Hurwitz[1,4]**

[1] Department of Biosystems Engineering, University of Arizona, Tucson, AZ, United States, [2] E.O. Lawrence Berkeley National Laboratory, Environmental Genomics and Systems Biology Division, Berkeley, CA, United States, [3] Daniel K. Inouye Center for Microbial Oceanography, University of Hawai'i, Honolulu, HI, United States, [4] BIO5 Institute, University of Arizona, Tucson, AZ, United States

OPEN ACCESS

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

Reviewed by:
Diogo Antonio Tschoeke,
Federal University of Rio de Janeiro, Brazil
Caroline Chénard,
National Research Council Canada (NRC-CNRC), Canada

***Correspondence:**
Bonnie L. Hurwitz
bhurwitz@email.arizona.edu

Specialty section:
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Microbiology

Received: 26 August 2021
Accepted: 16 November 2021
Published: 08 December 2021

Citation:
Blumberg KL, Ponsero AJ, Bomhoff M, Wood-Charlson EM, DeLong EF and Hurwitz BL (2021) Ontology-Enriched Specifications Enabling Findable, Accessible, Interoperable, and Reusable Marine Metagenomic Datasets in Cyberinfrastructure Systems. Front. Microbiol. 12:765268. doi: 10.3389/fmicb.2021.765268

Marine microbial ecology requires the systematic comparison of biogeochemical and sequence data to analyze environmental influences on the distribution and variability of microbial communities. With ever-increasing quantities of metagenomic data, there is a growing need to make datasets Findable, Accessible, Interoperable, and Reusable (FAIR) across diverse ecosystems. FAIR data is essential to developing analytical frameworks that integrate microbiological, genomic, ecological, oceanographic, and computational methods. Although community standards defining the minimal metadata required to accompany sequence data exist, they haven't been consistently used across projects, precluding interoperability. Moreover, these data are not machine-actionable or discoverable by cyberinfrastructure systems. By making 'omic and physicochemical datasets FAIR to machine systems, we can enable sequence data discovery and reuse based on machine-readable descriptions of environments or physicochemical gradients. In this work, we developed a novel technical specification for dataset encapsulation for the FAIR reuse of marine metagenomic and physicochemical datasets within cyberinfrastructure systems. This includes using Frictionless Data Packages enriched with terminology from environmental and life-science ontologies to annotate measured variables, their units, and the measurement devices used. This approach was implemented in Planet Microbe, a cyberinfrastructure platform and marine metagenomic web-portal. Here, we discuss the data properties built into the specification to make global ocean datasets FAIR within the Planet Microbe portal. We additionally discuss the selection of, and contributions to marine-science ontologies used within the specification. Finally, we use the system to discover data by which to answer various biological questions about environments, physicochemical gradients, and microbial

communities in meta-analyses. This work represents a future direction in marine metagenomic research by proposing a specification for FAIR dataset encapsulation that, if adopted within cyberinfrastructure systems, would automate the discovery, exchange, and re-use of data needed to answer broader reaching questions than originally intended.

Keywords: ontology, FAIR, metagenomics, marine microbiology, cyberinfrastructure (CI), next generation sequencing—NGS, omics

INTRODUCTION

Recently, efforts have been made to encourage scientific data producers and publishers to make their data Findable, Accessible, Reusable, and Interoperable (FAIR) (Wilkinson et al., 2016), referred to as the FAIR guiding principles for scientific data management and stewardship. These principles provide high-level suggestions for how to improve the digital ecosystem of data producers, publishers, and consumers (Wilkinson et al., 2016). The FAIR principles state that datasets should: (1) be annotated with metadata to allow for the discovery of datasets based on their metadata (Findability), (2) be freely accessible using standard protocols (Accessibility), (3) use standardized metadata that are able to work together in combination with other metadata (Interoperability), and (4) employ metadata which can thoroughly describe a plurality of accurate and relevant attributes from the dataset (Reusability). The authors specify that data should be FAIR, not only apply to humans but more importantly to automated web-discovery systems (referred to as machine-agents) (Wilkinson et al., 2016). The long-term objectives laid out by the authors of the FAIR principles are to facilitate the integration and reuse of published data to enable novel discoveries and innovations (Wilkinson et al., 2016). Long term ecological monitoring sites have data that are important for understanding the global ocean, however, these datasets (including metagenomic data) are often not consistently made available in a manner following using the FAIR data principles.

Advances in sequencing technologies have enabled the generation of unprecedented volumes of metagenomic data (Mitchell et al., 2016). This has resulted in a proliferation of publicly available metagenomic datasets, especially from marine sampling expeditions which have enabled novel insights into the taxonomic structure and functional capabilities of microbial communities from diverse ocean environments (Karl and Lukas, 1996; Rusch et al., 2007; Zinger et al., 2011; Sunagawa et al., 2015; Zheng et al., 2016; Mende et al., 2017; Biller et al., 2018). Better understanding the role of microbes in heterogeneous and dynamic marine ecosystems will require both longitudinal and temporal collections of large-scale and historical data. Therefore, these data will need to be integrated with one another, as well as linked with supplemental data detailing their broader physio-chemical context. Data will need not only to be made reusable within projects generated within a greater sampling campaign, but also across datasets generated by different campaigns as well as investigators.

Despite the proliferation of marine metagenomic and accompanying contextual data, the lack of commonly used FAIR principles by which to standardize the wide range of physicochemical attributes remains a barrier toward elucidating the biogeochemical and environmental drivers of community structure. Thus, efforts to compare across studies using meta-analyses have infrequently been attempted due to lack of interoperable data (Field et al., 2009). Further, as identified by the EarthCube Geoscience 2020 report, understanding critical changes to ecosystems requires appropriate historical context by which to detect long-term trends; a common need across geoscience domains[1]. By integrating disparate data types, such as 'omics and environmental data, across many sampling efforts to cover larger spatiotemporal ranges, we can begin to fill the contextual void that currently limits our understanding of ecosystem resilience. By applying the FAIR data principles to these 'omics datasets we can begin to realize this potential.

Today, metagenomic datasets are typically published within publicly available International Nucleotide Sequence Database Collaboration (INSDC) repositories such as the European Nucleotide Archive (ENA) and the National Center for Biotechnological Information (NCBI) (Nakamura et al., 2013). Thus, published metagenomic data are Findable, and after an optional holding period become Accessible as well. Additionally, the Genomic Standards Consortium (GSC) was established to standardize contextual data (a.k.a. metadata) requirements for genomic and metagenomic data (Yilmaz et al., 2011a). As a result of GSC efforts, several Minimum Information about any (x) Sequence (MIxS) checklists were created to establish a unified standard for the description of data commonly collected along with sequencing data (Yilmaz et al., 2011b). Contextual information about the environmental source, spatiotemporal location, and other environmental characteristics are essential to interpreting and analyzing metagenomic data (Yilmaz et al., 2011a). As such, the MIxS specifications are organized into a collection of environment-specific packages (e.g., water, soil), with discipline-specific metadata requirements (Yilmaz et al., 2011b).

In an effort to standardize their genomic and metagenomic contextual data, both the ENA and NCBI have integrated the MIxS checklists into their submission procedures (Karsch-Mizrachi et al., 2018; Mitchell et al., 2018); however, many metadata fields are not associated with ontology terms, thus

[1]https://www.researchgate.net/publication/281445479_Geoscience_2020_Cyberinfrastructure_to_reveal_the_past_comprehend_the_presentand_envision_the_future

hindering their Interoperability. Additionally, these datasets are not automatically reusable because MIxS metadata fields are entered as free text. Although the MIxS checklists provide a syntax for how data should be entered (as well as examples), users are expected to correctly type out ontology term identifiers and labels or provide both numeric values and free text units. Finally, there are no mechanisms to ensure data conforms to the expected data type and syntax (e.g., checks if a number is provided, if units are provided, or if a correct ontology term is provided). So, although there is more widespread use of the MIxS checklists in major genomic sequencing repositories, the resulting physicochemical data accompanying the genomic data sequences remains inconsistent.

To perform meta-analyses, it is not only important for datasets to use consistent vocabularies, but it's also important for those vocabularies to follow a rigorous semantic framework that enables computation and discovery. This can be accomplished by utilizing vocabularies which serve as the metadata conformant to the FAIR data principles (Wilkinson et al., 2016). All MIxS checklists mandate the use of terminology from a semantic resource called the Environment Ontology (ENVO) (Buttigieg et al., 2013, Buttigieg et al., 2016) for the annotation of the broad-scale environmental context, local-scale environmental context, and environmental medium of metagenomic data (Yilmaz et al., 2011b). ENVO provides semantic descriptions of environment types, environmental materials, and biomes, by which to annotate biological samples (Buttigieg et al., 2013, Buttigieg et al., 2016). Environmental information represented within ENVO is both human and machine-readable, hierarchically structured, and contains logical and machine-processable relationships to other represented entities. Such features make ENVO an ontology, rather than a controlled vocabulary (Buttigieg et al., 2016).

ENVO is not a standalone resource but is rather part of a larger consortium of ontologies called the Open Biomedical and Biological Ontologies (OBO) Foundry and Library (Smith et al., 2007). OBO ontologies use common design strategies to work together interoperably as a unified multidisciplinary knowledge representation model (Walls et al., 2014); each representing information from specific domains such as genomics via the Gene Ontology (GO) (Ashburner et al., 2000), and scientific investigative processes via the Ontology for Biomedical Investigations (OBI) (Bandrowski et al., 2016). In addition, ontologies are semantic web resources (Bechhofer et al., 2004), supporting querying features that enable the discovery of ontology terms based on input conditions (Prud'hommeaux and Seaborne, 2008). Ontologies also support reasoning, a process employing formal logic to determine the hierarchical placement of terms within an ontology, based on their relationships and logical statements (Kazakov et al., 2014). Due to their machine-readability and inclusion of machine-processable relationships between entities (Smith et al., 2007), ontologies are considered maximally robust semantic systems (McCreary, 2010).

Although ontologies are an important part of making data FAIR, simply annotating data with ontology terms is insufficient toward achieving the FAIR principles. Also required are standard practices for packaging, storing, and transferring datasets such that they can be made discoverable to machine agents via search routines on their ontology annotations. Other environmental science disciplines such as Oceanography and Meteorology achieve data interoperability by using the Network Common Data Form (NetCDF) standard (Rew and Davis, 1990; Brown et al., 1993). Although NetCDF has been used in genome-wide association studies based on precomputed Single Nucleotide Polymorphisms (Muñiz Fernandez et al., 2012), it is not intended to link genomic data to other sources such as physicochemical data.

Currently, a variety of web-based portals and cyberinfrastructure systems exist for metagenomic data. The Genomes Online Database (GOLD) hosted by the Joint Genome Institute (JGI) is an open access online portal that maintains contextual metadata associated with genomic and metagenome projects (Kyrpides, 1999; Mukherjee et al., 2019). GOLD makes use of the MIxS checklists to ensure that metagenomic datasets are consistent prior to analysis within the Integrated Microbial Genomes (IMG) system (Markowitz et al., 2007; Mukherjee et al., 2019). Another metagenomic portal is the widely used Meta Genomics Rapid Annotation using Subsystems Technology (MG-RAST) server, which also leverages the MIxS checklists and previous versions of ENVO (Meyer et al., 2008). The MGnify resource, hosted by the European Bioinformatics Institute (EBI), is another metagenomic analysis suite that aggregates metagenomic data from ENA (Mitchell et al., 2018), also linking to metadata sourced from MIxS checklists. Finally, a recent United States Department of Energy (DOE) initiative, the National Microbiome Data Collaborative (NMDC) was launched to support the data science ecosystem as well as community practices around making 'omics data FAIR (Eloe-Fadrosh et al., 2021; Vangay et al., 2021). Although these resources do leverage existing standards such as MIxS checklists, there are currently not commonly accepted workflows for connecting metagenomic data with accompanying physicochemical data to make them FAIR implemented within existing metagenomic web portals.

In their 2017 work Wilkinson et al. (2017) demonstrated an example of how several web and semantic web technologies can be brought together to create a reference implementation of an interoperability architecture to enhance the discovery, integration, and reuse of biological data in accordance with the FAIR principles. To unlock the full potential of metagenomic data to elucidate the interactions between physicochemical gradients and the structure and function of microbial communities, a standardized containerization system analogous to NetCDF or that demonstrated in by Wilkinson et al. (2017) is needed. Such a system will need to (1) semantically annotate marine metagenomic and contextual data with ontologies, (2) containerize data products and their semantic annotations such that they make the data discoverable via its constituents to machine searches within cyberinfrastructure systems, and (3) support the transfer of containerized data products between cyberinfrastructure systems.

Here we demonstrate a prototype for FAIR re-use of metagenomic datasets using ontology-enriched specifications in

Planet Microbe, a cyberinfrastructure system. Planet Microbe is the first data portal to offer FAIR data for large-scale analyses across multiple marine metagenomic studies (Ponsero et al., 2021). Current datasets encompass more than 2,300 samples collected from multiple projects around the world. These harmonized datasets fuel the search feature to allow users—including scientists, educators, and citizen scientists—to run computational tools on samples across systems. Planet Microbe is not a monolithic database and analysis system, nor a public data repository, but rather a collection of data products and tools that are grouped together under a common light-weight web-based interface. The approach to developing data packages we describe here is a way to package datasets as products in and of themselves and can be used by any person or system independently. In our working example at Planet Microbe, these data are containers that can be ported from system-to-system or cloud-to-cloud. Thus, all data packages and tools developed here are independent resources for the community. Moreover, the approach we propose is a foundation and set of standards that can be developed by the community over time, building upon and extending the data products currently available in Planet Microbe.

RESULTS AND DISCUSSION

Open Biomedical and Biological Ontologies-Frictionless Data Products

Using a combination of Frictionless Data and OBO ontologies, we encapsulate multi-part marine metagenomic datasets and their accompanying physicochemical and environmental contextual data into FAIR data products. The Planet Microbe Data Package specification is built upon Frictionless Data[2]. Frictionless Data is a technical standard for the containerization, publication, and mobilization of data. Frictionless Data provides specifications and software libraries for the construction and use of Frictionless Data Packages, including software tools for loading Data Packages into database systems. Frictionless Data Packages are JavaScript Object Notation Format (JSON) files in which metadata about multiple data resources such as tab-separated value (TSV) files can be encapsulated (see **Figure 1A**). Annotation metadata, such as ontology terms can be added within Frictionless Data Package JSON files to describe the resource files. The Data Packages enable comprehensive data validation and annotation. See **Figure 1B** for an example of this in which a single column from a dataset TSV file is annotated with a triad of OBO ontology terms to capture (1) what the data column is about, (2) the units of measure the data is reported in, and (3) the measurement device used to collect the data. When used together in systems like the Planet Microbe cyberinfrastructure web portal, the Data Packages can enable novel queries across integrated data sources. See the section on addressing biological questions. Additionally, the Data Packages can be used independently of the Planet Microbe portal, by anyone for any purpose. The Data Packages are available in github.[3]

Planet Microbe Implementation

In Planet Microbe, we implemented the following FAIR data properties to which the Planet Microbe OBO-Frictionless Data Package specification should conform to ameliorate the FAIR use of marine metagenomic data within cyberinfrastructure systems. We define the proposed properties as follows: (1) Machine-exposability: the ability for containerized data products to be unpacked into, understood by, and used within various cyberinfrastructure systems; (2) Complete transferability: the ability for multi-component datasets to be easily exchanged between informatic systems employed by data producers and consumers; (3) Reusability indicators: comprehensive data annotations including data type checks and machine-readable semantics enabling additional decisions to be made about whether or not data attributes should be reused in combination; and (4) Attribute-discoverability: the ability to discover individual constituents from datasets based upon their annotation with machine-readable semantics specifying the type of the attribute.

Machine-Exposability

Although not prescribing a specific solution to making data FAIR, the FAIR Principles emphasize machine-actionability, the ability for computational systems to find, access, interoperate, and reuse data in an automated manner. The current standard for annotating attributes accompanying metagenomic datasets, the MIxS checklists, are not machine-readable. Additionally, there is no standard way to encapsulate marine metagenomic datasets along with their annotation semantics within a machine-readable framework. Thus, to make marine metagenomic data FAIR, we developed an informatic system that can connect machine-readable annotation semantics for both data attributes and provenance information, with data. The ability for Frictionless Data Packages to enable the machine-readable annotation of variable resources, is highly amenable for the management of marine metagenomic datasets, which often contain multiple components each with a variable template.

Complete Transferability

Another consideration when designing the Planet Microbe Frictionless Data Package system was what we refer to as the property complete transferability, the ability to completely exchange multi-component data products between data producers and consumers. Complete transferability is also satisfied using Frictionless Data Packages, which enable multiple data resources to be described within a master metadata resource. All files including the master JSON file and resource files can be stored together within a common directory structure where the relative paths within the directory to the resource files are described within the master JSON file. The master JSON file contains the MD5 checksums of the resource files, thus the complete transferability of all data files can be machine-validated

[2]https://frictionlessdata.io/

[3]https://github.com/hurwitzlab/planet-microbe-datapackages

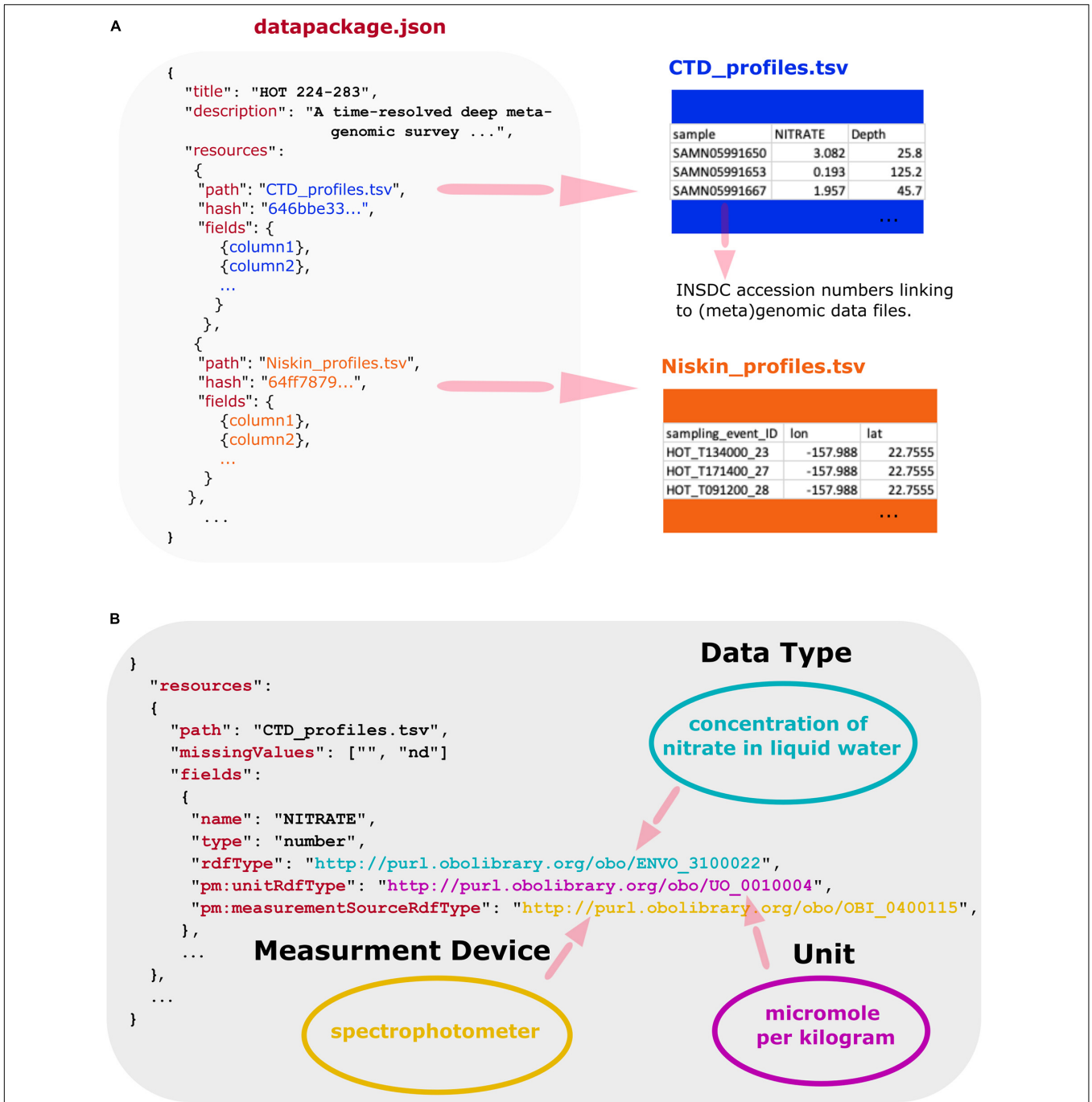

FIGURE 1 | (A) Overview of Planet Microbe Frictionless Data Packages that for any given project includes a "datapackage.json" file containing machine-readable descriptive information about the individual resource TSV files and their columns, as well as the original (or slightly modified) data files. Note that at least one data file per Data Package also includes data columns with International Nucleotide Sequence Database Collaboration (INSDC) accession numbers linking to (meta)genomic datasets. The machine readable JSON description enables data to be exposed to and processed by machine systems. Additionally, the JSON files contain information about the resource file path and md5 checksums (hashes) allowing for complete transferability of multi-component datasets between systems. **(B)** Simplified view of data annotations used to markup a dataset column in a Planet Microbe Frictionless Data Package for an individual data column from an example data set resource. Multiple reusability indicators are built into the system. This includes missing values checklists that describe datasets-specific strings specifying "NA" values or missing data. Additionally, the expected data type (e.g., "number") is specified and can be used to validate the data to make sure it is as expected (e.g., a numeric value not a string). Unit interoperability is specified by the pm:unitRdfType, annotating the units of measurement with a term [e.g., UO "*micromole per kilogram*" (UO:0010004)] from the Units Ontology. Collection instrumentation is specified by a term from the Ontology for Biomedical Investigations in the pm:measurmentSourceRdfType [e.g., OBI "*spectrophotometer*" (OBI:0400115)] to denote the type of device with which the measurement was taken. Finally, attribute types are made discoverable to machine systems via the rdfType, such as ENVO "*concentration of nitrate in liquid water*" (ENVO:3100022), specifying what the data attribute is about.

by checking if all the resource files exist where they should and if they have the correct MD5 checksum.

Reusability Indicators

For datasets to be reused in meta-analyses, it may be necessary to perform certain consistency checks or make decisions about whether or not data attributes should be reused in combination. We refer to these properties as reusability indicators and have incorporated three of them into the Planet Microbe Frictionless Data Packages. These include checks for (1) data inconsistencies, (2) unit interoperability: machine-readable semantics specifying the units attributes were measured in, and (3) measurement device indicators, semantics specifying what devices were used to measure data attributes.

Checks for Inconsistencies

Before data can be reused, we must ensure that it is reported in a way that is consistent with its intended use and purpose. Frictionless Data Packages enable this by allowing for data annotation at a fine level of resolution, namely specifying the expected type of each attribute from an original data source to be "string," "number," or "datetime." In addition, Frictionless Data provides software libraries for data validation, making sure it conforms to these user-specified constraints. Users can further specify expected formats to which data columns are expected to conform to, for example, specifying a column of collection dates to be of type "datetime," which follow a specific format such as "%Y-%m-%dT%H:%M:%SZ." These Frictionless Data validation features were leveraged by scripts we developed to validate the Planet Microbe Data Packages. This enabled us to ensure the original data sources were formatted correctly, while preserving the data's original configuration (e.g., datetime formatting). During validation of the original datasets a reasonable number of inconsistencies were found and were logged in the "README.md" files associated with each Data Package. There also exists community software tooling created by users within the Frictionless Data community, such as the Good Tables python library and command line tool for validating and transforming tabular data within Frictionless Data Packages (Heughebaert, 2020).[4] In addition to our own validation script, we used the Good Tables validation software to add numeric range constraint checks for latitude and longitude values in the Planet Microbe Data Packages. We did so by adding constraint code blocks to fields (parameters) within Planet Microbe Frictionless Data Packages to specify minimum and maximum acceptable values (e.g.,–90 to 90 and –180 to 180 for latitude and longitude, respectively). This provides an additional sanity check on the data and helps catch errors such as swapped latitude and longitude values. See **Supplementary Figure 1** showing an example of a Planet Microbe Data Package with range constraints for a latitude field from the Beyster Family Fund and Life Technologies Foundation-funded Global Ocean Sampling Expedition (GOS), 2009–2011 dataset. By building these types of consistency checks into the Planet Microbe Data Packages, we were able to future proof the data, protecting ourselves and future reusers of these data products against inconsistencies.

Unit Interoperability

Another key attribute required for data to work interoperably is that the units of measure in which individual data attributes were collected, a concept such as "nanograms per liter" must be understandable at a machine level. Although the MIxS checklist defines standard units in which 'omic accompanying data should be reported in, conformance cannot just be assumed. Thus, within the Planet Microbe Frictionless Data Package specification, semantic annotations for unit types are included. Many unit vocabularies and systems exist which would be fit for this purpose (Madin et al., 2007; Rijgersberg et al., 2011; FAIRsharing Team, 2015). Here we opted to make use of terminology from the Units Ontology (UO) (Gkoutos et al., 2012) for the units of measure annotations included in the Frictionless Data Package JSON files. This was done both to remain within the OBO knowledge modeling paradigm, as well as for the ease of use in importing UO, within the PMO application ontology, as UO is an OBO ontology following standard formatting conventions. See **Figure 1** for an example of an individual parameter within a Planet Microbe Data Package annotated with a UO unit term.

Measurement Device Terms

Another reusability indicator, which further enriches the descriptive value of annotated data, is the inclusion of machine-readable semantics describing the instrumentation used in the collection of individual data attributes. A variety of semantic systems for instrumentation exist such as the NERC P10, the Rolling Deck to Repository (R2R) device type vocabulary, and the Biological and Chemical Data Management Office (BCO-DMO)'s instruments vocabularies (Coburn et al., 2012; Chen et al., 2014; Moncoiffe and Kokkinaki, 2018). However, for the Planet Microbe Frictionless Data Package specification, we again opted to stay within the OBO paradigm for ease of use with the application ontology, and reused measurement device terminology from the OBO ontology OBI (the Ontology for Biomedical Investigations) (Bandrowski et al., 2016). Although the information about instrumentation was not as easy to discern from the original project data sources, several examples of annotations with OBI measurement devices terms have been included in the Planet Microbe Frictionless Data Packages. Information regarding the use of high-performance liquid chromatography instruments was most generally deciphered from the source repositories for the Planet Microbe Data Packages. As a result, 1593 annotations with "*high performance liquid chromatography instrument*" (OBI:0001057) were made. Other measurement devices for which a reasonable basis for annotation from the metadata in the source repositories included 9 measurement device annotations with the class "*flow cytometer*" (OBI:0400044), 11 annotations with "*fluorometer*" (OBI:0400143), 1 instance of the use of a "microscope" (OBI:0400169), and 7 clear uses of a "*spectrophotometer*" (OBI:0400115). See **Figure 1B**, for an example of an individual parameter from a Planet Microbe Data Package annotated with an OBI measurement device term.

[4]https://goodtables.readthedocs.io/en/latest/

Attribute-Discoverability

The final property we built the Planet Microbe Frictionless Data Package specification to conform to we call attribute attribute-discoverability. By this we mean the ability for machine systems to search for data attributes based on their annotation types; types which should themselves be machine-readable semantics that exist within a larger knowledge representation framework.

Planet Microbe Application Ontology

To ensure attribute-discoverability of Planet Microbe Frictionless Data Packages, all semantic annotations used in the Data Packages are included within the Planet Microbe application (PMO).[5] The PMO application ontology was built following OBO foundry tools[6] (Ponsero et al., 2021) and includes ontology imports from other relevant OBO foundry ontologies such as ENVO, UO, and OBI. OBO ontologies are built following common design practices including shared top-level terms, relations annotations properties, and design patterns to ensure interoperability. Thus, when available, we leveraged appropriate terminology from these existing ontologies to maximize interoperability with external projects that also reuse those ontologies. However, not all spatiotemporal and physicochemical concepts needed for the annotation of marine metagenomic data were available from relevant OBO ontologies. Thus, within PMO we include additional terminology [e.g., "*latitude coordinate measurement datum*" "*start*" and "*stop*" (PMO:00000076) and (PMO:00000079), respectively] to more comprehensively annotate the datasets encapsulated within Planet Microbe Frictionless Data Packages. In addition, PMO includes terminology added to ENVO as a result of this work. See "Materials and Methods" sections on ENVO contributions, as well as **Figure 1B** for an example of an individual parameter within a Planet Microbe Data Package annotated with an ENVO chemical concentration term.

Limits of Semantic Harmonization

Although this work, when possible, included annotation information about the measurement devices used to collect data, this is not always clear from the source data repositories. Future data submission frameworks might consider mandating or more strongly encouraging the inclusion of information about measurement devices as part of the data submission process. This could be facilitated by automated protocols by instrument producers to capture structured and semantically enhanced data. Clearly annotating these types of differences in measurement devices would enable future systems and users to make decisions about what data can be compared based on the methodology used. Even if the data are about the same measurement or object (e.g., "chlorophyll a concentration"), the process by which these measurements were collected may result in non-comparable data. Furthermore, it should be noted that the precision of a concept being annotated is also important. It is more difficult to compare measurements annotated with a more generic concept such as "*concentration of chlorophyll in liquid water*" (ENVO:3100036), than data annotated with a more precise concept such as "*concentration of chlorophyll a in liquid water*" (ENVO:3100008). While the latter is specific enough to refer to a particular molecule, the former could be used to refer more generally to any mixture of chlorophyll compounds. For example, the "CHLPIG" parameter from some the Hawaiian Ocean Time Series (HOT) datasets is fluorometrically collected "chloropigment" data. As there are a variety of pigments associated with chlorophyll or "chloropigments," we annotated this data with the more general term "*concentration of chlorophyll in liquid water*" (ENVO:3100036). Depending on the wavelengths used in the analysis protocol this parameter could be measuring a different set of compounds than other projects also reporting chlorophyll data using fluorometric methods. In such cases reporting the data type, unit type, and measurement device may not be sufficient to make data intercomparable. Additional efforts to intercalibrate methods and link protocols, perhaps using systems like Protocols.io, are also required.

Addressing Biological Questions Using Harmonized Data

The second section of this paper addresses a series of biological queries of data integrated using the newly proposed specifications to demonstrate the efficacy of validating results against known biological and biogeochemical relationships and distributions.

Revisiting the Redfield Ratio

The Redfield ratio describes the stoichiometric carbon to nitrogen to phosphate ratio occurring in marine phytoplankton and is fundamental to our understanding of marine biogeochemistry (Tyrrell, 2019). Here we explore how the proposed specification method can automate the discovery of data by which to build upon existing hypotheses. Drawing upon the harmonized data and accompanying metadata provided by the Planet Microbe Data Packages, we selected 1076 samples with both nitrate and phosphate values (**Figures 2A–C**). Performing a linear regression on the selected data, we found the coefficient for the phosphate to nitrate ratio to be 0.0623. This is very close to 0.0625, the inverse of the 16:1 nitrate to phosphate ratio reported by Redfield (1934) and reconfirmed in many follow up studies (Takahashi et al., 1985; Anderson and Sarmiento, 1994).

The source data (following the MIxS checklists) contains accompanying metadata about the environmental context of the samples. Specifically, "*biome*" (ENVO:00000428), "*environmental feature*" (ENVO:00002297) and "*environmental material*" (ENVO:00010483) terms from ENVO. This metadata, which was cleaned and harmonized within the Planet Microbe Data Packages can be used to provide additional views into the data (**Figures 2A,B**). For example, there are two observations at approximately 37–40 μM nitrate and 0 μM phosphate that clearly deviate from the Redfield ratio. This can be explained by the ENVO environmental feature and biome annotations of these samples which were "*freshwater lake*" (ENVO:00000021) and "*freshwater lake biome*" (ENVO:01000252), respectively. This is consistent with the fact that, unlike pristine marine systems, freshwater systems (especially those experiencing anthropogenic influence) are not necessarily expected to follow the Redfield ratio (They et al., 2017).

[5]https://github.com/hurwitzlab/planet-microbe-ontology

[6]https://doi.org/10.5281/zenodo.4973944

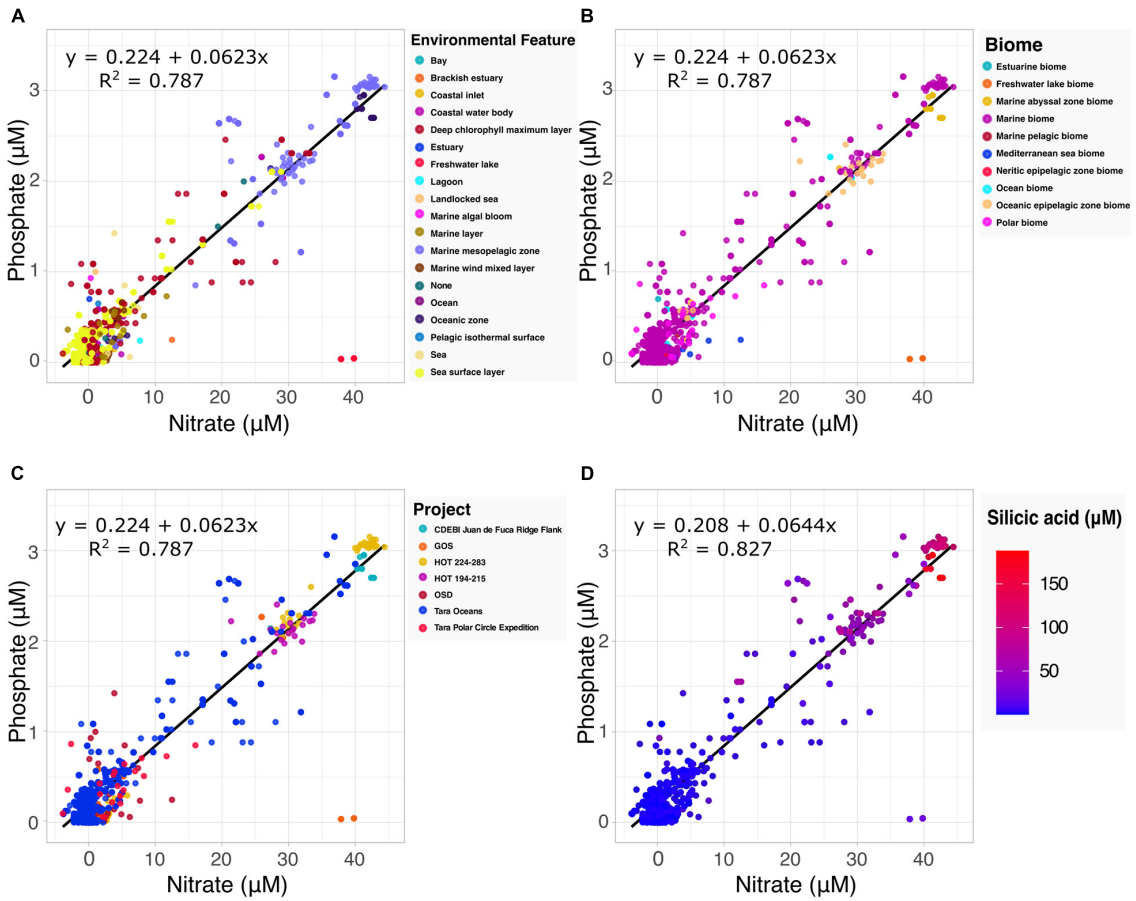

FIGURE 2 | Examination of the Redfield ratio, the relationship between phosphate and nitrate concentrations, faceted by the various harmonized metadata types included in the Planet Microbe Data Packages. **(A)** The Redfield ratio colored by the 18 ENVO environmental feature types [e.g., "*bay*" (ENVO:00000032)]. **(B)** The Redfield ratio colored by the 10 ENVO biome types [e.g., "*estuarine biome*" (ENVO:01000020)]. **(C)** The Redfield ratio colored by the project name for the seven projects that include both phosphate and nitrate data. **(D)** The correlation between silicic acid and the Redfield ratio for a subset of the data shown in **(A–C)** that also include silicate data. Linear equations modeling phosphate as a function of nitrate and R^2 are displayed in the panels. Nitrate, phosphate, and silicic acid are reported in micromolar concentrations.

Investigating the correlation between additional parameters and the Redfield ratio, we also searched for data with nitrate, phosphate, and silicic acid values within the Planet Microbe Data Packages. This narrowed the original dataset down from 1076 to 1063 samples. We used these 1063 samples, as well as the original 1076 samples with depth values to investigate the correlation between the Redfield ratio with silicic acid and depth. We found that silicic acid and depth have moderate positive correlations with the nitrate to phosphate ratio, with Spearman correlations of 0.420 and 0.413, respectively (p-values less than 2.2e-16). The individual Spearman correlations between phosphate and nitrate with silicic acid are higher, being 0.677 and 0.695, respectively (p-values less than 2.2e-16). See **Figure 2D** showing that the observed silicic acid concentration varies proportionally with increased nitrate and phosphate concentrations. This is consistent with the fact that diatoms, which require silica for their cell walls, can be an important source of sinking particulate organic matter and nutrient export.

Latitudinal Variation Across Environments and Physicochemical Gradients

We then leveraged the newly harmonized spatiotemporal and physicochemical variables, in combination with the descriptions of environment types provided with sequence data (see **Figure 3**). To showcase the discoveries that can result from the integration of the aforementioned data types, we asked the question "What variations in physicochemical parameters do we observe across environmental types and climate zones?" To examine this question, we searched the Planet Microbe Data Packages for data which included several physicochemical parameters including temperature, pH, and oxygen and nitrate concentrations. We analyzed 995 samples with temperature oxygen and nitrate values (**Figure 3**), and 488 samples with pH values (**Supplementary Figure 2**).

Results show that as expected, temperature increases in the "*deep chlorophyll maximum layer*" (ENVO:01000326) and

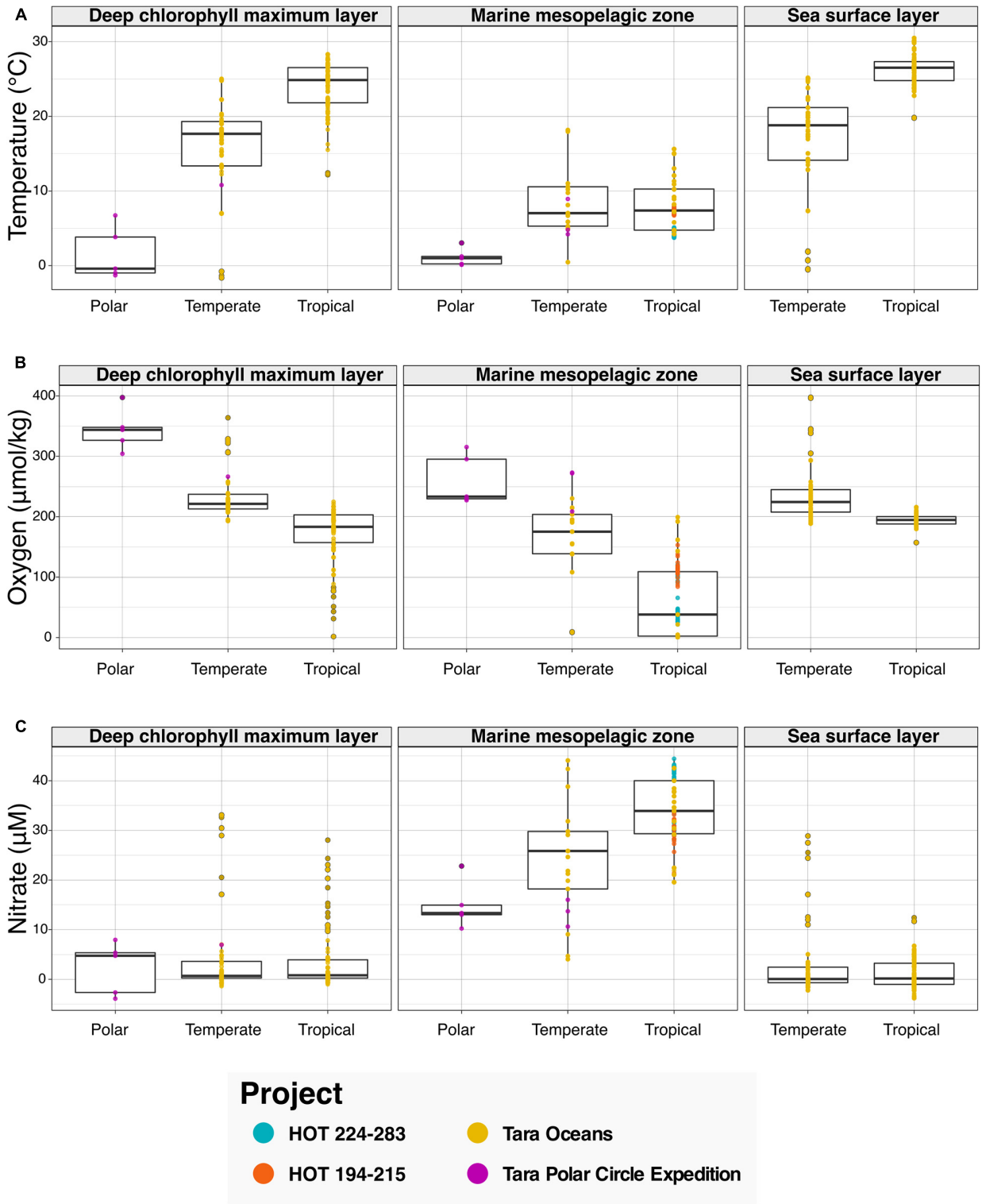

FIGURE 3 | Bar graphs showing physicochemical parameters faceted by annotation with various ENVO "*environmental feature*" (ENVO:00002297) terms serving as descriptors of environments. Data in this analysis are derived from various HOT and Tara projects that are annotated with the following ENVO terms: "*deep chlorophyll maximum layer*" (ENVO:01000326), "*marine mesopelagic zone*" (ENVO:00000213), and "*sea surface layer*" (ENVO:01001581). Additionally, the data are binned by latitude values into major climatic conditions "*polar*" (ENVO:01000238), "*temperate*" (ENVO:01000206), and "*tropical*" (ENVO:01000204). **(A–C)** The ENVO parameters: "*temperature of water*" (ENVO:09200014), "*concentration of dioxygen in liquid water*" (ENVO:3100011), and "*concentration of nitrate in liquid water*" (ENVO:3100022) in units of degree Celsius, and micromolar, respectively.

"*sea surface layer*" (ENVO:01001581) environments between polar, temperate, and tropical zones (**Figure 3A**). Additionally, there is not a significant difference between the temperature of temperate and tropical zones in the "*marine mesopelagic zone*" (ENVO:00000213). This result is consistent with the description of the mesopelagic zone being the uppermost region of a pelagic aphotic zone, bounded at the thermocline temperature transition zone (del Giorgio and Duarte, 2002)[7]. Although these results regarding the variation in temperature across environment types and climatic zones are not new per say, they help to sanity check that this system is working correctly by providing expected results.

In addition to temperature, other physicochemical factors such as oxygen and nitrate have been used to identify and differentiate distinct marine environments (Sayre et al., 2017; Sutton et al., 2017). Examining the results for oxygen, we see decreases in oxygen concentrations between polar, temperate, and tropical zones in all environment types (**Figure 3B**). It is notable that although the dissolved oxygen values for the "*deep chlorophyll maximum layer*" (ENVO:01000326) and "*sea surface layer*" (ENVO:01001581) are similar across temperate and tropical zones, there is a large ∼150 μM difference between the averages of temperate and tropical samples oxygen values in the analyzed "*marine mesopelagic zone*" (ENVO:00000213) samples. Many of the low oxygen mesopelagic tropical samples are sourced from 500 to 1,000 meter depths from the subtropical oceans near Hawai'i. This is consistent with the known oxygen minimum found at ∼800 m at Station ALOHA (Bingham and Lukas, 1996), as well as previous findings that low oxygen water masses are common in midwater depth range of 500–1,500 m in subtropical regions, due to the microbial remineralization of sinking organic matter (Jürgens and Taylor, 2018).

Analyzing the results for nitrate, we observe that the average of nitrate values in the "*marine mesopelagic zone*" (ENVO:00000213) are quite different from the other environment types (**Figure 3C**). We observe a similar pattern with pH values in tropical samples being much lower in the "*marine mesopelagic zone*" (ENVO:00000213) than in the "*sea surface layer*" (ENVO:01001581) (**Supplementary Figure 2**). It is notable that for the environmental descriptor "*marine mesopelagic zone*" (ENVO:00000213), we observe differences in the values of three physiological parameters: temperature, nitrate, and pH in tropical samples relative to the other ENVO environment types. This serves as an example of the expert information about specific concepts captured within an ontology (e.g., environment types in ENVO) being differentiable based on observed patterns in real world data.

Harmonization Across Time Series

Finally, we explored the benefits derived from data harmonization across time series. The Hawaiian Ocean Time Series (HOT) study is one of the longest running open ocean time-series surveys spanning more than 30 years of data (Karl and Church, 2014). Resulting from this work's data harmonization we were able to combine existing data from two separate HOT metagenomic projects one that used pyrosequencing (Bryant et al., 2016), and a second that leveraged Illumina sequencing (Mende et al., 2017) into a new dataset comprising 52 metagenomic samples with 19 common physicochemical variables. Shown in **Supplementary Figure 3** are the correlation coefficients between the 19 variables. Nine of the 19 variables have significant moderate correlations with depth (*p*-values less than 0.01 and Spearman correlation coefficients greater than 0.4). This along with the fact that depth is an important factor regulating the distribution of microbial species within marine systems, we performed a taxonomic analysis of those samples to investigate what microbial species vary most with depth in the North Pacific Subtropical Gyre (**Figure 4**).

The results of the dbRDA analysis shown in **Figure 4**, can be interpreted as follows, the coefficients indicate the extent to which the abundance of any given microbial species correlates with depth. Positive correlation values indicate species that increase in relative abundance with depth. Negative values indicate species that are anti-correlated with depth, i.e., species whose relative abundance decreases with depth. Near-zero values indicate no correlation with depth. This analysis shows that in the North Pacific Subtropical Gyre, of the top twenty correlations of species with depth, the five strongest, non-zero correlations are Cyanobacteria and Thaumarchaeota species. Whereas the majority of non-zero correlations are from the phylum Proteobacteria.

Our results showed that the Cyanobacteria related to *Prochlorococcus* sp. MIT 0801, and Thaumarchaeota most closely affiliated with *Candidatus Nitrosopelagicus brevis*, were more abundant with increased depth. The former is consistent with the findings reported by Biller et al. (2014), where the authors performed a phylogenetic analysis of *Prochlorococcus* strains which they paired with physiological studies on isolate representatives from a subset of the species in the analysis. The authors' results determined that the MIT 0801 strain is in a low light adapted *Prochlorococcus* subclade, thus validating our result that the MIT 0801 strain is correlated with depth as it prefers lower light levels found deeper in the water column (positive correlation value in **Figure 4**).

Regarding the Thaumarchaeota *Candidatus Nitrosopelagicus brevis*, our results show this archaeon to be most prevalent in upper water column samples from ∼125 m depth. Thaumarchaeota in general, and specifically *Candidatus Nitrosopelagicus brevis,* have been shown to be abundant microbes below ocean surface waters (Santoro et al., 2015). Furthermore, prior studies have shown that Thaumarchaeota begin to be detectable right around the deep chlorophyll maximum at ∼125 m at Station ALOHA (Mincer et al., 2007; Lincoln et al., 2014). Our results build upon these findings adding that *Candidatus Nitrosopelagicus brevis* are also likely low light adapted as they are most abundant at greater depths following the same pattern as the known low light adapted *Prochlorococcus* sp. MIT 0801 strain.

All three species most anti-correlated with depth are *Prochlorococcus* strains including *Prochlorococcus* sp. RS50, *Prochlorococcus* sp. MIT 0604, and *Prochlorococcus marinus*. The aforementioned work of Biller et al. (2014) classified the *Prochlorococcus* sp. MIT 0604 as belonging to a high light adapted

[7]https://oceanservice.noaa.gov/facts/thermocline.html

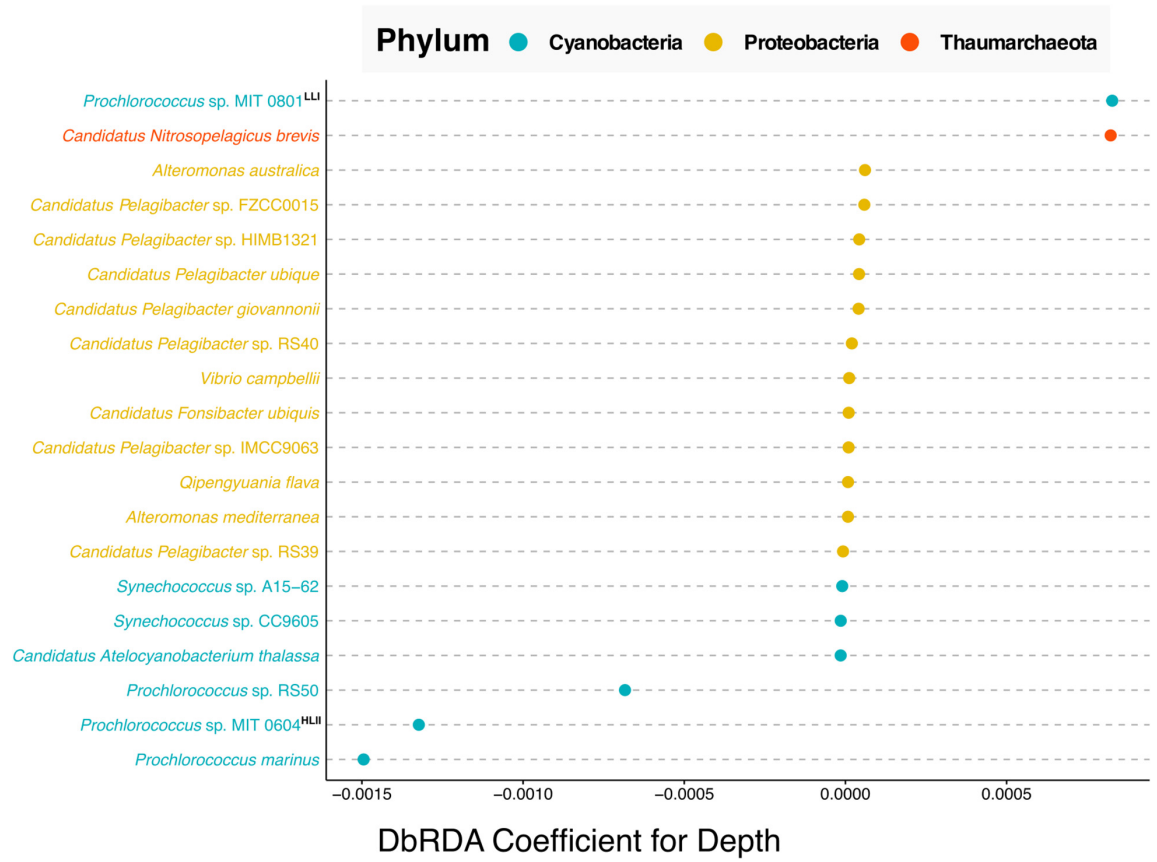

FIGURE 4 | Cleveland's dot plot showing the top 20 coefficients ordered by absolute value, derived from a distance-based redundancy analysis (dbRDA) of the relative abundance of bacterial and archaeal species against depth. Species are colored by taxonomic phylum with Cyanobacteria in blue, Proteobacteria in yellow and Thaumarchaeota in orange. LLI and HLII indicate "high light adapted" and "low light adapted," respectively based on the analysis of Biller et al. (2014). The depth of samples used in this analysis ranged between 25 and 125 m.

clade, which is congruent with our results showing MIT 0604 to be more abundant at shallower depths.

Our results about the *Prochlorococcus* sp. MIT 0604 and *Prochlorococcus* sp. MIT 0801 ecotype depth distributions are consistent with a newly proposed *Prochlorococcus* taxonomy, in which the authors identified five *Prochlorococcus* genera with distinct ecological attributes (Tschoeke et al., 2020). According to the proposed model, the *Eurycoliumgenus* genus, including *Prochlorococcus* sp. MIT 0604 are associated with high temperature and oligotrophic environments, whereas the genera *Prolificoccus* including *Prochlorococcus* sp. MIT 0801 are most abundant at low temperatures (Walter et al., 2017; Tschoeke et al., 2020). This is supported in our results by a −0.570 spearman correlation between depth and temperature, see **Supplementary Figure 3**, hence the noted similarity in *Prochlorococcus* temperature distribution patterns.

Regarding *Prochlorococcus marinus*, it is known that at the strain level there are a variety of both high and low light adapted ecotypes (Rocap et al., 2003; Tschoeke et al., 2020). Although our analysis was only able to determine the *Prochlorococcus marinus* at a species level, it is notable that the *Prochlorococcus marinus* sequences detected were the most strongly anti-correlated with depth. This indicates that high light adapted *Prochlorococcus marinus* ecotypes are more abundant than low light adapted ecotypes in the North Pacific Subtropical Gyre, consistent with prior oceanographic studies (Johnson et al., 2006; Thompson et al., 2018).

A final result derived from this analysis concerns the depth distribution of *Prochlorococcus* sp. RS50. The RS50 genome was assembled in 2017 based on surface samples collected from the Red Sea in 2014 but has not been published upon.[8] Currently very little is known about distribution and abundance of the *Prochlorococcus* sp. RS50 strain. Our results show that RS50 is one of the top three species most anti-correlated with depth, thus indicating that it is also high light adapted. This result demonstrates the utility of integrating and harmonizing metagenomic datasets to enable new discoveries from meta-analyses reusing previously published data. As updates are made to taxonomic databases (e.g., the inclusion of the RS50 genome), this will enable new results to be derived from reanalyzing published datasets together in meta-analyses.

[8]https://www.ncbi.nlm.nih.gov/biosample/SAMN06061931

MATERIALS AND METHODS

Creating Frictionless Data Packages

We implemented the new OBO-Frictionless specification to create Frictionless Data Packages in Planet Microbe. These OBO-Frictionless Data Packages were created as follows, all data resources from a given project including resources linking to metagenomic datasets as well as constituent environmental contextual information was encapsulated within a master Frictionless Data Package JSON file. Inconsistencies in MIxS environmental contextual fields (e.g., "broad-scale environmental context") were manually corrected. We also added new dataset columns containing links to the corresponding machine readable ENVO PURLs, e.g., "http://purl.obolibrary.org/obo/ENVO_01000253" for "freshwater river biome." Data were standardized by a triad of OBO Foundry ontology terms to enable Interoperability of datasets: by (1) standardizing the data type annotation, (2) specifying machine-readable unit types that can be automatically converted into standardized or desired units programmatically, and (3) enabling users to further select which data should be compared in meta-analyses based on the measurement devices from which the data was collected. Additional dataset provenance information such as links to sources and usage licenses, as well as missing values, data types and data value constraints are also specified with the Frictionless Data Package JSON files. The Planet Microbe Frictionless Data Packages were validated using both the Good Tables python library and command line tool (Heughebaert, 2020; see text footnote 2), as well as custom scripts available from the planet microbe scripts repository.[9] Planet Microbe Data Packages are available from the following repository (see text footnote 1). The protocol for generating Planet Microbe Data Packages is available from[10].

Environment Ontology Term Contributions

To standardize the representation of the large variety of physicochemical data attributes collected in accompaniment to the marine metagenomic datasets aggregated in Planet Microbe, new terminology was required. Addressing this issue in this work, we created a new ENVO module for chemical concentrations to create terms such as "*concentration of chlorophyll a in liquid water*" (ENVO:3100008). These new ENVO chemical concentration terms were created using a Dead Simple Ontology Design Pattern (DOSDP) (Osumi-Sutherland et al., 2017), in which input data in a tabular format is converted and compiled into Web Ontology Language (OWL) code. These new chemical concentration terms represent information about a chemical solvent, and material solute to machine systems by including an OWL equivalence axiom which links to terms from both from the Chemical Entities of Biological Interest (CHEBI) CHEBI "*chemical entity*" (CHEBI:24431) hierarchy for the former, and the ENVO "*environmental material*" (ENVO:00010483) hierarchy for the latter. For example, the equivalence class from the new ENVO term "*concentration of chlorophyll a in liquid water*" (ENVO:3100008) links it to both CHEBI "*chlorophyll a*" (CHEBI:18230) as well as ENVO "*liquid water*" (ENVO:00002006) in a machine searchable way. In addition, the equivalence axioms used in the new ENVO chemical concentration terms also include a linkage to the Phenotype and Trait Ontology (PATO) term "*concentration of*" (PATO:0000033). A total of 31 new chemical concentration terms were added to the new ENVO chemical concentration module[11] as a result of this work.

Measurement Device Annotation

In addition to semantic and unit types, we included the annotation of dataset parameters measurement device terms when possible. The following are examples of measurement device annotations that were included within the Planet Microbe Data Packages. Within the various HOT time series studies data about the concentration of chlorophyll a is reported twice using separate measurement devices. In one reported parameter the concentration of chlorophyll a is measured via a "*high performance liquid chromatography instrument*" (OBI:0001057) and in the other the data is measured using a "*fluorometer*" (OBI:0400143). Another example concerns organismal or cell count data, it is important to know if data were collected via flow cytometry vs. microscopy. Although it was possible to decipher various examples of cell counts which were measured using a "*flow cytometer*" (OBI:0400044) from the various Hawaiian Ocean Time Series projects as well as the Amazon continuum datasets. In contrast the single assignment to "microscope" (OBI:0400169) was made to the Ocean Sample Day (OSD) parameter: "Nanoplankton and microplankton aggregate" based on information in the OSD handbook.[12] An example of a source ambiguity of employed measurement devices comes from the OSD project. Although OSD is well documented and used the microb3 vocabulary to annotate parameter types as well as their units, information about the measurement devices could only be found in the "Description" comments. Ambiguities in the handbook's descriptions e.g., "Concentration of pigments (e.g., chlorophyll a) extracted and analyzed by fluorometry or HPLC" precluded the assignment of precise measurement type semantics.

Whole Genome Sequencing Taxonomic Analysis

Analysis of Whole Genome Sequencing (WGS) data was conducted as follows, after downloading the fastq files deposited in the Short Read Archive (SRA) for the selected samples, a quality filtering and removal of human sequences was performed using fastqc v0.11.9 and trimGalore v0.6.6 with default parameters. Quality-filtered sequences were screened to

[9]https://github.com/hurwitzlab/planet-microbe-scripts

[10]http://dx.doi.org/10.17504/protocols.io.bzsdp6a6

[11]https://github.com/EnvironmentOntology/envo/blob/master/src/envo/modules/chemical_concentration.csv

[12]https://www.microb3.eu/sites/default/files/osd/OSD_HandbooK_2016.pdf

remove human sequences using bowtie2 v2.4.2 against a non-redundant version of the Genome Reference Consortium Human Build 38, patch release 7.[13] After quality control and human read filtering, metagenomes containing less than 10 million paired-end reads were discarded. Taxonomic profiling of the metagenomic samples was performed using the k-mer-based taxonomic classification software Kraken2 (Wood et al., 2019). Finally, the Kraken2 taxonomic abundances were reassigned to more specific taxonomic ranks using Braken (Lu et al., 2017). Briefly Kraken2 v2.1.1 was run on the paired read using the PlusPF database[14] and Bracken v2.6.1 was run on the Kraken2 outputs. The code used for this analysis is available from the following repository.[15]

Addressing Biological Questions

Data used to address various biological questions was collected using the Planet Microbe search interface[16] and data download feature. Linear regression for Redfield ratios (**Figure 2**) was conducted using the "lm" R package and plotted using the "ggplot 2" package. Box plots for physicochemical variables (**Figure 3** and **Supplementary Figure 2**) were plotted using the "ggplot 2" R package. The correlation heatmap (**Supplementary Figure 3**) was created using code sourced from the following repository,[17] which leveraged the "corrplot" R library, using Spearman correlations as the method. Individual Spearman correlations of physicochemical variables to depth were done using the R "cor" package. Taxonomic profile data, generated by the WGS pipeline, was analyzed in R using the "phyloseq" and "microbiome" packages. A list of contaminant species was manually selected for removal prior to analysis. Shallow samples (under 100 k reads) were removed prior to normalization by relative abundance. Distance-based redundancy analysis (dbRDA) of relative taxonomic abundance against depth was performed using the "vegan" R package's adonis method (Anderson, 2001), using Bray-Curtis dissimilarities. DbRDA results were plotted in a Cleveland's dot plot (**Figure 4**) using the ggdotchart method from the "ggpubr" R package. Final versions of figures were edited with Inkscape.[18]

CONCLUSION

Motivated by the long-term vision of harmonizing marine 'omics and environmental data, in data products that are available to people and machines, we devised and implemented cyberinfrastructure specifications using OBO ontologies and Frictionless Data Packages to make data FAIR within the Planet Microbe web-portal. This new specification allows for marine 'omic and contextual environmental data to be (1) exposed within machine searches, (2) be completely transferable between CI systems, (3) have mechanisms for automated data validation, (4) use common vocabularies for measurement types, devices and units, as well as (5) enable discovery of individual attributes from datasets based on their vocabulary annotations.

Furthermore, we leveraged this system to discover data with given oceanographic measurements, features, and thresholds to synthesize and analyze global datasets in novel ways. This work promotes a new understanding of the infrastructure and data coordination requirements for performing global ocean analyses at unprecedented spatial and temporal resolution, including the distribution of microbes and responses to environmental drivers. Taken together, Frictionless Data Packages in Planet Microbe provide a much-needed resource for uniting 'omics data and associated data products from diverse ocean surveys with environmental data to allow the geosciences community to include federated oceanographic data in global models and analyses.

Finally, it should be noted that the cyberinfrastructure specifications presented here are not alone sufficient to ensure the reusability of existing and future global marine sampling projects. Importantly, future efforts will be needed to harmonize standard methods and protocols for dataset intercalibration.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://github.com/hurwitzlab/planet-microbe-frictionless-data-package-paper.

AUTHOR CONTRIBUTIONS

KB wrote the manuscript, contributed to ontology development, annotated, assembled the Data Packages, and performed the statistical analyses. AP performed the WGS taxonomic analyses. MB developed the Planet Microbe scripts for Data Package validation and database integration. KB, AP, and MB developed the Planet Microbe Data Package specifications. AP, EW-C, ED, and BH contributed to the revision and preparation of the manuscript. BH, EW-C, and ED designed the project and scope of research. BH directed the project. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the NSF (OCE-1639614 Planet Microbe to BH, ED, EW-C), NSF (CISE-1640775 Ocean Cloud Commons to BH, in part), Simons Foundation SCOPE (ID 329108 to ED), Gordon and Betty Moore Foundation (GBMF 8751 to BH), Gordon and Betty Moore Foundation (GBMF 3777 to ED), Simons Foundation (329108 to ED), and Simons Foundation (721223 to ED).

[13] https://genome-idx.s3.amazonaws.com/bt/GRCh38_noalt_as.zip

[14] https://benlangmead.github.io/aws-indexes/k2

[15] https://github.com/aponsero/readbased_metagenomes_snakemake

[16] https://www.planetmicrobe.org/#/search

[17] https://community.rstudio.com/t/correction-in-correlation-method-in-rquery-cormat/73031

[18] https://inkscape.org/

ACKNOWLEDGMENTS

We thank all members of the Environment Ontology development team, especially Chris Mungall and Pier L. Buttigieg for helping with ontology contributions. We would also like to thank Adam Shepherd and Danie Kinkade from the Biological and Chemical Oceanography Data Management Office (BCO-DMO) for their technical and scientific expertise and support. We would also like to thank Ramona Walls for her expertise and support. Finally, we thank all members of the Hurwitz Lab for their insights and discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.765268/full#supplementary-material

REFERENCES

Anderson, L. A., and Sarmiento, J. L. (1994). Redfield ratios of remineralization determined by nutrient data analysis. *Glob. Biogeochem. Cyc.* 8, 65–80. doi: 10.1029/93GB03318

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral. Ecol.* 26, 32–46. doi: 10.1111/j.1442-9993.2001.01070.pp.x

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Bandrowski, A., Brinkman, R., Brochhausen, M., Brush, M. H., Bug, B., Chibucos, M. C., et al. (2016). The ontology for biomedical investigations. *PLoS One* 11:e0154556. doi: 10.1371/journal.pone.0154556

Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D., Patel-Schneider, P., et al. (2004). *OWL Web Ontology Language Reference*. Available Online at: https://www.w3.org/TR/owl-ref/ [Accessed Apr 13, 2021].

Biller, S. J., Berube, P. M., Berta-Thompson, J. W., Kelly, L., Roggensack, S. E., Awad, L., et al. (2014). Genomes of diverse isolates of the marine cyanobacterium *Prochlorococcus*. *Sci. Data* 1:140034. doi: 10.1038/sdata.2014.34

Biller, S. J., Berube, P. M., Dooley, K., Williams, M., Satinsky, B. M., Hackl, T., et al. (2018). Marine microbial metagenomes sampled across space and time. *Sci. Data* 5:180176. doi: 10.1038/sdata.2018.176

Bingham, F. M., and Lukas, R. (1996). Seasonal cycles of temperature, salinity and dissolved oxygen observed in the Hawaii Ocean Time-series. *Deep Sea Res. Part II Top. Stud. Oceanogr.* 43, 199–213. doi: 10.1016/0967-0645(95)00090-9

Brown, S. A., Folk, M., Goucher, G., Rew, R., and Dubois, P. F. (1993). Software for Portable scientific data management. *Comput. Phys.* 7, 304–308. doi: 10.1063/1.4823180

Bryant, J. A., Aylward, F. O., Eppley, J. M., Karl, D. M., Church, M. J., and DeLong, E. F. (2016). Wind and sunlight shape microbial diversity in surface waters of the North Pacific Subtropical Gyre. *ISME J.* 10, 1308–1322. doi: 10.1038/ismej.2015.221

Buttigieg, P., Morrison, N., Smith, B., Mungall, C. J., Lewis, S. E., and The Envo Consortium (2013). The environment ontology: contextualising biological and biomedical entities. *J. Biomed. Semant.* 4:43. doi: 10.1186/2041-1480-4-43

Buttigieg, P. L., Pafilis, E., Lewis, S. E., Schildhauer, M. P., Walls, R. L., and Mungall, C. J. (2016). The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation. *J. Biomed. Semant.* 7:57. doi: 10.1186/s13326-016-0097-6

Chen, Y., Shepherd, A., Chandler, C., Arko, R., and Leadbetter, A. (2014). Ontology based vocabulary matching for oceanographic instruments. *EGU Gen. Assemb. Conf. Abstr.* 16:12909.

Coburn, E., Maffei, A. R., Chandler, C. L., and Raymond, L. M. (2012). Controlled vocabularies and ontologies for oceanographic data: the R2R eventlogger project. *AGU Fall Meet. Abstr.* 51:1699.

del Giorgio, P. A., and Duarte, C. M. (2002). Respiration in the open ocean. *Nature* 420, 379–384. doi: 10.1038/nature01165

Eloe-Fadrosh, E. A., Ahmed, F., Anubhav, Babinski, M., Baumes, J., Borkum, M., et al. (2021). The national microbiome data collaborative data portal: an integrated multi-omics microbiome data resource. *Nucleic Acids Res.* doi: 10.1093/nar/gkab990

FAIRsharing Team (2015). *FAIRsharing record for: Quantities, Units, Dimensions and Types*. Oxford: FAIRsharing. doi: 10.25504/FAIRSHARING.D3PQW7

Field, D., Friedberg, I., Sterk, P., Kottmann, R., Glöckner, F. O., Hirschman, L., et al. (2009). Meeting report: "Metagenomics, Metadata and Meta-analysis" (M3) special interest group at ISMB 2009. *Stand. Genomic Sci.* 1, 278–282. doi: 10.4056/sigs.641096

Gkoutos, G. V., Schofield, P. N., and Hoehndorf, R. (2012). The Units Ontology: a tool for integrating units of measurement in science. *Database* 2012:bas033. doi: 10.1093/database/bas033

Heughebaert, A. (2020). *Frictionless Data and Darwin Core*. Available Online at: https://data-blog.gbif.org/post/frictionless-data-and-darwin-core/ [Accessed July 9, 2021].

Johnson, Z. I., Zinser, E. R., Coe, A., McNulty, N. P., Woodward, E. M. S., and Chisholm, S. W. (2006). Niche partitioning among *prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* 311, 1737–1740. doi: 10.1126/science.1118052

Jürgens, K., and Taylor, G. (2018). *Microbial Ecology of the Oceans*. Hoboken: Wiley.

Karl, D. M., and Church, M. J. (2014). Microbial oceanography and the Hawaii Ocean Time-series programme. *Nat. Rev. Microbiol.* 12, 699–713. doi: 10.1038/nrmicro3333

Karl, D. M., and Lukas, R. (1996). The Hawaii Ocean Time-series (HOT) program: background, rationale and field implementation. *Deep Sea Res. Part II Top. Stud. Oceanogr.* 43, 129–156. doi: 10.1016/0967-0645(96)00005-7

Karsch-Mizrachi, I., Takagi, T., and Cochrane, G. (2018). The international nucleotide sequence database collaboration. *Nucleic Acids Res.* 46, D48–D51. doi: 10.1093/nar/gkx1097

Kazakov, Y., Krötzsch, M., and Simanèík, F. (2014). The Incredible ELK. *J. Autom. Reason.* 53, 1–61. doi: 10.1007/s10817-013-9296-3

Kyrpides, N. C. (1999). Genomes OnLine Database (GOLD 1.0): a monitor of complete and ongoing genome projects world-wide. *Bioinformatics* 15, 773–774. doi: 10.1093/bioinformatics/15.9.773

Lincoln, S. A., Wai, B., Eppley, J. M., Church, M. J., Summons, R. E., and DeLong, E. F. (2014). Planktonic Euryarchaeota are a significant source of archaeal tetraether lipids in the ocean. *Proc. Natl. Acad. Sci. U. S. A.* 111, 9858–9863. doi: 10.1073/pnas.1409439111

Lu, J., Breitwieser, F. P., Thielen, P., and Salzberg, S. L. (2017). Bracken: estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* 3:e104. doi: 10.7717/peerj-cs.104

Madin, J., Bowers, S., Schildhauer, M., Krivov, S., Pennington, D., and Villa, F. (2007). An ontology for describing and synthesizing ecological observation data. *Ecol. Inform.* 2, 279–296. doi: 10.1016/j.ecoinf.2007.05.004

Markowitz, V. M., Ivanova, N. N., Szeto, E., Palaniappan, K., Chu, K., Dalevi, D., et al. (2007). IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res.* 36, D534–D538. doi: 10.1093/nar/gkm869

Matentzoglu, N. (2021). *INCATools/ontology-development-kit: June 2020 Release*. Zenodo. doi: 10.5281/zenodo.4973944

McCreary, D. (2010). *Semantic Integration Patterns*. Available Online at: https://www.slideshare.net/dmccreary/semantic-integrationpatterns [Accessed Apr 13, 2021].

Mende, D. R., Bryant, J. A., Aylward, F. O., Eppley, J. M., Nielsen, T., Karl, D. M., et al. (2017). Environmental drivers of a microbial genomic transition zone in the ocean's interior. *Nat. Microbiol.* 2, 1367–1373. doi: 10.1038/s41564-017-0008-3

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., et al. (2008). The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9:386. doi: 10.1186/1471-2105-9-386

Mincer, T. J., Church, M. J., Taylor, L. T., Preston, C., Karl, D. M., and DeLong, E. F. (2007). Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre. *Environ. Microbiol.* 9, 1162–1175. doi: 10.1111/j.1462-2920.2007.01239.x

Mitchell, A., Bucchini, F., Cochrane, G., Denise, H., ten Hoopen, P., Fraser, M., et al. (2016). EBI metagenomics in 2016 - an expanding and evolving resource for the analysis and archiving of metagenomic data. *Nucleic Acids Res.* 44, D595–D603. doi: 10.1093/nar/gkv1195

Mitchell, A. L., Scheremetjew, M., Denise, H., Potter, S., Tarkowska, A., Qureshi, M., et al. (2018). EBI Metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies. *Nucleic Acids Res.* 46, D726–D735. doi: 10.1093/nar/gkx967

Moncoiffe, G., and Kokkinaki, A. (2018). The NERC Vocabulary Server (NVS) as a shared resource. *EGU Gen. Assemb.* 20:15770.

Mukherjee, S., Stamatis, D., Bertsch, J., Ovchinnikova, G., Katta, H. Y., Mojica, A., et al. (2019). Genomes OnLine database (GOLD) v.7: updates and new features. *Nucleic Acids Res.* 47, D649–D659. doi: 10.1093/nar/gky977

Muñiz Fernandez, F., Carreño Torres, A., Morcillo-Suarez, C., and Navarro, A. (2012). "Application of array-oriented scientific data formats (NetCDF) to genotype data, GWASpi as an example," in *Bioinformatics for Personalized Medicine Lecture Notes in Computer Science*, eds A. T. Freitas and A. Navarro (Berlin: Springer), 8–20. doi: 10.1007/978-3-642-28062-7_2

Nakamura, Y., Cochrane, G., Karsch-Mizrachi, I., and International Nucleotide Sequence Database Collaboration (2013). The International Nucleotide Sequence Database Collaboration. *Nucleic Acids Res.* 41, D21–D24. doi: 10.1093/nar/gks1084

Osumi-Sutherland, D., Courtot, M., Balhoff, J. P., and Mungall, C. (2017). Dead simple OWL design patterns. *J. Biomed. Semant.* 8:18. doi: 10.1186/s13326-017-0126-0

Ponsero, A. J., Bomhoff, M., Blumberg, K., Youens-Clark, K., Herz, N. M., Wood-Charlson, E. M., et al. (2021). Planet Microbe: a platform for marine microbiology to discover and analyze interconnected 'omics and environmental data. *Nucleic Acids Res.* 49, D792–D802. doi: 10.1093/nar/gkaa637

Prud'hommeaux, E., and Seaborne, A. (2008). *SPARQL Query Language for RDF*. Available Online at: https://www.w3.org/TR/rdf-sparql-query/ [Accessed Apr 20, 2021].

Redfield, A. (1934). "On the proportions of organic derivatives in sea water and their relation to the composition of plankton," in *James Johnstone Memorial Volume*, ed R. J. Daniel (Liverpool: University Press of Liverpool), 176–192.

Rew, R., and Davis, G. (1990). NetCDF: an interface for scientific data access. *IEEE Comput. Graph. Appl.* 10, 76–82. doi: 10.1109/38.56302

Rijgersberg, H., Wigham, M., and Top, J. L. (2011). How semantics can improve engineering processes: a case of units of measure and quantities. *Adv. Eng. Inform.* 25, 276–287. doi: 10.1016/j.aei.2010.07.008

Rocap, G., Larimer, F. W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N. A., et al. (2003). Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* 424, 1042–1047. doi: 10.1038/nature01947

Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooseph, S., et al. (2007). The Sorcerer II global ocean sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol.* 5:e77. doi: 10.1371/journal.pbio.0050077

Santoro, A. E., Dupont, C. L., Richter, R. A., Craig, M. T., Carini, P., McIlvin, M. R., et al. (2015). Genomic and proteomic characterization of " *Candidatus* Nitrosopelagicus brevis": an ammonia-oxidizing archaeon from the open ocean. *Proc. Natl. Acad. Sci. U. S. A.* 112, 1173–1178. doi: 10.1073/pnas.1416223112

Sayre, R., Dangermond, J., Wright, D., Breyer, S., Butler, K., Van Graafeiland, K., et al. (2017). *A New Map of Global Ecological Marine Units-an Environmental Stratification Approach*. Washington: Association of American Geographers.

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., et al. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.* 25, 1251–1255. doi: 10.1038/nbt1346

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., et al. (2015). Structure and function of the global ocean microbiome. *Science* 348, 1261359–1261359. doi: 10.1126/science.1261359

Sutton, T. T., Clark, M. R., Dunn, D. C., Halpin, P. N., Rogers, A. D., Guinotte, J., et al. (2017). A global biogeographic classification of the mesopelagic zone. *Deep Sea Res. Part Oceanogr. Res. Pap.* 126, 85–102. doi: 10.1016/j.dsr.2017.05.006

Takahashi, T., Broecker, W. S., and Langer, S. (1985). Redfield ratio based on chemical data from isopycnal surfaces. *J. Geophys. Res. Oceans* 90, 6907–6924. doi: 10.1029/JC090iC04p06907

They, N. H., Amado, A. M., and Cotner, J. B. (2017). Redfield ratios in Inland Waters: higher biological control of C:N:P Ratios in tropical semi-arid high water residence time lakes. *Front. Microbiol.* 8:1505. doi: 10.3389/fmicb.2017.01505

Thompson, A. W., van den Engh, G., Ahlgren, N. A., Kouba, K., Ward, S., Wilson, S. T., et al. (2018). Dynamics of *Prochlorococcus* diversity and photoacclimation during short-term shifts in water column stratification at station ALOHA. *Front. Mar. Sci.* 5:488. doi: 10.3389/fmars.2018.00488

Tschoeke, D., Salazar, V. W., Vidal, L., Campeão, M., Swings, J., Thompson, F., et al. (2020). Unlocking the genomic taxonomy of the *Prochlorococcus* collective. *Microb. Ecol.* 80, 546–558. doi: 10.1007/s00248-020-01526-5

Tyrrell, T. (2019). "Redfield ratio," in *Encyclopedia of Ocean Sciences (Third Edition)*, eds J. K. Cochran, H. J. Bokuniewicz, and P. L. Yager (Oxford: Academic Press), 461–472. doi: 10.1016/B978-0-12-409548-9.11281-3

Vangay, P., Burgin, J., Johnston, A., Beck, K. L., Berrios, D. C., Blumberg, K., et al. (2021). Microbiome metadata standards: report of the national microbiome data collaborative's workshop and follow-on activities. *mSystems* 6, e01194–20. doi: 10.1128/mSystems.01194-20

Walls, R. L., Deck, J., Guralnick, R., Baskauf, S., Beaman, R., Blum, S., et al. (2014). Semantics in support of biodiversity knowledge discovery: an introduction to the biological collections ontology and related ontologies. *PLoS One* 9:e89606. doi: 10.1371/journal.pone.0089606

Walter, J. M., Coutinho, F. H., Dutilh, B. E., Swings, J., Thompson, F. L., and Thompson, C. C. (2017). Ecogenomics and taxonomy of cyanobacteria phylum. *Front. Microbiol.* 8:2132. doi: 10.3389/fmicb.2017.02132

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., 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

Wilkinson, M. D., Verborgh, R., da Silva Santos, L. O. B., Clark, T., Swertz, M. A., Kelpin, F. D. L. et al. (2017). Interoperability and FAIRness through a novel combination of Web technologies. *PeerJ Comput. Sci.* 3:e110. doi: 10.7717/peerj-cs.110

Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20:257. doi: 10.1186/s13059-019-1891-0

Yilmaz, P., Gilbert, J. A., Knight, R., Amaral-Zettler, L., Karsch-Mizrachi, I., Cochrane, G., et al. (2011a). The genomic standards consortium: bringing standards to life for microbial ecology. *ISME J.* 5, 1565–1567. doi: 10.1038/ismej.2011.39

Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J. R., Amaral-Zettler, L., et al. (2011b). 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

Zheng, X., Dai, X., and Huang, L. (2016). Spatial variations of prokaryotic communities in surface water from india ocean to chinese marginal seas and their underlining environmental determinants. *Front. Mar. Sci.* 3:17. doi: 10.3389/fmars.2016.00017

Zinger, L., Amaral-Zettler, L. A., Fuhrman, J. A., Horner-Devine, M. C., Huse, S. M., Welch, D. B. M., et al. (2011). Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. *PLoS One* 6:e24570. doi: 10.1371/journal.pone.0024570

Conflict of Interest: BH holds concurrent appointments as an Associate Professor of Biosystems Engineering at the University Arizona and as an Amazon Scholar. This publication describes work performed at the University Arizona and is not associated with Amazon.

The remaining 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.

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.

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 07 September 2022
DOI 10.3389/fmicb.2022.999925

OPEN ACCESS

EDITED BY
Kelly D. Goodwin,
National Oceanic and Atmospheric Administration (NOAA), United States

REVIEWED BY
Nastassia Patin,
National Oceanic and Atmospheric Administration (NOAA), United States
Chloe Victoria Robinson,
Ocean Wise, Canada

*CORRESPONDENCE
Matthias Wietz
matthias.wietz@awi.de

SPECIALTY SECTION
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 21 July 2022
ACCEPTED 15 August 2022
PUBLISHED 07 September 2022

CITATION
Wietz M, Metfies K, Bienhold C, Wolf C, Janssen F, Salter I and Boetius A (2022) Impact of preservation method and storage period on ribosomal metabarcoding of marine microbes: Implications for remote automated samplings.
Front. Microbiol. 13:999925.
doi: 10.3389/fmicb.2022.999925

Impact of preservation method and storage period on ribosomal metabarcoding of marine microbes: Implications for remote automated samplings

Matthias Wietz[1,2]*, Katja Metfies[3,4], Christina Bienhold[1,2], Christian Wolf[3], Felix Janssen[1,2], Ian Salter[1,5] and Antje Boetius[1,2,6]

[1]Deep-Sea Ecology and Technology, Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany, [2]Max Planck Institute for Marine Microbiology, Bremen, Germany, [3]Polar Biological Oceanography, Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany, [4]Helmholtz Institute for Functional Marine Biodiversity at the University of Oldenburg, Oldenburg, Germany, [5]Faroe Marine Research Institute, Torshavn, Faroe Islands, [6]MARUM Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany

Automated sampling technologies can enhance the temporal and spatial resolution of marine microbial observations, particularly in remote and inaccessible areas. A critical aspect of automated microbiome sampling is the preservation of nucleic acids over long-term autosampler deployments. Understanding the impact of preservation method on microbial metabarcoding is essential for implementing genomic observatories into existing infrastructure, and for establishing best practices for the regional and global synthesis of data. The present study evaluates the effect of two preservatives commonly used in autosampler deployments (mercuric chloride and formalin) and two extraction kits (PowerWater and NucleoSpin) on amplicon sequencing of 16S and 18S rRNA gene over 50 weeks of sample storage. Our results suggest the combination of mercuric chloride preservation and PowerWater extraction as most adequate for 16S and 18S rRNA gene amplicon-sequencing from the same seawater sample. This approach provides consistent information on species richness, diversity and community composition in comparison to control samples (nonfixed, filtered and frozen) when stored up to 50 weeks at *in situ* temperature. Preservation affects the recovery of certain taxa, with specific OTUs becoming overrepresented (SAR11 and diatoms) or underrepresented (*Colwellia* and pico-eukaryotes) after preservation. In case eukaryotic sequence information is the sole target, formalin preservation and NucleoSpin extraction performed best. Our study contributes to the design of long-term autonomous microbial observations in remote ocean areas, allowing cross-comparison of microbiome dynamics across sampling devices (e.g., water and particle samplers) and marine realms.

KEYWORDS

seawater microbiome, sample preservation, DNA extraction, amplicon sequencing, 16S rRNA, 18S rRNA, autonomous sampling, time-series

Introduction

Microbial communities have fundamental ecological and biogeochemical roles in nutrient recycling and carbon sequestration (Jørgensen and Boetius, 2007; Fuhrman et al., 2015). Understanding the consequences of global change for marine ecosystems requires a robust assessment of microbial community dynamics over temporal and spatial scales (Sunagawa et al., 2015; Buttigieg et al., 2018). Automated sampling devices attached to observational platforms, e.g., ocean moorings, enable time-series observations of microbial dynamics (Herfort et al., 2016; Zhang et al., 2019, 2021). Typically combined with physical and chemical sensors, automated samplers are of particular value in remote and inaccessible areas, such as seasonally ice-covered environments (Liu et al., 2020; von Appen et al., 2021; Wietz et al., 2021; Ramondenc et al., 2022). For instance, autonomous sediment traps allow linking particle flux with microbial diversity over extended periods, based on microscopic counts (Salter et al., 2007, 2012, 2014; Nöthig et al., 2020; Zúñiga et al., 2021) and DNA sequencing (Metfies et al., 2017; Bachy et al., 2022; Valencia et al., 2022).

There is a growing toolbox and increasing application of automated water and particle sampling approaches (Supplementary Table S1). As *in situ* molecular analysis is still an emerging technology (Moore et al., 2021) and beyond the resource capacity of many observing programs, automated samplers mostly perform *in situ* preservation of sample material (Yamahara et al., 2019; Lindsay, 2021; Truelove et al., 2022). *In situ* preservation intends to minimize signal modification over the extended duration of device deployment and laboratory processing. Formalin and mercuric chloride are commonly used to preserve sinking particles in long-term monitoring programs (Lee et al., 1992; UNESCO-IOC, 1994; Bauerfeind et al., 2009; Lampitt et al., 2010; Fischer et al., 2016). Although these chemicals originally aimed to preserve tissues, particles and cells for bulk biogeochemical analyses, recent studies have demonstrated that downstream molecular analyses are feasible with both mercuric chloride- (Metfies et al., 2017) and formalin-fixed (Boeuf et al., 2019) sediment trap samples. Likewise, preservation with mercuric chloride (Liu et al., 2020; Wietz et al., 2021) and formalin (Stern et al., 2015) allows ribosomal metabarcoding of microbes in autonomously collected seawater. Also the nucleic acid stabilizers RNAlater and DNAgard can preserve environmental DNA (Gray et al., 2013; Rachel and Gieg, 2020), however requiring frozen storage in stabilizer solution or the concentration of microbial biomass on filters (Ottesen et al., 2011). Both reagents have been tested as preservative in automated microbial samplings (Boeuf et al., 2019; Formel et al., 2021; Poff et al., 2021), but can lead to DNA loss (Renshaw et al., 2015) and are likely unsuitable in remote regions where samples cannot be frozen immediately. Hence, although automated technologies – in particular comparative sampling across different regions – offer exciting perspectives, preservation method and storage time are challenging factors for microbial diversity studies (Sherr and Sherr, 1993; Rissanen et al., 2010; Metfies et al., 2017; Spens et al., 2017; Sano et al., 2020; Pratte and Kellogg, 2021).

In the present study, we examined how preservation and DNA extraction methods affect molecular microbial analyses after long-term storage of seawater samples. Specifically, we addressed DNA yields, PCR amplification efficiency and microbiome composition after sample storage for 10, 28, and 50 weeks (0°C) to mimic long-term autosampler deployments. The approach was chosen to match deployment conditions of autonomous samplers in polar waters, which are installed on moorings and typically serviced only once per year (e.g., von Appen et al., 2021). We evaluate the consistency of 16S and 18S rRNA sequence information obtained from samples after different periods of post-sampling storage. We focus on formalin and mercuric chloride as they are widely used preservatives (Supplementary Table S1) and functionally different, particularly with respect to long-term storage at *in situ* temperatures. Furthermore, we aimed to assess how results from freshly preserved samples align with those from legacy samples, and indeed allow decadal-scale characterization of ecosystem dynamics. Our results have implications for microbial time-series collected with automated samplers, both regarding short-term methodological aspects and long-term archiving of biodiversity information.

Materials and methods

Experimental design and sampling regime

Approx. 6 l of surface seawater were collected at the pier on Helgoland Island in the German Bight (54° 10′ 58.3″N, 7° 53′ 19.9″E) on March 30, 2017. The water sample was kept at 4°C in the dark for ~35 days, then well mixed and split into 40 ml subsamples. Five subsamples were directly filtered as reference. The following preservatives were added to four sets of five replicate subsamples: (i) saturated mercuric chloride ($HgCl_2$) solution (0.15% *w*/*v* final concentration per sample), (ii) 20% formalin (1.8% *v*/*v* final concentration per sample), (iii) RNAlater (1% final concentration per sample), and (iv) DNAgard (1% final concentration per sample). Preserved 40 ml subsamples were stored in the dark at 0°C to mimic conditions during high-latitude mooring deployments. After 10, 28, and 50 weeks, respectively (hereafter referred to as 10w, 28w, 50w), five replicates per preservation method were subjected to DNA extraction with two different kits after filtering each 20 ml onto Isopore membrane filters (Millipore, Burlington, MA, United States; 0.2 µm pore size, 47 mm diameter). Filters were stored frozen at −20°C for the same amount of time until DNA extraction with the NucleoSpin II (NS; Macherey-Nagel, Germany) or PowerWater (PW; QIAGEN, Germany) kit following the manufacturers' protocols. Filters from formalin-preserved samples were subjected to additional rinsing steps before DNA extraction following Bucklin and Allen (2004). DNA extracts were quantified using a Nanodrop 1000 photometer (Thermo Fisher Scientific, Germany) and stored frozen until library preparation.

Amplicon sequencing

Libraries were prepared according to the standard instructions of the 16S Metagenomic Sequencing Library Preparation protocol (Illumina, San Diego, CA, United States). The V4 region of eukaryotic 18S rRNA genes was amplified using PCR primers 528F (5′-GCGGTAATTCCAGCTCCAA-3′; Elwood et al., 1985) and 964iR (5′-ACTTTCGTTCTTGATYRR-3′; Balzano et al., 2015). The V4-5 region of bacterial and archaeal 16S rRNA genes was amplified using primers 515F (5′-GTGYCAGCMG CCGCGGTAA-3′) and 926R (5′-CCGYCAATTYMTTTRAGT TT-3′; Parada et al., 2016). All PCRs had a final volume of 25 µl and contained 12.5 µl KAPA HiFi HotStart ReadyMix (Roche, Basel, Switzerland), 2.5 µl of each primer (1 µM) and 2.5 µl template. Amplification included initial denaturation (95°C, 3 min) followed by 25 cycles of denaturation (95°C, 30 s), annealing (55°C, 30 s), and extension (72°C, 30 s) with a single final extension (72°C, 5 min). 18S rRNA PCR products were gel-purified using the AMPure XP PCR purification kit (Beckman Coulter, Pasadena, CA, United States) according to the manufacturer's protocol. All PCR products were quantified using a Quantus Fluorometer (Promega, Madison, WI, United States). Indices and sequencing adapters were attached *via* PCRs (final volume 50 µl), each containing 25 µl of KAPA HiFi HotStart ReadyMix (Roche), 5 µl of each Nextera XT Index Primer [1 µmol/l], 5 µl template (~5 ng DNA in total) and 10 µl PCR grade water. Amplification included initial denaturation (95°C, 3 min) followed by 8 cycles of denaturation (95°C, 30 s), annealing (55°C, 30 s), and extension (72°C, 30 s) with a single final extension (72°C, 5 min). 18S rRNA libraries were gel-purified using the AMPure XP PCR purification kit (Beckman Coulter). All libraries were quantified using a Quantus fluorometer (Promega) and sequenced using MiSeq and the MiSeq Reagent Kit V3 (2 × 300 bp) according to the manufacturer's protocol (Illumina).

Processing and analysis of amplicon reads

Reads were processed using Trimmomatic v0.38 (Bolger et al., 2014) by scanning each sequence from the 5′ to 3′ end, trimming the 3′ end if average Phred Q-score of < 8 in a sliding window of 3 bp. Paired ends were merged using VSEARCH v2.3.0 (Rognes et al., 2016), discarding pairs with <50 bp overlap and > 5 mismatches in the overlapping segment. To guarantee identical orientation, sequences were filtered so forward sequences occur before reverse complement sequences. If sequences did not match this pattern, their reverse complement was also scanned using cutadapt v1.17 (Martin, 2011), requiring minimum overlaps of 17 and 13 bp for forward and reverse primer sequences respectively, and only one mismatch. Primer sequences were truncated, and sequences feature-filtered using VSEARCH. Sequences were discarded if (i) < 300 bp or > 550 bp, (ii) containing ambiguous bases (assigned as RYSWKMBDHVN per IUPAC nomenclature), or (iii) having an expected error (sum of all base error probabilities) > 0.25. Each sample was independently dereplicated, and the abundance of each sequence added to the sequence header. Chimeras were sample-wise predicted *de novo* by VSEARCH with default settings and removed. Subsequently, only samples with at least 10,000 sequences were used. Cleaned sample files were pooled and dereplicated in total, keeping amplicon abundances in the sequence headers. The pooled file was used as input for OTU clustering with SWARM v2.2.2 (Mahé et al., 2014), using the most abundant amplicon of an OTU as representative for annotation. Sequences were annotated with the default classifier implemented in mothur v1.38.1 using the Protist Ribosomal database v4.11.1 (Guillou et al., 2013) and the Silva v132 database (Quast et al., 2013) for 18S and 16S rRNA amplicons respectively, with a confidence cut-off of 80. One representative sequence was used to annotate the full OTU cluster, discarding singletons as well as OTUs with < 0.005% relative abundance. Statistical evaluation was carried out with R v.4.1.1 in RStudio using packages phyloseq, ampvis2, iNEXT, vegan, ape, tidyverse and scico (McMurdie and Holmes, 2013; Oksanen et al., 2013; Hsieh et al., 2016; Andersen et al., 2018; Paradis and Schliep, 2019; Wickham et al., 2019; Crameri, 2021). As our 16S rRNA dataset contained almost no archaeal sequences, 16S results are only referred to as "bacteria". Relative abundances were Hellinger-transformed (the square root of the relative abundance per OTU and sample), an ecologically relevant transformation to correct for the compositionality of amplicon sequence data (Legendre and Gallagher, 2001).

Preliminary sequence analyses showed that only $HgCl_2$ and formalin performed well in our experimental design (Supplementary Figure S1). The nucleic acid stabilizers RNAlater and DNAgard were originally tested, since being used in some automated sampling approaches (Supplementary Table S1). However, as nucleic acid stabilizers are not designed for long-term sample storage without freezing, we omitted results from RNAlater and DNAgard from further analysis.

Data and code availability

The entire workflow from raw sequence processing to statistical evaluation is available at https://github.com/matthiaswietz/MicroPreserve. Sequence data have been deposited in the European Nucleotide Archive (ENA) under accession number PRJEB43307, using the data brokerage service of the German Federation for Biological Data (GFBio) in compliance with MIxS standards (Yilmaz et al., 2011).

Results and discussion

We evaluated microbial community composition in seawater samples following two different preservation methods, based on poisoning ($HgCl_2$) and fixation by protein cross-links (formalin). The concentrations of $HgCl_2$ and formalin, common preservatives to study water column biogeochemistry and microbiology, were at the higher end of the range typically used, aiming at the observation of the strongest preservative effect expected. $HgCl_2$

(0.15% *w*/*v*) and formalin (1.8% *v*/*v*) concentrations correspond to those used in particle traps (Bauerfeind et al., 2009; Lampitt et al., 2010). $HgCl_2$ concentrations in water sampler deployments can be tenfold lower (von Appen et al., 2021; Wietz et al., 2021) as biomass in seawater is commonly lower compared to particles.

DNA yields and PCR amplification

Preservation with $HgCl_2$ resulted in a higher proportion of successful DNA extractions compared to formalin (Table 1; Supplementary Figure S2) and approx. tenfold higher yields, despite pre-treatment of formalin-preserved samples (Bucklin and Allen, 2004). DNA-protein cross-linking through formalin may explain lower success rates and DNA yields. For both formalin and $HgCl_2$, approximately twofold higher DNA yields were observed with PW extraction, likely corresponding to the combined bead-beating and enzymatic lysis compared to only chemical lysis with NS extraction (Yuan et al., 2015). In general, preservation decreased DNA yields two to fourfold compared to non-preserved controls. The impact of preservatives on DNA yields was observed at the earliest experimental time-point (10w), with no significant further decreases over the experimental period (Supplementary Figure S2). Hence, the chemical effect of preservatives is the major determinant of DNA yields, without further impact of prolonged storage, at least for up to 50 weeks. Independent of extraction kit, PCR amplification failures were ~50% for formalin compared to <10% for $HgCl_2$ (Table 1). Although formalin-preserved samples allowed DNA extraction and amplification in several cases, our results hence support that formalin can impede downstream molecular analyses (Hoffman et al., 2015; Reid et al., 2017).

TABLE 1 DNA yields and successful PCRs after preservation in comparison to the unpreserved reference, when extracted with either PowerWater (PW) or NucleoSpin (NS).

	Extractions with detectable DNA yield/ sample number	DNA yield [ng µl^{-1}]	Successful PCRs (16S/18S rRNA)
Reference; PowerWater	5/5	0.85 ± 0.2	5/5
Reference; NucleoSpin	5/5	0.2 ± 0.1	5/5
Mercuric chloride; PowerWater	15/15	0.24 ± 0.17	15/14
Mercuric chloride; NucleoSpin	13/15	0.03 ± 0.03	12/15
Formalin, PowerWater	5/15	0.2 ± 0.08	5/2
Formalin; NucleoSpin	11/15	0.01 ± 0.02	9/14

In some cases, PCR was successful despite NanoDrop did not detect DNA, probably related to detection sensitivity of the instrument. Reference: directly filtered environmental sample without preservation, immediately frozen at −20°C.

Microbial community composition

We obtained a mean of 34,000 and 62,000 chimera-filtered 16S and 18S rRNA amplicon reads, respectively (Supplementary Table S2). Principal coordinates analysis revealed clear clustering of both eukaryotic and bacterial communities by preservation method (Figure 1; PERMANOVA, $p < 0.01$), with little effect of storage time

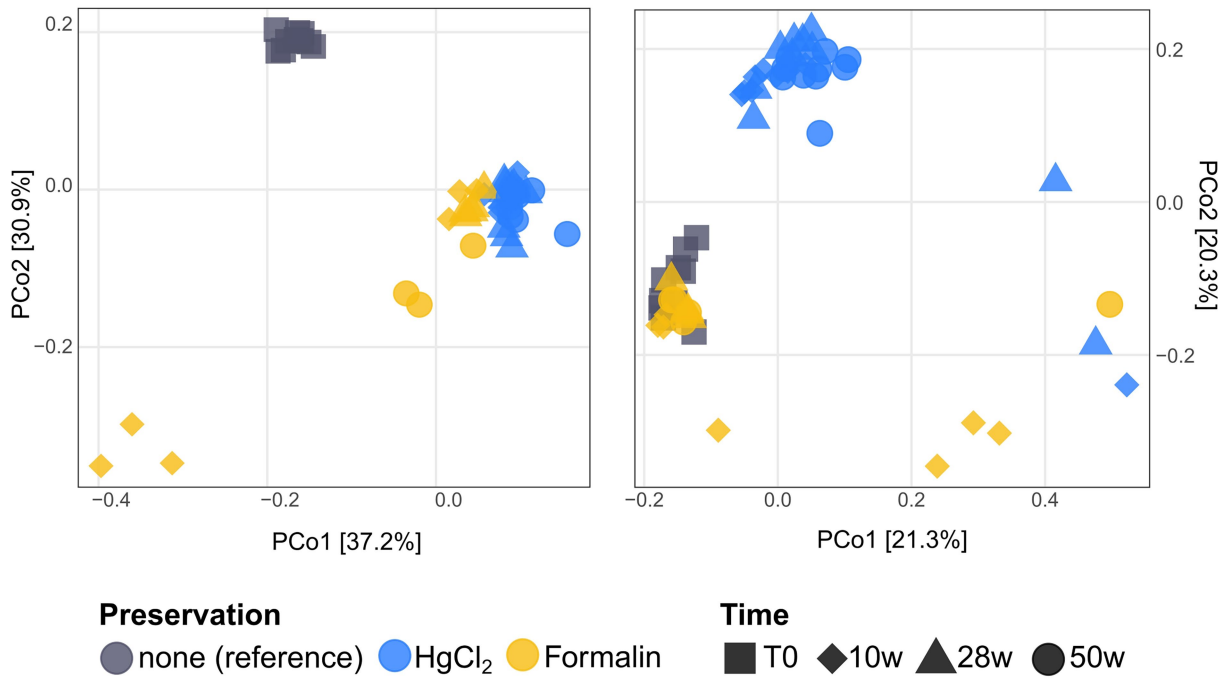

FIGURE 1
Principal Coordinates Analysis of Bray–Curtis dissimilarities of bacterial (left) and eukaryotic (right) communities after preservation and storage over different intervals. DNA extraction kits are not visually separated, as the influence of fixation significantly outweighs that of extraction.

or extraction kit. Hence, in line with DNA extraction and PCR results, preservation method is the major determinant of ribosomal metabarcoding results under the specific microbial community and storage conditions tested in this study.

Bacterial communities

Preservation significantly influenced bacterial community composition compared to the unpreserved reference (PERMANOVA, $p < 0.001$). However, differences to the unpreserved control were minor, with a taxonomic distance of ~0.2 particularly for $HgCl_2$ samples and little change over time. In addition to preservation method, an effect of the extraction kit was observed. While $HgCl_2$ + PW, $HgCl_2$ + NS and formalin + NS performed comparably for bacterial communities, communities obtained from formalin + PW clustered separately (Supplementary Figure S3). The inverse Simpson index, considering both evenness and richness to determine alpha-diversity, was elevated after $HgCl_2$ preservation (Figure 2; Kruskal–Wallis with Dunn's *post hoc* test, $p = 0.04$). This concurred with higher relative abundances of planctomycetes, Deltaproteobacteria, and Actinobacteria (Figure 3), indicating that preservation can overestimate the rare biosphere. Among the major classes, preservation influenced the representation of alphaproteobacterial and gammaproteobacterial abundances (Figure 3), mainly relating to SAR11 clade Ia (higher) and *Colwellia* (lower abundances) respectively (Figure 4A). These taxa are at the lower and higher size spectrum of pelagic marine bacteria, respectively (Bowman, 2014; Giovannoni, 2017), indicating that preservation might favor smaller bacterial cells. Alternatively, cell wall structure and glycosylation (Dadon-Pilosof et al., 2017) might influence preservation efficiency. Compositionality effects can amplify such observations, but can be alleviated by normalizing relative abundances (Legendre and Gallagher, 2001; Weiss et al., 2017). Indeed, Hellinger-transformed relative abundances

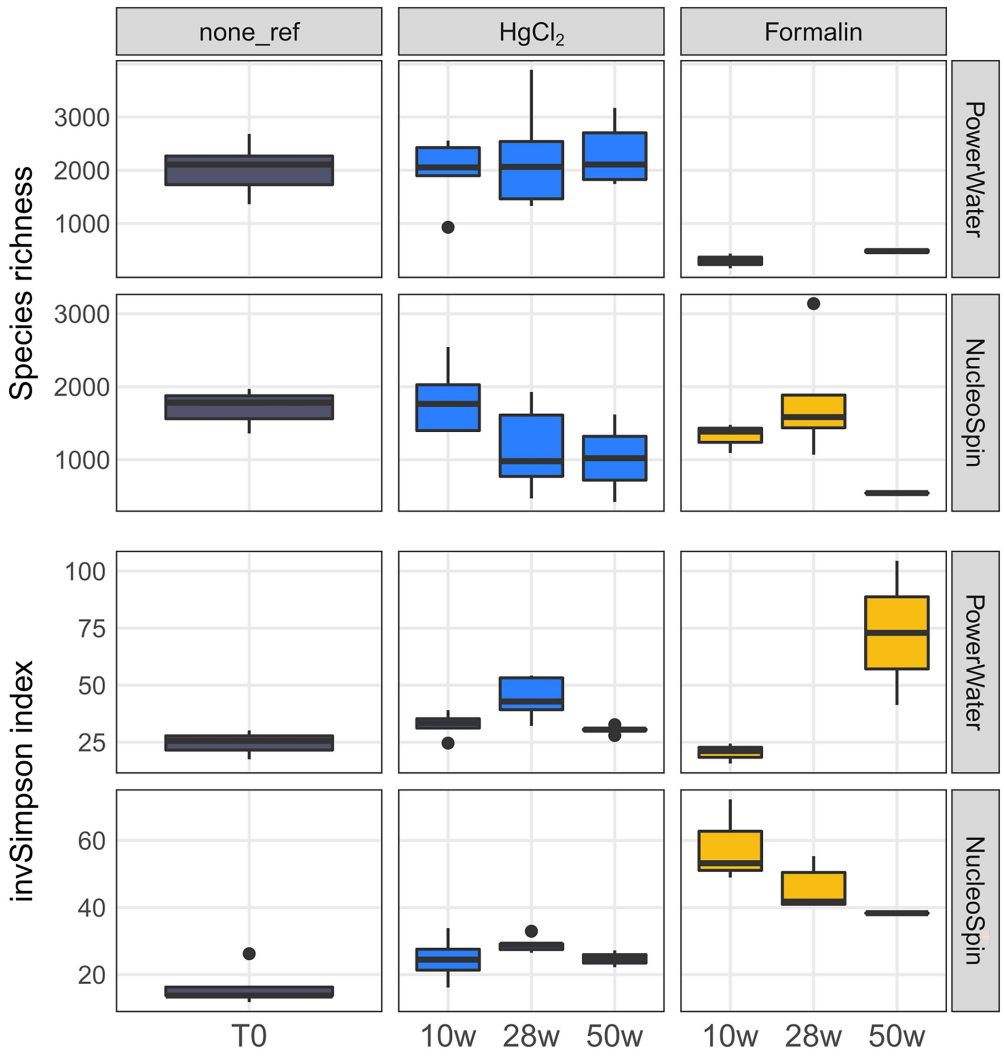

FIGURE 2
Bacterial species richness and inverse Simpson index by preservation, storage time, and DNA extraction. The number of samples per group is shown in Supplementary Table S2.

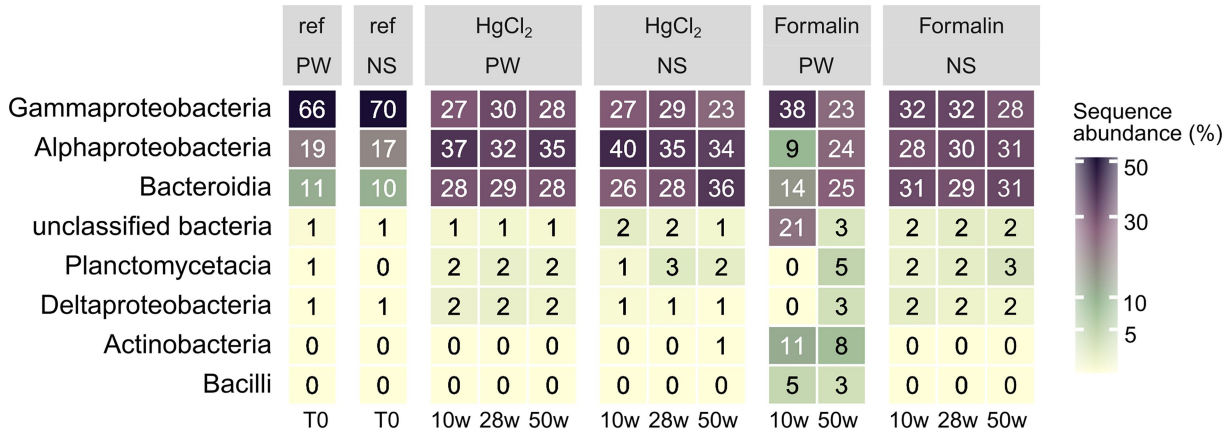

FIGURE 3
Relative abundances of major bacterial classes (average of all replicates per sampling event) by preservation, storage time, and DNA extraction. PW: PowerWater, NS: NucleoSpin.

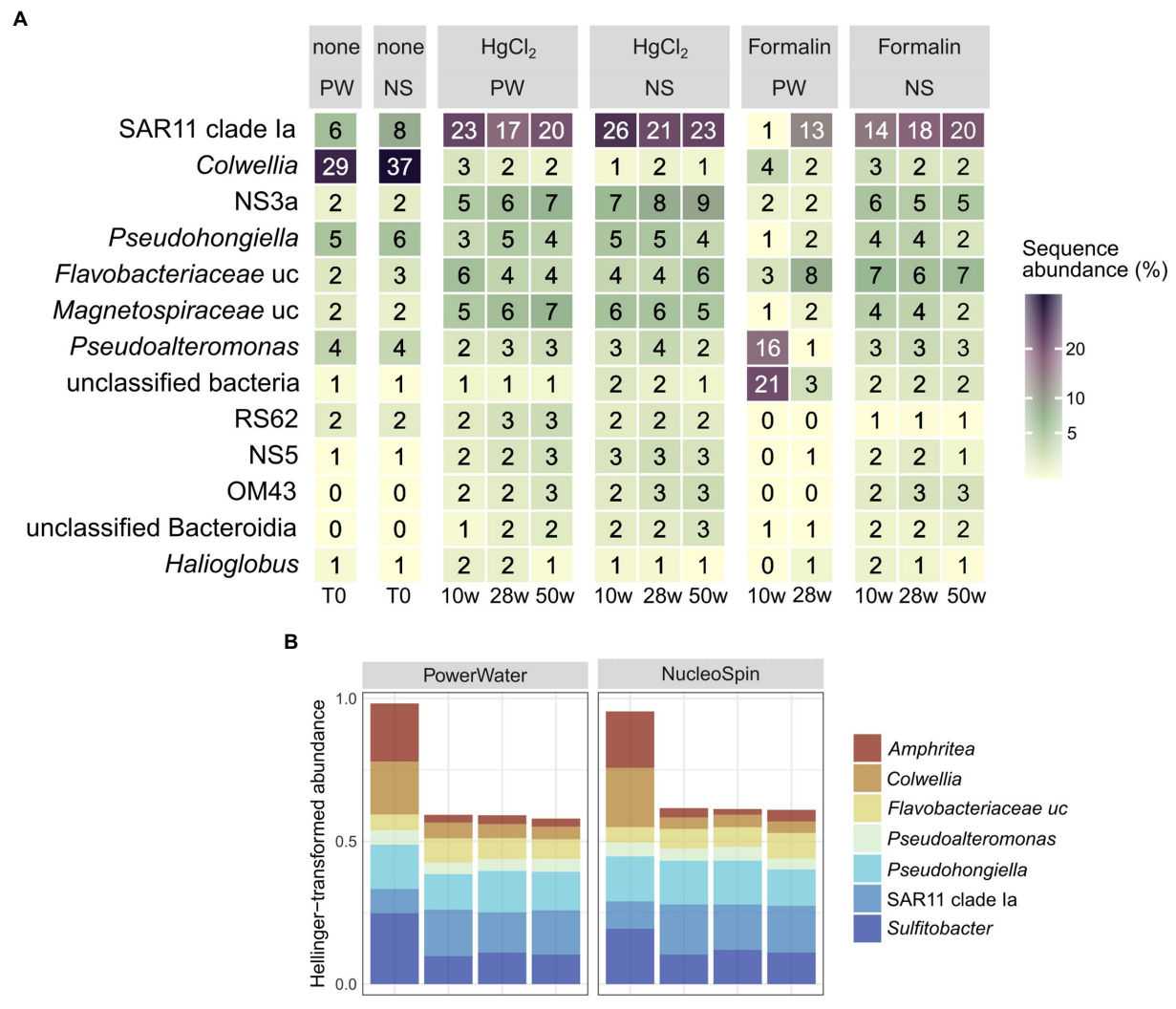

FIGURE 4
Relative abundances **(A)** and Hellinger-transformed relative abundances after $HgCl_2$ preservation **(B)** of major bacterial genera (average of all replicates per sampling event) in relation to preservation, storage time, and DNA extraction. PW: PowerWater, NS: NucleoSpin, uc: unclassified.

provided a more even picture of community structure (Figure 4B), with smaller differences for *Colwellia* while identifying highest variability *for Amphritea* (Gammaproteobacteria: Oceanospirillales). Previous studies have identified seasonal microbial dynamics in polar waters based on $HgCl_2$ + PW preserved, autonomously collected samples (Liu et al., 2020; Wietz et al., 2021). Our results indicate that detection of *Colwellia* in such samples (Wietz et al., 2021) represented a true ecological finding, supported by stable OTU numbers from Alpha- and Gammaproteobacteria in $HgCl_2$ + PW samples (Supplementary Figure S4).

Eukaryotic communities

For eukaryotes, PW extraction of formalin-preserved samples largely failed (Table 1). Hence, we restricted assessment of method performance to formalin + NS and $HgCl_2$ samples. Preservation significantly influenced eukaryotic community composition compared to the unpreserved reference (PERMANOVA, $p < 0.001$), albeit with minor differences to the unpreserved control (maximum taxonomic dissimilarities of ~0.3) comparable to bacteria. Formalin + NS, $HgCl_2$ + NS and $HgCl_2$ + PW performed similarly, providing comparable composition and diversity patterns compared to the reference (Figures 1, 5; Supplementary Figure S5). $HgCl_2$ and formalin resulted in higher proportions of *Bacillariophyta* (i.e., diatoms) in comparison to the unpreserved reference. In addition, *Filosa-Imbricatea* and unclassified stramenopiles were overrepresented in $HgCl_2$ + NS (Figure 6). The total number of OTUs detected within stramenopile groups was lower after both $HgCl_2$ and formalin preservation (Supplementary Figure S4), contributing to an overall lower species richness compared to the reference (Figure 5). The relative abundances of picoplankton classes Picozoa, MAST and Choanoflagellata were most similar between the reference and formalin + NS (Figure 6). As opposed to the overrepresentation of smaller bacterial cells, $HgCl_2$ preservation

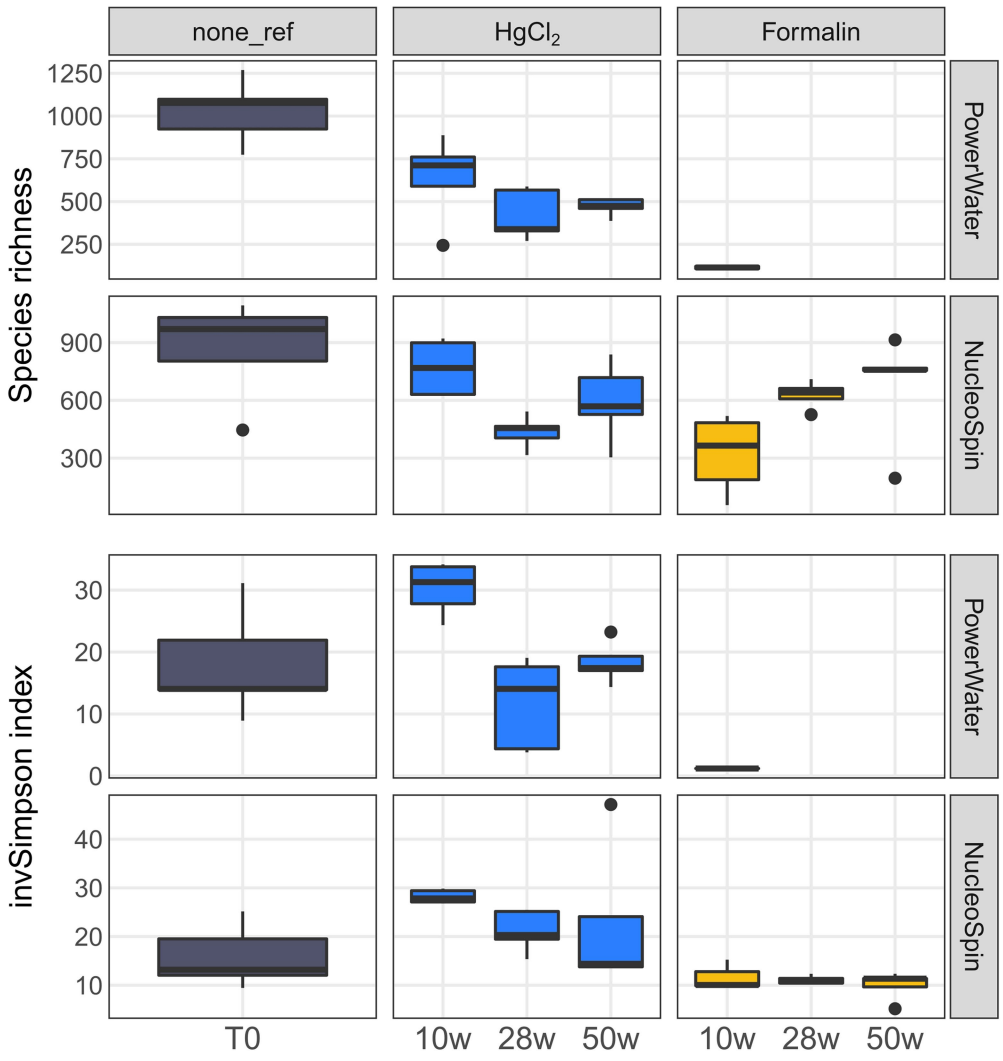

FIGURE 5
Eukaryotic species richness and inverse Simpson index by preservation, storage time, and DNA extraction. The number of samples per group is shown in Supplementary Table S2.

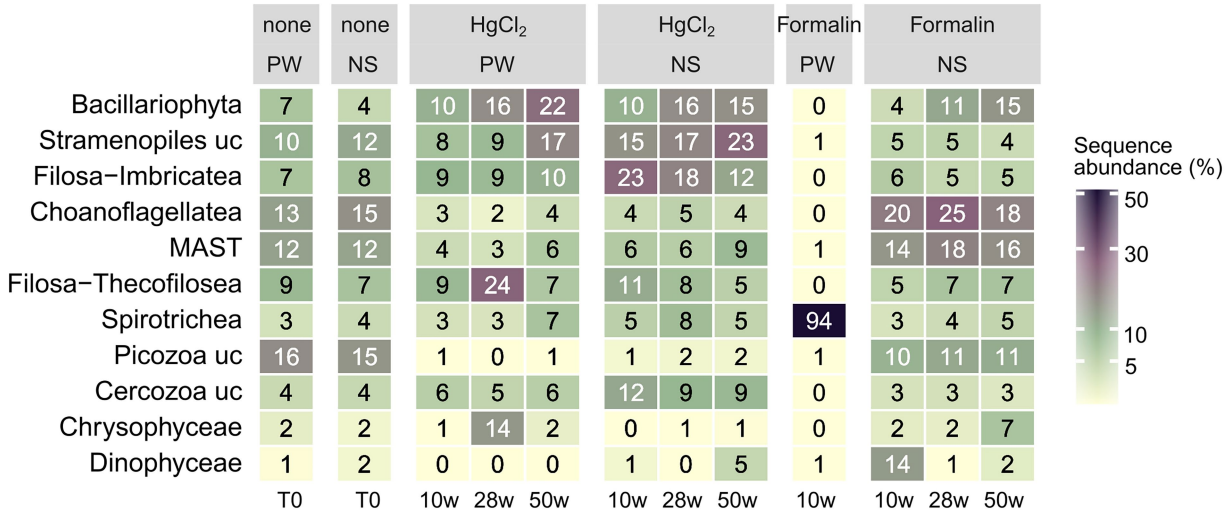

FIGURE 6
Relative abundances of major eukaryotic classes (average of all replicates per sampling event) by preservation, storage time, and DNA extraction. PW: PowerWater, NS: NucleoSpin, uc: unclassified.

favored larger-size eukaryotes such as centric diatoms, with higher abundances of especially unclassified *Mediophyceae* compared to the reference (Figure 7A). If resources allow, additional microscopy (Metfies et al., 2017), quantitative PCR or flow cytometry are advised to assess the effect of preservation on cell numbers and/or size classes. As for bacteria, Hellinger-transformed data provided a more even picture of community structure (Figure 7B).

A comparison of results from the five technical replicates per treatment and time point allowed assessing the variability introduced by extraction and sequencing. Community structure in technical replicates were was highly reproducible for bacteria, but varied more for eukaryotes (Supplementary Figure S6). This observation potentially corresponds to disproportional distribution of large eukaryotic cells in some replicates, suggesting sample volumes should be maximized whenever possible. However, volumes and replicate numbers often need to be balanced with the desired temporal resolution, which can be challenging in remote locations relying on autonomous sampling.

Conclusion

Understanding the ecological and biogeochemical roles of marine microbes substantially benefits from automated sampling in long-term ecological observatories. We herein assessed the combined effects of preservation, DNA extraction and storage time on ribosomal metabarcoding of bacterial and eukaryotic communities. These insights inform the design of automated microbial observation in remote waters, which rely on *in situ* preservation and *ex situ* extraction after extended storage between sample collection and retrieval of the sampler. We present four major conclusions:

1. $HgCl_2$ + PW provided the best representation of bacterial diversity and composition, even after 1 year of storage. Despite altering some patterns observed in the original community, abundances of the major taxa were overall reproducible and differences restricted to only few taxa.
2. Formalin + NS performed best for eukaryotes, despite low DNA yields. Although logistically demanding in (automated) field studies, sampling volumes should be as large as possible to maximize the robustness of analyses. Hellinger or centered-log ratio transformations can counteract the inherent compositionality of amplicon data and provide a more reasonable picture of microbial dynamics.
3. For parallel assessment of bacteria and eukaryotes, we recommend $HgCl_2$ + PW, as this provides good 16S and reasonable 18S rRNA sequence information from single DNA extracts. Our results indicate that the $HgCl_2$ shortcomings in eukaryotes outweigh the formalin shortcomings in bacteria, indicating $HgCl_2$ as most suitable for observatories aiming to study both groups based on DNA from the same samples. Nonetheless, individual time-series should perform similar benchmark studies, as the respective strengths and weaknesses might differ at other *in situ* temperatures and for other microbial communities.
4. In order to minimize bias, we recommend that the choice of preservation should also consider potentially desired comparisons with other sites, as well as other samples from the same observatory. For instance, in case of the FRAM observatory of the Alfred Wegener Institute, the use of $HgCl_2$ + PW facilitates cross-comparability with metabarcoding of sinking particles from sediment traps, including decade-old legacy samples that are treated similarly.

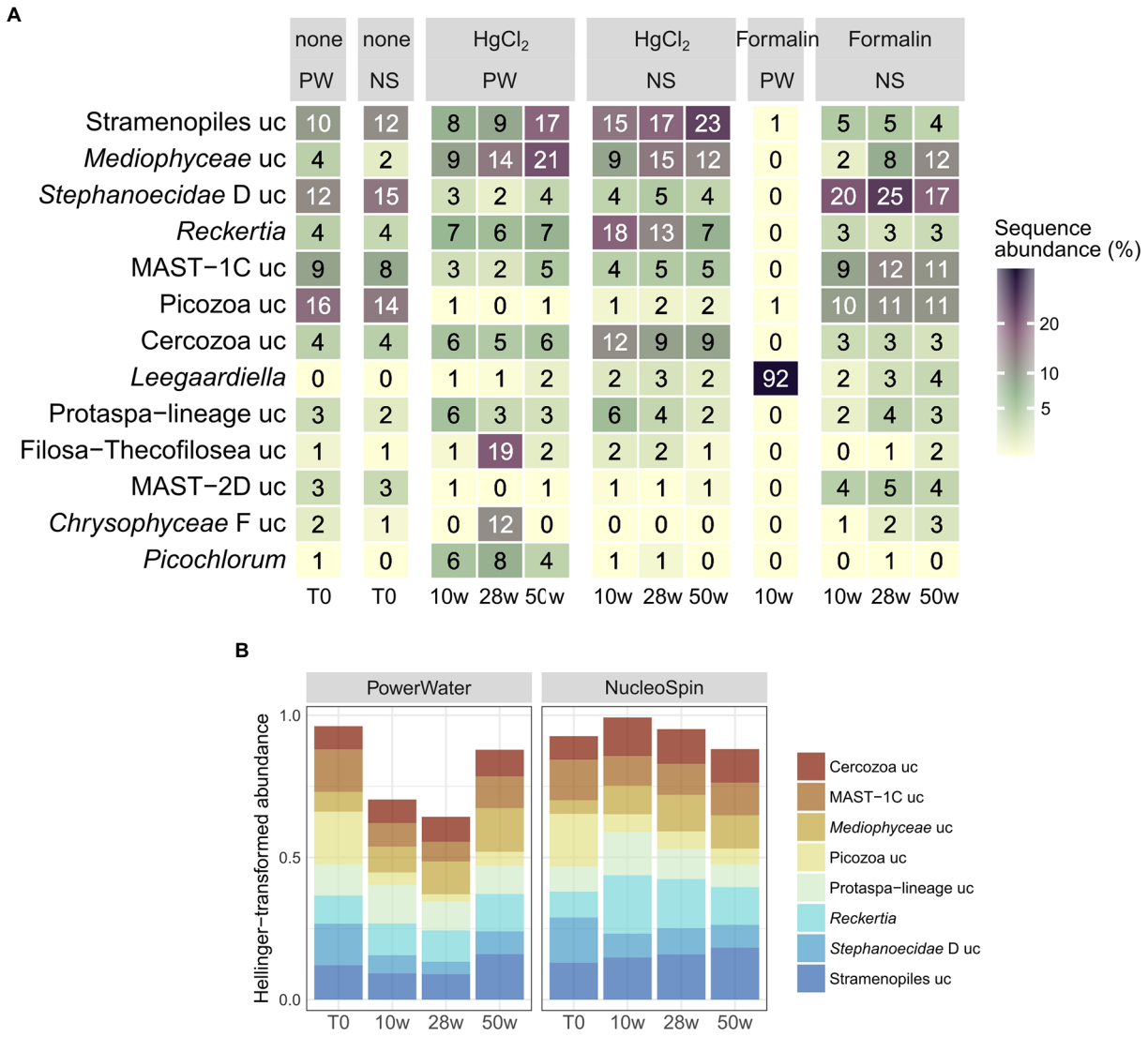

FIGURE 7
Relative abundances **(A)** and Hellinger-transformed relative abundances after $HgCl_2$ preservation **(B)** of major eukaryotic genera (average of all replicates per sampling event) by preservation, storage time, and DNA extraction. PW: PowerWater, NS: NucleoSpin, uc: unclassified.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ebi.ac.uk/ena, PRJEB43307.

Author contributions

MW and KM analyzed the data and wrote the paper. CW performed the experimental work. CB, FJ, IS, and AB co-designed the study and contributed to writing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by institutional funds of the Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, funds provided within the framework of the Helmholtz infrastructure program Frontiers in Arctic Marine Monitoring (FRAM), the Horizon 2020 project AtlantOS (grant ID 633211), and the ERC AdvGrant project ABYSS (grant ID 294757) to AB.

Acknowledgments

We thank Jana Bäger, Jakob Barz, Theresa Hargesheimer, Swantje Rogge, and Kerstin Korte for technical support, Halina

Tegetmeyer for 16S rRNA gene sequencing, Stefan Neuhaus for OTU processing, and Pier Luigi Buttigieg for valuable discussions.

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.

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.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.999925/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Principal Coordinates Analysis (PCoA) comparing community structure between all preservatives in relation to the unpreserved control, revealing marked separation of RNAlater and DNAgard samples despite similar read counts (**Supplementary Table S2**).

SUPPLEMENTARY FIGURE S2
DNA yields from reference (unpreserved, directly filtered) and preserved samples ($HgCl_2$ and formalin) by extraction method. The number of samples per group is shown in **Supplementary Table S2**.

SUPPLEMENTARY FIGURE S3
Hierarchical clustering (complete linkage) of bacterial community composition based on Bray–Curtis dissimilarities (left) and Jaccard presence–absence (right). PW: PowerWater, NS: NucleoSpin.

SUPPLEMENTARY FIGURE S4
OTU numbers within the major bacterial **(A)** and eukaryotic **(B)** classes by preservation, storage time and extraction method.

SUPPLEMENTARY FIGURE S5
Hierarchical clustering (complete linkage) of eukaryotic community composition based on Bray-Curtis dissimilarities (left) and Jaccard presence–absence (right). PW: PowerWater, NS: NucleoSpin.

SUPPLEMENTARY FIGURE S6
Relative abundance heatmaps for bacteria **(A)** and eukaryotes **(B)** showing the abundance of major classes across all replicates by preservation, storage time and DNA extraction. Selected, markedly deviating replicates are encircled in red. Labels on the *x*-axis correspond to sample_titles of raw fastq files as deposited at the European Nucleotide Archive.

SUPPLEMENTARY TABLE S1
Selected marine microbial sampling efforts based on continuous, autonomous techniques.

SUPPLEMENTARY TABLE S2
Read counts and statistics from Swarm-OTU processing, and ENA accession numbers of original fastq files. Samples labeled "P-Buffer" were treated with phosphate buffer (i.e. unpreserved) and not considered in this study. Samples marked in red did not pass Swarm quality thresholds and were excluded from further analysis.

References

Andersen, K. S. S., Kirkegaard, R. H., Karst, S. M., and Albertsen, M. (2018). ampvis2: an R package to analyse and visualise 16S rRNA amplicon data. *bioRxiv*. doi: 10.1101/299537 [Epub ahead of preprint].

Bachy, C., Sudek, L., Choi, C. J., Eckmann, C. A., Nöthig, E.-M., Metfies, K., et al. (2022). Phytoplankton surveys in the Arctic Fram Strait demonstrate the tiny eukaryotic alga micromonas and other Picoprasinophytes contribute to deep sea export. *Microorganisms* 10:961. doi: 10.3390/microorganisms10050961

Balzano, S., Abs, E., and Leterme, S. (2015). Protist diversity along a salinity gradient in a coastal lagoon. *Aquat. Microb. Ecol.* 74, 263–277. doi: 10.3354/ame01740

Bauerfeind, E., Nöthig, E.-M., Beszczynska, A., Fahl, K., Kaleschke, L., Kreker, K., et al. (2009). Particle sedimentation patterns in the eastern Fram Strait during 2000–2005: results from the Arctic long-term observatory HAUSGARTEN. *Deep-Sea Res. I Oceanogr. Res. Pap.* 56, 1471–1487. doi: 10.1016/j.dsr.2009.04.011

Boeuf, D., Edwards, B. R., Eppley, J. M., Hu, S. K., Poff, K. E., Romano, A. E., et al. (2019). Biological composition and microbial dynamics of sinking particulate organic matter at abyssal depths in the oligotrophic open ocean. *Proc. Natl. Acad. Sci.* 116, 11824–11832. doi: 10.1073/pnas.1903080116

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bowman, J. P. (2014). "The family *Colwelliaceae*" in *The Prokaryotes: Gammaproteobacteria*. eds. E. Rosenberg, E. F. DeLong, S. Lory, E. Stackenbrandt and F. Thompson (Berlin, Germany: Springer), 179–195.

Bucklin, A., and Allen, L. D. (2004). MtDNA sequencing from zooplankton after long-term preservation in buffered formalin. *Mol. Phylogenet. Evol.* 30, 879–882. doi: 10.1016/j.ympev.2003.11.002

Buttigieg, P. L., Fadeev, E., Bienhold, C., Hehemann, L., Offre, P., and Boetius, A. (2018). Marine microbes in 4D—using time series observation to assess the dynamics of the ocean microbiome and its links to ocean health. *Curr. Opin. Microbiol.* 43, 169–185. doi: 10.1016/j.mib.2018.01.015

Crameri, F. (2021). Scientific colour maps. doi: 10.5281/zenodo.5501399

Dadon-Pilosof, A., Conley, K. R., Jacobi, Y., Haber, M., Lombard, F., Sutherland, K. R., et al. (2017). Surface properties of SAR11 bacteria facilitate grazing avoidance. *Nat. Microbiol.* 2, 1608–1615. doi: 10.1038/s41564-017-0030-5

Elwood, H. J., Olsen, G. J., and Sogin, M. L. (1985). The small-subunit ribosomal RNA gene sequences from the hypotrichous ciliates *Oxytricha nova* and *Stylonychia pustulata*. *Mol. Biol. Evol.* 2, 399–410. doi: 10.1093/oxfordjournals.molbev.a040362

Fischer, G., Romero, O., Merkel, U., Donner, B., Iversen, M., Nowald, N., et al. (2016). Deep ocean mass fluxes in the coastal upwelling off Mauritania from 1988 to 2012: variability on seasonal to decadal timescales. *Biogeosciences* 13, 3071–3090. doi: 10.5194/bg-13-3071-2016

Formel, N., Enochs, I. C., Sinigalliano, C., Anderson, S. R., and Thompson, L. R. (2021). Subsurface automated samplers for eDNA (SASe) for biological monitoring and research. *Hardware X* 10:e00239. doi: 10.1016/j.ohx.2021.e00239

Fuhrman, J. A., Cram, J. A., and Needham, D. M. (2015). Marine microbial community dynamics and their ecological interpretation. *Nat. Rev. Microbiol.* 13, 133–146. doi: 10.1038/nrmicro3417

Giovannoni, S. J. (2017). SAR11 bacteria: The most abundant plankton in the oceans. *Annu. Rev. Mar. Sci.* 9, 231–255. doi: 10.1146/annurev-marine-010814-015934

Gray, M. A., Pratte, Z. A., and Kellogg, C. A. (2013). Comparison of DNA preservation methods for environmental bacterial community samples. *FEMS Microbiol. Ecol.* 83, 468–477. doi: 10.1111/1574-6941.12008

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., et al. (2013). The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* 41, D597–D604. doi: 10.1093/nar/gks1160

Herfort, L., Seaton, C., Wilkin, M., Roman, B., Preston, C. M., Marin, R. III, et al. (2016). Use of continuous, real-time observations and model simulations to achieve autonomous, adaptive sampling of microbial processes with a robotic sampler. *Limnol. Oceanogr. Methods* 14, 50–67. doi: 10.1002/lom3.10069

Hoffman, E. A., Frey, B. L., Smith, L. M., and Auble, D. T. (2015). Formaldehyde crosslinking: A tool for the study of chromatin complexes. *J. Biol. Chem.* 290, 26404–26411. doi: 10.1074/jbc.R115.651679

Hsieh, T. C., Ma, K. H., and Chao, A. (2016). iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods Ecol. Evol.* 7, 1451–1456. doi: 10.1111/2041-210X.12613

Jørgensen, B. B., and Boetius, A. (2007). Feast and famine — microbial life in the deep-sea bed. *Nat. Rev. Microbiol.* 5, 770–781. doi: 10.1038/nrmicro1745

Lampitt, R. S., Salter, I., de Cuevas, B. A., Hartman, S., Larkin, K. E., and Pebody, C. A. (2010). Long-term variability of downward particle flux in the deep Northeast Atlantic: causes and trends. *Deep-Sea Res. II Top. Stud. Oceanogr.* 57, 1346–1361. doi: 10.1016/j.dsr2.2010.01.011

Lee, C., Hegdes, J., Wakeham, S., and Zhu, N. (1992). Effectiveness of various treatments in retarding microbial activity in sediment trap material and their effects on the collection of swimmers. *Limnol. Oceanogr.* 37, 117–130. doi: 10.4319/lo.1992.37.1.0117

Legendre, P., and Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia* 129, 271–280. doi: 10.1007/s004420100716

Lindsay, D. J. (2021). Stealthy tracking of deep ocean organisms with Mesobot. *Sci. Robot.* 6:eabj3949. doi: 10.1126/scirobotics.abj3949

Liu, Y., Blain, S., Crispi, O., Rembauville, M., and Obernosterer, I. (2020). Seasonal dynamics of prokaryotes and their associations with diatoms in the Southern Ocean as revealed by an autonomous sampler. *Environ. Microbiol.* 22, 3968–3984. doi: 10.1111/1462-2920.15184

Mahé, F., Rognes, T., Quince, C., de Vargas, C., and Dunthorn, M. (2014). Swarm: robust and fast clustering method for amplicon-based studies. *Peer J.* 2:e593. doi: 10.7717/peerj.593

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

McMurdie, P. J., and Holmes, S. (2013). Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. doi: 10.1371/journal.pone.0061217

Metfies, K., Bauerfeind, E., Wolf, C., Sprong, P., Frickenhaus, S., Kaleschke, L., et al. (2017). Protist communities in moored long-term sediment traps (Fram Strait, Arctic)–preservation with mercury chloride allows for PCR-based molecular genetic analyses. *Front. Mar. Sci.* 4:301. doi: 10.3389/fmars.2017.00301

Moore, S. K., Mickett, J. B., Doucette, G. J., Adams, N. G., Mikulski, C. M., Birch, J. M., et al. (2021). An autonomous platform for near real-time surveillance of harmful algae and their toxins in dynamic coastal shelf environments. *J. Mar. Sci. Eng.* 9:336. doi: 10.3390/jmse9030336

Nöthig, E.-M., Lalande, C., Fahl, K., Metfies, K., Salter, I., and Bauerfeind, E. (2020). Annual cycle of downward particle fluxes on each side of the Gakkel ridge in the Central Arctic Ocean. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* 378:20190368. doi: 10.1098/rsta.2019.0368

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., et al. (2013). Vegan: community ecology package.

Ottesen, E. A., Mari, R. A., Preston, C. M., Young, C. R., Ryan, J. P., Scholin, C. A., et al. (2011). Metatranscriptomic analysis of autonomously collected and preserved marine bacterioplankton. *ISME J.* 5, 1881–1895. doi: 10.1038/ismej.2011.70

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

Paradis, E., and Schliep, K. (2019). Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528. doi: 10.1093/bioinformatics/bty633

Poff, K. E., Leu, A. O., Eppley, J. M., Karl, D. M., and DeLong, E. F. (2021). Microbial dynamics of elevated carbon flux in the open ocean's abyss. *Proc. Natl. Acad. Sci. U. S. A.* 118:e2018269118. doi: 10.1073/pnas.2018269118

Pratte, Z. A., and Kellogg, C. A. (2021). Comparison of preservation and extraction methods on five taxonomically disparate coral microbiomes. *Front. Mar. Sci.* 8:684161. doi: 10.3389/fmars.2021.684161

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219

Rachel, N. M., and Gieg, L. M. (2020). Preserving microbial community integrity in oilfield produced water. *Front. Microbiol.* 11:581387. doi: 10.3389/fmicb.2020.581387

Ramondenc, S., Nöthig, E.-M., Hufnagel, L., Bauerfeind, E., Busch, K., Knüppel, N., et al. (2022). Effects of Atlantification and changing sea-ice dynamics on zooplankton community structure and carbon flux between 2000 and 2016 in the eastern Fram Strait. *Limnol. Oceanogr.* doi: 10.1002/lno.12192 (in press).

Reid, K. M., Maistry, S., Ramesar, R., and Heathfield, L. J. (2017). A review of the optimisation of the use of formalin fixed paraffin embedded tissue for molecular analysis in a forensic post-mortem setting. *Forensic Sci. Int.* 280, 181–187. doi: 10.1016/j.forsciint.2017.09.020

Renshaw, M. A., Olds, B. P., Jerde, C. L., McVeigh, M. M., and Lodge, D. M. (2015). The room temperature preservation of filtered environmental DNA samples and assimilation into a phenol–chloroform–isoamyl alcohol DNA extraction. *Mol. Ecol. Resour.* 15, 168–176. doi: 10.1111/1755-0998.12281

Rissanen, A. J., Kurhela, E., Aho, T., Oittinen, T., and Tiirola, M. (2010). Storage of environmental samples for guaranteeing nucleic acid yields for molecular microbiological studies. *Appl. Microbiol. Biotechnol.* 88, 977–984. doi: 10.1007/s00253-010-2838-2

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *Peer J.* 4:e2584. doi: 10.7717/peerj.2584

Salter, I., Kemp, A. E. S., Moore, C. M., Lampitt, R. S., Wolff, G. A., and Holtvoeth, J. (2012). Diatom resting spore ecology drives enhanced carbon export from a naturally iron-fertilized bloom in the Southern Ocean. *Glob. Biogeochem. Cycles* 26:GB1014. doi: 10.1029/2010GB003977

Salter, I., Lampitt, R., Sanders, R., Poulton, A., Kemp, A., Boorman, B., et al. (2007). Estimating carbon, silica and diatom export from a naturally fertilised phytoplankton bloom in the Southern Ocean using PELAGRA: A novel drifting sediment trap. *Deep-Sea Res. I-Top. Stud. Oceanogr.* 54, 2233–2259. doi: 10.1016/j.dsr2.2007.06.008

Salter, I., Schiebel, R., Ziveri, P., Movellan, A., Lampitt, R., and Wolff, G. A. (2014). Carbonate counter pump stimulated by natural iron fertilization in the polar frontal zone. *Nat. Geosci.* 7, 885–889. doi: 10.1038/ngeo2285

Sano, M., Makabe, R., Kurosawa, N., Moteki, M., and Odate, T. (2020). Effects of Lugol's iodine on long-term preservation of marine plankton samples for molecular and stable carbon and nitrogen isotope analyses. *Limnol. Oceanogr. Methods* 18, 635–643. doi: 10.1002/lom3.10390

Sherr, E. B., and Sherr, B. F. (1993). "Preservation and Storage of Samples for Enumeration of Heterotrophic Protists," in *Handbook of Methods in Aquatic Microbial Ecology*. eds. P. F. Kemp, B. F. Sherr, E. B. Sherr and J. J. Cole (Boca Raton, FL: CRC Press), 207–212.

Spens, J., Evans, A. R., Halfmaerten, D., Knudsen, S. W., Sengupta, M. E., Mak, S. S. T., et al. (2017). Comparison of capture and storage methods for aqueous microbial eDNA using an optimized extraction protocol: advantage of enclosed filter. *Methods Ecol. Evol.* 8, 635–645. doi: 10.1111/2041-210X.12683

Stern, R. F., Picard, K. T., Hamilton, K. M., Walne, A., Tarran, G. A., Mills, D., et al. (2015). Novel lineage patterns from an automated water sampler to probe marine microbial biodiversity with ships of opportunity. *Prog. Oceanogr.* 137, 409–420. doi: 10.1016/j.pocean.2015.04.015

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., et al. (2015). Structure and function of the global ocean microbiome. *Science* 348:1261359. doi: 10.1126/science.1261359

Truelove, N. K., Patin, N. V., Min, M., Pitz, K. J., Preston, C. M., Yamahara, K. M., et al. (2022). Expanding the temporal and spatial scales of environmental DNA research with autonomous sampling. *Environ. DNA* 4, 972–984. doi: 10.1002/edn3.299

UNESCO-IOC (1994). *Protocols for the Joint Global Ocean Flux Study (JGOFS) Core Measurements*. Paris: UNESCO.

Valencia, B., Stukel, M. R., Allen, A. E., McCrow, J. P., Rabines, A., and Landry, M. R. (2022). Microbial communities associated with sinking particles across an environmental gradient from coastal upwelling to the oligotrophic ocean. *Deep-Sea Res. I Oceanogr. Res. Pap.* 179:103668. doi: 10.1016/j.dsr.2021.103668

von Appen, W.-J., Waite, A. M., Bergmann, M., Bienhold, C., Boebel, O., Bracher, A., et al. (2021). Sea-ice derived meltwater stratification slows the biological carbon pump: results from continuous observations. *Nat. Commun.* 12:7309. doi: 10.1038/s41467-021-26943-z

Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., et al. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* 5:27. doi: 10.1186/s40168-017-0237-y

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., et al. (2019). Welcome to the Tidyverse. *J. Open Source Softw.* 4:1686. doi: 10.21105/joss.01686

Wietz, M., Bienhold, C., Metfies, K., Torres-Valdés, S., von Appen, W.-J., Salter, I., et al. (2021). The polar night shift: seasonal dynamics and drivers of Arctic Ocean microbiomes revealed by autonomous sampling. *ISME Commun.* 1:76. doi: 10.1038/s43705-021-00074-4

Yamahara, K. M., Preston, C. M., 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. R., 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

Yuan, J., Li, M., and Lin, S. (2015). An improved DNA extraction method for efficient and quantitative recovery of phytoplankton diversity in natural assemblages. *PLoS One* 10:e0133060. doi: 10.1371/journal.pone.0133060

Zhang, Y., Ryan, J. P., Hobson, B. W., Kieft, B., Romano, A., Barone, B., et al. (2021). A system of coordinated autonomous robots for Lagrangian studies of microbes in the oceanic deep chlorophyll maximum. *Sci. Robot.* 6:eabb9138. doi: 10.1126/scirobotics.abb9138

Zhang, Y., Ryan, J. P., Kieft, B., Hobson, B. W., McEwen, R. S., Godin, M. A., et al. (2019). Targeted sampling by autonomous underwater vehicles. *Front. Mar. Sci.* 6:415. doi: 10.3389/fmars.2019.00415

Zúñiga, D., Sanchez-Vidal, A., Flexas, M. M., Carroll, D., Rufino, M. M., Spreen, G., et al. (2021). Sinking diatom assemblages as a key driver for deep carbon and silicon export in the Scotia Sea (Southern Ocean). *Front. Earth Sci.* 9:579198. doi: 10.3389/feart.2021.579198

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 09 September 2022
DOI 10.3389/fmars.2022.991126

OPEN ACCESS

EDITED BY
Catarina Magalhães,
University of Porto, Portugal

REVIEWED BY
Yong Zhang,
Fujian Normal University, China
Mitsuhide Sato,
Nagasaki University, Japan

*CORRESPONDENCE
Deli Wang
deliwang@xmu.edu.cn

SPECIALTY SECTION
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Marine Science

RECEIVED 12 July 2022
ACCEPTED 25 August 2022
PUBLISHED 09 September 2022

CITATION
Gu L, Yue X, Zhong H, Mei K and Wang D (2022) A new technique of quantifying protoporphyrin IX in microbial cells in seawater. *Front. Mar. Sci.* 9:991126. doi: 10.3389/fmars.2022.991126

A new technique of quantifying protoporphyrin IX in microbial cells in seawater

Lide Gu[1,2], Xinli Yue[1,2], Haowen Zhong[1,2], Kang Mei[1,2] and Deli Wang[1,2]*

[1]State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen, China, [2]College of Ocean and Earth Sciences, Xiamen University, Xiamen, China

Protoporphyrin IX (PPIX), a fundamental precursor in the synthesis of heme and chlorophyll, plays a vital role in the biological metabolism and biogeochemical cycling in the ocean. PPIX has previously been identified in humans, animals, and plants, while so far as we know, there is no measurements until now regarding its contents in microbes, and especially in marine phytoplankton and bacteria. Here, for the first time, we reported a method of determining PPIX in marine microbial cells *via* acetone extraction followed by reversed phase high-performance liquid chromatography quantification, in which acetone-acetonitrile/water-formic acid buffer was used as a gradient elution solvent. The method was optimized with the detection limit of 3.8 $\pm$ 1.0 pM, and recovery rate of 97.5 $\pm$ 1.9%. The structure of the extracted PPIX was further confirmed using tandem mass spectrometry as positively associated with specific protonated molecules $[M + H]^+$. The method was then successfully applied in the determination of PPIX in microbial cells in the water samples collected from a median-sized subtropical estuary (the Jiulong River Estuary, China). The results showed that PPIX existed widely and ranged from 20 – 170 ng/L in cells in the water samples. In the whole estuary, cellular PPIX generally decreased linearly with increasing salinity. A positive correlation of PPIX with particulate organic matter in the estuary suggested of sediment suspension and dissolution as its possible source. In addition, a general hyperbolic fitting pattern was observed for PPIX against dissolved inorganic nitrogen, PO_4^{3-}, and the bacterial abundance (10^4 – 10^6 cells/mL) in the estuary. Such results indicated that PPIX played a crucial role in linking nutrients and the microbial productivity. In summary, we developed a new technique of quantifying cellular PPIX in water samples and confirmed the wide existence of cellular PPIX in natural waters. The data from Jiulong River estuary further suggest that the contents of cellular PPIX be enhanced with the nutrient supply from riverine inputs and sediment suspensions, which thereafter dictate the productivity of phytoplankton and bacteria in coastal waters.

KEYWORDS

Protoporphyrin IX, HPLC, coastal ecosystem, essential metabolites, phytoplankton and bacteria

Introduction

Protoporphyrin IX (PPIX) is an essential precursor with some specific physiological functions such as oxygen transport and storage mediated by heme (iron/zinc-PPIX) or hemocyanin (copper-PPIX) and photosynthesis by chlorophyll (magnesium-PPIX). It is also a ubiquitous prosthetic group of proteins such as cytochrome, catalases, peroxidases, and nitrate reductase (Xiong et al., 2011; Senge et al., 2014; Sachar et al., 2016). Consequently, PPIX participates in fundamental metabolic processes in cells and plays a vital role in cell biology. PPIX is an endogenous fluorescent heterocyclic organic molecule, and its biosynthesis includes eight key enzymatic processes in the mitochondria and cytoplasm of cells (Sachar et al., 2016). Figure 1 shows the proposed biosynthetic pathway of PPIX in the cells of mammals, fungi, yeasts, and the α-subgroup of proteobacteria, which starts from source materials: glycine and succinyl-CoA (Sachar et al., 2016). Alternatively, PPIX is initiated from glutamate and tRNA in plants, green algae, cyanobacteria, archaea, and most bacteria (Senge et al., 2014). All the synthesizing processes are dictated from strict enzymatic reactions under specific environmental conditions (Bonkovsky et al., 2013). Thus, any disturbance in these steps might cause an accumulation of or deficiency in intracellular PPIX, and even lead to severe porphyria (Dailey and Meissner, 2013). It has been reported that extra supply of nitrogen nutrients might seriously affect the formation of proteins including synthetase, reductase, and aminomutase, as well as the porphyrin rings. On the other hand, lack of necessary carbon compounds hindered the tricarboxylic acid (TCA) cycle and thus affected the production of succinyl-CoA, and influenced the porphyrin synthesis (Kaneko, 2008). Similarly, lack of phosphorus also interfered the biological energy and material cycling in the marine ecosystem (Tian et al., 2021). In addition, endogenous metabolites could dictate the eukaryotic and prokaryotic members of a microbial community *via* complex ecological interaction between producers and auxotrophs (Chen et al., 2018). These specific biomolecules might play a critical role in modulating the ecological systems, and even linking the global biogeochemical cycles (Suffridge et al., 2018).

PPIX has previously been detected by using a series of techniques including the fluorometry (Datta et al., 1998), thin-layer chromatography (De Maere et al., 2014), capillary electrophoresis (Weinberger and Sapp, 1990), and the high-performance liquid chromatography (HPLC) coupled with mass spectrometry (MS) (Fyrestam et al., 2015; Fyrestam and Östman, 2017). Mass spectrometry provides structural information, and HPLC is used for precise quantification. HPLC-MS has become a reliable technique for measuring PPIX recently due to its high precision and accuracy (Lim; Zvezdanovic et al). Previous analyses of PPIX mainly focused on samples taken from human, such as the urines (Ausio et al., 2000), blood, and feces (Danton and Lim, 2006), and there were only a few published data available regarding microorganisms. For example, Fyrestam et al. firstly determined a porphyrin profile of *Aggregatibacter actinomycetemcomitans* in oral biofilms by using the HPLC/MS method (Fyrestam et al., 2015). In addition, the PPIX contents have also been reported in some plants and algae (Stillman LC, 1978; Weinstein JD, 1984; Espinas et al., 2012). Although important, the currently available PPIX studies focused primarily on its applications in human health, industrial development, or single-tissue testing. There is still no records of PPIX measurements in natural environments so far. We are still lack of studies on its ecological functioning in microbial communities, and especially in marine ecosystems. Therefore, there is a need of establishing a reliable and accurate method for extracting and quantifying PPIX in natural samples. This study aimed to develop such a method to determine cellular PPIX in estuarine and coastal waters. Along with its post-derivatives (chlorophyll, pheophytin, and bacteriochlorophyll), we expect that this new analytical protocol will allow for a preliminary understanding of the cycling of PPIX in natural environment, and its roles in mediating global marine biogeochemical cycles and the succession of microbial species.

Methods

Chemicals and materials

The protoporphyrin IX (purity ≥ 95%), chlorophyll-*a*, and bacteriochlorophyll-*a* were obtained from Merck (Darmstadt, Germany). The HPLC-grade acetonitrile, acetone, and formic acid (≥ 98%) were also purchased from Merck. The InertSustain C18 column (4.6 × 150 mm, 5 μm) was sourced from GL Science (Tokyo, Japan). A Milli-Q Element water purification system from Millipore (USA) was used to produce deionized water at 18.2 MΩ cm. The 0.45 μm Millipore organic needle filters were from Merck, and the 1.5 mL sample vials were from CNW Technologies GmbH (USA). The 0.7 μm and 47 mm GF/F filters were from Whatman (UK). All of the other inorganic chemicals and organic solvents used in the study were of analytical grade.

Environmental settings of study area

The Jiulong River is the 13th largest river in China with an average annual total freshwater discharge of 1.47 × 10^{10} m^3. The Jiulong River is located in a highly populated area in Southeast China with influences from agricultural, industrial, and municipal activities nearby (Wang et al., 2018). The Jiulong River estuary (JRE) is a typical subtropical macrotidal estuary, and subject to human perturbation within the whole estuary (Figure 2). The Jiulong River

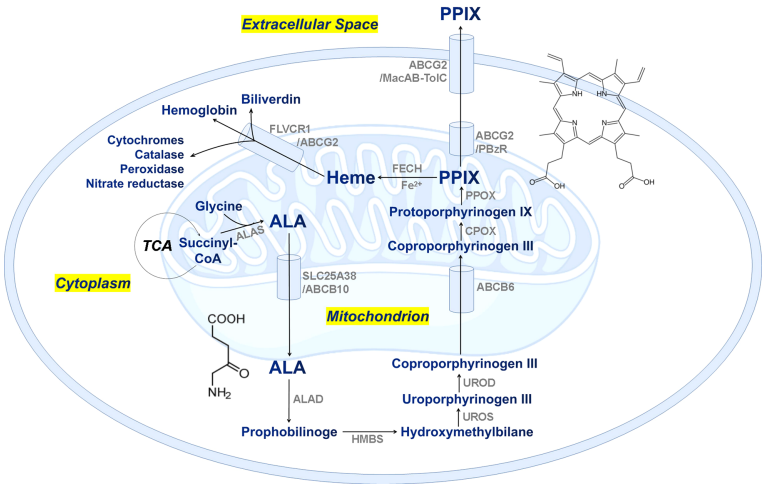

FIGURE 1
The proposed biosynthetic pathway of PPIX/heme in mammalian cells. TCA, tricarboxylic acid cycle; ALA, δ-aminolevulinic acid; ALAS, ALA synthase; SLC25A38, solute carrier family 25 member 38; ABCB, ATP-binding cassette subfamily B member; ALAD, ALA dehydratase; HMBS, hydroxymethylbilane synthase; UROS, uroporphyrinogen III synthase; UROD, uroporphyrinogen decarboxylase; CPOX, coproporphyrinogen oxidase; PPOX, protoporphyrinogen oxidase; ABCG2, ATP-binding cassette subfamily G member 2; FECH, ferrochelatase; FLVCR1, feline leukemia virus subgroup c receptor 1; MacAB-TolC, an efflux pump involved in *E. coli* and *Salmonella* for macrolide efflux.

estuary could be divided into three parts geochemically: the upper, lower estuaries and the bay areas. The upper estuary receives a large amount of inputs of terrestrial materials from the river, and the lower estuary is subject to strong tidal intrusion with sediment disturbances locally. The bay area is surrounding the Xiamen Island, with a low oxygen zone and even algal bloom occurring occasionally (Chen et al., 2019; Baohong et al., 2021).

Sample collection and processing

The water samples used for quantifying PPIX were collected from the surface (~ 0.5 m below) and bottom (~ 1 m above) in the JRE on 14th – 15th November 2021 (Figure 2). Totally, thirty-two water samples were taken at 16 stations, and the values of salinity increased gradually from 0 ~ 32 from the upper to lower estuaries (stations of A3-JY3), and afterwards remained

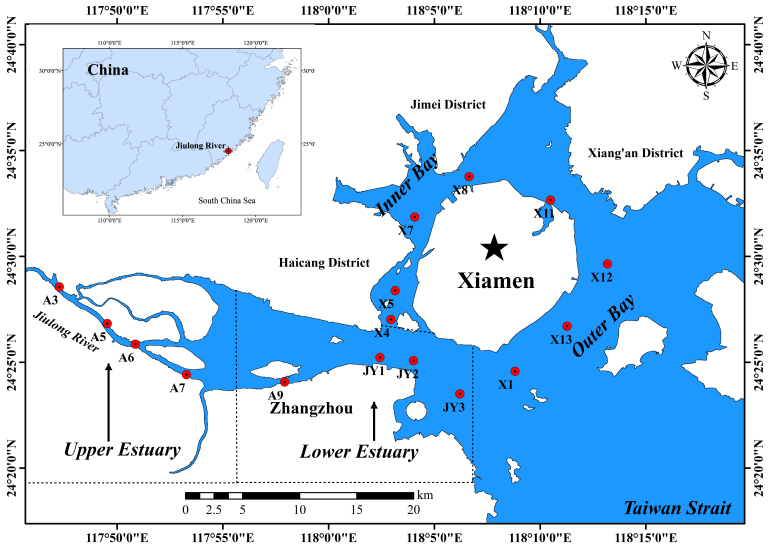

FIGURE 2
The geographic setting and sampling stations in the Jiulong River Estuary.

relatively stable (~ 32) in the bay area (stations of X1-X13). Water samples were stored in pre-cleaned polyethylene bottles once collected *via* using Niskin samplers with Tygon tubing. All the water samples for analyzing PPIX and its derivatives (chlorophyll-*a*, pheophytin and bacteriochlorophyll-*a*) were pre-screened through a 10 μm mesh to remove zooplankton and large particles, and then filtered through 0.7 μm GF/F filters. The filter membranes were immediately stored at −80°C once back in lab.

The water samples for analyzing dissolved inorganic nitrogen (DIN, including NO_3^-, NO_2^- and NH_4^+) and PO_4^{3-} were filtered using a 0.45 μm polycarbonate membrane (Millipore, USA), and analyzed later *via* the SEAL AutoAnalyzer 3 (Dai et al., 2008). The samples for analyzing particulate organic carbon (POC) and particulate nitrogen (PN) were collected using the GF/F filters (precombusted at 450°C for 4 h) and analyzed with a Vario EL cube elemental analyzer (Elementar, Hanau, Germany) after carbonates being removed with the fumes of concentrated HCl for 24 h (Liu et al., 2020). The abundances of autotrophic and heterotrophic microbial cells were determined *via* using the BD Accuri® C6 flow cytometry, once the samples were fixed by glutaraldehyde (final concentration 1%) at 4°C in the dark (Suffridge et al., 2018). The chlorophyll-*a* (Chl-*a*) and pheophytin concentrations in one aliquot of the water samples were determined using a Turner Designs Trilogy fluorometer according to the protocols in the Joint Global Ocean Flux Study (JGOFS) Core Measurements guide (Knap et al., 1996). Bacteriochlorophyll-*a* (BChl-*a*) was quantified at the wavelengths of 770 nm using HPLC according to the method of Goericke's (Goericke, 2002). Hydrological parameters, including salinity, water temperature, dissolved oxygen (DO), and pH, were measured *via* a portable water quality meter (WTW Multi 3430, Germany) on board.

The other aliquot of samples for PPIX were processed according to the procedure for chlorophyll *a* described in the JGOFS protocol, and all the samples were protected from light as much as possible during being processed (Knap et al., 1996). The PPIX molecules were extracted after the filters being soaked in 5.0 mL of 100% acetone and gently shaken for a few seconds. The samples in tubes were then wrapped in aluminium foil to avoid photo-degradation and then stored at −20°C for 20 h. Once done, 0.5 mL of the solvent was pipetted with 0.1 mL 1.2 M hydrochloric acid (v/v, 5/1) at pH < 2, and the acidification took place in the dark at room temperature for 24 h. The final solution was then quantified *via* HPLC after being filtered through a 0.45 μm polycarbonate membrane.

HPLC-MS analysis

PPIX in the samples was quantified *via* a Thermo Ultimate 3000 HPLC system (Dionex, California, U.S.A.) with a fluorescence detector. The chromatographic separation was performed through an InertSustain C18 column (4.6 × 150 mm, 5 μm, particle size), and the column oven temperature was set at 35 ± 1°C. A 12 min gradient flow was used with mobile phase A: 60% acetonitrile + 40% water + 0.1% formic acid and mobile phase B: 100% acetone + 0.1% formic acid (with all reagents of LC/MS grade). The flow rate was 1 mL/min and the injection volume was 100 μL. The linear gradient program was set as the following: 0 – 2 min: 20% B; 2 – 2.2 min: 100% B; 2 – 10 min: 100% B; 10 – 10.2 min: 20% B; and 10.2 – 12 min: 20% B. The excitation and emission wavelengths were set at 406 and 635 nm, respectively. The structure of PPIX in several samples was further confirmed *via* an Agilent 1290 UPLC binary pump and Agilent 6490 triple-quadrupole mass spectrometer equipped with an electrospray ionization source system (Agilent Technologies, California, U.S.A.). The detailed setup parameters for the HPLC and ESI–MS/MS were listed in the Table 1.

Method validation: Limits of detection, limits of quantitation, and quality control

The PPIX stock solution was prepared in 6.0 M formic acid and maintained at 4°C. A linear calibration curve was plotted using the chromographic values from ten standard calibration solutions (1.0 – 1500.0 nM/L diluted in acetone) with triplicate injections. An exemplary calibration curve is shown in the Figure 3D with standard peak areas (y) against the concentrations (x) $R^2 > 0.99$. Quality control samples were prepared with the following concentrations: 0.5, 1.0, 5.0, and 60.0 nM/L. To determine the limits of detection (LOD) and quantitation (LOQ), the signal-to-noise (S/N) ratio is calculated from triplicate injections of the 0.2 nM PPIX standard solution. Here, the noise is defined as the standard deviation of the peak areas, and the signal, as the mean of the peak areas of the triplicate injections. The LOD and LOQ are defined as 3 and 10 times of the S/N ratio, respectively. The results show that the signal-to-ratio (S/N) is 3.8, and the LOD and LOQ are 3.8 ± 1.0 pM and 12.8 ± 3.3 pM, respectively.

The intra- and inter-daily determinations of spiked PPIX samples were conducted for quality control. Here, six replicates (n = 6) at four concentrations were used. The accuracy is reflected by the percentage of measured concentration compared to the true value, and precision is expressed as the relative standard deviation (RSD). Table 2 shows the accuracy and precision of the PPIX in seawater (n = 6). The samples with less PPIX (0.5 nM) showed with a lower precision in both intra- and inter-daily measurements (Table 2). The samples with more PPIX (1.0 – 60.0 nM) showed with higher precision, with the RSD values of 0.4 – 8.3% (for the intra-daily check) and 1.7 – 2.1% (for the inter-daily check), respectively. Both sets of the values were below 15%, suggesting of an acceptable repeatability

TABLE 1 Experimental setups of HPLC and ESI–MS/MS for measuring PPIX.

Ultimate 3000-HPLC			Agilent 6490 Triple-Quadrupole Mass Spectrometer		
Chromatographic conditions			**Electrospray conditions**		
Column dimensions 4.6 × 150 mm			Ionization mode		ESI [+]
Stationary phase C-18			Capillary voltage		3000 V
Particle size 5 μm			Source temperature		250°C
Mobile phase A	60% Acetonitrile + 40% Water + 0.1% Formic Acid		Sheath gas temperature		350°C
			Sheath gas flow		11 L/min
Mobile phase B	100% Acetone + 0.1% Formic Acid		Nebulizer gas		30 psi
			MS1/MS2 heater temperature		100°C
Column temperature 35 ± 1°C			Dwell time		80 ms
Autosampler temperature 4 ± 1°C			MS/MS		MRM
Flow rate	1 mL/min				
Injection volume	100 μL		Compounds	Transition	Collision energy
Linear gradient program			PPIX	563 > 504	46
Time (min)	A	B		563 > 489	55
0	80%	20%		563 > 445	56
2	80%	20%	Pheophytin a	872 > 593	44
2.2	0%	100%		872 > 533	46
10	0%	100%			
10.2	80%	20%			
12	80%	20%			

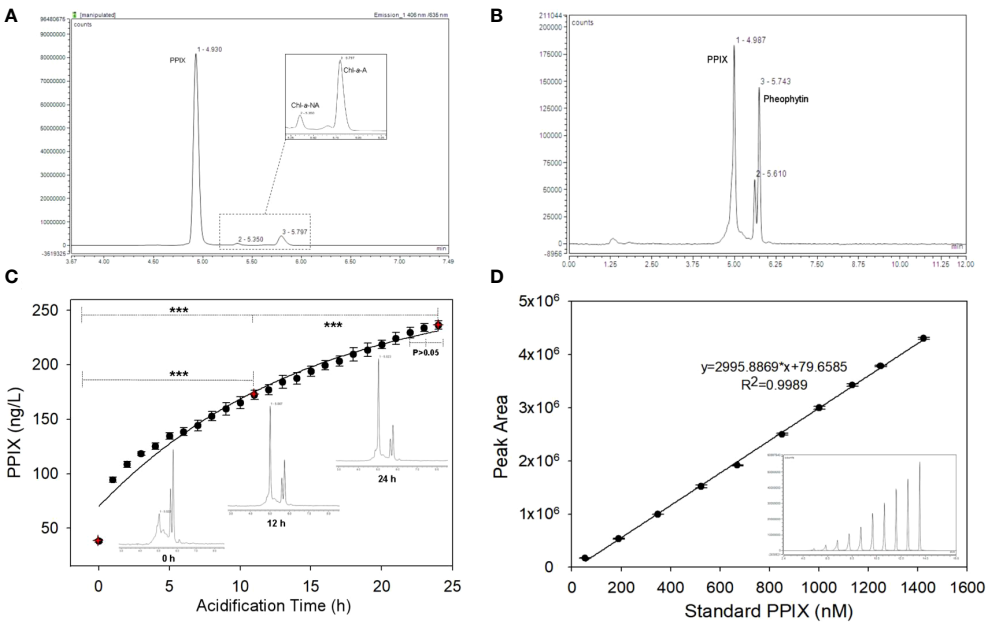

FIGURE 3
(A) The chromatogram of standard PPIX, Chl-*a* (pheophytin-*a*) and BChl in a mixed standard sample; **(B)** The HPLC chromatogram of PPIX in a seawater sample; **(C)** The variation of PPIX signals in a series of samples along with increasing acidification time (Inset: the chromatograms at 0, 12 and 24 h; *** denotes that $p < 0.001$ and the error bar represents the relative standard deviation of the analytes, n=5); **(D)** The exemplary calibration curve of PPIX (in peak area) at 10 different levels (Inset: the offset chromatogram of ten peaks; and the error bar refers to the relative standard deviation, n=3).

TABLE 2 The accuracy, precision, and stability of the determination of PPIX in seawater (n = 6).

Conc.(nM)	Intra-Day		Inter-Day		Stability (± SD%)			
	Accuracy (± SD%)	Precision (± SD%)	Accuracy (± SD%)	Precision (± SD%)	Day 4	Day 8	Day 14	Day 21
0.5	80.0 ± 4.6	14.4 ± 0.6	90.8 ± 11.7	4.6 ± 3.9	101.6 ± 4.4	108.3 ± 3.2	108.3 ± 6.3	108.9 ± 4.8
1	88.6 ± 3.1	3.6 ± 3.6	97.7 ± 10.6	2.1 ± 1.2	106.9 ± 6.8	99.9 ± 13.0	102.5 ± 6.9	104.8 ± 2.9
5	85.9 ± 2.2	0.4 ± 0.2	97.0 ± 11.3	2.1 ± 1.4	108.0 ± 3.0	106.1 ± 2.6	105.2 ± 2.1	99.6 ± 2.1
60	95.0 ± 11.0	8.3 ± 8.0	98.9 ± 8.8	1.7 ± 1.1	102.8 ± 0.6	102.3 ± 2.2	96.5 ± 0.8	95.3 ± 2.7

(intra-daily precision) and reproducibility (inter-daily precision) of the method. To evaluate the recovery, the collected filter membranes were spiked with 1 nM/mL PPIX standard (and additional triplicate filter membranes without spikes as control), and then processed according to our method. Thus, the average recovery was calculated as 97.5 ± 1.9% (n = 3). To evaluate the extraction efficiency of acetone, a set of samples (n = 5) were extracted consecutively (with acetone from one time until 4 times), and accumulative chromatographic values were calculated. According to the accumulative data, PPIX reached with 97.9 ± 0.7% after first extraction (coefficient of variation: ~ 0.64%), and remained relative constant with more extractions. Such results indicate of one-step extraction with acetone as efficiently but with low laborious cost. We further checked the stability of the samples (with 4 known concentrations as measured initially) along with time as we determined the PPIX contents in samples after 4, 8, 14, and 21 days (n = 6) (stored at 4°C). The data obtained at the later time were consistent with the original concentrations with the stabilities of 95 to 108% (the ratio of the later measurement to the original) (Table 2). Here, our results indicate that the collected filter membranes could be kept at 4°C for at least 21 days with no significant loss of PPIX.

Results

HPLC-ESI MS/MS analysis

The determination of PPIX extracted from water samples is summarized as follows. PPIX was quantified *via* HPLC at the excitation wavelength of 406 nm and emission wavelength of 635 nm, and the retention time of 4.7 to 5.2 min. The PPIX molecules are hydrophobic in structure with two peripherally ionizable propionate groups. The compound tends to be easily dissolved in organic solvent under acidic or basic conditions (Junga et al., 2010; Hoseini et al., 2019). Espinas et al. reported that the extraction efficiency of plant porphyrin strongly depended on the pH values as pH affects the solubility, the aggregation and fluorescence intensity of PPIX (Espinas et al., 2012). Here, we used acetone to extract porphyrin from the samples for it not being destroyed as much as possible. On the other hand, the pigments including Chls and BChls in the other aliquot of samples were extracted simultaneously for subsequent analyses. It has been reported that the absorption peak for monomer PPIX was at 406 nm at pH < 2; a broad absorption peak appeared as an intermediate at 352 ~ 450 nm at pH of 3 to 7; a absorption spectrum was at 382 nm at pH > 8 (with dimer PPIX formed) (Hoseini et al., 2019). In fact, PPIX in organisms exists as in free forms, or as covalently and non-covalently bound to other ligands. Therefore, the sufficient acidification is required for converting all the PPIX forms to monomers (Espinas et al., 2012; Fyrestam et al., 2015).

The target peak of PPIX in the chromatogram was also identified along with Chl-*a* and BChl-*a*. The chromatogram (Figure 3A) shows with signals of standard PPIX, acidic Chl-*a* (pheophytin-*a*) and BChl-*a*. Here, the peaks of PPIX, Chl-*a* and pheophytin-*a* were clearly demodulated, but the BChl-*a* signal cannot be measured. This might be due to the concentration being too low or the signal intensity of PPIX too high at the excitation wavelength, which obscured the signal. The chromatogram of organic compounds extracted from the samples (Figure 3B) shows with the three distinct peaks at retention times of 4.9, 5.6, and 5.7 min, respectively. The target and standard PPIX appeared at an identical retention time of 4.9 min, whereas that of pheophytin was at 5.7 min. The structure of the PPIX molecules in the extracted samples was further confirmed based on the corresponding MS/MS ion spectra of their fragmented products. Previous researchers suggested of a primary fragmentation pathway of PPIX (Fyrestam and Östman, 2017). Our MS/MS analyses shows that the PPIX $[M + H]^+$ ions at *m/z* 563 formed the product ions at *m/z* 504, 445, and 489 (see Figure S1). The PPIX was accordingly characterized with these specific MS/MS product ion spectra (see Figure S1). Benzylic cleavage, with the loss of one (563 – 59) and two (563 – 2 × 59) $–CH_2COOH$ groups, made PPIX forming the product ions at *m/z* 504 and 445,

respectively, which is again referred to as primary fragmentation pathway. The ion spectrum at *m/z* 489 here refers to the primary fragmentation product after the loss of a $-CH_2CH_2COOH$ group and proton (563 – 73 – 1), and the peak at *m/z* 489 might be derived from the ion at *m/z* 504 after the loss of a $-CH_3$ radical from later protonated ions (504 – 15). Accordingly, the results show that the PPIX targets could be detected according to their fragmentation products.

To evaluate the effects of acidification time on the measurements, a series of samples were acidified continuously for 24 h at ~ 25°C (lab temperature), and five samples were taken for analysis at each acidification time (Figure 3C). The chromatograph shows that the PPIX concentrations increased gradually with increasing acidification time, and the PPIX signals appeared to be well separated throughout (see the inserts in the Figure 3C). Apparently, the PPIX levels at different acidification time showed with significant differences (the levels at 12 h vs 0 h, $p < 0.001$; 24 h vs 12 h, $p < 0.001$). The kinetics of acidification was shown in the plot of all PPIX levels against acidification time, which was fitted with a first-order rate equation ($R^2 > 0.9699$; see Figure 3C). We therefore calculated that the acidification rate was ~ 0.0675 h^{-1} and the half time of forming PPIX monomers was ~ 4.4 h. No significant difference between signals at 22 h and 24 h ($p > 0.05$) indicates that measurements tends to be stable after ~ 24 h acidification. It should be noted that, based on the first-order acidification equation, 24 h acidification still rendered an underestimation of < 5% of the measurements. Finally, the newly developed protocol could be summarized as below: PPIX in cells is first extracted in acetone at -20°C for 20 h, and then adjusted to pH < 2, and the extracted samples are finally acidified in the dark for 24 h prior to being analyzed *via* HPLC.

The PPIX dynamics from the freshwater to saline waters

With the method, we measured cellular PPIX in the samples taken from the Jiulong River estuary. A total of 32 samples in 16 stations (surface and bottom) were collected from the upper to lower estuaries until the bay area. The results (Figure 4) show that the cellular PPIX concentrations in water samples ranged from 20 to 170 ng/L, with a decreasing pattern from the upper to lower estuaries until the bay waters. In the upper reaches of the upper estuary, PPIX (Figure 4A) was higher in the surface samples (140 ~ 170 ng/L) than in the bottom samples (60 ~ 140 ng/L) (Figure 4B), while in the lower estuary and inner bay waters, PPIX showed with higher concentrations in the bottom waters (40 ~ 100 ng/L) than in the surface waters (30 ~ 50 ng/L). In the outer bay waters, the concentrations of PPIX have no significant difference between the surface and bottom samples (except the station of X11).

PPIX has been previously measured in microbial pure cultures, with the average concentrations of 286, 214, and 2 ng/g in the cells of *Aggregatibacter actinomycetemcomitans*, *Porphyromonas gingivalis*, and *Saccharomyces cerevisiae*, respectively, accounting for 55%, 87%, and 1% of total porphyrins, respectively (Fyrestam et al., 2015). Our results show with the dynamics of PPIX in different waters, which could be attributable to the fact that microorganisms have differences in synthesizing and secreting PPIX. In particular, we for the first time quantified the cellular PPIX in the samples taken from natural waters with a high recovery rate and extraction efficiency, compared with the existing methods (Hur et al., 2014; Fyrestam et al., 2015; Fyrestam and Östman, 2017), our method is simpler and more efficient, and the method is expected to have a wider application in aquatic sciences.

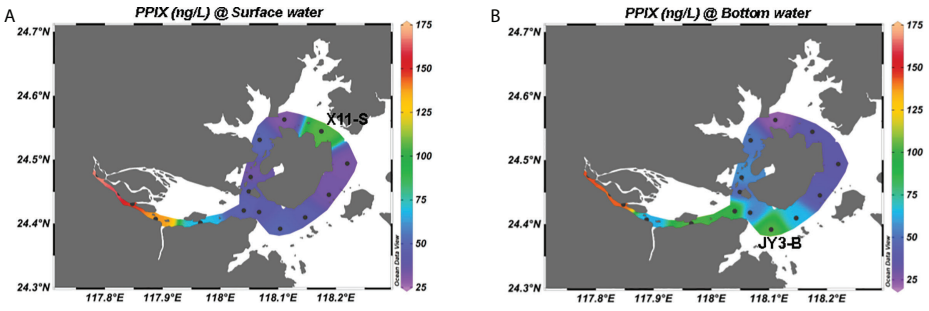

FIGURE 4
Spatial distributions of cellular PPIX in the Jiulong River estuary: **(A)** in the surface water; **(B)** in the bottom water.

Discussion and ecological relevance of PPIX

Tetrapyrroles are initially synthesized from a common precursor δ-aminolevulinic acid (ALA), and then form a series of metabolites including hemes, chlorophylls, vitamin B_{12}, siroheme, cofacter F_{430}, and porphyrins. As an essential metabolic intermediate, PPIX (with a cyclic tetrapyrrole core structure) is crucial in many biological processes including photosynthesis, respiration, methanogenesis, the assimilation and metabolism of inorganic nitrogen and sulphur (Senge et al., 2014). As PPIX is synthesized through multiple steps, the levels of PPIX in cells might be subject to vary depending on the cellular metabolic activities and ambient conditions. In a natural ecosystem, any disturbance in the PPIX levels might be further linked to changes in the ecological biodiversity and community structure.

Here, we measured cellular PPIX in the water samples taken from the Jiulng River estuary with acetone extraction and HPLC–MS/MS quantification. Firstly, our results show with different chromographic signals of PPIX at different acidification conditions. The sample (at pH < 2) has one single chromographic peak of PPIX; The sample (at pH of 3 ~ 7) has double peaks; The sample (at pH > 7) didn't show clear target peak (Figure S2). Secondly, we optimized the acidification time by conducting a time-series experiment. The separation of the PPIX signals was improved with increasing acidification time (Figure 3C). As there exist numerous metalloporphyrins (i.e., Fe-PPIX, Mg-PPIX and Zn-PPIX), the structures of these metalloporphyrins differ depending on their associated central metal ions, which might lead to a variable rate at which protons replace metal ions to form PPIX monomers (Fleischer, 1970; Yang et al., 2021). Besides, metalloporphyrins are also unstable according to their associated ions, likely following an order of stability as the following: medium–sized metals (like Fe^{2+}, Cu^{2+}, Zn^{2+}) > large sized metals (Hg^{2+}, Pb^{2+}, Cd^{2+}) > alkali metals (Na^{+}, K^{+}, Li^{+}) (Yang et al., 2021). For example, Cu^{2+} is more stable as a soft Lewis acid once tightly bound to porphyrin rings, which can only be desorbed or destructed under drastic conditions (Yang et al., 2021). The metalloporphyrins with large metals and alkali metals could be more easily transformed into the corresponding free porphyrin molecules in dilute acids (Barnes and Dorough, 1950; Yang et al., 2021). Here we confirmed that an acidification time of 24 h is sufficient for the dissociation of most metalloporphyrins and no destruction of their structures.

The Jiulong River is a typical subtropical macrotidal estuary with a depth of 3-16 m and its mean annual tidal range was 3.9 m with a maximum of 6.4 m (Yu et al., 2019). The freshwater inflow from the river and tidal intrusion from the Taiwan Strait together made the estuary with an obvious salinity gradient ranging from 0 to 32 in the upper to lower portions of the estuary (Figure S3) (Yu et al., 2020). Tidal intrusion leads to the water column more stratified in the lower estuaries (Yu et al., 2019). The plot of PPIX against salinity showed with a consistently decreasing pattern in the estuary ($p < 0.01$; Figure 5A). As described previously, the study area is divided into two sections: the estuary and the bay. The estuary includes the stations of A3 to JY1, with a salinity range of 2 to 26 (Figure S3). Here, the concentrations of PPIX decreased with increasing salinity, with a significantly negative correlation ($p < 0.01$). The bay area (JY2-X13) exhibited with a salinity range of 27 to 32, and no significant correlation was observed between PPIX and salinity ($p > 0.05$). It should be noted that a few outliers (at JY3-B and X11-S) occurred in the Bay waters. The Jiulong River received a large quantity of organic matter and nutrients from the upper streams, storm water, and even domestic sewages, which in turn triggered the growth and reproduction of microorganisms and phytoplankton locally (Chen et al., 2018; Yu et al., 2020). Higher levels of PPIX at the upper reaches of the upper estuary in this study could be attributable to the higher riverine nutrient inputs and associated higher algal and bacterial activities. In the lower estuary (with a maximum width of approximately 100 km^2, the water flow rate decreased and water retention time increased. Under such conditions, suspended particles are subject to deposit, and algae tend to grow easily with increased sunlight (Sullivan et al., 2001; Zeng et al., 2006). The higher concentrations of PPIX in the bottom waters of the lower estuary (Figure 4) might be associated with the resuspension and dissolution of the deposited particles and organic materials.

The ambient environment could be critical affecting the levels of cellular PPIX. We therefore statistically analyzed the samples taken from stations of JY2-X13 with similar salinity of ~ 30 to minimize the effects of salinity (excluding the stations [blue dots] with lower salinity in the Figure 5A). As PPIX is highly involved in the metabolic processes in microbes, we first plotted the cellular PPIX dynamics against the autotrophic and heterotrophic microbial cells. The abundances of autotrophic microbes were higher in the surface waters than in the bottom of the Jiulong River estuary (Figure S3), and especially in the upper estuary and inner bay waters, which might be attributable to the fact that the autotrophic organisms could receive sufficient sunlight in surface for photosynthesis (Kolber et al., 2000). The bacterial abundances were higher in the bottom waters than in the surface waters in the estuary, probably as a result of more organic matter accumulation there. We did not observe any consistent pattern of cellular PPIX against autotrophic and bacterial abundances (Figure 5B, $p > 0.05$), and here further experiments are needed for elucidating possible connections between PPIX with biological activities. Chlorophylls are the most abundant and noticeable biological pigments in virtually all the photosynthetic organisms, including all the higher plants, green algae and cyanobacteria (Grimm et al., 2006). Through photosynthesis in chlorophylls, the solar energy was converted into chemical energy. Therefore chlorophylls can be used as an indicator of the photosynthetic potential and primary production

(Senge et al., 2014). The biosynthetic pathway of chlorophyll involves a branch of non-metallized PPIX in which one Mg^{2+} ion is inserted in the center of the porphyrin ring catalyzed by ATP-dependent Mg-chelatase (Chen et al., 2018). Thereafter, the Mg-PPIX experiences a series of reactions leading to the syntheses of chlorophyll and bacteriochlorophyll. Pheophytin is a compound formed by the degradation of chlorophyll, by the action of weak acids or enzymes, where Mg^{2+} is replaced by two hydrogen atoms (pheophytin *a* and pheophytin *b* are derived from chlorophyll *a* and chlorophyll *b*, respectively) (Hsu et al., 2013). Therefore, the relationships between PPIX and the three biopigments (Chl-*a*, pheophytin and BChl-*a*) might be used as an indicator of the status of environmental microbial communities. We plotted the ratios of PPIX/Chl-*a*, PPIX/BChl-*a*, and PPIX/Pheophytin as biological variables against the concentrations of PPIX to exposit the variation of PPIX content per unit biomass. The results (Figure 5C) show a significant positive correlation ($p < 0.05$) between PPIX and the three pigments. Except for the surface sample at X11, all the three ratios increased with increasing PPIX concentrations in the high-salinity waters (Figure 5C). Such results indicated of a close interaction between PPIX and the three bio-pigments. Interestingly, the three ratios in the bottom samples were greater than those in the surface water. It is possibly related to the fact that the bottom waters contained higher levels of organic matter and nutrients. In addition, the surface waters might be exposed to excessive sunlight leading to a photoinhibition effect on photosynthetic microorganisms (as the samples at A9-JY3, X1-X7 were collected at 12:30 - 16:30 pm) (Jiao et al., 2010), with possible degradation of free PPIX under light. As a metabolite in the biosynthetic pathway of chlorophylls, PPIX could be an important indicator of the growth status of photosynthetic organisms.

Nutrients (including N and P as backbone elements in biology) are traditionally believed to be crucial in dictating planktonic biomass, taxonomic composition and primary production (Smith, 2003). Low concentrations of nutrients (below the threshold) restricted the growth rate of primary production and autotroph biomass; and conversely, excessive concentrations could lead to ecological disturbances including harmful algal blooms, hypoxia and the production of toxins (Wilkinson, 2017). The Jiulong River estuary receives a large amount of particles from both riverine inputs and adjacent urban sewages (Cao et al., 2005). As a result, the concentrations of POC and PN showed with a gradually decreasing pattern from the upper to lower estuaries (Figure S3). Specifically,

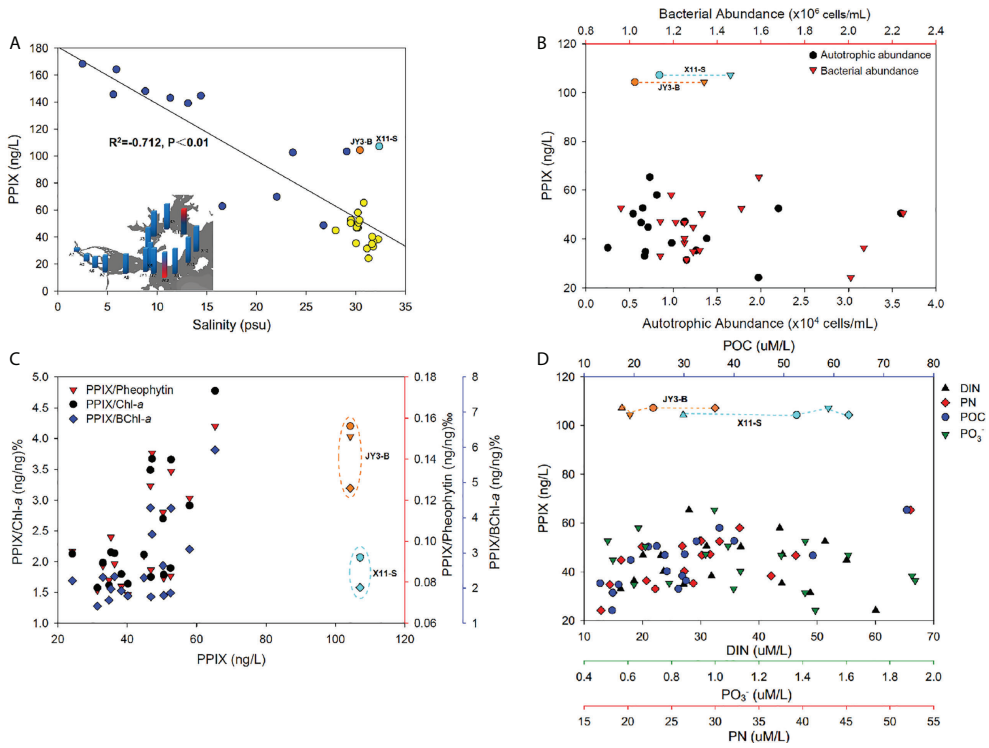

FIGURE 5
(A) The negative correlation of PPIX concentrations against salinity (the inset refers to the salinity distribution, the red marks denoted the outliers of PPIX that responded to JY3-B (orange dot) and X11-S (light blue dot), and the yellow dots referred to the high salinity samples). **(B)** The relationship of PPIX with the autotrophic and bacterial abundances. **(C)** The variations of PPIX/Chl-*a* (black dots), PPIX/BChl-*a* (blue diamonds) and PPIX/Pheophytin (red triangles) along with PPIX. **(D)** The synergistic changes of PPIX with the DIN (black triangles), PN (red diamonds), POC (blue dots) and PO_4 (3–) (green triangle down) concentrations.

increased levels of POC and PN occurred at the bottom near the river mouth as a result of particle inputs and resuspensions there. Similar decreasing trends were found for the nitrate, nitrite, ammonium, and phosphate levels, with extremely high levels at the river mouth (Figure S3). In particular, the cellular PPIX concentrations were significantly positively correlated ($p < 0.01$) with POC and PN in those high-salinity waters (see Figure 5D), but no significantly linear correlations of PPIX with DIN and phosphate ($p > 0.05$). Instead, PPIX varied dynamically along with DIN and phosphate: PPIX increased slightly with increasing DIN and phosphate, and then decreased (Figure 5D). This could be attributable to the biological quotas of PPIX were relatively stable, and increased nutrients will not necessarily trigger more PPIX synthesis beyond the quota. It should be noted here that more researches are required to confirm the conclusion. Consistently, the stations of JY3-B showed with unusually high PPIX values (Figures S3, 5A), which was located in the estuary-ocean interface with high POC and PN, but low DIN and phosphate.

The Jiulong River estuary is partially stratified particularly in the lower portion (Yu et al., 2019). The JY3 sample was collected at 16:30 pm, near a neap tide and the salinity in the bottom waters was at least > 1 higher than in the surface waters. Consistently, Zeng et al. (2006) reported that the phytoplankton abundances were significantly negatively correlated with the nitrate, nitrite and phosphate levels in the Xiangxi River (one of the upper branches in the Three Gorges, Yangtze River). Previous researchers reported that the abundances of primary plankton groups (diatoms, cryptophytes, chlorophytes and cyanobacteria) were negatively correlated with the nitrate and phosphate levels in the Jiulong River estuary (Xie et al., 2018). In addition, sediment suspension in the water column could provide extra nutrients for the growth of particle-attached microorganisms, while the degradation of organic matter could further provide logistical support for the growth of free-living microorganisms (Li et al., 2021). Consequently, microbial activities commonly become more intense in places with high levels of particulate organic matter, possibly leading to accelerated metabolism and increased PPIX syntheses. Figure 4 shows that PPIX were higher in the upper estuary (A3-A7 stations), while relatively lower in the lower estuary, and lowest at the outer bay waters (Figures S3, 5D). At the stations of A9-X1 near the river mouth, the PPIX contents in the bottom samples were significantly higher than in the surface samples, which was consistent with higher particle organic matter there. It is worthy to note that the station of X11 also revealed a high PPIX level at surface (Figure 4), as directly receiving materials' input from a lagoon nearby. The highest and lowest PPIX concentrations (168 and 24 ng/L) occurred in the samples taken from the stations of A3 and X8, respectively. The station A3 is located at the river mouth with the highest nutrient levels (Yu et al., 2019).

Conclusion

In this study, we developed a new technique for the determination of PPIX in natural waters, and then successfully applied it to quantify the contents of PPIX in seawater samples took from a subtrophic estuary-the Jiulong River estuary. We for the first time linked PPIX with natural ecosystems, and explored the intrinsic factors that affect the construction and diversity of microbial communities. Generally, we obtained a first dataset of PPIX in a coastal system and further discussed the potential significance of PPIX in the system. In particular, a significantly negative correlation was observed between PPIX and salinity, and a positive correlation of PPIX with downstream derivatives (the three biopigments: Chl-*a*, pheophytin and BChl-*a*) and particulate organic substances (PN and POC). In addition, we observed that PPIX varied dynamically with increasing levels of nutrients (DIN and phosphate), indicative of the association of PPIX with the growth of phytoplankton and bacteria. Our findings emphasized the ecological functioning of PPIX as a natural bioorganic molecule in natural waters. The technique and findings might shed light on our further understanding on the structure of microbial communities, and the cycling of nutrients in natural ecosystems. Future studies should focus on biological data such as those from high-throughput sequencing and metagenome or meta-transcriptomics analyses.

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.

Author contributions

The methods described in this paper were developed by LG with support and guidance from DW. All the experiments and the manuscript preparation were performed by LG. KM helped with the experiment and manuscript writing. XY helped collecting the samples and analyzing the data. HZ helped plotting the figures. The manuscript was reviewed by LG, and substantially revised by DW. All authors contributed to the article and approved the submitted version.

Funding

The work was partly supported by The National Natural Science Foundation of China (U1805242) and the PhD Fellowship of the State Key Laboratory of Marine Environmental Science at Xiamen University.

Acknowledgments

We would like to thank the State Key Laboratory of Marine Environmental Science, Fujian Provincial Key Laboratory for Coastal Ecology and Environmental Studies at Xiamen University for the comprehensive technical and equipment supports. We express our thanks to Ms. Shanshan Lin for her assistances with LC-MS/MS, and the captain, and crew members of the R/V Haiyang No.2 for their assistances in sample collection.

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.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2022.991126/full#supplementary-material

References

Ausio, X., Grimalt, J. O., Ozalla, D., and Herrero, C. (2000). On-line LC-MS analysis of urinary porphyrins. *Analytical Chem.* 72 (20), 4874–4877. doi: 10.1021/ac0005060

Baohong, C., Kang, W., Xu, D., and Hui, L. (2021). Long-term changes in red tide outbreaks in xiamen bay in China from 1986 to 2017. *Estuarine Coast. Shelf Sci.* 249, 107095. doi: 10.1016/j.ecss.2020.107095

Barnes, J. W., and Dorough, G. D. (1950). Exchange and replacement reactions of α,β,γ,δ-Tetraphenyl-metalloporphins1. *J. Am. Chem. Soc.* 72 (9), 4045–4050. doi: 10.1021/ja01165a057

Bonkovsky, H. L., Guo, J. T., Hou, W., Li, T., Narang, T., and Thapar, M. (2013). Porphyrin and heme metabolism and the porphyrias. *Compr. Physiol.* 3 (1), 365–401. doi: 10.1002/cphy.c120006

Cao, W., Hong, H., and Yue, S. (2005). Modelling agricultural nitrogen contributions to the jiulong river estuary and coastal water. *Global Planetary Change* 47 (2), 111–121. doi: 10.1016/j.gloplacha.2004.10.006

Chen, G. E., Canniffe, D. P., Barnett, S. F. H., Hollingshead, S., Brindley, A. A., Vasilev, C., et al. (2018). Complete enzyme set for chlorophyll biosynthesis in escherichia coli. *Sci. Adv.* 4 (1), eaaq1407. doi: 10.1126/sciadv.aaq1407

Chen, B., Kang, W., and Hui, L. (2019). Akashiwo sanguinea blooms in Chinese waters in 1998–2017. *Mar. pollut. Bull.* 149, 110652. doi: 10.1016/j.marpolbul.2019.110652

Chen, N., Krom, M. D., Wu, Y., Yu, D., and Hong, H. (2018). Storm induced estuarine turbidity maxima and controls on nutrient fluxes across river-estuary-coast continuum. *Sci. Total Environ.* 628-629, 1108–1120. doi: 10.1016/j.scitotenv.2018.02.060

Dailey, H. A., and Meissner, P. N. (2013). Erythroid heme biosynthesis and its disorders. *Cold Spring Harb. Perspect. Med.* 3 (4), a011676. doi: 10.1101/cshperspect.a011676

Dai, M., Wang, L., Guo, X., Zhai, W., Li, Q., He, B., et al. (2008). Nitrification and inorganic nitrogen distribution in a large perturbed river/estuarine system: The pearl river estuary, China. *Biogeosciences* 5 (5), 1227–1244. doi: 10.5194/bg-5-1227-2008

Danton, M., and Lim, C. K. (2006). Porphyrin profiles in blood, urine and faeces by HPLC/electrospray ionization tandem mass spectrometry. *BioMed. Chromatogr.* 20 (6-7), 612–621. doi: 10.1002/bmc.656

Datta, S., Loh, C., MacRobert, A., Whatley, S., and Matthews, P. (1998). Quantitative studies of the kinetics of 5-aminolaevulinic acid-induced fluorescence in bladder transitional cell carcinoma. *Brifsh J. Cancer* 78 (8), 1113–1118. doi: 10.1038/bjc.1998.637

De Maere, H., Jaros, M., Dziewięcka, M., De Mey, E., Fraeye, I., Sajewicz, M., et al. (2014). Determination of hemin, protoporphyrin ix, and Zinc(Ii) protoporphyrin ix in parma ham using thin layer chromatography. *J. Liquid Chromatogr. Related Technol.* 37 (20), 2971–2979. doi: 10.1080/10739149.2014.906995

Espinas, N. A., Kobayashi, K., Takahashi, S., Mochizuki, N., and Masuda, T. (2012). Evaluation of unbound free heme in plant cells by differential acetone extraction. *Plant Cell Physiol.* 53 (7), 1344–1354. doi: 10.1093/pcp/pcs067

Fleischer, E. B. (1970). Structure of porphyrins and metalloporphyrins. *Accounts Chem. Res.* 3 (3), 105–112. doi: 10.1021/ar50027a004

Fyrestam, J., Bjurshammar, N., Paulsson, E., Johannsen, A., and Ostman, C. (2015). Determination of porphyrins in oral bacteria by liquid chromatography electrospray ionization tandem mass spectrometry. *Anal. Bioanal Chem.* 407 (23), 7013–7023. doi: 10.1007/s00216-015-8864-2

Fyrestam, J., and Östman, C. (2017). Determination of heme in microorganisms using HPLC-MS/MS and cobalt(III) protoporphyrin IX inhibition of heme acquisition in escherichia coli. *Anal. Bioanal Chem.* 409, 6999–7010. doi: 10.1007/s00216-017-0610-5

Goericke, R. (2002). Bacteriochlorophyll a in the ocean: Is anoxygenic bacterial photosynthesis important? *Limnology Oceanog* 47 (1), 290–295. doi: 10.4319/lo.2002.47.1.0290

Grimm, B., Porra, R. J., Rüdiger, W., and Scheer, H. (2006). *Chlorophylls and bacteriochlorophylls: Biochemistry, biophysics, functions and applications* (The Netherlands: Springer). doi: 10.1007/1-4020-4516-6

Hoseini, M., Sazgarnia, A., and Sharifi, S. (2019). Effect of environment on protoporphyrin IX: Absorbance, fluorescence and nonlinear optical properties. *J. Fluoresc* 29 (3), 531–540. doi: 10.1007/s10895-019-02366-4

Hsu, C.-Y., Chao, P.-Y., Hu, S.-P., and Yang, C.-M. (2013). The antioxidant and free radical scavenging activities of chlorophylls and pheophytins. *Food Nutr. Sci.* 04 (08), 1–8. doi: 10.4236/fns.2013.48A001

Hur, Y., Han, S. B., Kang, S. W., and Lee, Y. (2014). Quantitative analysis of porphyrins in dried blood spots using liquid chromatography-tandem mass spectrometry with pre-column derivatization. *Bull. Korean Chem. Soc.* 35 (10), 3103–3106. doi: 10.5012/bkcs.2014.35.10.3103

Jiao, N., Zhang, F., and Hong, N. (2010). Significant roles of bacteriochlorophylla supplemental to chlorophylla in the ocean. *ISME J.* 4 (4), 595–597. doi: 10.1038/ismej.2009.135

Junga, C.-I., Yang, J.-I., Park, C.-H., Lee, J.-B., and Park, H.-R. (2010). *Formation of the metal complexes between protoporphyrin IX and divalent metal cations in the environment, molecular environmental soil science at the interfaces in the earth's critical zone, Berlin, Heidelberg, 2010.* Eds. J. Xu and Huang, (Berlin, Heidelberg: Springer Berlin Heidelberg) pp 97–99.

Jerry Kaneko, J. (1997). Chapter 8 - Porphyrins and the porphyrias. *Clin. Biochem. Domest. Anim. (Fifth Edition)*, Salt Lake City, Utah, USA, 1997. Eds. J. Jerry Kaneko, John W. Harvey and Michael L. Bruss, (Salt Lake City, Utah: American Academic Press) pp 205–221. doi: 10.1016/B978-012396305-5/50009-9

Knap, A. H., Michaels, A., Close, A. R., Ducklow, H., and Dickson, A. G. (1996). *Protocols for the joint global ocean flux study (JGOFS) core measurements* Vol. Measurement of Chlorophyll a and Phaeopigments by Fluorometric Analysis. Ed. A. G. Dickson (Paris, France: UNESCO-IOC) 119–122.

Kolber, Z. S., Van Dover, C. L., Niederman, R. A., and Falkowski, P. G. (2000). Bacterial photosynthesis in surface waters of the open ocean. *Nature* 407 (6801), 177–179. doi: 10.1038/35025044

Li, J., Gu, L., Bai, S., Wang, J., Su, L., Wei, B., et al. (2021). Characterization of particle-associated and free-living bacterial and archaeal communities along the water columns of the south China Sea. *Biogeosciences* 18 (1), 113–133. doi: 10.5194/bg-18-113-2021

Lim, C. K. (2009) High-performance liquid chromatography and mass spectrometry of porphyrins, chlorophylls and bilins. (United Kingdom: World Scientific Publishing, UK Ltd, UK). *Methods in Chromatography*. 2, 1–256.

Liu, Q., Liang, Y., Cai, W.-J., Wang, K., Wang, J., and Yin, K. (2020). Changing riverine organic C:N ratios along the pearl river: Implications for estuarine and coastal carbon cycles. *Sci. Total Environ.* 709, 136052. doi: 10.1016/j.scitotenv.2019.136052

Sachar, M., Anderson, K. E., and Ma, X. (2016). Protoporphyrin IX: The good, the bad, and the ugly. *J. Pharmacol. Exp. Ther.* 356 (2), 267. doi: 10.1124/jpet.115.228130

Senge, M., Ryan, A., Letchford, K., MacGowan, S., and Mielke, T. (2014). Chlorophylls, symmetry, chirality, and photosynthesis. *Symmetry* 6 (3), 781–843. doi: 10.3390/sym6030781

Smith, V. H. (2003). Eutrophication of freshwater and coastal marine ecosystems a global problem. *Environ. Sci. pollut. Res.* 10, 126–139. doi: 10.1065/espr2002.12.142

Stillman LC, G. M. (1978). Protoheme extraction from plant tissue. *Analytical Biochem.* 91 (1), 166–172. doi: 10.1016/0003-2697(78)90827-8

Suffridge, C. P., Gómez-Consarnau, L., Monteverde, D. R., Cutter, L., Arístegui, J., Alvarez-Salgado, X. A., et al. (2018). And their congeners as potential drivers of microbial community composition in an oligotrophic marine ecosystem. *J. Geophysical Research: Biogeosci.* 123 (9), 2890–2907. doi: 10.1029/2018JG004554

Sullivan, B. E., Prahl, F. G., Small, L. F., and Covert, P. A. (2001). Seasonality of phytoplankton production in the Columbia river: A natural or anthropogenic pattern? *Geochimica Cosmochimica Acta* 65 (7), 1125–1139. doi: 10.1016/S0016-7037(00)00565-2

Tian, J., Ge, F., Zhang, D., Deng, S., and Liu, X. (2021). Roles of phosphate solubilizing microorganisms from managing soil phosphorus deficiency to mediating biogeochemical p cycle. *Biology* 10 (2):158–177. doi: 10.3390/biology10020158

Wang, D., Lu, S., Chen, N., Dai, M., and Gueguen, C. (2018). Human perturbation increases the fluxes of dissolved molybdenum from land to ocean - the case of the jiulong river in China. *J. Environ. Manage* 210, 139–145. doi: 10.1016/j.jenvman.2017.12.074

Weinberger, R., and Sapp, E. (1990). Capillary electrophoresis of urinary porphyrins with absorbance and fluorescence detection. *J. Chromatogr.* 516, 271–285. doi: 10.1016/S0021-9673(01)90225-0

Weinstein JD, B. S. (1984). Biosynthesis of protoheme and heme a precursors solely from glutamate in the unicellular red alga cyanidium caldarium. *Plant Physiol.* 74 (1), 146–151. doi: 10.1104/pp.74.1.146

Wilkinson, G. M. (2017). "Eutrophication of freshwater and coastal ecosystems," in *Encyclopedia of sustainable technologies*, (Oxford, UK: Elsevier) pp 145–152.

Xie, Y., Wang, L., Liu, X., Li, X., Wang, Y., and Huang, B. (2018). Contrasting responses of intertidal microphytobenthos and phytoplankton biomass and taxonomic composition to the nutrient loads in the jiulong river estuary. *Phycological Res.* 67 (2), 152–163. doi: 10.1111/pre.12363

Xiong, J., Fu, G., Yang, Y., Zhu, C., and Tao, L. (2011). Tungstate: Is it really a specific nitrate reductase inhibitor in plant nitric oxide research? *J. Exp. Bot.* 63, 33–41. doi: 10.1093/jxb/err268

Yang, H., Zhou, Y., and Liu, J. (2021). Porphyrin metalation catalyzed by DNAzymes and nanozymes. *Inorganic Chem. Front.* 8 (9), 2183–2199. doi: 10.1039/D1QI00105A

Yu, D., Chen, N., Cheng, P., Yu, F., and Hong, H. (2020). Hydrodynamic impacts on tidal-scale dissolved inorganic nitrogen cycling and export across the estuarine turbidity maxima to coast. *Biogeochemistry* 151 (1), 81–98. doi: 10.1007/s10533-020-00712-4

Yu, D., Chen, N., Krom, M. D., Lin, J., Cheng, P., Yu, F., et al. (2019). Understanding how estuarine hydrology controls ammonium and other inorganic nitrogen concentrations and fluxes through the subtropical jiulong river estuary, S.E. China under baseflow and flood-affected conditions. *Biogeochemistry* 142 (3), 443–466. doi: 10.1007/s10533-019-00546-9

Zeng, H., Song, L., Yu, Z., and Chen, H. (2006). Distribution of phytoplankton in the three-gorge reservoir during rainy and dry seasons. *Sci. Total Environ.* 367 (2-3), 999–1009. doi: 10.1016/j.scitotenv.2006.03.001

Zvezdanovic, B. J., Petrovic, M. S., and Markovic, Z. D. (2017) Hematoporphyrin derivatives: The ultrahigh performance liquid chromatography: Diode array: Electrospray ionization: Mass spectrometry analysis. *Advanced technologies*. 6:26–30. doi: 10.5937/savteh1702026Z

frontiers | Frontiers in Microbiology

TYPE Review
PUBLISHED 12 January 2023
DOI 10.3389/fmicb.2022.1026596

OPEN ACCESS

EDITED BY
Marcelino T. Suzuki,
Sorbonne Universités,
France

REVIEWED BY
Francisco Rodriguez-valera,
Miguel Hernández University of Elche,
Spain
Kelle Cardin Freel,
University of Hawaii at Manoa,
United States

*CORRESPONDENCE
Nastassia Virginia Patin
nastassia.patin@noaa.gov

SPECIALTY SECTION
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 24 August 2022
ACCEPTED 22 November 2022
PUBLISHED 12 January 2023

CITATION
Patin NV and Goodwin KD (2023) Capturing marine microbiomes and environmental DNA: A field sampling guide. *Front. Microbiol.* 13:1026596. doi: 10.3389/fmicb.2022.1026596

Capturing marine microbiomes and environmental DNA: A field sampling guide

Nastassia Virginia Patin[1,2,3]* and Kelly D. Goodwin[1,3]

[1]Atlantic Oceanographic and Meteorological Laboratory, Ocean Chemistry and Ecosystems Division, National Oceanic and Atmospheric Administration, Miami, FL, United States, [2]Cooperative Institute for Marine and Atmospheric Studies, Rosenstiel School of Marine, Atmospheric, and Earth Science, University of Miami, Miami, FL, United States, [3]Stationed at Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, La Jolla, CA, United States

The expanding interest in marine microbiome and eDNA sequence data has led to a demand for sample collection and preservation standard practices to enable comparative assessments of results across studies and facilitate meta-analyses. We support this effort by providing guidelines based on a review of published methods and field sampling experiences. The major components considered here are environmental and resource considerations, sample processing strategies, sample storage options, and eDNA extraction protocols. It is impossible to provide universal recommendations considering the wide range of eDNA applications; rather, we provide information to design fit-for-purpose protocols. To manage scope, the focus here is on sampling collection and preservation of prokaryotic and microeukaryotic eDNA. Even with a focused view, the practical utility of any approach depends on multiple factors, including habitat type, available resources, and experimental goals. We broadly recommend enacting rigorous decontamination protocols, pilot studies to guide the filtration volume needed to characterize the target(s) of interest and minimize PCR inhibitor collection, and prioritizing sample freezing over (only) the addition of preservation buffer. An annotated list of studies that test these parameters is included for more detailed investigation on specific steps. To illustrate an approach that demonstrates fit-for-purpose methodologies, we provide a protocol for eDNA sampling aboard an oceanographic vessel. These guidelines can aid the decision-making process for scientists interested in sampling and sequencing marine microbiomes and/or eDNA.

KEYWORDS

microbiome, eDNA, field methods, sampling, molecular ecological methods

Introduction

The study of marine microbiomes is critical for understanding global biogeochemical cycles, ecosystem health, and microbial ecology and evolution. Advances in sequencing technology and computational approaches to analyze massive DNA sequence data sets have facilitated important discoveries of marine microbial physiology and interactions (Moran,

2015). Moreover, environmental DNA ("eDNA") is becoming an important tool for detection of marine macrofauna (e.g., Boussarie et al., 2018; Postaire et al., 2020; Aglieri et al., 2021). However, sampling protocols have yet to be standardized, due largely to an inability to gauge "correct" or "sufficient" data generation, which also depend on factors downstream of sampling such as sequencing strategy. Nevertheless, there are multiple effective methods for collecting eDNA with benefits and disadvantages under various circumstances.

Early oceanographic studies gave microbes little to no consideration. This neglect began to change about half a century ago, when evidence of microbial impacts on photosynthesis and organic matter transformation in the ocean led to new attention on microorganisms (Pomeroy, 1974). The "great plate count anomaly" (Staley and Konopka, 1985) highlighted the shortcomings of cultivation-based assessments of microbial communities, and techniques like microscopy and enzymatic activity assays began to reveal the massive role of marine bacteria in organic matter transformation (Hagström et al., 1979; Fuhrman and Azam, 1982; Azam et al., 1983) and cycling of global carbon (Azam, 1998) and nitrogen (Horrigan et al., 1988; Zehr and Kudela, 2011). With the advent of molecular methods, the discovery of diverse microbial assemblages (Giovannoni et al., 1990; Schmidt et al., 1991) and the recognition of archaea as consistent and important components of marine microbiomes (DeLong, 1992; Fuhrman et al., 1993; Fuhrman and Davis, 1997) further broadened perspectives of microbial evolution and ecology. The extent of marine microbial diversity revealed by molecular methods also highlighted the importance of cultivation efforts, which remain critically important for experimentally validating sequence-based hypotheses. Innovative cultivation strategies have led to the isolation of diverse marine microbial lineages including the ubiquitous heterotrophic bacterium SAR11 (Rappé et al., 2002), the major prokaryotic primary producer *Prochlorococcus* (Chisholm et al., 1992), and the ammonia-oxidizing archaeon *Nitrosopumilus* (Könneke et al., 2005). Nevertheless, it remains challenging to cultivate the marine microbiome, which include members of the Bacteria and Archaea as well as fungi, protists, unicellular phytoplankton, and viruses (Stulberg et al., 2016). Molecular methods, particularly genomics and metagenomics, have thus become a critical window into marine microbiome structure and function and, in some cases, can guide cultivation efforts (e.g., Carini et al., 2013).

Today, the study of marine microbial communities is facilitated by massively parallel DNA and RNA sequencing. The collection of environmental DNA and sequencing of marker genes (for community composition), metagenomes (for functional potential), and metatranscriptomes (for functional activity), collectively known as the field of 'omics, have become fundamental to improving our understanding of microbial ecology and biogeochemistry in the ocean. Marker gene amplification and Sanger sequencing gave the first insights into the uncultivated diversity of marine prokaryotes (Giovannoni et al., 1990; DeLong, 1992) and eukaryotes (Rappé et al., 1998; Diez et al., 2001; Moon-van der Staay et al., 2001). Metabarcoding studies using high-throughput sequencing provided orders of magnitude more data on planktonic communities compared to clone libraries or denaturing gradient gel electrophoresis (DGGE; Sogin et al., 2006; Stoeck et al., 2010), but functional characterization remained out of reach until the availability of shotgun metagenomic sequencing. The first major biochemical discovery from marine 'omics data was the gene for proteorhodopsin, an enzyme that allows cells to harvest energy from sunlight without photosynthetic machinery (Béjà et al., 2000). Proteorhodopsins were subsequently found to be widely distributed taxonomically and geographically, altering our understanding of light-based energy flow in the oceans (Rusch et al., 2007; DeLong and Beja, 2010; Olson et al., 2018). Since then, 'omics work has revealed incredible taxonomic and functional diversity of marine microbiomes (Moran, 2015). Interest in marine eDNA is reflected in the exponential rise in publications using this term (Figure 1) and has expanded to include surveys of higher organisms for biodiversity assessments. As we enter an era of rapidly changing climate and human-driven shifts in marine ecosystems, eDNA and other 'omics approaches will be critical tools for understanding and predicting community responses from local to global scales.

Even as eDNA-based analyses continue to reveal novel diversity, there are limitations to culture-independent methods. Results can vary widely with differences at every step in the process, from sample collection to preservation to processing and sequencing (Stewart, 2013; McCarthy et al., 2015; Padilla et al., 2015; Torres-Beltrán et al., 2019; Eble et al., 2020). Microbial life has been shown to exist along a physical continuum in the ocean, ranging from truly planktonic to particle biofilms (Stocker, 2012). Homogeneous sampling of this continuum is challenging even within a single study; when methods differ across studies, the chances of equitable sampling further decrease. Moreover, the importance of rare and temporally ephemeral microbial taxa (Galand et al., 2009; Campbell et al., 2011) highlights the need for standardized methods so that results can be compared across time, space, and environmental conditions.

Environmental DNA sequencing has benefited from steep decreases in sequencing costs (Mardis, 2017; Karst et al., 2021) and rapid advances in bioinformatic analysis methods. Amplicon sequencing of rRNA marker genes has been used extensively to characterize marine microbiomes; more recently, improved eukaryotic gene taxonomies now allow accurate surveys of organisms like teleost fish (Miya et al., 2020; Gold et al., 2021) and mammals (Closek et al., 2019). Such metabarcoding studies can provide more information with less sampling effort relative to traditional visual or trawl surveys and combining eukaryotic and microbial community survey data can reveal new linkages among trophic levels (e.g., Djurhuus et al., 2020). However, regardless of their magnitude and accuracy, sequencing data are only as reliable as their source material, i.e., the eDNA.

As the number of marine eDNA studies rapidly increases (Figure 1), the need to share and discuss best practices has grown (Pearlman et al., 2021). Harmonized methods are critical if data are to be compared over space and time for the purposes of

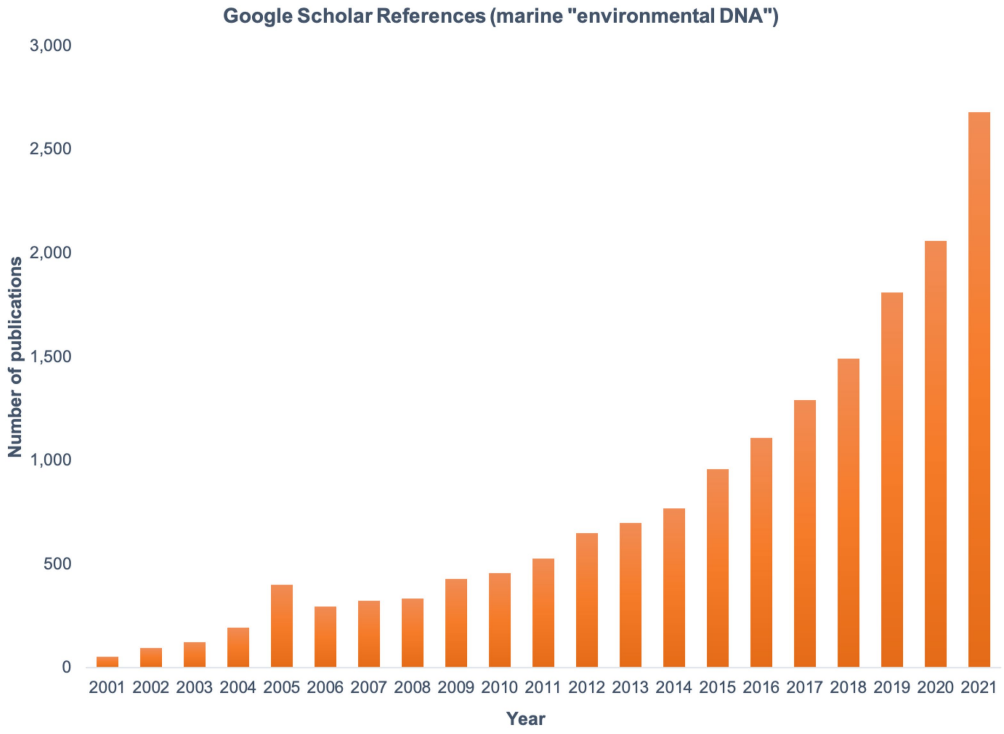

FIGURE 1
The number of publications found by searching the term marine "environmental DNA" in Google Scholar, shown by year from 2001 to 2021.

making trusted management decisions. Multiple programs are working to establish common guidelines and provide the resources and training to facilitate standardized approaches worldwide (Table 1). Such standardization is especially critical as we consider how global microbiomes will respond to a changing climate and impact broader oceanographic processes.

The recognition of this need is illustrated in overarching programs endorsed by the United Nations (UN) Decade for Sustainable Ocean Development that call for harmonized practices, such as the Ocean Biomolecular Observing Network (OBON; Leinen et al., 2022) and Marine Life 2030 (Canonico et al., 2022). There are multiple large-scale efforts to standardize marine microbiome sampling methodologies, and several provide open access resources that can be valuable to those interested in eDNA sampling (Table 1). The Marine Biodiversity Observation Network (MBON) is one of the first biodiversity monitoring programs to perform method comparisons to establish standardized workflows for sampling and analysis of eDNA, including for prokaryotic targets (e.g., Djurhuus et al., 2017; McElroy et al., 2020). Several current efforts feature protocols of the Tara Oceans expedition (Pesant et al., 2015), such as AtlantECO Mission Microbiomes (Pesant et al., 2022) and EMO BON (Santi et al., 2021; Table 1). Others, such as the Bio-GO-SHIP program (Clayton et al., 2022), have expanded to incorporate a subset of these protocols (Pesant et al., 2022).

Nonetheless, a consensus has yet to be reached on a standard sampling workflow, and sequence data continue to be produced at an accelerated pace using a wide range of sampling techniques. The array of available programs working on harmonization is a testament to the importance of and commitment to developing standard practices to ensure that results can be compared and interlaced across studies. For example, the Ocean Best Practices System (Samuel et al., 2021) is a program run by the International Oceanographic Data and Information Exchange and supported by the UN with the stated vision of having "agreed and broadly adopted methods across ocean research, operations and applications" (Pearlman et al., 2019). The project Planet Microbe seeks to associate standardized environmental data with marine 'omics data sets and has called for large-scale intercalibration efforts to develop community-accepted methods across the full sample-to-data pipeline. Their stated goal of creating "cross-comparable 'omics datasets[…]to better elucidate global questions on microbial driven biogeochemical processes in the ocean" (Ponsero et al., 2020) captures the growing need for microbiome and eDNA researchers to integrate their results with those of other disciplines and develop a holistic understanding of Earth systems.

Approach

Literature review is a first step on the path toward standard methods and best sampling practices. This approach is challenging because no one study covers the multitude of possible parameter

TABLE 1 A list of programs working toward standardization of marine eDNA sampling and 'Omics methods.

Program	Geographic scope	Primary aims and/or mission	Program website	Associated protocols/ manuals
Marine Biodiversity Observation Network (MBON)	Global	Mission statement: "Foster and coordinate a global community of practice for collecting, curating, analyzing, good management, and communicating marine biodiversity data and related services to the scientific community, policymakers, the public, and other stakeholders."	https://marinebon.org/	MBON protocols
Ocean Best Practices System	Global	Vision: "To have agreed and broadly adopted methods across ocean research, operations and applications." Supported by UNESCO, IODE, and GOOS.	https://www.oceanbestpractices.org/	None; documents from Better Biomolecular Ocean Practices (BeBOP) forthcoming.
Bio-GO-SHIP	Global	"To quantify the molecular diversity, size spectrum, chemical composition, and abundances of plankton communities across large spatial, vertical, and eventually temporal scales. This will be achieved through systematic, high-quality, and calibrated sampling of 'omics, plankton imaging, particle chemistry, and optical techniques as operational oceanographic tools."	https://biogoship.org/overview/	None
AtlantECO	Atlantic and Southern Oceans	"[…]AtlantECO…aims to develop and apply a novel, unifying framework that provides knowledge-based resources for a better understanding and management of the Atlantic Ocean and its ecosystem services." European Union-led with collaborators in Brazil and South Africa. Protocols are from *Tara* Oceans.	https://www.atlanteco.eu/	AtlantECO Protocols
Ocean Biomolecular Observing Network (OBON)	Global	"[A] global programme[…]that uses techniques to analyse biomolecules such as DNA, RNA, and proteins (e.g., eDNA analysis, metabarcoding, omics) to greatly enhance coastal and open ocean biodiversity observations."	https://www.obon-ocean.org/	None; documents forthcoming
European Marine Omics Biodiversity Observation Network (EMO BON)	Europe	Primary aim: "To ensure steady, continuous generation of 'baseline' data on biodiversity at EMBRC [European Marine Biological Resource Center] sites following FAIR (Findable, Accessible, Interoperable, and Reusable) data principles."	https://www.embrc.eu/emo-bon	EMO BON Handbook
The eDNA Society	Japan	Aims: "[F]ostering and developing eDNA science as a discipline that contributes to the human well-being, such as sustainable use of ecosystems and environmental conservation."	https://ednasociety.org/en/	eDNA Society Manuals
Southern eDNA Society	Australia and New Zealand	Mission statement: "As a society, we aim to promote science and industry collaboration across Australia and New Zealand to advance best practice eDNA methods and adoption in government, private and community sectors."	https://sednasociety.com/	Protocol Development Guide
Nansen Legacy Sampling Protocols	Polar oceans	"Ten major Norwegian research institutions[…]are sharing their resources, competence and infrastructure in an unprecedented endeavor to provide a cross-disciplinary scientific basis for long-term, holistic, and sustainable management of marine ecosystems and human presence in the northern Barents Sea and adjacent Arctic Ocean."	https://arvenetternansen.com/	Sampling Protocol Collection
DNAqua-Net	Europe	"The goal of DNAqua-Net is to nucleate a group of researchers across disciplines with the task to identify gold-standard genomic tools and novel eco-genomic indices and metrics for routine application for biodiversity assessments and biomonitoring of European water bodies."	https://dnaqua.net/	DNAqua-Net Handbook

If there are documents or protocols associated with the program's efforts, those are provided in the final column. If no documents are provided, the efforts are still in progress or protocols are not (yet) publicly available.

10.3389/fmicb.2022.1026596

variations, even for a single molecular target. The variety of targets (viruses to vertebrates) and molecular methods (qPCR to metaomics) confounds intercomparisons. Extrapolating findings to applications other than those in a study should proceed with caution. Moreover, factors like DNA state (e.g., intra-vs. extracellular), decay rates, and transport in the environment are poorly constrained (Andruszkiewicz Allan et al., 2021; Barnes et al., 2021; Jo and Minamoto, 2021; Mauvisseau et al., 2022). Overall, the issues surrounding microbiome and eDNA analyses are complex, particularly for macro-organisms, and a number of reviews and perspectives tackle the challenge of summarizing the state of the science (e.g., Dickie et al., 2018; Harrison et al., 2019; Ruppert et al., 2019; Beng and Corlett, 2020; Kumar et al., 2020; Pawlowski et al., 2020; Bowers et al., 2021; Rodriguez-Ezpeleta et al., 2021).

Although by no means exhaustive, we summarize data from the literature and offer examples from our own experience to guide experimental design. The literature considered here focuses on sample filtration and preservation of marine eDNA using readily available supplies and reagents, although some impactful freshwater studies were included. We address sampling for analysis of prokaryotic and microeukaryotic organisms and include considerations for multi-trophic sampling. We recognize that guidelines are ecosystem specific (Harrison et al., 2019), and recommendations drawn from this analysis may not apply to other marine biomes, such as sediments or host-associated systems. In Supplementary Table S1, we provide an annotated list of studies in which methodological tests were performed. We note where conflicting results necessitate further investigation. As cost is an important consideration for many researchers when designing surveys, we provide a table comparing the cost of different filters and housings to aid study design (Supplementary Table S2).

The main field sampling elements addressed here include water filtration, filter type, sample storage, and DNA extraction. While the latter is rarely performed in the field, it is closely linked to the choice of filter and storage method and thus was an important parameter to include in a guide to overall sampling design. In the final section, we provide an example field sampling workflow that has been used effectively for collecting and preserving marine microbiome eDNA aboard an oceanographic vessel. Filtration based on size fraction is typical for marine eDNA surveys; therefore, eDNA capture methods such as precipitation (Ficetola et al., 2008; Deiner et al., 2015; Eichmiller et al., 2016; Hinlo et al., 2017), centrifugation (Klymus et al., 2015; Eichmiller et al., 2016), or tangential flow capture (Bruno et al., 2017) are not a primary focus. We do not discuss protocols for downstream sample processing such as library preparation, sequencing, bioinformatic data analysis, or standardized metadata. These important topics are outside the scope of this review but are considered elsewhere (e.g., ten Hoopen et al., 2017; Grey et al., 2018; Kelly et al., 2019; Zinger et al., 2019; Mathieu et al., 2020; Berube et al., 2022).

In an effort to build capacity and share knowledge broadly, this overview is appropriate for those new to the field, consistent with goals in programs supported by the UN Decade for Sustainable Ocean Development (Claudet et al., 2020; Canonico et al., 2022; Leinen et al., 2022). We provide both scientific and logistical background to better equip researchers to choose methods appropriate for their needs and circumstances. In accordance with the theme of this special issue, we seek to provide recommendations but do so with humility given the rapid pace of advancement in the field of microbiome science. The aim is to provide a resource for scientists looking to incorporate eDNA into their research or venturing into the field for the first time.

Logistic and resource considerations

Resources and facilities available for field sampling can vary widely and should be considered during sampling design. Factors like filtration time and sample storage temperature can affect microbiome community composition and diversity results (Rochelle et al., 1994; Oldham et al., 2019), but optimal conditions are often dependent on electrical power, space, and other resources. As with all field work, the protocols for microbiome sampling will differ substantially from those used for sampling in a well-equipped scientific laboratory. In the next sections we will cover the specific aspects of sampling that require particular attention relative to work performed at a typical institutional biological laboratory.

Clean technique in the field

Microbiome studies are subject to contamination at multiple points along the sample collection and processing pipeline due to the sensitivity of PCR amplification. Standard molecular laboratory practices are designed to reduce the risk of contamination, and systemic contamination appears rare (Sepulveda et al., 2020). However, it can be challenging to protect against contamination during field operations for a variety of reasons. For example, common equipment and supplies available in the lab, such as laminar flow hoods, UV lighting, and supplies of molecular grade water are often unavailable in the field. The level of care needed is dependent on the study; for example, collection of eDNA to detect an invasive species during operations in which the species itself was handled would require the highest level of containment. Moreover, contamination can be mitigated somewhat during sequencing library preparation. Metabarcoding protocols should minimize the number of PCR cycles to prevent exponential amplification of contaminant sequences, while shotgun metagenomic sequencing requires only minimal amplification and is thus less susceptible to low levels of contaminant DNA. In general, microbiome researchers should have a working knowledge of the hygiene practices used in laboratories working with pathogens or radiolabeled chemicals because these specialize in eliminating the spread of trace amounts of target material.

Despite the challenges, simple steps can be applied in field settings to minimize potential contamination. A variety of detailed marine microbiome sampling protocols are available to guide researchers (e.g., Santi et al., 2021; Pesant et al., 2022), with a few key highlights provided here. In general, aseptic practices form the foundation of clean molecular technique. This starts with a clean workspace physically separated from other biological workstations. As with any molecular laboratory, clean gloves and lab coats should be donned upon entering the workspace. Staff that work directly with the target of interest (e.g., fish processing) must ensure that their clothes, shoes, and hair are free of contamination before entering the molecular workspace. The workspace and equipment should be decontaminated prior to initiating work to ensure that exogenous DNA does not contaminate samples. Items coming in and out of the workspace (e.g., coolers, containers) should be minimized and frequent disinfection of the external surfaces can ensure such items do not become a source of sample contamination. A more detailed discussion of common methods for chemical decontamination are provided in the addendum below.

Negative controls should be generated to account for potential sources of contamination according to risk tolerance. Determining the tolerance for risk requires defining the number of samples one is willing to discard if a control shows gross contamination. To minimize costs, a subset of the negative controls should routinely be processed, leaving others as insurance. Although eDNA can travel through the air (Klepke et al., 2022), standard aseptic technique is designed to minimize such contamination. Pilot studies should be conducted to evaluate whether or not collection of routine air blanks is warranted. A variety of reviews can be consulted for additional information (e.g., Goldberg et al., 2016; Mathieu et al., 2020; Sepulveda et al., 2020).

Approaches to minimize sample cross-contamination vary widely, including rinsing with sample water, using various regiments of chemical decontamination, and usage of sterile, single-use consumables. The circumstances are too varied to provide a single recommendation in this regard other than to invest in initial method validation and thereafter practice quality assurance testing routinely. If testing indicates the need for additional work practices or engineering controls, those actions should be taken to ensure robust data generation.

Hold and filtration times

Water samples should be filtered and processed as soon as possible to prevent changes in microbiome composition and/or eDNA degradation (Rochelle et al., 1994; Oldham et al., 2019). Filtration time may be dictated by logistical issues and/or the stability of the target molecule. If only DNA is targeted, a general guide is to limit filtration time to a maximum of 1 hour. However, there is little rigorous data for filtration of marine water column samples to support specific time limitations. Studies on human microbiome samples (Gorzelak et al., 2015; Song et al., 2016) and eukaryotic eDNA (Thomsen et al., 2012; Maruyama et al., 2014; Yamanaka et al., 2016) have shown changes in DNA yield and composition over relatively short time scales; however, other reports offer conflicting results on microbiome stability (e.g., Lauber et al., 2010). Environmental RNA is more prone to degradation than eDNA (Marshall et al., 2021; Zaiko et al., 2022); thus, shorter filtration times are needed. If RNA is targeted for collection, limit filtration to 20 minutes to provide a practical timeframe to process multiple samples (Zaiko et al., 2022) or 15 minutes to be in accordance with AtlantECO protocols (Pesant et al., 2022).

If immediate filtration is not possible, water should be maintained at <10°C (e.g., in a refrigerator or in a cooler with ice packs) to slow microbial growth. Subjecting the water to freezing temperatures is not recommended, in accordance with guidance provided by the United States Environmental Protection Agency [USEPA; Wymer et al., 2010; United States Environmental Protection Agency (USEPA), 2022] due to the possibility of prokaryotic cell lysis resulting in increased loss of DNA. This recommendation also is supported by results from an eukaryotic eDNA study (Hinlo et al., 2017); however, we recognize that sample freezing is practiced in some laboratories that use eDNA to detect invasive animal species (Hunter et al., 2019). Ideally, validation experiments should be conducted to determine the effect of storage conditions on various taxa. In lieu of application-specific information, we recommend limiting hold times to 6 h, consistent with validated methods to assess recreational water quality using molecular methods [e.g., United States Environmental Protection Agency (USEPA), 2022]. If longer hold times cannot be avoided, the use of a DNA preservative for the intended application should be validated. For example, protocols used by the United States Geological Survey (USGS) employ a combination of 3 M sodium acetate and 95% ethanol to preserve the eDNA of animals in freshwater samples (Ladell et al., 2019).

Sampling volume

Filtration time and sampling volume co-vary (Zaiko et al., 2022), representing fundamental tradeoffs. Concentrations of target, non-target, and PCR-inhibiting molecules affect sampling design. The amount of microbial biomass and non-microbial particulates affect filtering speed, total collected DNA, and ratio of microbial: eukaryotic eDNA. Coastal water features higher cell concentrations than pelagic water, with factors like proximity to river mouths, time period since last precipitation, and water temperature all affecting microbial biomass and water chemistry. In contrast, pelagic environments, deep water, and oligotrophic biomes such as coral reefs are more likely to feature lower levels of both microbes and particulate matter and may require filtration of higher volumes to capture target species (Kumar et al., 2022). For example, the typical estimate of microbial cell density in the upper 200 m of marine water columns is 5×10^5 cells/ml (Cho and Azam, 1990; Whitman et al., 1998); however, this number can vary widely with depth (Schattenhofer et al., 2009), season (Malone and Ducklow, 1990;

Buck et al., 1996; Carlson et al., 1996; Fuhrman and Ouverney, 1998; Whitman et al., 1998; Wigington et al., 2016), and time of day (Gilbert et al., 2010; Weber and Apprill, 2020). Nutrient loads from runoff into the nearshore environment can result in an order of magnitude higher cell concentration than is typical for coastal habitats in both temperate (Palumbo et al., 1984) and tropical (Yeo et al., 2013) locations. Thus, while filtering 100 ml or less may be sufficient from a coastal site next to a river mouth, water from a middle ocean basin may require 2 L or more to capture enough microbial DNA for molecular studies (Kumar et al., 2022).

PCR inhibition causes another fundamental tradeoff in sampling design (Wilson, 1997). Water chemical parameters associated with eutrophication, such as total suspended solids and pH, can impact DNA yield and detection sensitivity (Liang and Keeley, 2013; Tsuji et al., 2017). PCR inhibitors can prevent efficient amplification even in samples with high DNA yields, and inhibitors are often found in systems containing humic acids (Schrader et al., 2012; Cox and Goodwin, 2013; Williams et al., 2017), such as wetland environments and host-associated systems like coral mucus (Sunagawa et al., 2009; Weber et al., 2017). Prior to finalizing a sampling design, researchers should verify the absence of inhibition or the ability to overcome it through DNA dilution or additional purification. For targets available in sufficient quantities, filtration volume can be reduced to avoid later dilution of DNA to combat inhibition. Overcoming inhibition is more challenging for rare targets in samples that contain PCR inhibitors. In one study of an invasive animal species, water filtration across multiple filters and extraction of the combined filters improved eDNA yield; however, extracting from filters individually and combining the resulting products resulted in overall eDNA loss (Hunter et al., 2019). Such results may apply to marine systems but would require verification.

Assuming PCR inhibition is not a limiting factor, a benefit to filtering large volumes of water is that samples will be less sensitive to contamination. One consideration is the availability of sample water, which can be a limited commodity depending on the facilities, equipment, and demands of other teams in the field. In addition, filtering large volumes in a timely manner may require infrastructure that is outside the reach of a typical field program. For example, filtering 20 L of water in 15 minutes or less is feasible aboard *Tara* by employing a dedicated system of sterile carboys, 142-mm filters, and large peristaltic pumps. AtlantECO is developing a filtration kit to address the need to build capacity and promote adoption of this sampling method across the Atlantic Ocean (Pesant et al., 2022).

Our analysis did not reveal a simple answer to the question of the sample volume needed to generate "enough" DNA. Recommendations ranged from 1 L to adequately capture the diversity of phytoplankton (Cermeño et al., 2014) or metazoan eDNA (Stoeckle et al., 2022) to needing 20 L or more to characterize prokaryotic diversity (Pesant et al., 2015). Furthermore, the optimal balance of tradeoffs may depend on the subsequent molecular processing method. For example, the DNA required to capture microbiome diversity (Schmidt et al., 2022) from marker gene amplicon (metabarcoding) or shotgun sequencing likely differs from that needed to successfully conduct a qPCR assay for a specific target (e.g., Andruszkiewicz et al., 2020; Roux et al., 2020; Rourke et al., 2022).

To run a simple test on the effect of sample volume on observed microbiome prokaryotic diversity, we plotted rarefaction curves (sequencing depth vs. observed taxa, either amplicon sequence variants (ASVs) or OTUs depending on the study) for three publicly available marine 16S rRNA gene amplicons data sets with a range of sample water volumes filtered (Figure 2). The 1 L sample curve shows the highest tendency toward reaching a plateau (Figure 2A), although this may be due to differences in study design; most of the 1 L samples came from the shallow water column and were not pre-filtered (Truelove et al., 2022) while the larger volumes were filtered from a deeper oxygen minimum zone environment with pre-filtration steps (Padilla et al., 2015; Torres-Beltrán et al., 2019). Nevertheless, while community richness varied widely among samples (Figure 2B), there was no apparent advantage to filtering 5 L over 1 L or even 500 ml. Analysis of additional molecular targets was outside the scope of this review; however, a recent paper suggested that up to 40 L of water may be needed for metazoan eDNA targets (Govindarajan et al., 2022).

Filtration strategy

Filtration strategy varies widely in the literature, with differences in the mode of filtration and in filter pore size, diameter, material, and housing (Supplementary Table S1). These differences reflect considerations for speed, contamination risk, waste generation, demand for clean water, storage space requirements, convenience of subsequent sample processing, costs, desire to maintain past practice, and overall performance. Those attributes are impacted by the type of target(s) to be captured and the concentration of those target(s) relative to non-target components that cause filters to clog and/or inhibit downstream sample processing. Although we do not discuss targeting viral communities in depth, we note that the most widely used protocols involve chemical flocculation of filtrate from the smallest size fraction followed by filtration to capture aggregated viral particles (John et al., 2011; Zhang et al., 2013; Pesant et al., 2022). The ultimate choice of filtration strategy should optimize trade-offs and match the goals and objectives of the eDNA survey.

Filtration mode

A common way to filter water is *via* vacuum filtration, in which a vacuum pump pulls water through a disc filter. Vacuum filtration using a manifold equipped with multiple filter funnels units is a standard practice in ambient water quality monitoring. If required filtration volumes are small, self-contained sterile filter units exist that are easily transportable, designed for one-time use, and can be attached to a hand pump, making them well-suited to field work on small boats or primitive field stations. However, sample volumes used for eDNA surveys can be >1 L,

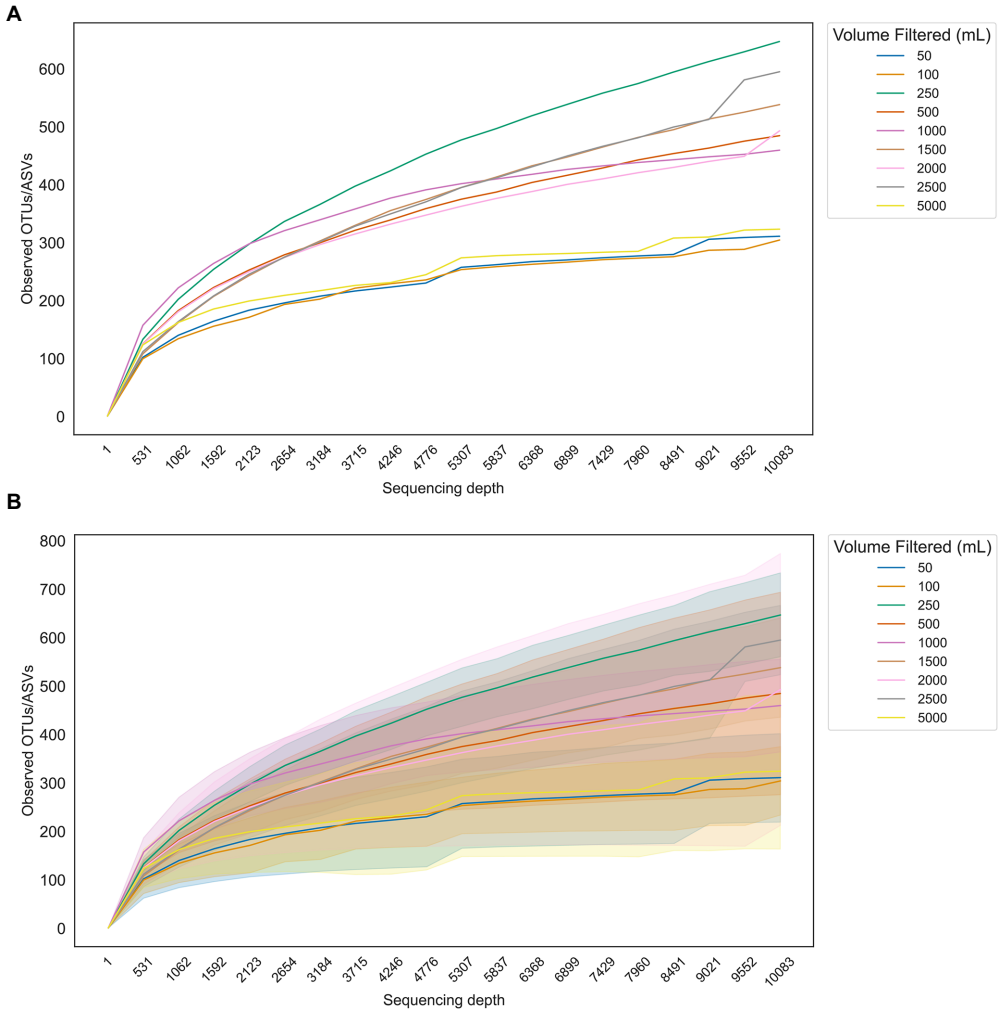

FIGURE 2
Rarefaction curves showing the number of taxa observed in samples with varying filtration volumes as a function of sequencing depth do not show a clear advantage to filtering larger volumes for downstream sequencing of 16S rRNA gene amplicons. **(A)** Average number of observed taxa for each set of samples corresponding to a different volume of water filtered. **(B)** Curves from **(A)** with standard deviations overlaid.

which may exceed the funnel volume and require an operator to monitor and re-fill the funnel. This is inconvenient and increases the risk of contamination and human error (e.g., mixing of samples) as well as spillage, particularly on a rocking ship. In such cases, a more enclosed system for the sample water is needed, such as carboys fitted with quick-connect vacuum connections.

Another common mode to filter water is *via* peristaltic filtration, in which a peristaltic pump moves water from the sample container through sample tubing and through the filter. Sample volumes are monitored by measuring the amount of water flowing out to waste. Pump speed is easily adjusted allowing finer control of filtration rates than vacuum filtration. Peristaltic filtration can be used with enclosed filters, such as Sterivex™ cartridges, or with membrane filters housed in enclosed containers (e.g., Swinnex™), with filter choices discussed further below. To minimize contamination risk, some approaches attach a sterile serological pipette (e.g., 10 ml) to the tubing to avoid placing the tubing directly into the sample. Approaches to avoid cross-contamination vary and include using new sterilized tubing for every sample, instituting rigorous cleaning procedures between samples, and simply flushing with sample water before sample filtration. All choices balance convenience, cost, and contamination risk. At this time, data appear insufficient to provide clear guidelines other than to ensure that sample contamination is avoided.

Passive filtration has recently emerged as a lower-effort approach to eDNA collection (Bessey et al., 2022; Chen et al., 2022). With this method, filters are directly submerged in the water column and eDNA adheres to a membrane over time, bypassing any water pumping altogether. This approach was recently used to characterize fish community eDNA and was proposed as a method to increase sample replication with subsequent benefits to data analysis (Bessey et al., 2021). Current data are insufficient to assess the potential for passive filtration as a standard method. Nonetheless, the promise of the approach to circumvent filtration provides a clear example of the need to pursue method harmonization while simultaneously embracing innovation.

Filter type

Filter type determines water flow rate and effective filtration area, and thus is a key component of sampling design. For example, Sterivex™ capsule membranes are available in both polyethersulfone (PES) and polyvinylidene difluoride (PVDF) formats, but PES provides faster water flow rates (Table 2). The pore sizes used depend on the application. For example, 0.45 µm pore size filters are typical in water quality monitoring of fecal indicators [e.g., United States Environmental Protection Agency (USEPA), 2022]. In contrast, marine microbial ecology usually employs 0.2 µm pore size filters to capture the full complement of marine microbial cell sizes (Cottrell and Kirchman, 2004; Kirchman et al., 2007), including picoeukaryotes (Vaulot et al., 2008).

Multi-trophic studies are severely challenged by the need to capture the full array of possible cell sizes (Pesant et al., 2022). The challenge is compounded by the fact that metazoan eDNA (e.g., from fish) exists in a range of states, ranging from dissolved to packaged inside an organelle – see Mauvisseau et al., 2022 for a review of this topic. Ultimately, the tradeoffs between filter type, filter time, and sample volume are rooted in the particle size distributions of target eDNA and other non-target suspended particulate matter (Turner et al., 2014; Shogren et al., 2016; Zhao et al., 2021). In the future, perhaps novel capture methods will be devised free from size fractionation. For now, best practices remain reliant on traditional filters. To that end, researchers can conduct tests to estimate the filtration area required (Table 2) for a given set of filtration conditions (e.g., maximum time at constant pressure).

If target cells are particle-associated, a small volume of turbid water may yield sufficient high-quality DNA for the needs of the study (e.g., investigating microbial taxa attached to eukaryotic phytoplankton). Other surveys require more volume to adequately capture the community profile (Kumar et al., 2022). In this case, it can be beneficial to pre-filter and/or size-fractionate the water sample. The term "pre-filter" often refers to removing unwanted large particles before the water passes through the target filter; this method can cut down time required to filter the target volume (Robson et al., 2016), improve PCR amplification due to removal of inhibitors, and reduce variability among replicates (Takasaki et al., 2021). As with the target sample filter, a new pre-filter should be used for each sample to avoid cross-contamination. The pre-filter, with a pore size anywhere from 3 to 20 µm (or even larger, such as a coffee filter or fine mesh), can then be discarded. However, if the pre-filter is preserved, it becomes a second size fraction containing the particle-associated microbial community.

TABLE 2 Example filter types, water flow rates, and filtration areas.

Filter material	Pore size (µm)	Water flow rate (ml/min/cm²)[a]
Supor®, PES	0.22	26
Durapore®, PVDF	0.22	5
Mixed cellulose esters	0.45	>65
Supor®, PES	0.45	58
Durapore®, PVDF	0.45	26
Filter type	**Filter size (mm)**	**Effective filtration area (cm²)**
Membrane	25	3.4
Capsule (Sterivex)	17	10
Membrane	47	13.8
Membrane	142	127

[a]at 10 psi, obtained from Pall.com.

Size-fractionation can both facilitate more effective filtering and be highly informative for comparisons of particle-associated and free-living microbial communities (DeLong et al., 1993; Ganesh et al., 2014; Orsi et al., 2015; Byappanahalli et al., 2021). The research vessel *Tara*, which performs standardized global ocean sampling, including the Tara Oceans expedition from 2009 to 2013, fractionates onto 3 µm and 0.2 µm filters and additionally uses the 0.2 µm filtrate for viral precipitation and final collection on to a 0.8 µm filter (Pesant et al., 2015, 2022). Results from multi-omics analyses provided information on understudied planktonic organisms and their associated microbes that would have been missed with one size fraction (Sunagawa et al., 2020). At the smallest end of the size spectrum, recent work has shown that marine microbes in the Candidate Phyla Radiation (CPR) and the Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota (DPANN) lineages can have ultra-small cells (<0.1 µm diameter) and would thus evade 0.2 µm filters (Castelle et al., 2018). Methods for capturing such microbial cells have not been widely studied but may include ultracentrifugation (e.g., Boström et al., 2004; Eichmiller et al., 2016) or ethanol precipitation from unfiltered water (Ficetola et al., 2008; Eichmiller et al., 2016; Spens et al., 2017). Jeunen et al. (2019) found that when smaller volumes (500 ml) were filtered, 0.2 µm pore size filters captured significantly more DNA than larger pore size filters, with no differences among filter materials; however, when filtration was performed until filters clogged (up to 5,000 ml), 0.2 µm cellulose nitrate (CN) filters outperformed polycarbonate (PC) and glass fiber (GF) filters with the same pore size in DNA yield. Other studies have shown that GF filters capture more biomass compared to PC or PES filters, but their higher yield may be counteracted by lower sensitivity (Eichmiller et al., 2016; Takahashi et al., 2020).

As with all sampling components, sampling efficacy must be balanced by resource availability. As several studies have shown comparable results in DNA yield and quality among different filters, we do not provide a single recommendation; however, we provide an overview of commonly available filters and their costs in Supplementary Table S2. In the addendum, we describe a sampling protocol we successfully employed aboard an operational fisheries vessel to help illustrate how topics discussed in the main text may manifest in practical situations. We hope the combination of these resources will aid those looking to develop a microbiome sampling routine.

Filter housing

Filter types can be divided into two major categories: disc and cartridge, or enclosed. Disc filters are flat circular membranes that are generally used with a vacuum filtration setup. After filtration is complete, the filter must be transferred aseptically to a separate container for storage, which poses a risk of contamination. In contrast, enclosed filters (e.g., Sterivex™, Millipore, Billerica, MA) feature membranes in a cylindrical capsule. "Enclosed" cartridges are self-contained throughout the filtration process and can be purchased with luer lock inlets for easy use with syringes and efficient sealing with compatible male adapters. These cartridges are amenable to a peristaltic pump setup. After filtration, the entire capsule is stored while awaiting nucleic acid extraction. This approach typically requires more storage space which needs to be considered in the context of available resources.

Despite offering convenience in the field, Sterivex™ cartridges present challenges for downstream nucleic acid extraction because the filter is bound to the plastic capsule. It must either be physically removed from the housing or reagents must be added to cover the filter area, which can lead to a more dilute nucleic acid product. In any case, most commercially available extraction kits are designed for liquid cultures or tissues that fully dissolve. This circumstance has resulted in a wide array of "standard" modifications to manufacturers' protocols (e.g., Cruaud et al., 2017; Ushio, 2019; Anderson and Thompson, 2022), further complicating the quest for harmonized practice.

One approach that combines elements of both disc and cartridge filters is a disc membrane housed in a Swinnex™ (Millipore, Billerica, MA). These housings are enclosed like a Sterivex™ during the filtration process and can easily be used with a peristaltic pump, but the Swinnex™ housings are reusable and can be opened to remove the disc. The disc filter-Swinnex™ method is a cheaper and less wasteful alternative to Sterivex™ that still provides the advantages of peristaltic pumping and flexible DNA extraction methods.

Spens et al. (2017) found that Sterivex™ cartridges captured more macrobial (fish) eDNA and fared better at room temperature storage for 2 weeks than membrane filters, particularly with the addition of Longmire's buffer or ethanol storage buffer. Takahashi et al. (2020) found Sterivex™ cartridges performed better than GF disc filters with low amounts of eDNA, while GF fared better when eDNA levels were high. Their small size, ease of use, and higher resistance to clogging compared to discs make Sterivex™ appealing for many marine microbial eDNA collection applications; however, their higher cost and associated extraction challenges may outweigh their advantages under some circumstances (see section below on DNA extraction).

Sample preservation

For many field studies, it is impractical or impossible to perform DNA extraction and amplification immediately following sample collection. Safely and efficiently preserving samples is therefore critical for study success. In all cases, filtration procedures should include briefly continuing to operate the pump after all sample water has been filtered to remove excess water before storage. A variety of buffers exist for preventing DNA degradation over time; however, dry preservation can be effective and has the advantage of reducing the required reagents (Majaneva et al., 2018; Sunagawa et al., 2020; Allison et al., 2021). The decision on whether or not to use a buffer will depend on the temperature at which filters can be stored, anticipated number of freeze–thaw cycles, and space capacity in the field. Sample preservation is less critical when target DNA is abundant and quantification is not a primary study goal. Otherwise, the study design may require that all possible precautions are taken to prevent DNA degradation and allow for community assessment.

Immediate freezing of samples after filtration is ideal for nucleic acid preservation. Flash freezing in liquid nitrogen or immediate storage at −80°C are both robust methods of preservation; however, unlike freezers, liquid nitrogen dewars are not prone to electrical issues or breakdowns and are thus preferable for field sampling. Provided immediate freezing is available and filters stay frozen until extraction, samples can be stored dry (i.e., without preservative) with minimal loss of nucleic acid material. Repeated freeze–thaw cycles have been shown to affect host-associated microbiome composition (Sergeant et al., 2012; Gorzelak et al., 2015) and should be avoided. If neither of those approaches are an option, freezing at −20°C as soon as possible offers more protection over 4°C or warmer conditions. However, at these warmer temperatures, additional measures to ensure sample integrity are important, particularly for long-term (>1 week) storage. In one study, the addition of 100% ethanol and room temperature storage for 4 days resulted in DNA yields similar to those from a dry −20°C storage protocol (Hinlo et al., 2017). Other studies found that macrobial DNA remained stable for 1 week in cetyltrimethyl ammonium bromide (CTAB) (Renshaw et al., 2015; Hunter et al., 2019) or up to 150 days at 20°C in Longmire's buffer (Wegleitner et al., 2015) at room temperature. Longmire's buffer (3:1 water to buffer) protected bony fish DNA in whole water samples stored for months both frozen and at ambient temperatures (Cooper et al., 2022). Multiple studies found increased yield over time, suggesting that extended lysis buffer storage releases more DNA than a brief incubation.

An increasingly popular method to aid long-term DNA preservation is the addition of silica gel or beads to dry filters. Allison et al. (2021) showed that silica gel preserved DNA integrity for up to 1 year at −20°C and performed better than 95% ethanol at 23°C storage temperature, and Majaneva et al. (2018) found highest consistency in metazoan community composition from filters preserved with silica relative to lysis buffer or ethanol (>99%, molecular grade). This efficacy, along with the important advantage of limiting large volumes of liquid reagents in the field, make silica desiccation a highly appealing option for preservation of environmental DNA.

TABLE 3 Three of the most common preservation buffers and their chemical composition.

Buffer name	Composition	Reference
CTAB	1.4 M NaCl, 2% (w/v) cetyltrimethyl ammonium bromide, 100 mM Tris, 20 mM EDTA and 0.25 mM polyvinylpyrrolidone	Dempster et al. (1999)
Longmire's buffer	0.1 M Tris, 0.1 M EDTA, 10 mM NaCl, 0.5% (w/v) SDS	Longmire et al. (1997)
Sucrose lysis solution	20 mM EDTA, 200 mM NaCl, 0.75 M sucrose, 50 mM Tris–HCl, pH 9.0	Giovannoni et al. (1990)

Along with ethanol, multiple buffers can be used to preserve environmental DNA on water filters with varying levels of efficacy. Common solutions include DNA lysis buffers and reagents containing guanidinium thiocyanate, such as DNAzol® (DN127, Molecular Research Center, Cincinnati, OH), or high levels of ammonium sulfate, like RNAlater™ (Catalog no. AM7021; Thermo Fisher Scientific, Waltham, MA). "Lysis buffer" can include kit reagents such as Qiagen ATL Lysis Buffer or in-house recipes like CTAB, Longmire's buffer, or sucrose lysis solution (Table 3). Storage buffers can be divided into those that will inhibit DNA extraction and/or purification and must be removed before DNA extraction (i.e., RNAlater, ethanol) and those that can be incorporated into the extraction protocol (lysis buffers). Commercially available buffers like RNAlater and DNA/RNA Shield (Catalog no. R1100-250; Zymo Research, Irvine, CA) can protect all nucleic acids and are often recommended for broadly applicable sampling practices; for example, the European Marine Omics Biodiversity Observation Network (EMO BON) Handbook (Santi et al., 2021).

In addition to impacting overall DNA yield, there is evidence that storage method can affect taxonomic composition results, presumably due to preferential cell lysis and/or preservation of DNA for certain taxa. Both ethanol and RNAlater yielded communities with lower alpha diversity and higher variability compared to samples preserved with lysis buffer or dried with silica gel (Majaneva et al., 2018; Oldham et al., 2019), and dry preservation provided higher levels of eDNA and more consistent community composition than filters stored with RNAlater in a freshwater lake study (McCarthy et al., 2015). Longmire's and CTAB lysis buffers have been shown to preserve DNA well at room temperature (Renshaw et al., 2015; Wegleitner et al., 2015) while sucrose lysis buffer performed better than a guanidine thiocyanate buffer (similar to DNAzol®) while frozen (Mitchell and Takacs-Vesbach, 2008).

We recommend prioritizing the freezing of samples with minimal delay, and desiccating filters before freezing if possible. If samples will remain at room temperature for an extended period of time or undergo freeze–thaw cycle, storage in Longmire's buffer or sucrose lysis buffer is an option, depending on the planned DNA extraction protocol.

DNA extraction

eDNA extraction efficiency from preserved filters depends on the filter type and, to a lesser extent, what type of preservation buffer (if any) was used during filter storage. Extraction protocols can be categorized as highly standardized (e.g., the use of a commercially available kit) or reliant on in-house reagents, which are subject to greater variability among protocols and labs. The reproducibility and accessibility of kits make them an appealing option for comparative studies, large sample numbers, and collaborative efforts. However, there is ample evidence that more involved protocols, such as those with a phenol:chloroform:isoamyl alcohol (25:24:1; hereafter PCI) purification step, result in higher eDNA yields and/or higher levels of eDNA purity (Urakawa et al., 2010; Djurhuus et al., 2017; Schiebelhut et al., 2017; Hunter et al., 2019). These tradeoffs must be weighed while also considering filter types and resource availability, including access to fume hoods and staff willing to work with toxic chemicals. Below we overview these different approaches.

Unlike with enclosed filters (see above), commercial kits are generally available for the extraction of membrane filters with little to no protocol modifications. Many kits provide screw-top vials in which filters can be combined with lysis buffer and stored until extraction, helpfully eliminating the time and effort required to transfer filters between storage and extraction containers. Many studies have produced high-quality microbial eDNA from marine samples using commercial kits. Large scale sampling efforts like Ocean Sampling Day (Tragin and Vaulot, 2018) and the Earth Microbiome Project (Thompson et al., 2017) employed commercial kits that included a bead-beating step, which aids in lysing cells with more resilient cell membranes like Gram-positive bacteria (de Boer et al., 2010) or certain phytoplankton (Mäki et al., 2017). In comparisons of three tested kits offered by Qiagen (Qiagen GmbH, Hilden, Germany), the DNeasy Blood & Tissue kit provided the best results for DNA yield (Djurhuus et al., 2017; Hinlo et al., 2017) and equaled or outperformed the PowerWater kit in metabarcoding data quality and consistency (Jeunen et al., 2019), particularly when modified with a bead-beating step (Djurhuus et al., 2017). Other less commonly used kits described in the literature we reviewed (Supplementary Table S1) include the ZymoBIOMICS 96 DNA/RNA MagBead (e.g., Anderson and Thompson, 2022), Epicentre MasterPure DNA purification (e.g., Geerts et al., 2018), Qiagen AllPrep DNA/RNA mini (e.g., Cruaud et al., 2017), and Presto Mini gDNA (e.g., Jeunen et al., 2019), but we found too few studies describing the performance of these kits to comment further.

The two most common non-commercial approaches for extracting DNA are protocols using CTAB or PCI; in some cases, both reagents are used (Needham and Fuhrman, 2016; Hunter et al., 2019). These protocols rely on chemical cell lysis and generally do not include a bead-beating step, although one can be added (e.g., Urakawa et al., 2010; Biller et al., 2018). A PCI

protocol was used for high-throughput sampling (>600 marine microbial metagenomes) in the bioGEOTRACES study (Biller et al., 2018) in a testament to its efficacy and lower cost. A combination CTAB/PCI was effectively used to extract microbial and phytoplankton DNA for a multi-trophic metabarcoding study (Needham and Fuhrman, 2016). These protocols can be particularly useful for enclosed or capsule filters, which can be challenging to integrate with kits. However, they rely on hazardous chemicals and require significantly more bench time than kits, limiting their usefulness for high-throughput sample processing. Several studies have found that PCI yielded higher microbial and macrobial DNA copy numbers (Urakawa et al., 2010; Renshaw et al., 2015; Djurhuus et al., 2017; Schiebelhut et al., 2017) and higher DNA purity (Urakawa et al., 2010; Schiebelhut et al., 2017) compared to kits. Other studies found higher detection rates for some, but not all, kits compared to PCI (Deiner et al., 2015). Due to better performance but higher effort, PCI protocols are recommended for rare taxa and/or if the number of samples to be processed is relatively small.

To extract nucleic acids from enclosed filters such as Sterivex™, storage buffer must be completely removed from the cartridge unless it can be integrated into the extraction protocol (e.g., Ganesh et al., 2014; Padilla et al., 2015). Removal can be performed by flushing with air using a syringe or a vacuum manifold. Lysis buffer must then be added to the cartridge, or the filter can be removed from the plastic housing. Cruaud et al. (2017) found that removing the Sterivex™ housing, cutting up the filter, and adding filter pieces to tubes with lysis buffer yielded 4 times as much DNA as an internal extraction protocol and produced similar alpha diversity and community composition results. Kawato et al. (2021) also recommended this method for detection of deep-sea fish. However, this process is susceptible to contamination, negating one of the main advantages of the enclosed cartridge.

Alternatively, beads can be added to the Sterivex™ cartridge housing. This mechanical lysis step improved microbial community DNA yield over both an internal extraction protocol without beads and a protocol based on opening the cartridge, while maintaining the enclosed filter environment (Ushio, 2019). Anderson and Thompson (2022) adapted this protocol in several useful ways, including adding beads after sample collection, optimization of bead composition to maximize recovery of hard-to-lyse organisms, and demonstrating high-throughput extractions performed on a magnetic bead-handling robot [KingFisher™ Flex Purification System (Thermo Fisher Scientific, Waltham, MA)].

All the protocols described above are well suited for subsequent manual benchtop library preparation and short read sequencing, which remain the predominant approaches for metabarcoding or metagenomic sequencing. However, technological advances are rapidly changing the DNA sequencing landscape and deserve consideration. Long read sequencing, which can be performed by PacBio® or Oxford Nanopore Technology® platforms, requires higher molecular weight DNA than short read platforms like Illumina (Jones et al., 2021). Thus, protocols designed to minimize shearing (e.g., Mayjonade et al., 2016; Russo et al., 2022) should be considered if long read sequencing is planned. Trigodet et al. (2022) found that column-based commercial kits with enzyme modifications were equally as, if not more, effective than PCI protocols for generating high-quality long read sequences. In addition, robotic sampling and processing instruments, such as the KingFisher™ Flex Purification System mentioned above, can reduce the time and effort required to process eDNA samples (Marotz et al., 2017; Anderson and Thompson, 2022). Autonomous samplers, like the Monterey Bay Aquarium Research Institute Long-Range Autonomous Underwater Vehicle containing an Environmental Sample Processor (3G ESP-LRAUV), can filter marine water *in situ* and preserve filters equivalent to traditional shipboard sampling for eDNA analysis by qPCR (Yamahara et al., 2019) or DNA metabarcoding (Truelove et al., 2022). These and similar technological advances indicate the field of marine microbiology has entered an exciting new era of discovery. Many of the field sampling considerations presented in this review may thus become irrelevant in the not-too-distant future. However, until such instruments become mass-produced and cost-effective, most labs will continue to rely on manual sampling and extraction protocols.

Discussion

The need to standardize sampling approaches for marine microbiomes and eDNA is widely recognized (Canonico et al., 2019; Pearlman et al., 2019; Hörstmann et al., 2021; Samuel et al., 2021). The wide ranges of physical eDNA states, applications, and test conditions available in the literature make finding consensus difficult. The quest for consensus is further confounded by the desire to maintain consistent methods once a particular study or time series has begun, a limited ability to unify the large community of researchers spread across the globe, and the fact that most marine eDNA surveys try to develop a fit for purpose protocol while actually attempting to meet multiple goals (e.g., a "microbes to mammals" approach).

Despite the surplus of data, protocols, and guidelines available, pilot studies should be designed, executed, and analyzed prior to committing to a protocol. Therefore, programs need to budget adequate time and money to verify that protocols meet the needs of the study. Protocol development should be undertaken in the context of environmental variability with the aim of elucidating which parameters are truly important to control. For example, investing in more PCR technical replicates is not warranted if the sample collection volume is insufficient for the purposes of the project.

Despite the lack of consensus on microbiome sample collection and processing methods, we recognize certain overarching guidelines here. Rigorous decontamination of work surfaces and equipment, sample replication, and controls included throughout the sample collection and processing workflow are

baseline practices. The need to rapidly collect, filter, and freeze samples with minimal delay is balanced by the requirement to obtain sufficient DNA to detect and characterize the species of interest. Partnerships between marine science and industry should be pursued to develop fast and reliable means to collect, preserve, and reproducibly extract large quantities of high-quality DNA from aquatic samples.

We recognize that differences in microbial and eukaryotic DNA prevalence (ubiquitous vs. patchy), concentration (high vs. low), and state (associated with different size fractions) lead to conflicting approaches for ideal sampling strategies; nevertheless, it is possible to design a versatile protocol that allows multiple sampling goals to be achieved. Ideally such a design can be implemented on a minimal budget and executed by readily available crew with little specialized training as they transverse waters from coastal to open ocean. An overarching authority with regard to best practices is needed but it is unclear how it might emerge. Emerging efforts such as the Better Biomolecular Ocean Practices (BeBOP; Table 1) are promising in this regard, although a disconnect in time scales between those processes and the pace of technological advancement is a concern (Trujillo-González et al., 2021). We recommend continued efforts in this area with an emphasis on including a diverse array of marine microbiome researchers who collectively hold a vast amount of knowledge, much of which remains anecdotal or unpublished.

The ultimate goal of harmonization efforts is to bring the field closer to a consensus on best practices (Pearlman et al., 2021) for sampling marine microbiomes. Harmonized approaches are critically needed as the pace of eDNA studies accelerates and the call to understand microbiome responses to climate change increases (Cavicchioli et al., 2019; Tiedje et al., 2022). Method harmonization will enable the current patchwork of observations to be stitched into a global network of observations to produce baselines by which ecosystem impacts can be accessed. Producing data that is interoperable over space and time is the first step to building a trusted time series upon which to base management decisions.

Addendum

Example protocol: eDNA sampling during a fisheries survey

Here we describe a marine eDNA sampling protocol to illustrate designing a fit for purpose "best practice" to balance the advantages and disadvantages described above. This protocol (Figure 3) is based on the experimental goals and available resources of our particular field sampling vessel and is based on the MBON protocols.[1]

Goals

We collected marine eDNA in conjunction with midwater trawls conducted during the annual NMFS Rockfish Recruitment and Ecosystem Assessment Survey (RREAS). eDNA sequencing and analysis included metabarcoding of microbial (16S and 18S rRNA) and eukaryotic/metazoan (12S rRNA and COI) gene amplicons, as well as shotgun metagenomics.

Field sampling facility

The eDNA lab aboard the vessel was separate from the main wet lab, which was frequently exposed to high levels of fish and microbial DNA from the trawls. A −80°C freezer was situated in a separate part of the ship. There was no Milli-Q water system available, but we collected freshly distilled water from the vessel's distillation system in the engine room using containers cleaned thoroughly with bleach and Milli-Q water ahead of the cruise. Sample water was collected using 10-L Niskin bottles mounted on a CTD rosette.

Preparation

Peristaltic pumps were set up on a bench space that was thoroughly cleaned with a 5%–10% bleach solution. Tubing was rinsed with ~500 ml bleach by running the pump, followed by a rinse with Milli-Q water. When possible, up to 1 L of Milli-Q water was used to thoroughly remove bleach; however, to compensate for the Milli-Q water limitation we incorporated a rinse with 1–2 L sample water per tubing (see below).

Notes regarding decontamination protocols

Complete and efficient decontamination of surfaces can be performed with various chemicals, with two common approaches described here. Sodium hypochlorite (bleach) has long been recognized as a low-cost, highly effective way of degrading DNA (Prince and Andrus, 1992). Typical protocols call for a 10% bleach solution; however, it is a common misconception that this solution refers to a 1:10 dilution of commercially available household bleach products, which are usually 6%–8% sodium hypochlorite (Santi et al., 2021). In fact, final w/v of sodium hypochlorite should be 1%–5%, requiring at most a 1:8 dilution with most commercial bleach products (Kemp and Smith, 2005; Goldberg et al., 2016; Wilcox et al., 2016). To fully decontaminate surfaces and equipment, bleach should be applied and sit for ~15 min. Complete removal of the bleach solution is then crucial to prevent residual sodium hypochlorite in the water from causing sample degradation. A thorough rinse of all tubing and filtration equipment with sterile water (fresh or saline) or with extra sample water after application of chemicals for decontamination is recommended. High bleach concentrations can be challenging to remove, especially if left sitting for a long time; we therefore recommend a 1% w/v bleach solution left for at most 20 min on any surface.

Many protocols suggest RNase AWAY™ (a sodium hydroxide solution) or hydrochloric acid in addition to or as an alternative to bleach (e.g., the MBON water sampling protocol). For small

1 https://mbari-bog.github.io/MBON-Protocols/WaterFilteringProtocol.html

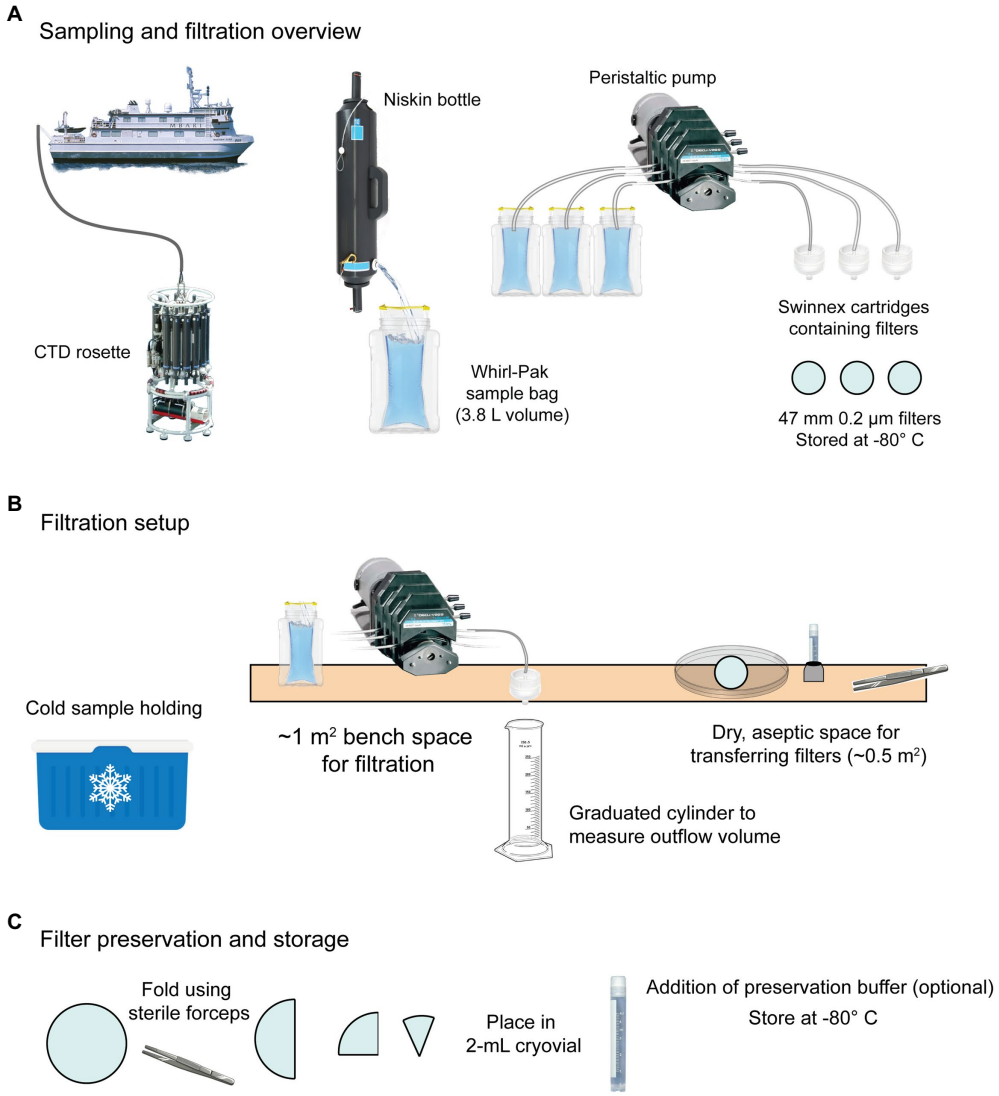

FIGURE 3
Example shipboard sampling workflow for eDNA based on MBON recommendations. **(A)** Water is collected using Niskin bottles mounted on a CTD rosette. Once aboard, water is transferred to sterile Whirl-Pak® bags and filtered using a peristaltic pump and membrane filters housed in Swinnex cartridges. **(B)** Whirl-Paks with sample water are stored cold (4°C) until filtration, which is performed on a bench with a peristaltic pump and Swinnex™ housed disc filters (47 mm diameter). Outflow is measured using a graduated cylinder and the filtration duration and volume is recorded. Filters are transferred from the housing to a 2-ml cryovial for storage. **(C)** Filters are loosely folded in cryovials. Preservation buffer can optionally be added at this point. Cryovials should be immediately stored at −80°C or in liquid nitrogen.

volume applications (e.g., sterilizing forceps) RNase AWAY™ is highly effective but due to its substantially higher cost is not a practical reagent for large-scale decontamination. These solutions tend to be gentler than bleach on stainless steel surfaces. However, all chemicals are toxic and require cautious handling.

Sampling

Water from Niskin bottles was collected in sterile 3.8-L Whirl-Pak bags by direct transfer (allowing a "clean stream" to run briefly before beginning collection). From each sampled Niskin, a separate 2-L Whirl-Pak bag was filled to provide water for rinsing. Bags were stored in a cooler with wet ice during the filtration process. Before filtration began, bags were inverted 3–5 times to thoroughly mix sample water. The first round of pumping was done without attached filters and using water from the 2-L "rinse bags" to rinse the tubing with ~1 L water from the sample Niskin bottle. To capture eDNA from sample water, 47-mm, 0.22-µm nitrocellulose filters in a Swinnex housing were attached to the tubing outflow end and a new serological pipet was attached to the intake end. The pipet end was placed in the sample bag (with tubing remaining outside) and the pump was run until 2 L of water was filtered or the filter clogged, whichever came first. Before stopping the pump, air was briefly run through the tubing to dry the filter.

Sample preservation

When filtration was complete, the pump was turned off and the Swinnex housing was disconnected from the tubing. The housing was opened and the filter was folded with sterile forceps and transferred to a cryovial. This container was placed inside a secondary container (e.g., a 250-ml Whirl-Pak bag) containing silica beads. Biological replicates were stored together in one bag, and full bags were transferred to the −80°C as soon as possible. Samples were stored on the ship and transferred on dry ice to the laboratory upon mission completion, where they were stored at −80°C until processed in bulk at a facility capable of high-throughput DNA extraction.

Methods

Figure 1 was generated by searching Google Scholar for the terms 'marine "environmental DNA"' so that eDNA from a marine environment would likely be a major component of the study. The filter was set to limit results from each year sequentially and results were plotted in Microsoft Excel.

Alpha rarefaction curves were generated from three publicly available marine water column 16S rRNA gene amplicon data sets generated by the following three studies: 1. Truelove et al. (2022), 2. Padilla et al. (2015), and 3. Torres-Beltrán et al. (2019). Raw sequences from [1] and [2] were run through DADA2 in QIIME2 with the following parameters: The processed OTU table from [3] was combined with the processed sequences from [1] and [2] and rarefaction curves were generated from the resulting feature tables using 'qiime alpha rarefaction' with a maximum sequence depth of 10,083 and 20 sub-sampling steps. The resulting rarefaction table was exported as a csv file and run through a custom Python script to generate rarefaction curves with and without standard deviations for each line. The script is available at the first author's GitHub repo: https://github.com/nvpatin/Amplicon-visualizations.

Author contributions

NP performed the analyses and created the figures. NP and KG wrote the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgments

We thank Francisco Chavez, Kobun Truelove, and Kathleen Pitz at MBARI for productive collaborations that led to the protocol presented in the addendum. We are also grateful to the captain and crew of the NOAA Ship Reuben Lasker for facilitating safe and effective water sampling during protocol development. We thank Frank Stewart, Cory Padilla, and Steven Hallam for providing data for the test of filtered water volume effect on microbial community diversity. This work was made possible by a grant from the National Oceanographic Partnership Program (NOPP) and the NOAA 'Omics Program and was carried out in part under the auspices of the Cooperative Institute of Marine and Atmospheric Studies (CIMAS), a Cooperative Institute of the University of Miami and the National Oceanic and Atmospheric Administration, cooperative agreement number NA20OAR4320472.

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.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.1026596/full#supplementary-material

References

Aglieri, G., Baillie, C., Mariani, S., Cattano, C., Calò, A., Turco, G., et al. (2021). Environmental DNA effectively captures functional diversity of coastal fish communities. *Mol. Ecol.* 30, 3127–3139. doi: 10.1111/mec.15661

Allison, M. J., Round, J. M., Bergman, L. C., Mirabzadeh, A., Allen, H., Weir, A., et al. (2021). The effect of silica desiccation under different storage conditions on filter-immobilized environmental DNA. *BMC. Res. Notes* 14:106. doi: 10.1186/s13104-021-05530-x

Anderson, S. R., and Thompson, L. R. (2022). Optimizing an enclosed bead beating extraction method for microbial and fish environmental DNA. *Environ. DNA* 4, 291–303. doi: 10.1002/edn3.251

Andruszkiewicz Allan, E., Zhang, W. G., Lavery, A., and Govindarajan, A. (2021). Environmental DNA shedding and decay rates from diverse animal forms and thermal regimes. *Environ. DNA* 3, 492–514. doi: 10.1002/edn3.141

Andruszkiewicz, E. A., Yamahara, K. M., Closek, C. J., and Boehm, A. B. (2020). Quantitative PCR assays to detect whales, rockfish, and common murre environmental DNA in marine water samples of the northeastern Pacific. *PLoS One* 15:e0242689. doi: 10.1371/journal.pone.0242689

Azam, F. (1998). Microbial control of oceanic carbon flux: the plot thickens. *Science* 280, 694–696.

Azam, F., Fenchel, T., Field, J. G., Gray, J. S., Meyer-Reil, L. A., and Thingstad, F. (1983). The ecological role of water-column microbes in the sea. *Mar. Ecol. Prog. Ser.* 10, 257–263. doi: 10.3354/meps010257

Barnes, M. A., Chadderton, W. L., Jerde, C. L., Mahon, A. R., Turner, C. R., and Lodge, D. M. (2021). Environmental conditions influence eDNA particle size distribution in aquatic systems. *Environ. DNA* 3, 643–653. doi: 10.1002/edn3.160

Béjà, O., Aravind, L., Koonin, E. V., Suzuki, M. T., Hadd, A., Nguyen, L. P., et al. (2000). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* 289, 1902–1906. doi: 10.1126/science.289.5486.1902

Beng, K. C., and Corlett, R. T. (2020). Applications of environmental DNA (eDNA) in ecology and conservation: opportunities, challenges and prospects. *Biodivers. Conserv.* 29, 2089–2121. doi: 10.1007/s10531-020-01980-0

Berube, P. M., Gifford, S. M., Hurwitz, B., Jenkins, B. D., Marchetti, A., Santoro, A. E., et al. (2022). "Roadmap towards community-wide Intercalibration and standardization of ocean nucleic acids 'omics measurements," in *Ocean Nucleic Acids Intercalibration and Standardization Workshop*. (Woods Hole, MA, USA: Woods Hole Oceanographic Institution), pp. 50.

Bessey, C., Gao, Y., Truong, Y. B., Miller, H., Jarman, S. N., and Berry, O. (2022). Comparison of materials for rapid passive collection of environmental DNA. *Mol. Ecol. Resour.* 22, 2559–2572. doi: 10.1111/1755-0998.13640

Bessey, C., Neil Jarman, S., Simpson, T., Miller, H., Stewart, T., Kenneth Keesing, J., et al. (2021). Passive eDNA collection enhances aquatic biodiversity analysis. *Commun Biol* 4:236. doi: 10.1038/s42003-021-01760-8

Biller, S. J., Berube, P. M., Dooley, K., Williams, M., Satinsky, B. M., Hackl, T., et al. (2018). Marine microbial metagenomes sampled across space and time. *Sci. Data* 5:180176. doi: 10.1038/sdata.2018.176

Boström, K. H., Simu, K., Hagström, Å., and Riemann, L. (2004). Optimization of DNA extraction for quantitative marine bacterioplankton community analysis. *Limnol. Oceanogr. Methods* 2, 365–373. doi: 10.4319/lom.2004.2.365

Boussarie, G., Bakker, J., Wangensteen, O. S., Mariani, S., Bonnin, L., Juhel, J. B., et al. (2018). Environmental DNA illuminates the dark diversity of sharks. *Sci. Adv.* 4:eaap9661. doi: 10.1126/sciadv.aap9661

Bowers, H. A., Pochon, X., von Ammon, U., Gemmell, N., Stanton, J. A., Jeunen, G. J., et al. (2021). Towards the optimization of eDNA/eRNA sampling technologies for marine biosecurity surveillance. *Water* 13:1113. doi: 10.3390/w13081113

Bruno, A., Sandionigi, A., Galimberti, A., Siani, E., Labra, M., Cocuzza, C., et al. (2017). One step forwards for the routine use of high-throughput DNA sequencing in environmental monitoring. An efficient and standardizable method to maximize the detection of environmental bacteria. *MicrobiologyOpen* 6:e00421. doi: 10.1002/mbo3.421

Buck, K. R., Chavez, F. P., and Campbell, L. (1996). Basin-wide distributions of living carbon components and the inverted trophic pyramid of the central gyre of the North Atlantic Ocean, summer 1993. *Aquat. Microb. Ecol.* 10, 283–298. doi: 10.3354/ame010283

Byappanahalli, M. N., Nevers, M. B., Shively, D., Nakatsu, C. H., Kinzelman, J. L., and Phanikumar, M. S. (2021). Influence of filter pore size on composition and relative abundance of bacterial communities and select host-specific MST markers in coastal waters of southern Lake Michigan. *Front. Microbiol.* 12:665664. doi: 10.3389/fmicb.2021.665664

Campbell, B. J., Yu, L., Heidelberg, J. F., and Kirchman, D. L. (2011). Activity of abundant and rare bacteria in a coastal ocean. *Proc. Natl. Acad. Sci. U. S. A.* 108, 12776–12781. doi: 10.1073/pnas.1101405108

Canonico, G., Buttigieg, P. L., Montes, E., Muller-Karger, F. E., Stepien, C., Wright, D., et al. (2019). Global observational needs and resources for marine biodiversity. *Front. Mar. Sci.* 6:367. doi: 10.3389/fmars.2019.00367

Canonico, G., Duffy, J. E., and Muller-Karger, F. E. (2022). Marine life 2030: building global knowledge of marine life for local action in the ocean decade. *Mar. Technol. Soc. J.* 3, 112–113. doi: 10.4031/MTSJ.56.3.15

Carini, P., Steindler, L., Beszteri, S., and Giovannoni, S. J. (2013). Nutrient requirements for growth of the extreme oligotroph 'Candidatus Pelagibacter ubique' HTCC1062 on a defined medium. *ISME J.* 7, 592–602. doi: 10.1038/ismej.2012.122

Carlson, C. A., Ducklow, H. W., and Sleeter, T. D. (1996). Stocks and dynamics of bacterioplankton in the northwestern Sargasso Sea. *Deep-Sea Res. Part Ii-Topical Stud. Oceanogr.* 43, 491–515.

Castelle, C. J., Brown, C. T., Anantharaman, K., Probst, A. J., Huang, R. H., and Banfield, J. F. (2018). Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. *Nat. Rev. Microbiol.* 16, 629–645. doi: 10.1038/s41579-018-0076-2

Cavicchioli, R., Ripple, W. J., Timmis, K. N., Azam, F., Bakken, L. R., Baylis, M., et al. (2019). Scientists' warning to humanity: microorganisms and climate change. *Nat. Rev. Microbiol.* 17, 569–586. doi: 10.1038/s41579-019-0222-5

Cermeño, P., Teixeira, I. G., Branco, M., Figueiras, F. G., and Marañón, E. (2014). Sampling the limits of species richness in marine phytoplankton communities. *J. Plankton Res.* 36, 1135–1139. doi: 10.1093/plankt/fbu033

Chen, X., Kong, Y., Zhang, S., Zhao, J., Li, S., and Yao, M. (2022). Comparative evaluation of common materials as passive samplers of environmental DNA. *Environ. Sci. Technol.* 56, 10798–10807. doi: 10.1021/acs.est.2c02506

Chisholm, S. W., Frankel, S. L., Goericke, R., Olson, R. J., Palenik, B., Waterbury, J. B., et al. (1992). Prochlorococcus-Marinus Nov gen-Nov Sp – an Oxyphototrophic marine prokaryote containing Divinyl chlorophyll-a and chlorophyll-B. *Arch. Microbiol.* 157, 297–300. doi: 10.1007/BF00245165

Cho, B. C., and Azam, F. (1990). Biogeochemical significance of bacterial biomass in the ocean's euphotic zone. *Mar. Ecol. Prog. Ser.* 63, 253–259.

Claudet, J., Bopp, L., Cheung, W. W. L., Devillers, R., Escobar-Briones, E., Haugan, P., et al. (2020). A roadmap for using the UN decade of ocean science for sustainable development in support of science, policy, and action. *One Earth* 2, 34–42. doi: 10.1016/j.oneear.2019.10.012

Clayton, S., Alexander, H., Graff, J. R., Poulton, N. J., Thompson, L. R., Benway, H., et al. (2022). Bio-GO-SHIP: the time is right to establish global repeat sections of ocean biology. *Front. Mar. Sci.* 8:767443. doi: 10.3389/fmars.2021.767443

Closek, C. J., Santora, J. A., Starks, H. A., Schroeder, I. D., Andruszkiewicz, E. A., Sakuma, K. M., et al. (2019). Marine vertebrate biodiversity and distribution within the Central California current using environmental DNA (eDNA) Metabarcoding and ecosystem surveys. *Front. Mar. Sci.* 6:732. doi: 10.3389/fmars.2019.00732

Cooper, M. K., Villacorta-Rath, C., Burrows, D., Jerry, D. R., Carr, L., Barnett, A., et al. (2022). Practical eDNA sampling methods inferred from particle size distribution and comparison of capture techniques for a critically endangered elasmobranch. *Environ. DNA* 4, 1011–1023. doi: 10.1002/edn3.279

Cottrell, M. T., and Kirchman, D. L. (2004). Single-cell analysis of bacterial growth, cell size, and community structure in the Delaware estuary. *Aquat. Microb. Ecol.* 34, 139–149. doi: 10.3354/ame034139

Cox, A. M., and Goodwin, K. D. (2013). Sample preparation methods for quantitative detection of DNA by molecular assays and marine biosensors. *Mar. Pollut. Bull.* 73, 47–56. doi: 10.1016/j.marpolbul.2013.06.006

Cruaud, P., Vigneron, A., Fradette, M.-S., Charette, S. J., Rodriguez, M. J., Dorea, C. C., et al. (2017). Open the SterivexTM casing: an easy and effective way to improve DNA extraction yields. *Limnol. Oceanogr. Methods* 15, 1015–1020. doi: 10.1002/lom3.10221

de Boer, R., Peters, R., Gierveld, S., Schuurman, T., Kooistra-Smid, M., and Savelkoul, P. (2010). Improved detection of microbial DNA after bead-beating before DNA isolation. *J. Microbiol. Methods* 80, 209–211. doi: 10.1016/j.mimet.2009.11.009

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

DeLong, E. F. (1992). Archaea in coastal marine environments. *Proc. Natl. Acad. Sci. U. S. A.* 89, 5685–5689.

DeLong, E. F., and Beja, O. (2010). The light-driven proton pump proteorhodopsin enhances bacterial survival during tough times. *PLoS Biol.* 8:e1000359. doi: 10.1371/journal.pbio.1000359

DeLong, E. F., Franks, D. G., and Alldredge, A. L. (1993). Phylogenetic diversity of aggregate-attached vs free-living marine bacterial assemblages. *Limnol. Oceanogr.* 38, 924–934.

Dempster, E. L., Pryor, K. V., Francis, D., Young, J. E., and Rogers, H. J. (1999). Rapid DNA extraction from eerns for PCR-based analyses. *BioTechniques* 27, 66–68. doi: 10.2144/99271bm13

Dickie, I. A., Boyer, S., Buckley, H. L., Duncan, R. P., Gardner, P. P., Hogg, I. D., 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

Diez, B., Pedrós-Alió, C., and Massana, R. (2001). Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. *Appl. Environ. Microbiol.* 67, 2932–2941. doi: 10.1128/AEM.67.7.2932-2941.2001

Djurhuus, A., Closek, C. J., Kelly, R. P., Pitz, K. J., Michisaki, R. P., Starks, H. A., et al. (2020). Environmental DNA reveals seasonal shifts and potential interactions in a marine community. *Nat. Commun.* 11:254. doi: 10.1038/s41467-019-14105-1

Djurhuus, A., Port, J., Closek, C. J., Yamahara, K. M., Romero-Maraccini, O., Walz, K. R., et al. (2017). Evaluation of filtration and DNA extraction methods for environmental DNA biodiversity assessments across multiple trophic levels. *Front. Mar. Sci.* 4:314. doi: 10.3389/fmars.2017.00314

Eble, J. E., Daly-Engel, T. S., DiBattista, J. D., Koziol, A., and Gaither, M. R. (2020). "Chapter two: marine environmental DNA: approaches, applications, and opportunities," in *Advances in Marine Biology*. ed. C. Sheppard (Academic Press), 141–169.

Eichmiller, J. J., Miller, L. M., and Sorensen, P. W. (2016). Optimizing techniques to capture and extract environmental DNA for detection and quantification of fish. *Mol. Ecol. Resour.* 16, 56–68. doi: 10.1111/1755-0998.12421

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

Fuhrman, J. A., and Azam, F. (1982). Thymidine incorporation as a measure of heterotrophic bacterioplankton production in marine surface waters: evaluation and field results. *Mar. Biol.* 66, 109–120.

Fuhrman, J. A., and Davis, A. A. (1997). Widespread archaea and novel bacteria from the deep sea as shown by 16S rRNA gene sequences. *Mar. Ecol. Prog. Ser.* 150, 275–285.

Fuhrman, J. A., Mccallum, K., and Davis, A. A. (1993). Phylogenetic diversity of subsurface marine microbial communities from the Atlantic and Pacific oceans. *Appl. Environ. Microbiol.* 59, 1294–1302.

Fuhrman, J. A., and Ouverney, C. C. (1998). Marine microbial diversity studied via 16S rRNA sequences: cloning results from coastal waters and counting of native archaea with fluorescent single cell probes. *Aquat. Ecol.* 32, 3–15.

Galand, P. E., Casamayor, E. O., Kirchman, D. L., and Lovejoy, C. (2009). Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc. Natl. Acad. Sci. U. S. A.* 106, 22427–22432. doi: 10.1073/pnas.0908284106

Ganesh, S., Parris, D. J., DeLong, E. F., and Stewart, F. J. (2014). Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone. *ISME J.* 8, 187–211. doi: 10.1038/ismej.2013.144

Geerts, A. N., Boets, P., van den Heede, S., Goethals, P., and van der heyden, C. (2018). A search for standardized protocols to detect alien invasive crayfish based on environmental DNA (eDNA): a lab and field evaluation. *Ecol. Indic.* 84, 564–572. doi: 10.1016/j.ecolind.2017.08.068

Gilbert, J. A., Field, D., Swift, P., Thomas, S., Cummings, D., Temperton, B., et al. (2010). The taxonomic and functional diversity of microbes at a temperate coastal site: a 'multi-omic' study of seasonal and diel temporal variation. *PLoS One* 5:e15545. doi: 10.1371/journal.pone.0015545

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

Giovannoni, S. J., DeLong, E. F., Schmidt, T. M., and Pace, N. R. (1990). Tangential flow filtration and preliminary phylogenetic analysis of marine picoplankton. *Appl. Environ. Microbiol.* 56, 2572–2575. doi: 10.1128/aem.56.8.2572-2575.1990

Gold, Z., Sprague, J., Kushner, D. J., Zerecero Marin, E., and Barber, P. H. (2021). eDNA metabarcoding as a biomonitoring tool for marine protected areas. *PLoS One* 16:e0238557. doi: 10.1371/journal.pone.0238557

Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thomsen, P. F., Murphy, M. A., et al. (2016). Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods Ecol. Evol.* 7, 1299–1307. doi: 10.1111/2041-210X.12595

Gorzelak, M. A., Gill, S. K., Tasnim, N., Ahmadi-Vand, Z., Jay, M., and Gibson, D. L. (2015). Methods for improving human gut microbiome data by reducing variability through sample processing and storage of stool. *PLoS One* 10:e0134802. doi: 10.1371/journal.pone.0134802

Govindarajan, A. F., McCartin, L., Adams, A., Allan, E., Belani, A., Francolini, R., et al. (2022). Improved biodiversity detection using a large-volume environmental DNA sampler with in situ filtration and implications for marine eDNA sampling strategies. *Deep Sea Res.* Part I: *Oceanogr. Res. Pap.* 189:103871.

Grey, E. K., Bernatchez, L., Cassey, P., Deiner, K., Deveney, M., Howland, K. L., et al. (2018). Effects of sampling effort on biodiversity patterns estimated from environmental DNA metabarcoding surveys. *Sci. Rep.* 8:8843. doi: 10.1038/s41598-018-27048-2

Hagström, U., Larsson, U., Hörstedt, P., and Normark, S. (1979). Frequency of dividing cells, a new approach to the determination of bacterial growth rates in aquatic environments. *Appl. Environ. Microbiol.* 37, 805–812. doi: 10.1128/aem.37.5.805-812.1979

Harrison, J. B., Sunday, J. M., and Rogers, S. M. (2019). Predicting the fate of eDNA in the environment and implications for studying biodiversity. *Proc. R. Soc. B Biol. Sci.* 286:20191409. doi: 10.1098/rspb.2019.1409

Hinlo, R., Gleeson, D., Lintermans, M., and Furlan, E. (2017). Methods to maximise recovery of environmental DNA from water samples. *PLoS One* 12:e0179251. doi: 10.1371/journal.pone.0179251

Horrigan, S. G., Hagstrom, A., Koike, I., and Azam, F. (1988). Inorganic nitrogen utilization by assemblages of marine bacteria in seawater culture. *Mar. Ecol. Prog. Ser.* 50, 147–150. doi: 10.3354/meps050147

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

Hunter, M. E., Ferrante, J. A., Meigs-Friend, G., and Ulmer, A. (2019). Improving eDNA yield and inhibitor reduction through increased water volumes and multi-filter isolation techniques. *Sci. Rep.* 9:5259. doi: 10.1038/s41598-019-40977-w

Jeunen, G. J., Knapp, M., Spencer, H. G., Taylor, H. R., Lamare, M. D., Stat, M., et al. (2019). Species-level biodiversity assessment using marine environmental DNA metabarcoding requires protocol optimization and standardization. *Ecol. Evol.* 9, 1323–1335. doi: 10.1002/ece3.4843

Jo, T., and Minamoto, T. (2021). Complex interactions between environmental DNA (eDNA) state and water chemistries on eDNA persistence suggested by meta-analyses. *Mol. Ecol. Resour.* 21, 1490–1503. doi: 10.1111/1755-0998.13354

John, S. G., Mendez, C. B., Deng, L., Poulos, B., Kauffman, A. K. M., Kern, S., et al. (2011). A simple and efficient method for concentration of ocean viruses by chemical flocculation. *Environ. Microbiol. Rep.* 3, 195–202. doi: 10.1111/j.1758-2229.2010.00208.x

Jones, A., Torkel, C., Stanley, D., Nasim, J., Borevitz, J., and Schwessinger, B. (2021). High-molecular weight DNA extraction, clean-up and size selection for long-read sequencing. *PLoS One* 16:e0253830. doi: 10.1371/journal.pone.0253830

Karst, S. M., Ziels, R. M., Kirkegaard, R. H., Sørensen, E. A., McDonald, D., Zhu, Q., et al. (2021). High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. *Nat. Methods* 18, 165–169. doi: 10.1038/s41592-020-01041-y

Kawato, M., Yoshida, T., Miya, M., Tsuchida, S., Nagano, Y., Nomura, M., et al. (2021). Optimization of environmental DNA extraction and amplification methods for metabarcoding of deep-sea fish. *MethodsX* 8:101238. doi: 10.1016/j.mex.2021.101238

Kelly, R. P., Shelton, A. O., and Gallego, R. (2019). Understanding PCR processes to draw meaningful conclusions from environmental DNA studies. *Sci. Rep.* 9:12133. doi: 10.1038/s41598-019-48546-x

Kemp, B. M., and Smith, D. G. (2005). Use of bleach to eliminate contaminating DNA from the surface of bones and teeth. *Forensic Sci. Int.* 154, 53–61. doi: 10.1016/j.forsciint.2004.11.017

Kirchman, D. L., Elifantz, H., Dittel, A. I., Malmstrom, R. R., and Cottrell, M. T. (2007). Standing stocks and activity of archaea and bacteria in the western Arctic Ocean. *Limnol. Oceanogr.* 52, 495–507. doi: 10.4319/lo.2007.52.2.0495

Klepke, M. J., Sigsgaard, E. E., Jensen, M. R., Olsen, K., and Thomsen, P. F. (2022). Accumulation and diversity of airborne, eukaryotic environmental DNA. *Environ. DNA* 4, 1323–1339. doi: 10.1002/edn3.340

Klymus, K. E., Richter, C. A., Chapman, D. C., and Paukert, C. (2015). Quantification of eDNA shedding rates from invasive bighead carp Hypophthalmichthys nobilis and silver carp Hypophthalmichthys molitrix. *Biol. Conserv.* 183, 77–84. doi: 10.1016/j.biocon.2014.11.020

Könneke, M., Bernhard, A. E., de la Torre, J. R., Walker, C. B., Waterbury, J. B., and Stahl, D. A. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* 437, 543–546. doi: 10.1038/nature03911

Kumar, G., Eble, J. E., and Gaither, M. R. (2020). A practical guide to sample preservation and pre-PCR processing of aquatic environmental DNA. *Mol. Ecol. Resour.* 20, 29–39. doi: 10.1111/1755-0998.13107

Kumar, G., Farrell, E., Reaume, A. M., Eble, J. A., and Gaither, M. R. (2022). One size does not fit all: tuning eDNA protocols for high-and low-turbidity water sampling. *Environ. DNA* 4, 167–180. doi: 10.1002/edn3.235

Ladell, B. A., Walleser, L. R., McCalla, S. G., Erickson, R. A., and Amberg, J. J. (2019). Ethanol and sodium acetate as a preservation method to delay degradation of environmental DNA. *Conserv. Genet. Resour.* 11, 83–88. doi: 10.1007/s12686-017-0955-2

Lauber, C. L., Zhou, N., Gordon, J. I., Knight, R., and Fierer, N. (2010). Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. *FEMS Microbiol. Lett.* 307, 80–86. doi: 10.1111/j.1574-6968.2010.01965.x

Leinen, M., Chavez, F., Meyer, R., Buttigieg, P. L., Davies, N., Casotti, R., et al. (2022). The ocean biomolecular observing network (OBON). *Mar. Technol. Soc. J.* 56, 106–107. doi: 10.4031/MTSJ.56.3.20

Liang, Z., and Keeley, A. (2013). Filtration recovery of extracellular DNA from environmental water samples. *Environ. Sci. Technol.* 47, 9324–9331. doi: 10.1021/es401342b

Longmire, J., Maltbie, M., and Baker, R. (1997). Use of "lysis buffer" in DNA Isolation and its Implication for Museum Collections. *Occasional Papers*, Museum of Texas Tech University. 163, 1–3.

Majaneva, M., Diserud, O. H., Eagle, S. H. C., Boström, E., Hajibabaei, M., and Ekrem, T. (2018). Environmental DNA filtration techniques affect recovered biodiversity. *Sci. Rep.* 8:4682. doi: 10.1038/s41598-018-23052-8

Mäki, A., Salmi, P., Mikkonen, A., Kremp, A., and Tiirola, M. (2017). Sample preservation, DNA or RNA extraction and data analysis for high-throughput phytoplankton community sequencing. *Front. Microbiol.* 8:1848. doi: 10.3389/fmicb.2017.01848

Malone, T. C., and Ducklow, H. W. (1990). Microbial biomass in the coastal plume of Chesapeake Bay-phytoplankton-Bacterioplankton relationships. *Limnol. Oceanogr.* 35, 296–312. doi: 10.4319/lo.1990.35.2.0296

Mardis, E. R. (2017). DNA sequencing technologies: 2006-2016. *Nat. Protoc.* 12, 213–218. doi: 10.1038/nprot.2016.182

Marotz, C., Amir, A., Humphrey, G., Gaffney, J., Gogul, G., and Knight, R. (2017). DNA extraction for streamlined metagenomics of diverse environmental samples. *BioTechniques* 62, 290–293. doi: 10.2144/000114559

Marshall, N. T., Vanderploeg, H. A., and Chaganti, S. R. (2021). Environmental (e)RNA advances the reliability of eDNA by predicting its age. *Sci. Rep.* 11:2769. doi: 10.1038/s41598-021-82205-4

Maruyama, A., Nakamura, K., Yamanaka, H., Kondoh, M., and Minamoto, T. (2014). The release rate of environmental DNA from juvenile and adult fish. *PLoS One* 9:e114639. doi: 10.1371/journal.pone.0114639

Mathieu, C., Hermans, S. M., Lear, G., Buckley, T. R., Lee, K. C., and Buckley, H. L. (2020). A systematic review of sources of variability and uncertainty in eDNA data for environmental monitoring. *Front. Ecol. Evol.* 8:135. doi: 10.3389/fevo.2020.00135

Mauvisseau, Q., Harper, L. R., Sander, M., Hanner, R. H., Kleyer, H., and Deiner, K. (2022). The multiple states of environmental DNA and what is known about their persistence in aquatic environments. *Environ. Sci. Technol.* 56, 5322–5333. doi: 10.1021/acs.est.1c07638

Mayjonade, B., Gouzy, J., Donnadieu, C., Pouilly, N., Marande, W., Callot, C., et al. (2016). Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. *BioTechniques* 61, 203–205. doi: 10.2144/000114460

McCarthy, A., Chiang, E., Schmidt, M. L., and Denef, V. J. (2015). RNA preservation agents and nucleic acid extraction method bias perceived bacterial community composition. *PLoS One* 10:e0121659. doi: 10.1371/journal.pone.0121659

McElroy, M. E., Dressler, T. L., Titcomb, G. C., Wilson, E. A., Deiner, K., Dudley, T. L., et al. (2020). Calibrating environmental DNA Metabarcoding to conventional surveys for measuring fish species richness. *Front. Ecol. Evol.* 8:8. doi: 10.3389/fevo.2020.00276

Mitchell, K. R., and Takacs-Vesbach, C. D. (2008). A comparison of methods for total community DNA preservation and extraction from various thermal environments. *J. Ind. Microbiol. Biotechnol.* 35, 1139–1147. doi: 10.1007/s10295-008-0393-y

Miya, M., Gotoh, R. O., and Sado, T. (2020). MiFish metabarcoding: a high-throughput approach for simultaneous detection of multiple fish species from environmental DNA and other samples. *Fish. Sci.* 86, 939–970. doi: 10.1007/s12562-020-01461-x

Moon-van der Staay, S. Y., De Wachter, R., and Vaulot, D. (2001). Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* 409, 607–610. doi: 10.1038/35054541

Moran, M. A. (2015). The global ocean microbiome. *Science* 350:6266. doi: 10.1126/science.aac8455

Needham, D. M., and Fuhrman, J. A. (2016). Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. *Nat. Microbiol.* 1:16005. doi: 10.1038/nmicrobiol.2016.5

Oldham, A. L., Sandifer, V., and Duncan, K. E. (2019). Effects of sample preservation on marine microbial diversity analysis. *J. Microbiol. Methods* 158, 6–13. doi: 10.1016/j.mimet.2019.01.006

Olson, D. K., Yoshizawa, S., Boeuf, D., Iwasaki, W., and DeLong, E. F. (2018). Proteorhodopsin variability and distribution in the North Pacific subtropical gyre. *ISME J.* 12, 1047–1060. doi: 10.1038/s41396-018-0074-4

Orsi, W. D., Smith, J. M., Wilcox, H. M., Swalwell, J. E., Carini, P., Worden, A. Z., et al. (2015). Ecophysiology of uncultivated marine euryarchaea is linked to particulate organic matter. *ISME J.* 9, 1747–1763. doi: 10.1038/ismej.2014.260

Padilla, C. C., Ganesh, S., Gantt, S., Huhman, A., Parris, D. J., Sarode, N., et al. (2015). Standard filtration practices may significantly distort planktonic microbial diversity estimates. *Front. Microbiol.* 6:547. doi: 10.3389/fmicb.2015.00547

Palumbo, A. V., Ferguson, R. L., and Rublee, P. A. (1984). Size of suspended bacterial cells and association of heterotrophic activity with size fractions of particles in estuarine and coastal waters. *Appl. Environ. Microbiol.* 48, 157–164.

Pawlowski, J., Apotheloz-Perret-Gentil, L., and Altermatt, F. (2020). Environmental DNA: What's behind the term? Clarifying the terminology and recommendations for its future use in biomonitoring. *Mol. Ecol.* 29, 4258–4264. doi: 10.1111/mec.15643

Pearlman, J., Bushnell, M., Coppola, L., Karstensen, J., Buttigieg, P. L., Pearlman, F., et al. (2019). Evolving and Sustaining Ocean best practices and standards for the next decade. *Front. Mar. Sci.*:6. doi: 10.3389/fmars.2019.00277

Pearlman, J., Buttigieg, P. L., Bushnell, M., Delgado, C., Hermes, J., Heslop, E., et al. (2021). Evolving and Sustaining Ocean best practices to enable interoperability in the UN decade of ocean science for sustainable development. *Front. Mar. Sci.*:8. doi: 10.3389/fmars.2021.619685

Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., le Bescot, N., Gorsky, G., et al. (2015). Open science resources for the discovery and analysis of Tara oceans data. *Sci. Data* 2:150023. doi: 10.1038/sdata.2015.23

Pesant, S., Muxagatan, E., Johns, D., Poulain, J., Elineau, A., Huete, M., et al. (2022). Handbook of standards and best practices (version V2), in deliverable 4.2. AtlantECO: Zenodo.

Pomeroy, L. R. (1974). The Ocean's food web, a changing paradigm. *BioScience* 24, 499–504.

Ponsero, A. J., Bomhoff, M., Blumberg, K., Youens-Clark, K., Herz, N. M., Wood-Charlson, E. M., et al. (2020). Planet microbe: a platform for marine microbiology to discover and analyze interconnected 'omics and environmental data. *Nucleic Acids Res.* 49, D792–D802. doi: 10.1093/nar/gkaa637

Postaire, B. D., Bakker, J., Gardiner, J., Wiley, T. R., and Chapman, D. D. (2020). Environmental DNA detection tracks established seasonal occurrence of blacktip sharks (*Carcharhinus limbatus*) in a semi-enclosed subtropical bay. *Sci. Rep.* 10:11847. doi: 10.1038/s41598-020-68843-0

Prince, A. M., and Andrus, L. (1992). Pcr-how to kill unwanted DNA. *BioTechniques* 12:358.

Rappé, M. S., Connon, S. A., Vergin, K. L., and Giovannoni, S. J. (2002). Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* 418, 630–633. doi: 10.1038/nature00917

Rappé, M. S., Suzuki, M. T., Vergin, K. L., and Giovannoni, S. J. (1998). Phylogenetic diversity of ultraplankton plastid small-subunit rRNA genes recovered in environmental nucleic acid samples from the Pacific and Atlantic coasts of the United States. *Appl. Environ. Microbiol.* 64, 294–303. doi: 10.1128/AEM.64.1.294-303.1998

Renshaw, M. A., Olds, B. P., Jerde, C. L., McVeigh, M. M., and Lodge, D. M. (2015). The room temperature preservation of filtered environmental DNA samples and assimilation into a phenol-chloroform-isoamyl alcohol DNA extraction. *Mol. Ecol. Resour.* 15, 168–176. doi: 10.1111/1755-0998.12281

Robson, H. L., Noble, T. H., Saunders, R. J., Robson, S. K. A., Burrows, D. W., and Jerry, D. R. (2016). Fine-tuning for the tropics: application of eDNA technology for invasive fish detection in tropical freshwater ecosystems. *Mol. Ecol. Resour.* 16, 922–932. doi: 10.1111/1755-0998.12505

Rochelle, P. A., Cragg, B. A., Fry, J. C., John Parkes, R., and Weightman, A. J. (1994). Effect of sample handling on estimation of bacterial diversity in marine sediments by 16S rRNA gene sequence analysis. *FEMS Microbiol. Ecol.* 15, 215–225. doi: 10.1111/j.1574-6941.1994.tb00245.x

Rodriguez-Ezpeleta, N., Morissette, O., Bean, C. W., Manu, S., Banerjee, P., Lacoursière-Roussel, A., et al. (2021). Trade-offs between reducing complex terminology and producing accurate interpretations from environmental DNA: comment on "environmental DNA: What's behind the term?" by Pawlowski et al. (2020). *Mol. Ecol.* 30, 4601–4605. doi: 10.1111/mec.15942

Rourke, M. L., Fowler, A. M., Hughes, J. M., Broadhurst, M. K., DiBattista, J. D., Fielder, S., et al. (2022). Environmental DNA (eDNA) as a tool for assessing fish biomass: a review of approaches and future considerations for resource surveys. *Environ. DNA* 4, 9–33. doi: 10.1002/edn3.185

Roux, L.-M. D., Giblot-Ducray, D., Bott, N. J., Wiltshire, K. H., Deveney, M. R., Westfall, K. M., et al. (2020). Analytical validation and field testing of a specific qPCR assay for environmental DNA detection of invasive European green crab (*Carcinus maenas*). *Environ. DNA* 2, 309–320. doi: 10.1002/edn3.65

Ruppert, K. M., Kline, R. J., and Rahman, M. S. (2019). Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: a systematic review in methods, monitoring, and applications of global eDNA. *Global Ecol. Conserv.*:17. doi: 10.1016/j.gecco.2019.e00547

Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooseph, S., et al. (2007). The sorcerer II Global Ocean sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* 5:e77. doi: 10.1371/journal.pbio.0050077

Russo, A., Mayjonade, B., Frei, D., Potente, G., Kellenberger, R. T., Frachon, L., et al. (2022). Low-input high-molecular-weight DNA extraction for long-read sequencing from plants of diverse families. *Front. Plant Sci.* 13:883897. doi: 10.3389/fpls.2022.883897

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. *Front. Mar. Sci.* 8:1488. doi: 10.3389/fmars.2021.758694

Santi, I., Casotti, R., Comtet, T., Cunliffe, M., Koulouri, Y., Macheriotou, L., et al. (2021). *European Marine Omics Biodiversity Observation Network (EMO BON) – Handbook*, Ed. E.M.B.R. Centre.

Schattenhofer, M., Fuchs, B. M., Amann, R., Zubkov, M. V., Tarran, G. A., and Pernthaler, J. (2009). Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. *Environ. Microbiol.* 11, 2078–2093. doi: 10.1111/j.1462-2920.2009.01929.x

Schiebelhut, L. M., Abboud, S. S., Gómez Daglio, L. E., Swift, H. F., and Dawson, M. N. (2017). A comparison of DNA extraction methods for high-throughput DNA analyses. *Mol. Ecol. Resour.* 17, 721–729. doi: 10.1111/1755-0998.12620

Schmidt, P. J., Cameron, E. S., Müller, K. M., and Emelko, M. B. (2022). Ensuring that fundamentals of quantitative microbiology are reflected in microbial diversity analyses based on next-generation sequencing. *Front. Microbiol.* 13:728146. doi: 10.3389/fmicb.2022.728146

Schmidt, T. M., DeLong, E. F., and Pace, N. R. (1991). Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J. Bacteriol.* 173, 4371–4378.

Schrader, C., Schielke, A., Ellerbroek, L., and Johne, R. (2012). PCR inhibitors – occurrence, properties and removal. *J. Appl. Microbiol.* 113, 1014–1026. doi: 10.1111/j.1365-2672.2012.05384.x

Sepulveda, A. J., Hutchins, P. R., Forstchen, M., Mckeefry, M. N., and Swigris, A. M. (2020). The elephant in the lab (and field): contamination in aquatic environmental DNA studies. *Front. Ecol. Evol.* 8:8. doi: 10.3389/fevo.2020.609973

Sergeant, M. J., Constantinidou, C., Cogan, T., Penn, C. W., and Pallen, M. J. (2012). High-throughput sequencing of 16S rRNA gene amplicons: effects of extraction procedure, primer length and annealing temperature. *PLoS One* 7:e38094. doi: 10.1371/journal.pone.0038094

Shogren, A. J., Tank, J. L., Andruszkiewicz, E. A., Olds, B., Jerde, C., and Bolster, D. (2016). Modelling the transport of environmental DNA through a porous substrate using continuous flow-through column experiments. *J. R. Soc. Interface* 13:20160290. doi: 10.1098/rsif.2016.0290

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., et al. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl. Acad. Sci. U. S. A.* 103, 12115–12120. doi: 10.1073/pnas.0605127103

Song, S. J., Amir, A., Metcalf, J. L., Amato, K. R., Xu, Z. Z., Humphrey, G., et al. (2016). Preservation methods differ in fecal microbiome stability, affecting suitability for field studies. *mSystems* 1:e00021-16. doi: 10.1128/mSystems.00021-16

Spens, J., Evans, A. R., Halfmaerten, D., Knudsen, S. W., Sengupta, M. E., Mak, S. S. T., et al. (2017). Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter. *Methods Ecol. Evol.* 8, 635–645. doi: 10.1111/2041-210X.12683

Staley, J. T., and Konopka, A. (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu. Rev. Microbiol.* 39, 321–346.

Stewart, F. J. (2013). Preparation of microbial community cDNA for metatranscriptomic analysis in marine plankton. *Methods Enzymol.* 531, 187–218. doi: 10.1016/B978-0-12-407863-5.00010-1

Stocker, R. (2012). Marine microbes see a sea of gradients. *Science* 338, 628–633. doi: 10.1126/science.1208929

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D. M., Breiner, H. W., et al. (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* 19, 21–31. doi: 10.1111/j.1365-294X.2009.04480.x

Stoeckle, M. Y., Adolf, J., Ausubel, J. H., Charlop-Powers, Z., Dunton, K. J., and Hinks, G. (2022). Current laboratory protocols for detecting fish species with environmental DNA optimize sensitivity and reproducibility, especially for more abundant populations. *ICES J. Mar. Sci.* 79, 403–412. doi: 10.1093/icesjms/fsab273

Stulberg, E., Fravel, D., Proctor, L. M., Murray, D. M., LoTempio, J., Chrisey, L., et al. (2016). An assessment of US microbiome research. *Nat. Microbiol.* 1:15015. doi: 10.1038/nmicrobiol.2015.15

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

Sunagawa, S., DeSantis, T. Z., Piceno, Y. M., Brodie, E. L., DeSalvo, M. K., Voolstra, C. R., et al. (2009). Bacterial diversity and white plague disease-associated community changes in the Caribbean coral Montastraea faveolata. *ISME J.* 3, 512–521. doi: 10.1038/ismej.2008.131

Takahashi, S., Sakata, M. K., Minamoto, T., and Masuda, R. (2020). Comparing the efficiency of open and enclosed filtration systems in environmental DNA quantification for fish and jellyfish. *PLoS One* 15:e0231718. doi: 10.1371/journal.pone.0231718

Takasaki, K., Aihara, H., Imanaka, T., Matsudaira, T., Tsukahara, K., Usui, A., et al. (2021). Water pre-filtration methods to improve environmental DNA detection by real-time PCR and metabarcoding. *PLoS One* 16:e250162. doi: 10.1371/journal.pone.0250162

ten Hoopen, P., Finn, R. D., Bongo, L. A., Corre, E., Fosso, B., Meyer, F., et al. (2017). The metagenomic data life-cycle: standards and best practices. *GigaScience* 6:gix047. doi: 10.1093/gigascience/gix047

Tiedje, J. M., Bruns, M. A., Casadevall, A., Criddle, C. S., Eloe-Fadrosh, E., Karl, D. M., et al. (2022). Microbes and climate change: A research prospectus for the future. *mBio* 13:e00800-22. doi: 10.1128/mbio.00800-22

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

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

Torres-Beltrán, M., Mueller, A., Scofield, M., Pachiadaki, M. G., Taylor, C., Tyshchenko, K., et al. (2019). Sampling and processing methods impact microbial community structure and potential activity in a seasonally anoxic flord: Saanich inlet. *FMARS* 6:132. doi: 10.3389/fmars.2019.00132

Tragin, M., and Vaulot, D. (2018). Green microalgae in marine coastal waters: the ocean sampling day (OSD) dataset. *Sci. Rep.* 8:14020. doi: 10.1038/s41598-018-32338-w

Trigodet, F., Lolans, K., Fogarty, E., Shaiber, A., Morrison, H. G., Barreiro, L., et al. (2022). High molecular weight DNA extraction strategies for long-read sequencing of complex metagenomes. *Mol. Ecol. Resour.* 22, 1786–1802. doi: 10.1111/1755-0998.13588

Truelove, N. K., Patin, N. V., Min, M., Pitz, K. J., Preston, C. M., Yamahara, K. M., et al. (2022). Expanding the temporal and spatial scales of environmental DNA research with autonomous sampling. *Environ. DNA* 4, 972–984. doi: 10.1002/edn3.299

Trujillo-González, A., Villacorta-Rath, C., White, N. E., Furlan, E. M., Sykes, M., Grossel, G., et al. (2021). Considerations for future environmental DNA accreditation and proficiency testing schemes. *Environ. DNA* 3, 1049–1058. doi: 10.1002/edn3.243

Tsuji, S., Yamanaka, H., and Minamoto, T. (2017). Effects of water pH and proteinase K treatment on the yield of environmental DNA from water samples. *Limnology* 18, 1–7. doi: 10.1007/s10201-016-0483-x

Turner, C. R., Barnes, M. A., Xu, C. C. Y., Jones, S. E., Jerde, C. L., and Lodge, D. M. (2014). Particle size distribution and optimal capture of aqueous macrobial eDNA. *Methods Ecol. Evol.* 5, 676–684. doi: 10.1111/2041-210X.12206

United States Environmental Protection Agency (USEPA) (2022). "Method 1696: Characterization of human fecal pollution in water by HF183/BacR287 TaqMan® quantitative polymerase chain reaction (qPCR) assay," in *EPA 821-R-19-002*.

Urakawa, H., Martens-Habbena, W., and Stahl, D. A. (2010). High abundance of ammonia-oxidizing archaea in coastal waters, determined using a modified DNA extraction method. *Appl. Environ. Microbiol.* 76, 2129–2135. doi: 10.1128/AEM.02692-09

Ushio, M. (2019). Use of a filter cartridge combined with intra-cartridge bead-beating improves detection of microbial DNA from water samples. *Methods Ecol. Evol.* 10, 1142–1156. doi: 10.1111/2041-210X.13204

Vaulot, D., Eikrem, W., Viprey, M., and Moreau, H. (2008). The diversity of small eukaryotic phytoplankton (< or =3 microm) in marine ecosystems. *FEMS Microbiol. Rev.* 32, 795–820. doi: 10.1111/j.1574-6976.2008.00121.x

Weber, L., and Apprill, A. (2020). Diel, daily, and spatial variation of coral reef seawater microbial communities. *PLoS One* 15:e0229442. doi: 10.1371/journal.pone.0229442

Weber, L., DeForce, E., and Apprill, A. (2017). Optimization of DNA extraction for advancing coral microbiota investigations. *Microbiome* 5:18. doi: 10.1186/s40168-017-0229-y

Wegleitner, B. J., Jerde, C. L., Tucker, A., Chadderton, W. L., and Mahon, A. R. (2015). Long duration, room temperature preservation of filtered eDNA samples. *Conserv. Genet. Resour.* 7, 789–791. doi: 10.1007/s12686-015-0483-x

Whitman, W. B., Coleman, D. C., and Wiebe, W. J. (1998). Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. U. S. A.* 95, 6578–6583.

Wigington, C. H., Sonderegger, D., Brussaard, C. P. D., Buchan, A., Finke, J. F., Fuhrman, J. A., et al. (2016). Re-examination of the relationship between marine virus and microbial cell abundances. *Nat. Microbiol.* 1:15024. doi: 10.1038/nmicrobiol.2015.24

Wilcox, T. M., McKelvey, K. S., Young, M. K., Sepulveda, A. J., Shepard, B. B., Jane, S. F., et al. (2016). Understanding environmental DNA detection probabilities: a case study using a stream-dwelling char Salvelinus fontinalis. *Biol. Conserv.* 194, 209–216. doi: 10.1016/j.biocon.2015.12.023

Williams, K. E., Huyvaert, K. P., and Piaggio, A. J. (2017). Clearing muddied waters: capture of environmental DNA from turbid waters. *PLoS One* 12:e0179282

Wilson, I. G. (1997). Inhibition and facilitation of nucleic acid amplification. *Appl. Environ. Microbiol.* 63, 3741–3751. doi: 10.1128/aem.63.10.3741-3751.1997

Wymer, L. J., et al., (2010). *Effects of Holding Time, Storage, and the Preservation of Samples on Sample Integrity for the Detection of Fecal Indicator Bacteria by Quantitative Polymerase Chain Reaction (qPCR)-Based Assays*. U.S. Environmental Protection Agency: Washington, DC.

Yamahara, K. M., Preston, C. M., Birch, J., Walz, K., Marin, R. III, 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

Yamanaka, H., Motozawa, H., Tsuji, S., Miyazawa, R. C., Takahara, T., and Minamoto, T. (2016). On-site filtration of water samples for environmental DNA analysis to avoid DNA degradation during transportation. *Ecol. Res.* 31, 963–967. doi: 10.1007/s11284-016-1400-9

Yeo, S. K., Huggett, M. J., Eiler, A., and Rappé, M. S. (2013). Coastal bacterioplankton community dynamics in response to a natural disturbance. *PLoS One* 8:e56207. doi: 10.1371/journal.pone.0056207

Zaiko, A., von Ammon, U., Stuart, J., Smith, K. F., Yao, R., Welsh, M., et al. (2022). Assessing the performance and efficiency of environmental DNA/RNA capture methodologies under controlled experimental conditions. *Methods Ecol. Evol.* 13, 1581–1594. doi: 10.1111/2041-210X.13879

Zehr, J. P., and Kudela, R. M. (2011). Nitrogen cycle of the Open Ocean: from genes to ecosystems. *Annu. Rev. Mar. Sci.* 3, 197–225. doi: 10.1146/annurev-marine-120709-142819

Zhang, Y., Riley, L. K., Lin, M., Purdy, G. A., and Hu, Z. (2013). Development of a virus concentration method using lanthanum-based chemical flocculation coupled with modified membrane filtration procedures. *J. Virol. Methods* 190, 41–48. doi: 10.1016/j.jviromet.2013.03.017

Zhao, B., van Bodegom, P. M., and Trimbos, K. (2021). The particle size distribution of environmental DNA varies with species and degradation. *Sci. Total Environ.* 797:149175. doi: 10.1016/j.scitotenv.2021.149175

Zinger, L., Bonin, A., Alsos, I. G., Bálint, M., Bik, H., Boyer, F., et al. (2019). DNA metabarcoding-need for robust experimental designs to draw sound ecological conclusions. *Mol. Ecol.* 28, 1857–1862. doi: 10.1111/mec.15060

frontiers | Frontiers in Microbiology

TYPE Review
PUBLISHED 20 February 2023
DOI 10.3389/fmicb.2022.1007877

OPEN ACCESS

EDITED BY
Catarina Magalhães,
University of Porto,
Portugal

REVIEWED BY
Francisco P. Chávez,
University of Chile,
Chile
Julie L. Meyer,
University of Florida,
United States

*CORRESPONDENCE
Denise P. Silva
✉ dapiedad@oregonstate.edu

SPECIALTY SECTION
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 31 July 2022
ACCEPTED 30 December 2022
PUBLISHED 20 February 2023

CITATION
Silva DP, Epstein HE and Vega Thurber RL (2023) Best practices for generating and analyzing 16S rRNA amplicon data to track coral microbiome dynamics. *Front. Microbiol.* 13:1007877. doi: 10.3389/fmicb.2022.1007877

Best practices for generating and analyzing 16S rRNA amplicon data to track coral microbiome dynamics

Denise P. Silva*, Hannah E. Epstein and Rebecca L. Vega Thurber

Department of Microbiology, Oregon State University, Corvallis, OR, United States

Over the past two decades, researchers have searched for methods to better understand the relationship between coral hosts and their microbiomes. Data on how coral-associated bacteria are involved in their host's responses to stressors that cause bleaching, disease, and other deleterious effects can elucidate how they may mediate, ameliorate, and exacerbate interactions between the coral and the surrounding environment. At the same time tracking coral bacteria dynamics can reveal previously undiscovered mechanisms of coral resilience, acclimatization, and evolutionary adaptation. Although modern techniques have reduced the cost of conducting high-throughput sequencing of coral microbes, to explore the composition, function, and dynamics of coral-associated bacteria, it is necessary that the entire procedure, from collection to sequencing, and subsequent analysis be carried out in an objective and effective way. Corals represent a difficult host with which to work, and unique steps in the process of microbiome assessment are necessary to avoid inaccuracies or unusable data in microbiome libraries, such as off-target amplification of host sequences. Here, we review, compare and contrast, and recommend methods for sample collection, preservation, and processing (e.g., DNA extraction) pipelines to best generate 16S amplicon libraries with the aim of tracking coral microbiome dynamics. We also discuss some basic quality assurance and general bioinformatic methods to analyze the diversity, composition, and taxonomic profiles of the microbiomes. This review aims to be a generalizable guide for researchers interested in starting and modifying the molecular biology aspects of coral microbiome research, highlighting best practices and tricks of the trade.

KEYWORDS

coral, microbiology, methods, high-throughput sequencing (HTS), 16S rRNA amplicon, microbiome

1. Introduction on coral microbiomes and measures to track their dynamics

Coral-associated microorganisms are critical in the maintenance of animal health, especially in the face of environmental stressors (Bourne et al., 2008; Morrow et al., 2018; Pootakham et al., 2018). The community of these microorganisms (referred to as the coral microbiome) has been identified as a lead indicator of coral health (see Box 1 for more details), with diagnostic signatures that predict coral bleaching, disease, and mortality (Bourne et al., 2016; Zaneveld et al., 2017; Glasl et al., 2019). In the last several years, culture-independent methods for interrogating the coral microbiome have become essential for exploring the impacts of microbiome variability. Specifically, High-Throughput Sequencing (HTS) of 16S ribosomal RNA (rRNA) genes has been widely adopted to understand bacterial and archaeal diversity more generally, making the profiling of existing microbiomes in different host species (Kamke et al., 2010; Ford et al., 2021) a common analysis in the coral field

(Siboni et al., 2008; Wada et al., 2016; Ziegler et al., 2017). These studies have advanced our understanding of the role microbes play in coral health, and have produced novel approaches for maintaining or enhancing coral resilience to environmental change, including microbiome engineering (e.g., the manipulation of microorganisms for the benefit of coral health) that can bolster the remediation and protection of corals against rising marine temperatures (Reshef et al., 2006; Peixoto et al., 2017; Epstein et al., 2019; Rosado et al., 2019; Santoro et al., 2021) or contamination from pollutants (Fragoso Ados Santos et al., 2015; Silva et al., 2021).

Despite the widespread adoption of HTS methods, the use of culturing methods has not lost its importance in microbiome research. Culture-dependent methods remain crucial for understanding the physiology and metabolism of coral-associated taxa and the interaction between the microbiome, the host, and the environment (see review by Schultz et al., 2022). In addition, culturomics methods can be complementary to HTS, allowing for a complete assessment of the coral-associated microbiome. However, corals are challenging to work with and require additional steps to isolate, extract, generate, and curate microbiome data. Thus, the careful collection, preservation, and processing of coral samples must be optimized to characterize and assess the coral microbiome using different HTS techniques accurately and precisely (Vega Thurber et al., 2022). Thus far, little guidance has been formalized on these best practices in corals due, in part, to the rapid expansion of HTS techniques and the influx of new researchers who aim to conduct them. We envision this article to be a practical guide of well adopted practices for readers who hope to use 16S rRNA gene-based data for conducting analyses of coral microbiomes. Specifically, we describe (1) strategies for collection, preservation, and processing samples to assess the different compartments of coral; (2) comparative extraction methods to isolate and preserve microbial DNA; (3) strategies to avoid host and off-target contamination during PCR and HTS library construction; (4) common approaches and issues surrounding different sequencing platforms; (5) basic quality control, analytical pipelines, and software that can be used to access some measures of coral microbiome diversity, composition, and stability.

Box 1: A Primer on Coral Microbiome Research.

Microorganisms are crucial biological components of all living organisms and influence ecological processes (Fraune and Bosch, 2010; Gibbons and Gilbert, 2015). Assessment of their ecology and evolution within hosts has been significantly advanced during the sequencing revolution of the 2000s and today. Due to early adopters (Wegley et al., 2007; Vega Thurber et al., 2009; Littman et al., 2011; Pollock et al., 2011; Sato et al., 2013), corals themselves were a touchstone of using HTS advances in interrogating microbiome features and dynamics within hosts and the environment. As a result, we know a significant amount about coral and reef microbiomes and their dynamics. Since the late 1990s and early 2000s the interrogation of what lives on and in corals, and how they change in response to numerous perturbations, has led to several hypotheses about the role of the myriad members of the coral holobiont.

The coral microbiome is composed of endosymbiotic algae (Symbiodinaceae), microeukaryotes, bacteria, archaea, viruses, fungi, and protozoa (Rosenberg, 2009; Sunagawa et al., 2010; Garren and Azam, 2012; Hernandez-Agreda et al., 2017; van Oppen and Blackall, 2019). Spatially and taxonomically distinct communities colonize all anatomical compartments of corals, which are most commonly split into three for comparative microbial analyses: the surface mucus layer, coral tissue, and coral skeleton (e.g., Sweet et al., 2011; Li et al., 2014; Pollock et al., 2018; Ricci et al., 2022). These compartments each provide unique environmental and physical conditions that can select for specific microbial communities depending on what resources are available (Sweet et al., 2011). Clear differentiation of microbiome structure and function have been found across these three major components in both individual coral species (Bourne and Munn, 2005; Sweet et al., 2011; Li et al., 2014) and among the scleractinian tree of life (32 coral species; Pollock et al., 2018). According to Ricci et al. (2022) the coral microhabitat niche and the phylogenetic characteristics of the host, shape the presence and relative abundance of symbiotic bacterial microorganisms.

Interactions among corals and their microorganisms can be mutualistic, antagonistic, commensal, and competitive. As is true of all symbiosis, these relationships can shift dramatically due to alterations in host or symbiont physiology, the environment, or both (for review see Maher et al., 2022). Bacteria, the most diverse taxonomic and metabolic lineage within the coral microbiome (Rohwer et al., 2001; Rosenberg et al., 2007; Rosenberg, 2009), can serve several essential functions that benefit the holobiont as mutualists, including host protection from pathogens or opportunists (e.g., *via* occlusion and/or antibiotic production; Ritchie, 2006; Bythell and Wild, 2011; Krediet et al, 2013), nutritional supplementation (e.g., vitamins and amino acids; Shinzato et al., 2011) metabolic expansion (e.g., sulfur and nitrogen cycling; Cai et al., 2018; Robbins et al., 2019) and increased growth, survival, and health maintenance through other mechanisms yet untested (Brown and Bythell, 2005; Rädecker et al., 2015; Hartmann et al., 2017; Webster and Reusch, 2017). However, if the coral is environmentally or physically stressed, both resident and/or transient bacteria can become opportunistically pathogenic and cause serious damage, infection, and/or disease (for review see Vega Thurber et al., 2020).

2. Use of high-throughput sequencing in tracking microbiome dyamics

Due to the limitations of culture-based methods, culture-independent techniques have progressed overwhelmingly in the past 20 years. Early studies utilizing culture-independent methods relied on low-throughput sequencing technology (e.g., Sanger sequencing) and finger-printing methods (e.g., TRFLP and DGGE) that led to many foundational inferences about the coral microbiome (Rohwer et al., 2002; Klaus et al., 2005; Sunagawa et al., 2010). Because these techniques result in relatively low numbers of sequences for fairly high costs, along with issues in poorly curated databases for comparative analysis, these techniques fell out of favor as HTS gained traction. The major advantages of HTS methods are the (1) high yield of data resulting from millions to billions of sequencing reads in a single run, (2) the low cost per base, and (3) comparable genetic data for cross-system compatibility due to wide adoption across the microbial ecology field. With HTS technology, it is possible to design and implement experiments with many more samples and replication (i.e., increased statistical power) that can provide advantages when assessing changes in microbial composition of different coral species and exploring spatiotemporal variability (Haydon et al., 2022) while decreasing the likelihood of type I and type II errors.

Currently, the most used microbiome HTS methods is amplicon sequencing, or the amplification of a single or multiple gene sequences [e.g., 16S rRNA and *recA*, *gryA* genes and the internal transcribed spacer (ITS) region]. Most of these target genes are present across specific domains or clades of organisms, and many are well studied molecular clocks that are useful for phylogenetics (Yang et al., 2016). HTS technology has itself evolved with many iterations and platforms including: 454 pyrosequencing (Margulies et al., 2005), Ion Torrent PGM (Personal Genome machine; Rothberg et al., 2011), PacBio (Pacific Biosciences; Eid et al., 2009), MinIon (nanopore sequencing; Mikheyev and Tin, 2014), and Illumina (Bentley et al., 2008). Over the past 5 years, Illumina Miseq and Hiseq platforms have been the most widely used

although the HiSeq platform is now being decommissioned and replaced by the Illumina NextSeq and NovaSeq platforms.

These sequencers can produce single reads of varying lengths, and platforms can also generate linked or 'paired end reads' that represent the forward and reverse portions of longer amplicons that may not reach across the sequenced reads. Along with length, importantly, the number of resulting reads also can vary significantly. For example, where NextSeq 550 has generated between 260 and 800 million reads, and MiSeq (with the Reagent Kit V3) can currently result in 30–40 million read pairs. Other platforms in place of the HiSeq, such as NextSeq 1,000 and 2000, where the maximum read length is 2×300 bp, can generate between 100 million–2.4 billion reads as of December 2022. Higher read depth can increase clustering and longer read length can increase the frequency of accurate taxonomic calls using reference libraries (see data analytics section below). Currently, MiSeq is commonly used for amplicon sequencing due to its long read length (2×300 bp), low cost, and high accuracy. However, the newest and most advanced Illumina sequencing platform, NovaSeq, is capable of generating up to 40 billion paired-end reads (reads lengths up to 2×250 bp) at a low cost. Thus, NovaSeq can be used for large-scale projects. Therefore, the optimal platform will depend on the nature and the objective of the study. Further, sequencing platforms are a rapidly evolving technology, and we encourage readers to compare platforms to inform such decisions (e.g., Singer et al., 2019).

2.1. The increasing number of options for the use of HTS in coral microbiome analysis

Due to the increase in the diversity of genetic tools and analysis pipelines, many researchers entering the coral microbiology field struggle to determine the best and most adopted techniques to answer specific questions. At the same time, methods rapidly change, with new techniques constantly pushing the boundaries of what we can do with HTS data. New concepts in how samples should be processed and analyzed are constantly changing. Thus, in addition to the experimental design considerations, we must consider factors that may influence the choice and application of both processing and analytical methods to elucidate different aspects of coral microbiology.

Identifying the best sample processing methods and molecular techniques to apply in a study can be laborious, and manuscripts can suffer in peer-review if the methods are not up-to-date or fully benchmarked. That said, no method is a panacea, and appropriate methods of collection, preservation, processing, and molecular tools must be tailored depending on the focus or question of the study, the source of materials, and the samples' history and provenance. For example, each coral species, individual, or even compartment may require optimization of techniques. These study-specific details will matter when designing and optimizing the HTS approach, but fundamentally the steps are similar: collect the sample, extract microbial DNA, amplify the target gene using PCR, generate sequencing libraries, and finally *in silico* analysis of microbiome communities.

2.2. Choosing in-house HTS library preparation vs. commercial or institutional sequence providers

There are now several companies and university core facilities that will conduct many or all of the below steps as paid services. Often the services are itemized and can be adjusted and personalized to best suit any one project. With economies of scale this can be an affordable, standardized, and reliable means to get samples and data back quickly. Both approaches have different benefits, and whether one chooses to conduct the work in-house or through such a provider is entirely up to the researcher's needs, goals, and finances. For example, DNA extraction and HTS library preparation services can reduce time and/or financial costs and lower levels of contamination due to the use of robotic preparations. However, the 'black box' nature of these providers makes scientific transparency difficult and reduces the opportunity for students/researchers to learn the process. For truly comparable datasets, the methods conducted by a service team or set of researchers must be as identical in their protocols as possible, making transparency critical. Even small deviations from any of the major steps of the process can cause extraction, amplification, and sequencing biases that may be revealed in downstream microbiome analyses as differences in taxonomy and composition (see below).

3. Best practices and options in methods for assessment of coral microbiome features

In the following sections, we describe several considerations and methodologies for generating accurate and precise assessments of coral microbiomes, which have unique requirements compared to many other host-associated microbiomes. We envision this as a 'how-to-guide,' but recognize that not all methods will be suitable for every study. Adoption of the methodologies below should always be considered and adapted in light of each research group's specific questions, system needs, and available resources.

3.1. Collection, preservation, and processing methods

Stony corals have only 2 tissue layers (gastroderm and ectoderm), a mesoglea, and an aragonite skeleton (Muscatine, 1969; Grottoli, 2001). Many corals produce a mucus coat for protection from environmental shifts and potential disease-causing agents (Allen, 1983; Shnit-Orland and Kushmaro, 2009). Differences in microbial assemblages between coral compartments (Sweet et al., 2011; Li et al., 2014; Pollock et al., 2018) and the functional role each plays makes the selection, proper sampling and processing of these compartments critical to research outcomes.

Several collection methods and/or sampling techniques can be applied to each coral compartment of interest (Figure 1). These sampling techniques can vary by their invasiveness and potential for negative effects on the holobiont; a consideration that must be ethically and experimentally weighed by the researcher before sampling begins. For example, removal of large portions of a colony can alter host physiology and/or cause extensive damage that may lead to colony mortality. Adjusting sampling to simultaneously reduce negative consequences on the colony or individual while also accurately analyzing the specimen for microbiome features is critical. Below, we describe available methods for sampling, preserving, and processing coral samples for microbiome work and briefly discuss when to use each.

Whole coral specimen – Small fragments (can range between 1 and 8 cm^2) of corals can be used for microbiome research evaluating the

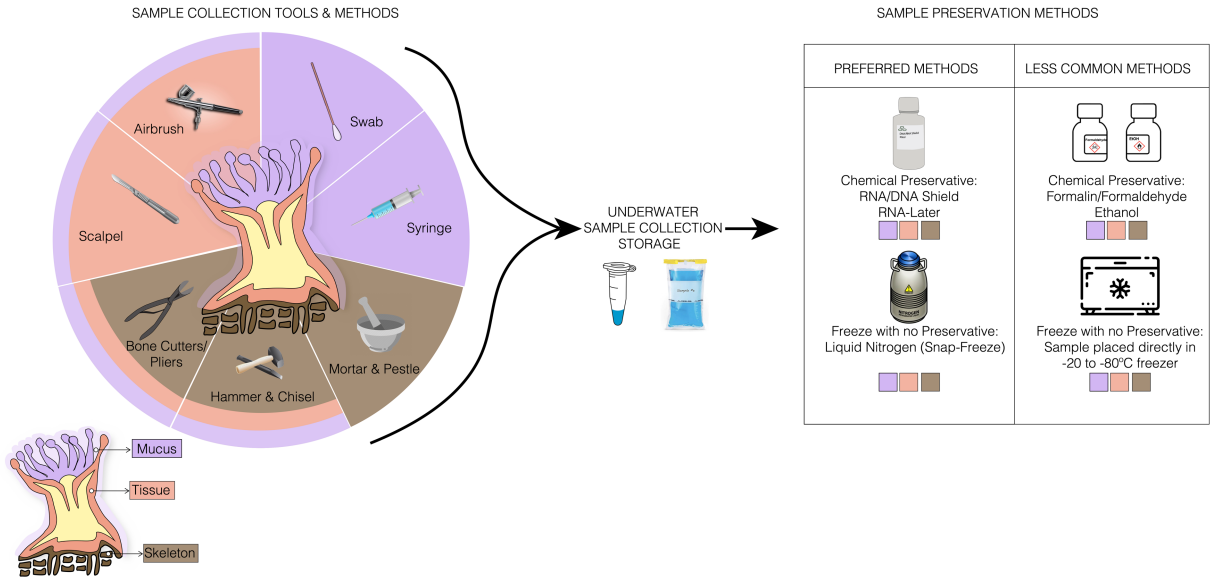

FIGURE 1
Illustration of ideal sampling strategies for each coral compartment, collection storage, and the preferred and less common sample preservation methods used for any coral samples. In samples collection methods, there are specific tools for a compartment of the coral, such as the syringe used only for the collection of mucus, and tools, such as the bone cutter, used to collect any sample from the corals.

dynamics of the holobiont. Experiments or surveys might require repeated sampling of individuals over time or the complete removal of a specimen during experimentation, so it is important to consider size at the beginning of your sampling period. For 16S amplicon analysis, only a very small fragment is required for accurate microbiome characterization, and amplicons can be generated from as little as 2 mm of diameter. However, to our knowledge there have been no studies that systematically examined the efficacy of DNA extraction from different fragment sizes. Typically, if other downstream analyses (e.g., other 'omics or physiological analyses) are conducted alongside 16S analysis, specimen sizes may need to be much larger, depending on the analyses of interest and the size and shape of the colony. During sampling, scleractinian corals are often collected using a hammer and chisel for massive or very thick branched corals, or bone cutters and snips for more delicate and smaller branching or plating corals (Apprill et al., 2016; Neave et al., 2017; Roitman et al., 2020). In the case of branching corals, collections can be made using needle-nose pliers. Underwater work makes it extremely difficult to maintain sterile technique, but wearing gloves and changing tools between sampling, especially if investigating disease, should be done to avoid cross-contamination among samples. See Box 2 for details on PCR contamination and mitigation strategies).

After collection, the specimen should be stored in sterile tubes or together with seawater in sterile, hermetically sealed bags, such as Whirl-Pak bags (Nasco, Salida, CA, United States) filled with local seawater (Kellogg et al., 2016; Neave et al., 2017). These bags are recommended because they are durable, leak-resistant, freezer-safe, and can be ordered with "write-on" labels that ensure permanent pen will not rub off in the freezer.

After collection, it is necessary to rapidly remove any excess liquid and place samples in as cold of conditions as possible (ideally ultra-frozen in liquid nitrogen), within preservatives (e.g., DNA/RNA Shield, RNAlater, or salt buffered DMSO), or fixatives (e.g., aldehydes) to prevent the microbiome from changing in composition, total abundance, and function (Gaither et al., 2011; Hernandez-Agreda et al., 2018b; Gardner et al., 2019; Pratte and Kellogg, 2021). The ultra-freezing method is widely used, as it preserves the sample instantly and leaves it free of artifacts present in chemical preservatives (Vega Thurber et al., 2022). However, this method may not be readily accessible under field conditions. Thus, the other methods available to preserve the integrity of microbial DNA or RNA include DNA/RNA shield (Zymo Research Corporation, Irvine, CA), which can stabilize nucleic acids at room temperature for up to 24 h (after which they must be placed in fridge or freezer according to the manufacturer's guidelines), and RNAlater at 4°C overnight to allow the buffer to infiltrate the samples before being transferred to −20°C or −80°C (Kellogg et al., 2016; Carradec et al., 2021). However, the efficiency of the preservation method depends on the next steps in the nucleic acid extraction methods. In the case of RNAlater, DNA extraction methods based on alcohol exclusion steps are not ideal because the high concentration of salt that is present in this solution can precipitate along with the DNA and can further inhibit later steps in the protocol (Athanasio et al., 2016). The use of different stabilizers might also limit what downstream kits can be used. For example, RNA/DNA Shield is highly compatible with its manufacturer's extraction protocols but is not optimized for other kit-based extraction methods. Always consult with the manufacturer when adapting sampling steps that may necessitate alterations to downstream molecular biology processes.

For post sample processing, fragments of the whole coral are usually either subsampled and/or placed directly into sterile tubes or tubes from DNA extraction kits that contain preservatives, macerated using a mortar and pestle while keeping the sample dry and cold with liquid nitrogen (Santos et al., 2012; Li et al., 2013; Zhang et al., 2015; Kwong et al., 2019), or homogenized using a bead beater (e.g., FastPrep24, MP Biomedicals, Irvine, CA; Klaus et al., 2005; Sekar et al., 2009; Sato et al., 2013; Kellogg et al., 2016; Biagi et al., 2020). Each of these methods can result in enough high-quality material for 16S amplicon library generation.

Mucus – Coral mucus can be used to investigate the role of the microbial assemblage and the interactions between coral and

Box 2: PCR contamination.

PCR is a highly efficient, processive, and relatively insensitive molecular process. Although these are typically advantageous attributes, there can be downsides that require us to use excessive precautions to avoid contamination from exogenous sources. Contamination can occur between and among samples (cross-contamination) or from an exogenous DNA source. Cross-contamination can be due to the mishandling of samples and/or materials as well as imprecise sterile technique. For example, practices such as keeping sample tubes open during PCR setup, pipetting reagents quickly that can generate aerosols, or inadequate disposal of tips and tubes can contaminate nearby samples and surfaces which can cause future contamination. The exogenous contamination source is related to the improper handling and storage of PCR reagents (e.g., primers, Taq polymerase, and water) and contaminated working environment. For these reasons, if at all possible, reagents should never be stored together with DNA samples or amplicons. The use of non-sterile materials such as pipettes, tubes, tips, laminar flow hood, and the incorrect or inappropriate use of PPE such as non-sterile gloves can also introduce exogenous DNA. To control for contamination, it is now standard to conduct, and sequence replicate negative control PCRs. In the event of library contamination, in silico removal of the sequences in the negative control libraries will improve the accuracy of the study.

Once introduced into a lab or system, PCR products or exogenous DNA can lead to a cascade of contamination throughout the laboratory, making it difficult to reestablish sterility. Therefore, preventing and if necessary, removing DNA contamination is a significant challenge that must be done effectively as contaminants can remain on surfaces for an extended period. For this reason, numerous methods of decontaminating DNA from laboratory surfaces have been developed including UV radiation. The UV radiation of the laminar flow hood and autoclave (Gefrides et al., 2010; Ziubrii, 2019), enzymatic method with exonuclease III (Zhu et al., 1991), use of Uracil-N-glycosylase (Longo et al., 1990), endonucleases (DNAse; Eshleman and Smith, 2001; Klaschik et al., 2002) and chemical methods such as hydroxylamine and hydrochloride (Aslanzadeh, 1993).

The most used methods are UV radiation, alcohols (ethyl alcohol and isopropyl alcohol), sodium hypochlorite, and DNase treatment. UV radiation damages the double strand of DNA, forming products, such as pyrimidine-pyrimidine and cyclobutyl pyrimidine dimers that impede the action of Taq polymerase (Cadet et al., 1986; Giussani et al., 2018). The efficiency of UV radiation will depend on the distance from the decontaminated surface, the molecular weight of the DNA, and exposure time. It is recommended that the UV radiation decontamination process take approximately 15min. PCR reaction reagents are sensitive to UV radiation. Therefore, it is not recommended to add reagents to the laminar flow while the UV light is on because this practice can affect the amplification of the DNA of interest. Alcohols such as ethyl and isopropyl alcohol help precipitate DNA, but they also denature proteins and inhibit enzymatic reaction when diluted (Wu et al., 2018b). Alcohols can be used for surface decontamination in concentrations between 60 and 70%. The most used is ethyl alcohol. However, there is no difference in effectiveness between the two. Pure sodium hypochlorite at a concentration between 1.0–1.5% is also widely used as a surface decontaminant (Fischer et al., 2016). This reactant can damage the cell membrane, inhibit enzymatic reactions, and directly damage the carbon-hydrogen bonds of DNA through oxidative cleavage (Prince and Andrus, 1992; Kampmann et al., 2017). Commercially-available sodium hypochlorite (bleach solution) in concentrations between 5.25–6.15% can also be used but should be used at a concentration of 10% (sodium hypochlorite 0.5–1%; Goodyear, 2012). When using sodium hypochlorite on the surface, or a 10% bleach solution, wait 10min and then remove excess bleach with ultrapure water or DNase, as prolonged use of hypochlorite can cause corrosion to laboratory surfaces.

DNase treatments have also been widely used on equipment and surfaces without the risk of material degradation. The most commercially used solutions are DNA away (Thermo Scientific, Wilmington, DE), and DNAzap (Invitrogen Corp., Carlsbad, CA). Even with several decontaminant options, these techniques may not eliminate all contamination. It is difficult to carry out the decontamination of DNA molecules with a low molecular weight (less than 200bp). Therefore, it is recommended the combination of techniques to have efficiency in the sterilization process. Champlot et al. (2010) combined strategies, such as UV radiation and DNAse treatment. For surface and equipment decontamination combined methods such as 75% ethyl alcohol, UV light, and hypochlorite solution ca be used (Wu et al., 2018b).

A major concern for coral microbiome research is that the PCR process is highly susceptible to contamination, leading to significant accuracy problems downstream. This is because a single PCR can produce thousands of amplifiable DNA molecules even if extremely rare in a sample. Thus, any foreign DNA can be amplified and contaminate your PCR and your resulting microbiome library. To ensure sterility during the pre-PCR process, wear clean gloves and, where possible, prepare PCR reactions inside sterile or laminar flow hoods. Use sterile tubes, tips, and keep all materials inside the hood decontaminated with 70% ethanol or bleach 10% and UV light for 15min (Aslanzadeh, 2004). Use sterile tips, pipettes, tubes, and racks exclusively stored inside the PCR hood and always use DNA-free reagents. All PCR reagents should be reviewed regularly and exchanged for new stock reagents if contaminated. It is also recommended to use special care when making stocks and then aliquot 'working stock' small volumes of reagents to ensure no new contamination of expensive and hard to replace highly concentrated stocks. Further post-PCR amplicon libraries should be stored safely and, if possible, never returned to the site of pre-PCR steps as they can contaminate all your materials and future studies. As a result of this well-known issue (Fox et al., 1991; Roux, 1995; Scherczinger et al., 1999), every lab should treat PCR products as a potential source of contamination. To avoid this, many labs separate the physical PCR setup phase from the actual amplification stage (Aslanzadeh, 2004). We suggest that, if at all possible, materials used for PCR setup are designated to a biological safety cabinet that has full UV decontamination capabilities and all PCR amplification steps, and all resulting PCR products and materials are kept in a separate room.

Contamination can also occur during the process steps that precede PCR, such as *via* DNA extraction kits (Salter et al., 2014; Weiss et al., 2014) Known as the 'kitome', contamination of DNA extraction kits can occur during the processing and preparation of kit reagents. Bacterial components from contamination may vary between kits (see Salter et al., 2014), and removal of contaminants can be difficult. Therefore, it is essential to use extraction and PCR negative controls (blanks) in parallel with real samples throughout the process. Sequencing negative controls from each stage of the extraction and PCR can help to identify specific contaminating bacterial taxa or sequences that arise erroneously and provides a confirmation that the coral microbiome profile is accurate.

environment. Many of the initial experiments on corals used mucus as a way to track microbiomes overtime without extensive damage to the host and have been used as a diagnostic tool for coral health (Carlos et al., 2013; Glasl et al., 2018). However, mucus sampling tends to result in more variable assemblages of microbes as these communities tend to have more transient microbiome members. It is important to note that both the amount produced and the age of the mucus can have major impacts on microbiome community composition (Glasl et al., 2016), which may limit the comparative power of this technique.

Mucus collection is typically carried out underwater using a sterile syringe (without the needle) and negative pressure. Sometimes minor abrasion is necessary to induce the coral to generate mucus (Hadaidi et al., 2017). The mucus is aspirated carefully from the coral surface without causing excessive damage and immediately after collection, the syringe can be inverted, allowing the mucus to accumulate at the base of the syringe due to its higher density. Ideally as much of the excess seawater should be expelled prior to transporting and/or transferring the mucus. Mucus can also be collected by sterile swab that is rolled or slid along the coral surface lightly collecting visible mucus *via* adhesion (Engelen et al., 2018; Weiler et al., 2018). This exposure method, in principle, reduces seawater contamination but is complicated by removing the animal from its natural environment and the unreliability of all coral species to produce mucus in this way.

Once collected, mucus can be transferred from the syringe to sterile tubes and be quickly placed on ice or dry ice, frozen in liquid nitrogen, transported to the laboratory and stored at −20°C or −80°C, or placed

in stabilizing buffers or preservatives (see section above) until DNA extraction.

Tissues – Coral tissue is a primary target for evaluating coral microbiome structure, function, and evolution. Given the intimate nature of hosting intra-and extracellular microbes in the tissue, the physiological and evolutionary interpretation of changes in coral tissue microbiomes are generally more straightforward than mucus-associated microbiomes, which are more variable and highly influenced by the external environment (Pollock et al., 2018).

To collect tissue samples, a coral fragment is usually collected as reported above for the whole coral specimen and then fractionated using a variety of methods that remove the tissue from the skeleton, such as airbrushing or water-picking with sterile fluids like phosphate buffered saline (PBS) or 0.22 μm filtered seawater. The use of PBS is a particularly effective strategy since this solution is cheap, isotonic, can come in sterile forms, and can be diluted with samples without generally interfering with any downstream molecular biology or chemistry in the samples (Hester et al., 2016; Weber et al., 2017). Tissues can also be dissected from the skeleton with a scalpel or razor blade, although the skeleton would almost certainly be present in any sample using this method (Littman et al., 2010; Kvennefors et al., 2012; Sudek et al., 2012). Another means to acquire exclusively tissue would be to add preservative and/or fixative that would allow downstream DNA extraction and then decalcify the coral using salt buffers or a mixture of formic acid and sodium nitrate (Berzins et al., 2011; Hernandez-Agreda et al., 2018a; Bergman et al., 2022). Tissue samples can be stored in ultra-freezers, in 100% molecular grade ethanol, or depending on the subsequent microbiological analysis, it can be fixed with 4% paraformaldehyde at 4°C for 12 h (Staley et al., 2017; Hernandez-Agreda et al., 2018a). In addition, the preservation of the tissues can be done with liquid nitrogen or salt-saturated dimethyl sulfoxide [salt-saturated DMSO (Gaither et al., 2011; Zhang et al., 2015)]. These substances are recommended for distant collections where options, such as freezers are unavailable, as salt-saturated DMSO and liquid nitrogen can remain viable for a long time and still result in accurate 16S library generation.

Skeleton – Although typically thought of as acellular, the skeleton contains a diverse and interesting collection of microbes. Collection and preservation methods can be done in the same way as with whole fragments but with additional steps to remove the mucus and tissues. During processing the skeleton can be separated from mucus and tissue through airbrushing with a sterile solution as discussed above (Neave et al., 2017; Weber et al., 2017; Marchioro et al., 2020). After this procedure, the skeleton samples can be preserved in liquid nitrogen and/or macerated with a sterilized mortar and pestle. Although we know of no papers that discuss this, it is also likely that bleached and/or dried coral specimens may contain internal DNA that could be used for coral microbiome studies. Future investigations on benchmarking such methods are necessary.

3.2. Nucleic acid extractions for coral microbiome analysis

The generation of 16S amplicons to track coral microbiomes requires efficient DNA extraction of both bacterial and host cells. Extraction protocols include three main steps: cell lysis (also called cell disruption or cell digestion), precipitation, and purification. Whether using a commercially available kit or an in-house method, these three steps are necessary for effective and high-quality DNA extractions. While it is difficult to standardize the extraction process to a single method due to the diversity of coral species and different sample types (e.g., coral compartment), several methods are commonly used that rely on readily available DNA extraction kits with different protocols (see below). Kits optimized for soil microbe samples are often good choices for coral DNA extractions because, like corals, soils contain high levels of inhibiting compounds, such as humic matter, that require additional DNA purification steps, making these kits more thorough in eliminating biological inhibitors.

The first crucial step in coral microbiome extraction is cell lysis, which is used to make microbial DNA accessible (Santos et al., 2012). In coral tissues, lysis can be challenging due to the presence of the mesoglea, a gelatinous layer between the epidermis and gastrodermis that is rich in collagen fibers and that are difficult to break, impeding to access to the microbial community contained within internal tissue. Without adequate cell lysis, extracted DNA may not accurately represent the microbial community. While some kits come with mechanical lysis tubes included (e.g., Qiagen PowerSoil, ZymoBiomics, etc.), the size and type of lysing matrix (often made from garnet, zirconia/silica, and/or glass beads) can affect both the efficiency of lysis and the amount of microbial DNA obtained. According to Weber et al. (2017), smaller beads may target the smaller microbial cells, whereas larger beads can also lyse eukaryotic cells in the coral and produce a flood of eukaryotic DNA in the sample. To account for variations in cell size, you can also use a combination of different types of beads, such as Lysing Matrix "A" bead-beating tubes (MP Biomedicals, LLC, Santa Ana, CA, United States) which combine garnets with large 1/4-inch ceramic spheres. These lysing matrices can be added to preservative collection tubes to stabilize nucleic acids and prepare for mechanical lysis at the same time. Mechanical lysis (aka "bead-beating") can be performed using commercial bead-beaters, such as the FastPrep24 (MP Biomedicals, Irvine, CA), the PowerLyzer24 Homogenizer (Qiagen Inc., Valencia, CA, EUA), or simply using a vortexer.

For effective breakdown of cells, most DNA extraction kits and protocols also use a chemical lysis, which is performed with a buffer that contains either an ionic detergent such as sodium dodecyl sulfate (SDS), which solubilizes, denatures, and breaks down cell membrane proteins to release DNA (Brown and Audet, 2008), or an enzyme. One of the enzymes used to lyse bacterial cells is lysozyme, which breaks down the glycosidic bonds in bacterial cell walls (i.e., the peptidoglycan layer; Shehadul Islam et al., 2017). In coral microbiome studies, it may be useful to use more than one type of chemical lysis to ensure the lysis of both gram-positive and gram-negative bacterial cells occurs, as gram-negative bacteria contain an outer membrane that can prevent lysozyme from accessing the peptidoglycan cell wall (Salazar and Asenjo, 2007; Ketchum et al., 2018). For example, enzymes such as proteinase K can be applied in an incubated digestion step (37–70°C) to increase yield and inactivate nucleases that could degrade DNA or RNA during the purification process. Proteinase K, when combined with chemicals such as SDS, Ethylenediaminetetraacetic acid (EDTA), enzymes such as RNAse, trypsin, and others, can improve DNA cleaning efficiency (Banaszak, 2007). Given the possibilities of combining methods, mechanical and chemical cell lysis can be optimized according to the specificity of the sample and the DNA to be extracted. However, it is important to note that some methods of cell lysis can increase PCR interferences due to the disruption of eukaryotic cells whose chemical composition (e.g., humic acid in tissues and calcium ions in skeleton) may affect the quality and quantity of bacterial and archaeal DNA and its amplification through inhibition of chemical reactions (Lorenz, 2012).

Following lysis of microbial cells, DNA is precipitated to separate it from cell debris and off target macromolecules. Alcohol (isopropanol or ethanol) and salt solutions are typically used to make the DNA insoluble. After precipitating the DNA and eliminating cellular debris, purification is conducted, again using alcohol as its main agent. Until recently, the conventional DNA extraction technique was called phenol: chloroform:Isoamyl alcohol method. This approach can be a cheap and efficient option, but any residual phenol can contaminate samples and make them difficult to work with downstream. Furthermore, these methods use caustic and volatile compounds (i.e., phenol) and must be carried out inside a chemical safety cabinet. Advancements in commercial kits have reduced reliance on this technique.

The effectiveness and accuracy of recovering high-quality and purity DNA can vary according to the extraction kits (Galkiewicz and Kellogg, 2008; Weber et al., 2017). A variety of kits have been used for the extraction of DNA from different coral species and different parts of the coral (Rodriguez-Lanetty et al., 2013; Glasl et al., 2019; Weber, 2020). For example, Santos et al. (2012) tested the efficiency of DNA extraction from fragments of the *Mussismilia hispida* coral by comparing 4 different DNA extraction kits (ZR Soil Microbe DNA Kit, Zymo Research, PowerSoil DNA Isolation Kit, FastDNA SPIN Kit for soil). They showed that the FastDNA SPIN Kit for soil was more efficient for high DNA yield compared with the other kits. Also, Weber et al. (2017) carried out a study comparing other DNA extraction kits from MoBio laboratories (PowerSoil®, PowerPlant® Pro, PowerBiofilm®, and UltraClean® Tissue & Cells) in 7 different species of coral and evaluated the amplification efficiency of 16S rRNA. PowerBiofilm® produced higher DNA yield and a more diverse microbial community when compared with other kits. These studies suggest that no specific DNA extraction kit must be used with coral samples (as exists for soils and plants studies). Bergman et al. (2022) also compared the output of microbial community analysis from 2 different coral species with 3 different kits (see citation for details) and found that, at least for the same coral species, each kit resulted in similar alpha and beta diversity estimates. Given these data, we recommend that research be carried out on the lysis and methods that each DNA extraction kit uses to determine the most suitable kit for a given sample type.

3.3. Amplicon sequence amplification

Amplification of microbial DNA sequences to create 'amplicons' is conducted *via* PCR. Each reaction consists of a mastermix that includes Taq polymerase, magnesium ions, free deoxynucleotide triphosphates (dNTPs), primers for an especific target gene region and DNA template. Below we discuss the steps of PCR and the considerations for choosing Taq polymerases and primers that will ensure effective amplification.

Primers – Primers should be selected to cover the ends of the specific rRNA gene region of interest, such as the forward primer that attaches in the $3' \rightarrow 5'$ direction (the antisense strand) and the reverse primer that attaches to the last nucleotide of the region to be amplified in the sense $5' \rightarrow 3$, direction (the sense strand). Primers are commercially synthesized and generally have a size of around 15–30 nucleotides with guanine-cytosine (G and C, respectively) content that can range between 40 and 60% (Lorenz, 2012). Furthermore, it is recommended that the primers have a CG clip at their ends, that is, the presence of C or G in one of the last 5 sequences to ensure primer binding to the complementary sequence.

While several target genes can be used for microbial taxonomic analysis, such as 23S rRNA (Pei et al., 2009), *rpoB* (Ogier et al., 2019), and others, typically, the 16S rRNA gene is used due to its presence across all bacterial and archaeal lineages and its slow evolutionary rate of change. The 16S rRNA gene makes up one component of the small subunit of the bacterial ribosome and is highly conserved due to its essential function of aligning mRNA to the ribosome for accurate and processive protein production. Interspersed with conserved regions of this gene are highly variable regions (V1-V9), which provide smaller, unique sections of gene sequence for comparison (Caporaso et al., 2011; Bukin et al., 2019). Which variable region to use for amplicon sequencing is hotly debated in the field, and the choice of primers is an extremely important consideration for any study. According to Kim et al. (2011), different regions of the 16S rRNA gene can produce different results regarding species richness and diversity of the microbial community. Primers that target the V4 region of 16S rRNA, in particular 515F and 806R and 806Rb, are currently the most commonly used for analyzing the taxonomic diversity of Bacteria and Archaea in corals (Apprill et al., 2015; Walters et al., 2016). Although widely used, this primer set has considerable downsides for the study of coral microbiomes (see section on host off-target contamination) and caution must be taken when using this popular primer set.

While performing PCR, some problems associated with the use of any primer set may arise. For instance, the creation of "primer-dimers" occurs during the annealing process where the primers may anneal with each other rather than the template DNA. This annealing occurs because primers are complementary and can bind at the 3' end. This failure can be seen in agarose gel electrophoresis images as intensely illuminating low molecular weight bands (<100 bp). For this problem, Lorenz (2012) suggests optimizing the amount of primer for the amount of template DNA in the reaction, although dimers can also be removed during cleaning steps.

A primer pair is considered ideal during amplification when they can achieve amplification efficiency and specificity, maximize coverage of the microbial community, and minimize PCR bias (Sambo et al., 2018). These optimal characteristics are attributed to (1) the position of the nucleotides compatible with the template DNA, avoiding amplifying other target sequences that are not selected; (2) amount of nucleotides in the primer; (3) GC (guanine-cytosine) content which should contain about <60% so that it does not interfere with successful amplification (> 60% tends to increase hydrogen bonds between GC and generate secondary structures such as hairpins and formation of dimers; Assal and Lin, 2021); (4) avoid sequences with dinucleotides (such as CGCGCG or ATATAT) so that there is no formation of secondary structures; (5) use of primers or degenerate primers to minimize PCR bias.

The efficiency of target gene amplification can be compromised and generate PCR artifacts as well. These artifacts can result from errors such as chimera formation during amplification or uneven distribution of PCR product amplification, also called "PCR bias" (Acinas et al., 2005). PCR bias can be attributed to primer incompatibility with some targets that can occur even for a single base. Thus, to avoid bias and cover the community of interest, primers can be modified using nucleotide sequences corresponding to variation between homologs (called "degenerate primers"). For this reason, Apprill et al. (2015) used primer 515F and 806RB with degeneracy to reduce bias and, consequently, resolve the underestimation of the SAR11 clade in marine samples. Walters et al. (2016) compared the performance of the original 806R primer and the 806RB degenerate primer for detecting the SAR11 clade

(a ubiquitous and abundant marine bacterial group) and observed that the degenerate primer not only increased the detection of the SAR11 clade but also interfered with the performance of taxa amplified by the original primer.

Taq polymerase – Taq polymerase is a thermostable enzyme that can synthesize DNA only when given a primer that provides a starting point for synthesizing a DNA region of interest. There are different types of Taq polymerase with each having a unique fidelity (i.e., accuracy) and processivity (i.e., how quickly it synthesizes) that can help you choose the most appropriate Taq polymerase for your study. Due to the diversity of Taq polymerase brands, there is no single Taq polymerase that is best suited for every study. Instead, the preferences of some researchers depends on the sample type, the efficiency of the Taq polymerase, the cost, and the practicality of use. As an enzyme, Taq polymerase requires the presence of a cofactor during the PCR reaction, such as Mg^{2+} ions. Some manufacturers offer Taq polymerase in a buffer containing this cofactor at a standard concentration, but others provide it as an aside or as an addition. However, magnesium chloride ($MgCl_2$; Markoulatos et al., 2002) if used in high concentrations, can lower the specificity of Taq and create spurious primer pairings (i.e., matches between primers and unwanted sites in the template DNA). Not only can excessive addition of Mg^{2+} cause problems during the action of Taq polymerase, but some inhibitors that come from DNA extraction or poor DNA purification can directly affect Taq polymerase. These inhibitors can prevent the interaction of Taq with Mg^{2+} ions (e.g., Ca^{+} ions from the skeleton), thereby preventing the action of Taq polymerase in the DNA amplification process. Furthermore, other contaminants can interact directly with Mg^{2+} ions, reducing their concentration and preventing the catalytic action with Taq polymerase.

dNTPs (deoxynucleotides 5′-triphosphates) – dNTPs are used in PCR to provide nucleotides that will be added to the growing oligonucleotide chain during the synthesis of new DNA amplicons (Markoulatos et al., 2002; Paul and Yee, 2010). Some manufacturers will add dNTPs to a buffer that includes both the Taq polymerase and Mg^{2+} ions in effective ratios, while others will provide them as an aside. If adding dNTPs separately, it is important to note that high concentrations can chelate Mg^{2+} ions reducing the effective function of Taq polymerase (Roux, 1995); it is thus necessary to work with small volume aliquots so that there is no loss of oligonucleotide yield.

DNA template – The purity of the DNA in the PCR technique is essential for accurate and effective microbiome analyses to be carried out. Thus, the DNA sample must be free of any inhibitors (see section on DNA extraction) and free of exogenous or contaminant DNA (see below for discussion). To check for inhibitors, DNA quantification performed by UV spectrophotometer can differentiate DNA from inhibitors through wavelength analysis (Boesenberg-Smith et al., 2012). In addition, an excessive amount of DNA can inhibit the amplification process.

3.3.1. Coral host off-target PCR contamination

Another major challenge in coral microbiome work is the efficient amplification of 'off-target' coral DNA sequences alongside microbial genes. In many coral species, several popular primers used for 16S rRNA amplification (e.g., 515F-806RB) have high similarity to coral mitochondria and chloroplast genes due to their shared ancestry with bacteria (Lopez et al., 2003). Non-specific or off-target amplification of coral host DNA can create multiple PCR products that result in a pool of eukaryotic amplicons mixed with bacterial amplicons (Galkiewicz and Kellogg, 2008). Without separation, the resulting libraries will contain both amplicons and reduce the sequencing depth of the target amplicon, potentially leading to an underestimate of the true diversity and/or taxonomic profile of the microbial community.

Steps to minimize, eliminate, or sidestep the off-target amplification issue are available, however. As eukaryotic DNA becomes available during the cell lysis step of DNA extraction, downstream optimization of the PCR protocol or purification methods can be used to minimize the amplification of non-microbial DNA during PCR. For example, Galkiewicz and Kellogg (2008) used an alternative primer set (63F/1542R) to separate eukaryotic from bacterial rRNA genes during the PCR technique. However, according to the authors, care should be taken when selecting the 63F primer, as this can develop a bias in the bacterial profile generated. Ten years later, Pollock et al. (2018) reported that primers that amplify the V4 region of the 16S rRNA, 515F-806R, also amplify the mitochondrial 12S rRNA gene from coral. However, these amplicons are slightly different lengths and can be removed or annotated separately *in silico*. The 16S rRNA gene amplicon is shorter ~300 bp, while the 12S rRNA off target coral gene amplicon is longer ~400 bp. As such, off target amplicons can be removed *via* gel-based size selection purification methods (e.g., using a BluePipin machine) and/or a 2-step PCR where only the proper size band is excised and barcoding is conducted on exclusively the targeted 16S band (see Figure 2 for details). Explicitly, after the first PCR with only the locus-specific primers, agarose gel electrophoresis is applied for the separation of the 12S and 16S amplicons. Given the band sizes are similar this can require a slow and long gel separation step. Next, the 16S rRNA amplicons are chemically purified (i.e, PCR clean up kits) or physically removed (i.e., excised with a sterile tip or razor blade) from the electrophoresis gel, and used for the DNA template in a second step of PCR (Caporaso et al., 2011). This technique is efficient and, when conducted properly, can eliminate a majority of the off-target amplicon. It is possible to also use peptide nucleic acid (PNA) clamps to bind to target DNA, preventing host DNA amplification and increasing bacterial DNA amplification (Reigel et al., 2020). This can provide a cheap and efficient alternative method for the decontamination of microbial DNA without underestimating the rare biosphere.

3.4. HTS library construction

For sequencing to be successful, it is necessary to prepare individually identifiable 'sequencing libraries' for each coral microbiome sample. As HTS platforms sequence many samples simultaneously ('multiplex'), each coral microbiome library must contain a unique coded set of nucleic acid markers or 'barcodes' that indicate which sample is which within the final pool that is sequenced. Barcodes are small oligonucleotide sequences (usually 8–12 nucleotides in length) used to identify sequences from a given sample that allows the pooling, or multiplexing, of several samples into a single library that can be sequenced on a single sequencing lane or run (Head et al., 2014; Lebonah and Chandrasekhar, 2014; Zhang et al., 2020). At the same time, individually barcoded samples must be purified, ensuring that they are free of extraneous nucleic acids including any remaining forward and reverse primers and/or primer dimers.

The complete process can require a different number of steps depending on whether one is conducting 1 or 2 step PCR (Figure 2; see above discussion on host 12S contamination). A 1-step protocol includes attaching specific forward and reverse primers containing the 16S rRNA region, and a sequence tail called overhang linker sequences, barcodes

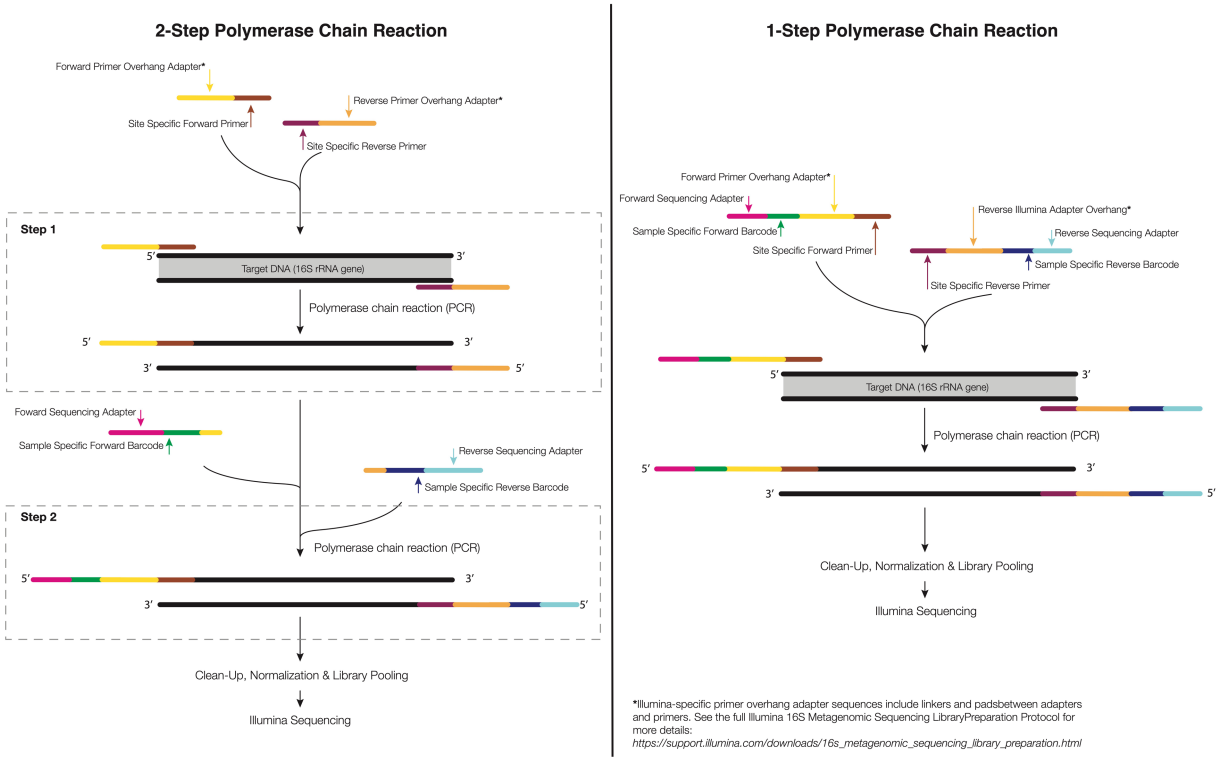

FIGURE 2
1-step and 2-step PCR amplification approach. In the 1-step PCR (right), the target gene is amplified using primers sequence composed of forward and reverse target gene primer, overhang linkers, barcodes, and Illumina adapters to bind to a flow cell in Illumina sequencing. In the 2-step PCR amplification (left), the primer of the first-step PCR contains a specific forward and reverse primer and an overhang adapter. For the second-step PCR, the primers have an overhang linker, barcodes, and the Illumina flow-cell linker sequence.

and adapters to allow binding to a flow cell in Illumina sequencing (currently, the most used sequencing; see Figure 2). In the 2-step protocol, the same initial PCR step is completed including specific forward and reverse primers, and overhang linker sequence. The second PCR step includes a small overhang linker sequence, barcodes, and Illumina adapters (Figure 2). The advantage of a 2-step PCR is the flexibility to amplify the gene target of low-biomass samples when compared with the 1-step PCR as well as the ability to ensure off-target sequences are avoided. However, the downside is that an additional PCR step must be completed which can increase the financial and opportunity cost of library generation. Further care must be taken when conducting multiple step PCR as any additional rounds of amplification can increase the risk of producing artifacts (Kozich et al., 2013).

For corals, library preparation can also be 1 or 2 steps (Figure 2). Two step PCR approaches generally require: (1) PCR for the separation of the 12S rRNA genes from the 16S rRNA and (2) validation by gel electrophoresis (a 1% agarose). The 16S rRNA amplicons derived from the electrophoresis gel must be purified or the reaction used as the target DNA to perform the second-step PCR where the indices/barcodes are added (Figure 2).

After amplification, amplicons must undergo purification to build a refined library since sequencing is a highly sensitive technique. At this stage, a more efficient method of purification is used, such as the use of magnetic beads in which the amplicons bind reversibly and undergo a simple washing process to remove the primers, primer-dimers, nucleotides, salts, and enzymes (Watson and Blackwell, 2000). An elution reagent (e.g., TE Buffer) or nuclease-free water is used to elute the amplicons for a purified final product. There are some commercially available library preparation kits that can streamline this process, such as Illumina DNA Prep and TruSeq DNA PCR free. Many perform purification by eliminating both short and long fragments through a two-step process. The long fragments first bind to the magnetic beads, then the supernatant is removed and purified to remove the short fragments. During bead purification, it is possible to size select the amplicons based on the proportion of beads to a sample volume. Most commercial kits are designed to capture amplicons >100 bp and eliminate <50 bp, but these values can be changed according to the size of the library of interest. It is important to note that the proportion of beads to sample can affect the final library and performance.

3.4.1. Quantifying and combining amplicons for multiplexed library sequencing

After purification of the amplicons, libraries must be quantified and mixed in similar proportions, so that the library will be equally represented in the final pool. Otherwise, samples that are amplified better than others may be over-represented in the dataset while others will have read levels so low that they cannot be used in the final analysis. Quantification and sizing of the gene library are performed by spectrophotometry (e.g., UV/Vis or Nanodrop), fluorometry (Qubit, Picogreen), quantitative PCR (qPCR), or droplet digital PCR (ddPCR). However, at this stage, care must be taken when using spectrophotometric quantification, as any impurity present can contribute to the absorbance. For the final library step, amplicons should be pooled at similar molar concentrations. Typically, an optimal initial library concentration is at

least 4 nM (Illumina, 2019). It is noteworthy that libraries with values less than 1 nM will have very low yields. The quality of the final library can be checked on an agarose gel or more accurately using a Bioanalyzer (e.g., Agilent 2,100). A workflow for HTS library preparation is shown in Figure 3, including all the steps mentioned above.

3.5. High-throughput sequencing steps

In preparation for cluster generation and sequencing on the standard Miseq sequencing platform, double-stranded libraries are denatured using sodium hydroxide (NaOH) in concentrations between 0.1–0.2 N, respectively, for samples with amplicon concentrations between 0.5 and 4 nM. High concentrations of NaOH can inhibit the hybridization of the library in the flow cell and, thus, decrease the cluster density (Wu et al., 2018a). Then, amplicons are diluted with an HT1 buffer (hybridization buffer) at the picomolar level l for final loading into flow cells.

It is noteworthy that some biases can affect the construction of a refined library, such as cross-contamination of indexed primers producing chimeras by recombination of different molecules (Kircher et al., 2012). Also, in 16S rRNA and almost all other amplicon libraries, sequences exhibit low base diversity, or an imbalance in the number and order of bases in a set of sequences. This imbalance can negatively impact the cluster model formed during sequencing. To expand the diversity and enrich this library with unique sequences, it is recommended to add a shotgun library to the pool. Typically, this is a PhiX library (the genome of the ΦX174 bacteriophage that is cut into small random segments) in Illumina sequencing. The phiX library is a ready-made library that provides quality control for the alignment and sequencing of clusters due to its diverse composition of bases (45% GC and 55% AT) and can be applied to increase confidence in your results. The concentration of PhiX to be added will depend on fragment length and sequencer software (Kozich et al., 2013). In some instances, you may be able to provide DNA from a diverse sample of your own (e.g., a coral microbiome sample) in place of PhiX, which can generate between 2 and 12 million bases of metagenome that may be used in downstream metagenomic analyses. The use of a PhiX replacement should be discussed with your sequence provider.

4. Considerations for bioinformatics

The sequencing analysis of 16S rRNA amplicons is based on software and algorithms that convert this sequencing data into biologically meaningful results. Bioinformatic pipelines from a variety of software programs can quickly and efficiently perform these analyses. While most software programs and pipelines include a similar sequence of steps for denoising, merging, grouping and taxonomy assignment to 16S rRNA sequences, they can vary by quality control parameters and clustering algorithms. Commonly used software programs to build 16S bioinformatics pipelines include Quantitative Insights into Microbial Ecology 2 (QIIME 2; Bolyen et al., 2019) with options for DADA2 or Deblur ASV-picking algorithms, mothur (Schloss et al., 2009; Schloss, 2020), DADA2 (Callahan et al., 2016) in R (Team R.C, 2020), and USEARCH (including UPARSE and UNOISE; Edgar, 2010). Although USEARCH is widely used, it is not open-source software and therefore has limitations for its use and redistribution. VSEARCH (Rognes et al., 2016) can be used as an open-source alternative to USEARCH.

These software programs have options for analysis at both the OTU (Operational Taxonomic Unit) and ASV (Amplicon Sequence Variant) levels (detailed in the Merge reads and Clustering section below). Previous studies have examined the sensitivity and consensus differences in several of these pipelines using default settings to mimic what most users have likely implemented (see Plummer et al., 2015; Prodan et al., 2020); however, customization can improve the performance of any pipeline. Below is a summary of each step in the 16S rRNA bioinformatic pipeline using the most widely used bioinformatic tools in coral microbiome studies, including a discussion on the differences between pipelines. The bioinformatics analysis steps used to process the data are cited below and shown in Figure 3.

4.1. Demultiplexing

Demultiplexing is the first '*in silico*' step after sequencing, in which the barcode sequences are used to identify and group sequences that come from the same sample. In some cases, the sequence provider will complete this step prior to returning sequence data to the user given that a spreadsheet identifying the sample barcodes is provided. When samples are returned multiplexed, demultiplexing can be done using most bioinformatic pipelines. QIIME 2 uses the "q2-demux" plug-in that can demultiplex both single and paired-end sequence reads. The barcodes are read as a reverse complement of the original sequence through the script "-p-rev-comp-mapping-barcodes" in the demultiplexing of paired readings. If adapters and primers are still present on the sequences, a cutadapt plug-in (Martin, 2011) for QIIME2 called "q2-cutadapt" can be used. Mothur uses a command called "make.contigs" for demultiplexing, where paired-end reads are also merged at the same time. This mothur command has the option to add an "oligos" parameter for removing primers and barcodes, and a "check orient" parameter to search for the reverse complements when primer and barcode sequences cannot be found.

While both QIIME2 and mothur have the ability to perform demultiplexing, OTU and ASV picking requires the input sequence data to be demultiplexed and trimmed (adapters, primers, and barcodes removed). This can be done using other pipelines or software (e.g., cutadapt, trimmomatic; Bolger et al., 2014) or even command line computation (e.g., using Python or Biopython).

4.2. Quality control

It is essential to check the quality of the sequences to avoid overestimating microbial diversity. Quality filtering is often used to truncate or discard overlapping matched reads to minimize the presence of any sequencing errors. The accuracy of sequencing is assessed by the Phred quality score (Q-score) provided for each nucleotide, which indicates the probability of an incorrect base call (Nilakanta et al., 2014); the higher the Q-score, the lower the probability of an incorrect base call. Pipelines such as QIIME2, DADA2, and other standalone software, such as FastQC (Wingett and Andrews, 2018), have a graphical user interface, which can visualize the quality scores for either the entire library or each forward and reverse read. These graphs can be used to set the

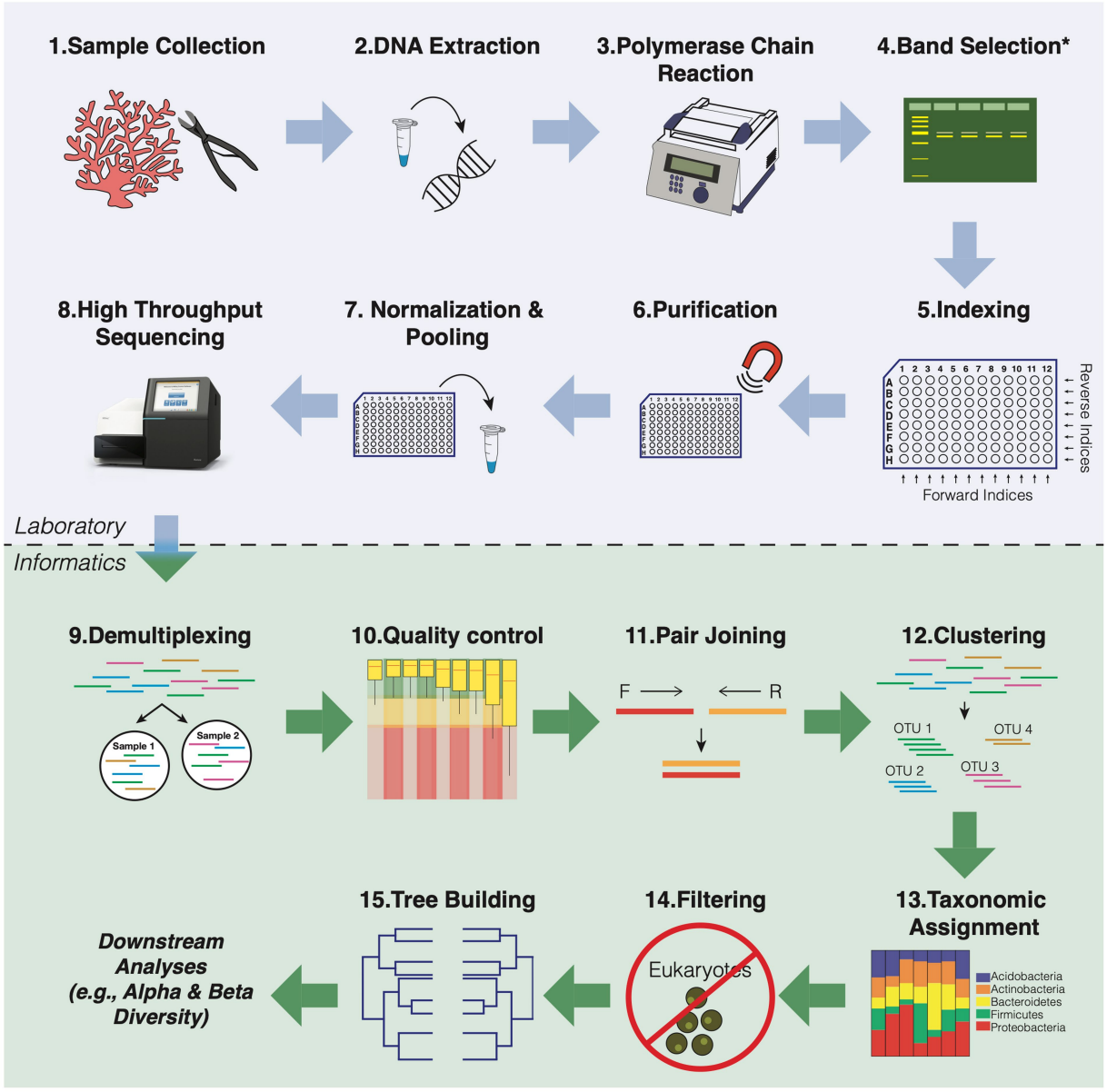

FIGURE 3
General workflow overview for 16S rRNA amplicon sequencing of coral microbial communities. The blue box indicates all the steps for preparing the library. The green box shows all the steps of bioinformatic data analysis that refer to the preparation of the readings for the downstream analyses.

parameters for trimming and noise reduction. It is worth mentioning that these parameters will differ for each dataset and read direction, as reverse reads are often of lower quality than forward reads, and should be optimized to avoid data loss through excessive reduction of the read length. Parameters for quality filtering include (1) primer removal, (2) off target outlier sequence read removal, (3) removal of poor-quality reads with Phred scores <4 and >60, and (5) removal of any reads that exceed a defined maximum number of "expected errors" (maxEE). QIIME2 incorporates either DADA2's quality control steps through the "dada2" plug-in, using the"dada2 denoise" command with the parameters "--p-trim-left" and "--p-trunc-len," or with Deblur through the "deblur denoise-16S" command with the parameter "--p-trim-length." In mothur, quality filtering is performed using the "screen.seqs" command. In DADA2, the filtering is done through the command "filterAndTrim." Both commands allow parameters to be defined for each forward and reverse read.

4.3. Merge reads and clustering

Clustering sequences based on similarity allows for accurate downstream identification of putative microbial species for taxonomic assignment and statistical analysis. For paired end sequencing, merging forward and reverse reads must occur prior to clustering, and is often incorporated into the clustering commands

for simplification of the pipeline. There are a number of different sequence similarity thresholds that have been used, and the choice for which threshold to pick should be dependent on the biological or ecological question that is posed. Sequences with a 97% or greater similarity can be grouped into what are called operational taxonomic units (OTUs). Historically, OTUs have been treated as similar to observing a species (Callahan et al., 2017). As sequencing technologies and bioinformatic analyses have improved our ability to identify sequencing errors, we can now group sequences based on higher thresholds, such as 99% or 100%, that represent amplicon sequence variants (ASVs), oligotypes, exact sequence variants (ESVs), or zero radius operational taxonomic units (zOTUs). Clustering methods based on higher thresholds, such as ASVs, infer biological sequences before amplification and sequencing errors and distinguish variants by only 1 nucleotide (Callahan et al., 2017). Clustering can be performed "*de novo*" without reference sequences, which creates the sequence clusters only through observed similarity and not based on a database. In contrast, closed reference clustering methods require a reference database to compare to the observed sequences, and any sequences that are not present in the reference databases may be lost. QIIME 2 provides two options for *de novo* ASV-picking: DADA2 using the plug-in "q2-dada2" (Callahan et al., 2016) or Deblur using the plug-in "q2-deblur" (Amir et al., 2017). With both methods in QIIME2, the joining of paired reads will be performed automatically during denoising. OTU-picking in QIIME2 utilizes VSEARCH *via* the "q2-vsearch" plug-in (Rognes et al., 2016). In DADA2, the readings are duplicated, and the ASVs inferred. In this case, DADA2 works by retaining a summary of quality scores associated with each sequence and thus performs the inference of ASVs. Then the forward and reverse ASVs are merged using the "mergepairs" command. In mothur, sequences are assigned to OTUs *via* the "Cluster" command. This command can be based on different clustering methods, including Search (does not require distance matrix), but commonly used methods are based on percentage distance between sequences. Furthermore, ASVs can be identified in mothur through the "pre-cluster" command.

4.4. Taxonomic assignment

Once sequences have been clustered, taxonomy can then be assigned. The taxonomic nomenclature is based on reference databases, of which the most popular in 16S rRNA-based phylogeny analysis include SILVA (Pruesse et al., 2007; Yilmaz et al., 2013), Greengenes (McDonald et al., 2012) and the Ribosomal Database Project (RDP; Wang et al., 2007). One of the main differences between these taxonomic databases is the origin of taxonomic rank information. For instance, the taxonomic classification for the RDP is obtained from the International Nucleotide Sequence Database Collaboration (INSDC). SILVA is based on Bergey's Taxonomic Outlines, List of Prokaryotic Names with Standing in Nomenclature (LPSN) and is manually curated. Greengenes is based on the National Center for Biotechnology Information (NCBI). It is important to note that taxonomic compositions of a dataset will depend on which taxonomic reference database is used (Sierra et al., 2020). Most bioinformatic pipelines offer a default taxonomic classifier; mothur uses RDP and QIIME2 uses Greengenes. However, these can be manually replaced by any other database. The algorithm for taxonomic classification can also differ. For instance, both QIIME 2 and DADA2 use a naïve Bayesian trained classifier, where the classifier is first trained on the specific region of the target sequences. These taxonomic classifiers are prepared based on specific sequencing parameters and target sequence compliance, which creates a new, dataset-specific taxonomic attribution repository.

4.5. Removal of unwanted taxa

After taxonomic assignment, a filtering step can be performed to remove any unwanted taxa (e.g., any eukaryotic contamination) or optimize the feature table. In QIIME2, some parameters can be optimized in this step, such as removing ASVs present only in 1 sample that may not represent the true biological diversity (perhaps errors during PCR amplification and sequencing). In addition, libraries can be curated to contain a minimum total number of reads to normalize the analysis across the datasets. This process called 'rarefaction' is however controversial. For additional reading on this topic see the works by Hughes and Hellmann (2005) and Willis (2019). It is noteworthy that even after the separation of 12S rRNA host during 2 step PCR, we often still find reads from the host, making the removal of chloroplasts and mitochondria from taxonomic attributions a critical step before undertaking downstream analyses. A supplementary database called Metaxa2 (Bengtsson-Palme et al., 2015) was created to assess the effects of mitochondrial sequences in the analysis of bacterial diversity and this database might underreport the existence of mitochondrial sequences in coral microbiome samples. That databases might underreport the existence of mitochondrial sequences in coral microbiome samples. The inclusion of mitochondrial reference sequence databases such as Metaxa2 is recommended for coral microbiome samples.

In QIIME 2, the removal of chimeras, where two or more sequences have been incorrectly joined together during sequencing, should also be done. This filtering is performed by the plug-in "q2-feature-table". In mothur, mitochondria exclusion and the removal of chimeras are performed before taxonomic attribution through the command "remove.lineage": and "chimera.Vsearch." In DADA2, the removal of chimeras is done with the function "removeBimeraDenovo."

5. Downstream analysis

After the computational treatment of the OTUs/ASVs, output files are generated that are available for taxonomic analysis, alpha and beta diversity estimation, measurement of dispersion, and even estimates of functional pathways. Output files generated by bioinformatic pipelines that are necessary for downstream statistical analyses include an OTU or ASV feature table of raw sequence counts (biom format file), a taxonomic reference file for each OTU or ASV (csv or txt format), and a phylogenetic tree file (newick format). With these data in hand, researchers can begin to unravel the microbiome dynamics of their individual systems. There is a diverse range of analytical and statistical tests that can be explored for 16S rRNA amplicon data that are not discussed in depth in this paper (Figure 4). These include, but are not limited to, functional prediction based on taxonomy (e.g., Picrust2; Douglas et al., 2020), network analyses that infer community

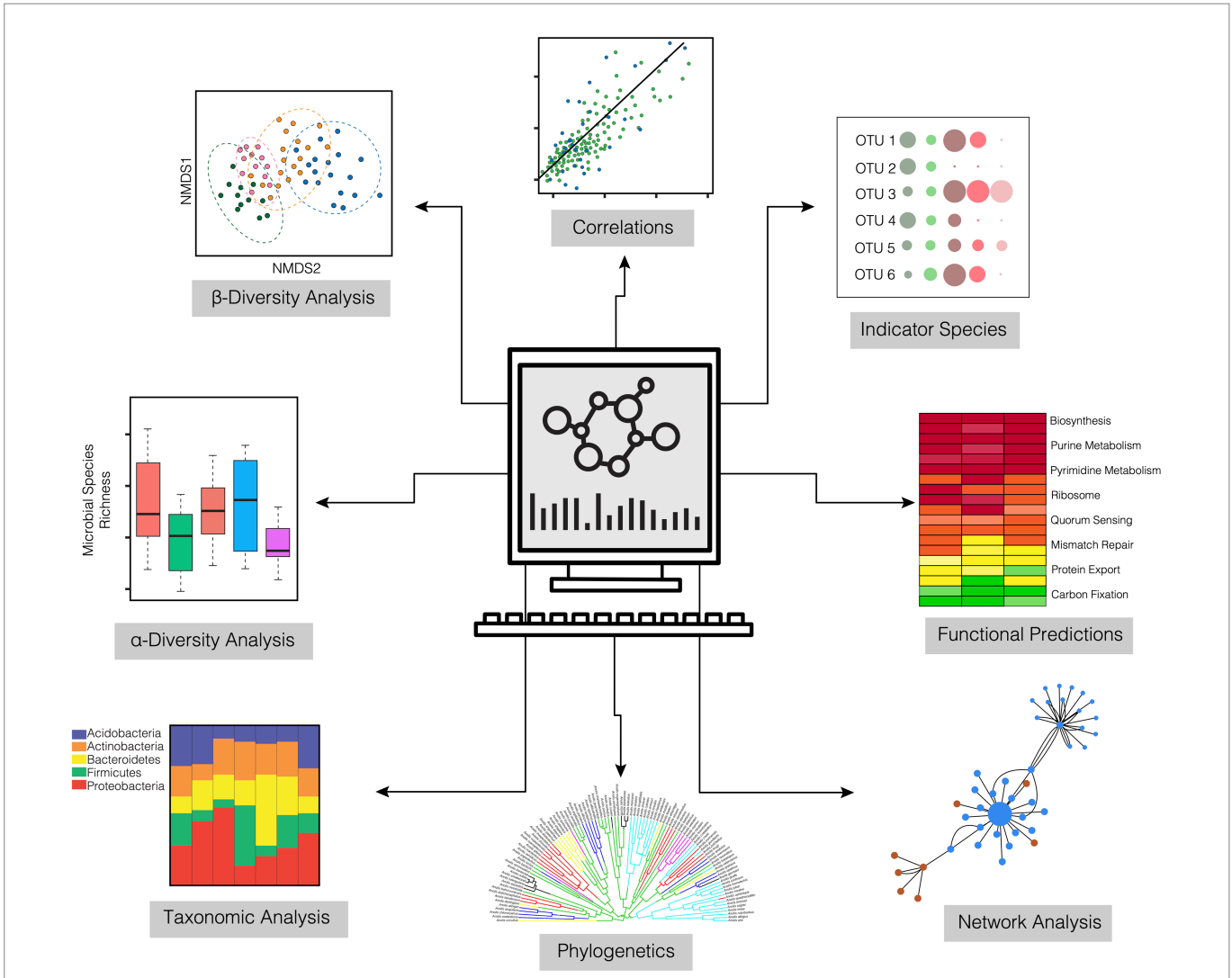

FIGURE 4
Downstream analyses for characterizing the abundance, diversity, and composition of coral microbiomes derived from 16S amplicon sequencing.

co-occurrence (Barberán et al., 2012), and multi-level pattern analyses that can identify bacterial indicator species (Dufrêne and Legendre, 1997).

5.1. Alpha and beta diversity metrics

Many ecological paradigms (e.g., resistance, resilience, and stable state dynamics) are built around how diversity changes in response to a given disturbance. Yet metrics of biodiversity come in many forms; knowing the difference among these is critical to understanding biological patterns in a study system. Alpha diversity is a collection of measures that characterize several aspects of the number of different taxa and their uniformity in a community. Alpha diversity can include metrics such as 'species' richness (i.e., the exact observed number of OTUs or ASVs of a given taxon), Chao1 (predicted richness based on species accumulation curves), evenness (the numerical distribution of different taxa relative to one another within a community), Shannon Index (an index that incorporates aspects of both richness and evenness), and inverse Simpson index (value between 0 and 1 represents the increase in diversity based on the average proportional abundance and the number of species). These metrics can be visually expressed, for example, through rarefaction curves. Furthermore, scatter plots of alpha diversity metrics against environmental measures (e.g., temperature, pH, dissolved oxygen, nutrients) can provide important insights into drivers of diversity within an ecosystem. These relationships can be tested using univariate statistical models, such as least-squares regression.

Beta diversity measures, unlike alpha diversity, assess the variety and relative abundance of different species that make up the microbiome. Beta-diversity is typically reported as either between variable beta-diversity or within variable beta-diversity, a measure also referred to as 'dispersion' which we discuss below. Beta-diversity measures are usually constructed from matrices that include all the taxa and their comparative abundances. Differences among samples or locations based on a given variable (e.g., host species, sampling time points, some experimental or environmentally altered variable like temperature) are typically tested through permutational multivariate analyses (e.g., PERMANOVA) and visualized using

ordination methods such as non-metric multidimensional scaling (NMDS), Principal Coordinate Analysis (PCoA), or Canonical Correspondence Analysis (CCA).

Differences in diversity between samples can provide insights into how microbiomes may change through time, between host health states, or among any target variable. These changes can be either deterministic (shift in the same way) or stochastic (shift in different ways). It has been suggested in the literature that environmental and health stressors of corals cause microbiome destabilization that is represented by stochastic changes in microbial community structure (see Zaneveld et al., 2016) and can be visualized by dispersion effects in ordination space (i.e., how close microbiomes of different samples cluster). As a result, highly dispersed microbiomes have been associated with negative impacts such as disease (e.g., Rosales et al., 2019; Becker et al., 2022) and anthropogenic disturbance (e.g., Zaneveld et al., 2016; Maher et al., 2019).

Different methods can estimate microbial community variances, such as methods based on dispersion estimation of individual taxa or communities of taxa using a weighted conditional probability [e.g., EdgeR see (Robinson and Smyth, 2007; Chen and McCarthy, 2015)]; or methods that model the dispersion of individual taxa by averaging the heterogeneity of the dispersion values for different taxa using a Bayesian approach (e.g., DSS, see Wu et al., 2013). We recommend the use and exploration of various multivariate analysis techniques that are required by the multidimensional nature of microbiome community data.

5.2. Differential abundance analysis

Depending on the research question and/or experimental design, it is commonly of interest to determine how the abundance of certain microbes varies among treatments or environments. These differentially abundant taxa may represent important biomarkers for coral health or other factors that may impact coral resilience. Multiple types of differential abundance (DA) analytical tools exist for use in microbiome studies, including traditional statistical tests (e.g., t-tests or Kruskal-Wallis rank sum tests), those originally designed for differential gene expression [e.g., DESeq2 (Love et al., 2014)], and those developed specifically for microbiome studies [e.g., Analysis of Composition of Microbiomes (ANCOM; Mandal et al., 2015)] and ANCOM with Bias Correction (Lin and Peddada, 2020). However, many of these tools face challenges associated with the treatment of microbial count data (see Swift et al., 2022), making this an active area of method development and care should be taken when choosing the appropriate test for your data.

6. Conclusion

Here, we reviewed current methods to sample, extract, and analyze the coral microbiome based on 16S rRNA gene sequencing. Coral represents a complex animal-symbiont holobiont whose molecular and *in silico* 16S rRNA pipelines may require technical adaptations that, in some cases, may differ in methodology from those described in manufacturer's documents or in the literature for more simple or well-studied host systems. Furthermore, careful consideration must be made of appropriate methods that meet study objectives while also accounting for differences in methods required for the various compartments and needs specific to different corals. Nevertheless, with an expanded number of field, laboratory, and computer techniques and tools available, reduced costs of analysis, and increased applicability, conducting coral microbiome research is increasingly available to new and established investigators. Due to the complexity of bioinformatic methods, the sections describing HTS and considerations represent a basic starting point for those pursuing 16S rRNA amplified sequence studies. However, more in-depth reading of the subject is recommended according to one's study aims. We hope that this review provides a condensed platform of knowledge and a set of methodologies to those initiating research in this area. Together we can advance and accelerate coral microbiome research and ideally the management and conservation of coral reefs worldwide.

Author contributions

DS and RVT conceptualized the manuscript. DS collected literature data and wrote the first draft of the manuscript. DS, RVT, and HE wrote text, provided revisions, created figures and editing to the manuscript. All authors contributed to the article and approved the submitted version.

Funding

Financial support from National Science Foundation (NSF) Rules of Life: Microbiome Grant number 2025457 to RVT and a NSF PFRB Award #2006244 to HE.

Acknowledgments

The authors thank NSF for funding the Moorea Coral Reef (MCR) LTER Site (NSF OCE 16-37396). We also would like to thank the Vega Thurber Lab for the fundamental and useful comments and discussion.

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.

Publisher's note

References

Acinas, S. G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., and Polz, M. F. (2005). PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl. Environ. Microbiol.* 71, 8966–8969. doi: 10.1128/AEM.71.12.8966-8969.2005

Allen, A. (1983). Mucus — a protective secretion of complexity. *Trends Biochem. Sci.* 8, 169–173. doi: 10.1016/0968-0004(83)90166-4

Amir, A., McDonald, D., Navas-Molina, J. A., Kopylova, E., Morton, J. T., Zech Xu, Z., et al. (2017). Deblur rapidly resolves single-nucleotide community sequence patterns. *mSystems* 2:2. doi: 10.1128/mSystems.00191-16

Apprill, A., McNally, S., Parsons, R., and Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat. Microb. Ecol.* 75, 129–137. doi: 10.3354/ame01753

Apprill, A., Weber, L. G., and Santoro, A. E. (2016). Distinguishing between microbial habitats unravels ecological complexity in coral microbiomes. *mSystems* 1:143. doi: 10.1128/msystems.00143-16

Aslanzadeh, J. (1993). Application of hydroxylamine hydrochloride for post-PCR sterilization. *Mol. Cell. Probes* 7, 145–150. doi: 10.1006/mcpr.1993.1020

Aslanzadeh, J. (2004). Preventing PCR amplification carryover contamination in a clinical laboratory. *Ann. Clin. Lab. Sci.* 34, 389–396.

Assal, N., and Lin, M. (2021). PCR procedures to amplify GC-rich DNA sequences of Mycobacterium bovis. *J. Microbiol. Methods* 181:106121. doi: 10.1016/j.mimet.2020.106121

Athanasio, C. G., Chipman, J. K., Viant, M. R., and Mirbahai, L. (2016). Optimisation of DNA extraction from the crustacean daphnia. *PeerJ* 4:2004. doi: 10.7717/peerj.2004

Banaszak, A. T. (2007). Optimization of DNA extraction from a scleractinian coral for the detection of thymine dimers by immunoassay. *Photochem. Photobiol.* 83, 833–838. doi: 10.1111/j.1751-1097.2007.00083.x

Barberán, A., Bates, S. T., Casamayor, E. O., and Fierer, N. (2012). Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* 6, 343–351. doi: 10.1038/ismej.2011.119

Becker, C. C., Brandt, M., Miller, C. A., and Apprill, A. (2022). Microbial bioindicators of stony coral tissue loss disease identified in corals and overlying waters using a rapid field-based sequencing approach. *Environ. Microbiol.* 24, 1166–1182. doi: 10.1111/1462-2920.15718

Bengtsson-Palme, J., Hartmann, M., Eriksson, K. M., Pal, C., Thorell, K., Larsson, D. G. J., et al. (2015). METAXA2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. *Mol. Ecol. Resour.* 15, 1403–1414. doi: 10.1111/1755-0998.12399

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456, 53–59. doi: 10.1038/nature07517

Bergman, J. L., Shaw, T., Egan, S., and Ainsworth, T. D. (2022). Assessing the coral microbiome at the scale of tissue-specific habitats within the coral meta-organism. *Front. Mar. Sci.* 9:496. doi: 10.3389/fmars.2022.985496

Berzins, I. K., Smolowitz, R. M., and Lewbart, G. A. (2011). Diagnostic techniques and sample handling. *Invertebrate Medicine*, 389–400. doi: 10.1002/9780470960806.ch22

Biagi, E., Caroselli, E., Barone, M., Pezzimenti, M., Teixido, N., Soverini, M., et al. (2020). Patterns in microbiome composition differ with ocean acidification in anatomic compartments of the Mediterranean coral Astroides calycularis living at CO2 vents. *Sci. Total Environ.* 724:138048. doi: 10.1016/j.scitotenv.2020.138048

Boesenberg-Smith, K. A., Pessarakli, M. M., and Wolk, D. M. (2012). Assessment of DNA yield and purity: an overlooked detail of PCR troubleshooting. *Clin. Microbiol. Newsl.* 34, 1–6. doi: 10.1016/j.clinmicnews.2011.12.002

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857. doi: 10.1038/s41587-019-0209-9

Bourne, D., Iida, Y., Uthicke, S., and Smith-Keune, C. (2008). Changes in coral-associated microbial communities during a bleaching event. *ISME J.* 2, 350–363. doi: 10.1038/ismej.2007.112

Bourne, D. G., Morrow, K. M., and Webster, N. S. (2016). Insights into the coral microbiome: underpinning the health and resilience of reef ecosystems. *Annu. Rev. Microbiol.* 70, 317–340. doi: 10.1146/annurev-micro-102215-095440

Bourne, D. G., and Munn, C. B. (2005). Diversity of bacteria associated with the coral Pocillopora damicornis from the great barrier reef. *Environ. Microbiol.* 7, 1162–1174. doi: 10.1111/j.1462-2920.2005.00793.x

Brown, R. B., and Audet, J. (2008). Current techniques for single-cell lysis. *J. R. Soc. Interface* 5, S131–S138. doi: 10.1098/rsif.2008.0009.focus

Brown, B. E., and Bythell, J. C. (2005). Perspectives on mucus secretion in reef corals. *Mar. Ecol. Prog. Ser.* 296, 291–309. doi: 10.3354/meps296291

Bukin, Y. S., Galachyants, Y. P., Morozov, I. V., Bukin, S. V., Zakharenko, A. S., and Zemskaya, T. I. (2019). The effect of 16S rRNA region choice on bacterial community metabarcoding results. *Scientific Data* 6:190007. doi: 10.1038/sdata.2019.7

Bythell, J. C., and Wild, C. (2011). Biology and ecology of coral mucus release. *J. Exp. Mar. Biol. Ecol.* 408, 88–93. doi: 10.1016/j.jembe.2011.07.028

Cadet, J., Berger, M., Decarroz, C., Wagner, J. R., van Lier, J. E., Ginot, Y. M., et al. (1986). Photosensitized reactions of nucleic acids. *Biochimie* 68, 813–834. doi: 10.1016/s0300-9084(86)80097-9

Cai, L., Tian, R. M., Zhou, G., Tong, H., Him Wong, Y., Zhang, W., et al. (2018). Exploring coral microbiome assemblages in the South China Sea. *Sci. Rep.* 8:2428. doi: 10.1038/s41598-018-20515-w

Callahan, B. J., McMurdie, P. J., and Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* 11, 2639–2643. doi: 10.1038/ismej.2017.119

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

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. U. S. A.* 108, 4516–4522. doi: 10.1073/pnas.1000080107

Carlos, C., Torres, T. T., and Ottoboni, L. M. M. (2013). Bacterial communities and species-specific associations with the mucus of Brazilian coral species. *Sci. Rep.* 3:1624. doi: 10.1038/srep01624

Carradec, Q., Poulain, J., Boissin, E., Hume, B. C. C., Voolstra, C. R., Ziegler, M., et al. (2021). Author correction: a framework for in situ molecular characterization of coral holobionts using nanopore sequencing. *Sci. Rep.* 11:2076. doi: 10.1038/s41598-021-81544-6

Champlot, S., Berthelot, C., Pruvost, M., Bennett, E. A., Grange, T., and Geigl, E.-M. (2010). An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. *PLoS One* 5:e13042. doi: 10.1371/journal.pone.0013042

Chen, Y., and McCarthy, D. (2015). edgeR: differential expression analysis of digital gene expression data User's guide. Bioconductor. Available at: https://bioconductor.statistik.tu-dortmund.de/packages/3.10/bioc/vignettes/edgeR/inst/doc/edgeRUsersGuide.pdf. (Accessed in May 3, 2022).

Douglas, G. M., Maffei, V. J., Zaneveld, J. R., et al. (2020). PICRUSt2 for prediction of metagenome functions. *Nat. Biotechnol.* 38, 685–688. doi: 10.1038/s41587-020-0548-6

Dufrêne, M., and Legendre, P. (1997). Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecol. Monogr.* 67, 345–366. doi: 10.1890/0012-9615(1997)067[0345:SAAIST]2.0.CO;2

Edgar, R. (2010). Usearch. Available at: https://www.osti.gov/biblio/1137186.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., et al. (2009). Real-time DNA sequencing from single polymerase molecules. *Science* 323, 133–138. doi: 10.1126/science.1162986

Engelen, A. H., Aires, T., Vermeij, M. J. A., Herndl, G. J., Serrão, E. A., and Frade, P. R. (2018). Host differentiation and compartmentalization of microbial communities in the Azooxanthellate Cupcorals Tubastrea coccinea and Rhizopsammia goesi in the Caribbean. *Front. Mar. Sci.* 5:391. doi: 10.3389/fmars.2018.00391

Epstein, H. E., Smith, H. A., Torda, G., and Oppen, M. J. H. (2019). Microbiome engineering: enhancing climate resilience in corals. *Front. Ecol. Environ.* 17, 100–108. doi: 10.1002/fee.2001

Eshleman, J., and Smith, D. G. (2001). Use of DNase to eliminate contamination in ancient DNA analysis. *Electrophoresis* 22, 4316–4319. doi: 10.1002/1522-2683(200112)22:20<4316::AID-ELPS4316>3.0.CO;2-V

Fischer, M., Renevey, N., Thür, B., Hoffmann, D., Beer, M., and Hoffmann, B. (2016). Efficacy assessment of nucleic acid decontamination reagents used in molecular diagnostic laboratories. *PLoS One* 11:e0159274. doi: 10.1371/journal.pone.0159274

Ford, A. K., Visser, P. M., van Herk, M. J., Jongepier, E., and Bonito, V. (2021). First insights into the impacts of benthic cyanobacterial mats on fish herbivory functions on a nearshore coral reef. *Sci. Rep.* 11:7147. doi: 10.1038/s41598-021-84016-z

Fox, J. C., Ait-Khaled, M., Webster, A., and Emery, V. C. (1991). Eliminating PCR contamination: is UV irradiation the answer? *J. Virol. Methods* 33, 375–382. doi: 10.1016/0166-0934(91)90037-z

Fragoso Ados Santos, H., Duarte, G. A. S., Rachid, C. T. D. C., Chaloub, R. M., Calderon, E. N., Marangoni, L. F. D. B., et al. (2015). Impact of oil spills on coral reefs can be reduced by bioremediation using probiotic microbiota. *Sci. Rep.* 5:18268. doi: 10.1038/srep18268

Fraune, S., and Bosch, T. C. G. (2010). Why bacteria matter in animal development and evolution. *BioEssays* 32, 571–580. doi: 10.1002/bies.200900192

Gaither, M. R., Szabó, Z., Crepeau, M. W., Bird, C. E., and Toonen, R. J. (2011). Preservation of corals in salt-saturated DMSO buffer is superior to ethanol for PCR experiments. *Coral Reefs* 30, 329–333. doi: 10.1007/s00338-010-0687-1

Galkiewicz, J. P., and Kellogg, C. A. (2008). Cross-kingdom amplification using bacteria-specific primers: complications for studies of coral microbial ecology. *Appl. Environ. Microbiol.* 74, 7828–7831. doi: 10.1128/AEM.01303-08

Gardner, S. G., Camp, E. F., Smith, D. J., Kahlke, T., Osman, E. O., Gendron, G., et al. (2019). Coral microbiome diversity reflects mass coral bleaching susceptibility during the 2016 El Niño heat wave. *Ecol. Evol.* 9, 938–956. doi: 10.1002/ece3.4662

Garren, M., and Azam, F. (2012). New directions in coral reef microbial ecology. *Environ. Microbiol.* 14, 833–844. doi: 10.1111/j.1462-2920.2011.02597.x

Gefrides, L. A., Powell, M. C., Donley, M. A., and Kahn, R. (2010). UV irradiation and autoclave treatment for elimination of contaminating DNA from laboratory consumables. *Forensic Sci. Int. Genet.* 4, 89–94. doi: 10.1016/j.fsigen.2009.06.008

Gibbons, S. M., and Gilbert, J. A. (2015). Microbial diversity — exploration of natural ecosystems and microbiomes. *Curr. Opin. Genet. Dev.* 35, 66–72. doi: 10.1016/j.gde.2015.10.003

Giussani, A., Conti, I., Nenov, A., and Garavelli, M. (2018). Photoinduced formation mechanism of the thymine–thymine (6–4) adduct in DNA; a QM(CASPT2//CASSCF):MM(AMBER) study. *Faraday Discuss.* 207, 375–387. doi: 10.1039/c7fd00202e

Glasl, B., Bourne, D. G., Frade, P. R., Thomas, T., Schaffelke, B., and Webster, N. S. (2019). Microbial indicators of environmental perturbations in coral reef ecosystems. *Microbiome* 7:94. doi: 10.1186/s40168-019-0705-7

Glasl, B., Bourne, D. G., Frade, P. R., and Webster, N. S. (2018). Establishing microbial baselines to identify indicators of coral reef health. *Microbiol. Aust.* 39, 42–46. doi: 10.1071/MA18011

Glasl, B., Herndl, G. J., and Frade, P. R. (2016). The microbiome of coral surface mucus has a key role in mediating holobiont health and survival upon disturbance. *ISME J.* 10, 2280–2292. doi: 10.1038/ismej.2016.9

Goodyear, N. (2012). Effectiveness of five-day-old 10% bleach in a student microbiology laboratory setting. *Clin. Lab. Sci.* 25, 219–223. doi: 10.29074/ascls.25.4.219

Grottoli, A. G. (2001). "Past climate from corals" in *Encyclopedia of ocean sciences*. ed. J. H. Steele (Oxford: Academic Press), 2098–2107.

Hadaidi, G., Röthig, T., Yum, L. K., Ziegler, M., Arif, C., Roder, C., et al. (2017). Stable mucus-associated bacterial communities in bleached and healthy corals of Porites lobata from the Arabian seas. *Sci. Rep.* 7:362. doi: 10.1038/srep45362

Hartmann, A. C., Baird, A. H., Knowlton, N., and Huang, D. (2017). The paradox of environmental symbiont Acquisition in Obligate Mutualisms. *Curr. Biol.* 27, 3711–3716.e3. doi: 10.1016/j.cub.2017.10.036

Haydon, T. D., Suggett, D. J., Siboni, N., Kahlke, T., Camp, E. F., and Seymour, J. R. (2022). Temporal variation in the microbiome of tropical and temperate Octocorals. *Microb. Ecol.* 83, 1073–1087. doi: 10.1007/s00248-021-01823-7

Head, S. R., Komori, H. K., LaMere, S. A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D. R., et al. (2014). Library construction for next-generation sequencing: overviews and challenges. *BioTechniques* 56, 61–64. doi: 10.2144/000114133

Hernandez-Agreda, A., Gates, R. D., and Ainsworth, T. D. (2017). Defining the Core microbiome in corals' microbial soup. *Trends Microbiol.* 25, 125–140. doi: 10.1016/j.tim.2016.11.003

Hernandez-Agreda, A., Leggat, W., and Ainsworth, T. D. (2018a). A comparative analysis of microbial DNA preparation methods for use with massive and branching coral growth forms. *Front. Microbiol.* 9:2146. doi: 10.3389/fmicb.2018.02146

Hernandez-Agreda, A., Leggat, W., Bongaerts, P., Herrera, C., and Ainsworth, T. D. (2018b). Rethinking the coral microbiome: simplicity exists within a diverse microbial biosphere. *MBio* 9:e00812-18. doi: 10.1128/mBio.00812-18

Hester, E. R., Barott, K. L., Nulton, J., Vermeij, M. J., and Rohwer, F. L. (2016). Stable and sporadic symbiotic communities of coral and algal holobionts. *ISME J.* 10, 1157–1169. doi: 10.1038/ismej.2015.190

Hughes, J. B., and Hellmann, J. J. (2005). The application of rarefaction techniques to molecular inventories of microbial diversity. *Methods Enzymol.* 397, 292–308. doi: 10.1016/S0076-6879(05)97017-1

Illumina. (2019). MiSeq System denature and dilute libraries guide (Document # 15039740 v10). Available at: https://support.illumina.com/downloads/prepare_libraries_for_sequencing_miseq_15039740.html (Accessed June 10, 2022).

Kamke, J., Taylor, M. W., and Schmitt, S. (2010). Activity profiles for marine sponge-associated bacteria obtained by 16S rRNA vs 16S rRNA gene comparisons. *ISME J.* 4, 498–508. doi: 10.1038/ismej.2009.143

Kampmann, M.-L., Børsting, C., and Morling, N. (2017). Decrease DNA contamination in the laboratories. *Forensic Science International: Genetics Supplement Series* 6, e577–e578. doi: 10.1016/j.fsigss.2017.09.223

Kellogg, C. A., Ross, S. W., and Brooke, S. D. (2016). Bacterial community diversity of the deep-sea octocoral Paramuricea placomus. *PeerJ* 4:e 2529. doi: 10.7717/peerj.2529

Ketchum, R. N., Smith, E. G., Vaughan, G. O., Phippen, B. L., McParland, D., Al-Mansoori, N., et al. (2018). DNA extraction method plays a significant role when defining bacterial community composition in the marine invertebrate Echinometra mathaei. *Front. Mar. Sci.* 5:255. doi: 10.3389/fmars.2018.00255

Kim, M., Morrison, M., and Yu, Z. (2011). Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *J. Microbiol. Methods* 84, 81–87. doi: 10.1016/j.mimet.2010.10.020

Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* 40:e3. doi: 10.1093/nar/gkr771

Klaschik, S., Lehmann, L. E., Raadts, A., Hoeft, A., and Stuber, F. (2002). Comparison of different decontamination methods for reagents to detect low concentrations of bacterial 16S DNA by real-time-PCR. *Mol. Biotechnol.* 22, 231–242. doi: 10.1385/MB:22:3:231

Klaus, J. S., Frias-Lopez, J., Bonheyo, G. T., Heikoop, J. M., and Fouke, B. W. (2005). Bacterial communities inhabiting the healthy tissues of two Caribbean reef corals: interspecific and spatial variation. *Coral Reefs* 24, 129–137. doi: 10.1007/s00338-004-0447-1

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., and Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* 79, 5112–5120. doi: 10.1128/AEM.01043-13

Krediet, C. J., Ritchie, K. B., Paul, V. J., and Max, T. (2013). Coral-associated micro-organisms and their roles in promoting coral health and thwarting diseases. *Proc. R. Soc. B.*280:20122328. doi: 10.1098/rspb.2012.2328

Kvennefors, E. C. E., Sampayo, E., Kerr, C., Vieira, G., Roff, G., and Barnes, A. C. (2012). Regulation of bacterial communities through antimicrobial activity by the coral holobiont. *Microb. Ecol.* 63, 605–618. doi: 10.1007/s00248-011-9946-0

Kwong, W. K., Del Campo, J., Mathur, V., Vermeij, M. J. A., and Keeling, P. J. (2019). A widespread coral-infecting apicomplexan with chlorophyll biosynthesis genes. *Nature* 568, 103–107. doi: 10.1038/s41586-019-1072-z

Lebonah, Dileep, and Chandrasekhar, . (2014). DNA barcoding on bacteria: A review. Adv. Eng. Educ. Available at: https://downloads.hindawi.com/archive/2014/541787.pdf.

Li, J., Chen, Q., Long, L.-J., Dong, J.-D., Yang, J., and Zhang, S. (2014). Bacterial dynamics within the mucus, tissue and skeleton of the coral Porites lutea during different seasons. *Sci. Rep.* 4:7320. doi: 10.1038/srep07320

Li, J., Chen, Q., Zhang, S., Huang, H., Yang, J., Tian, X.-P., et al. (2013). Highly heterogeneous bacterial communities associated with the South China Sea reef corals Porites lutea, Galaxea fascicularis and Acropora millepora. *PLoS One* 8:e71301. doi: 10.1371/journal.pone.0071301

Lin, H., and Peddada, S. D. (2020). Analysis of compositions of microbiomes with bias correction. *Nat. Commun.* 11:3514. doi: 10.1038/s41467-020-17041-7

Littman, R. A., Bourne, D. G., and Willis, B. L. (2010). Responses of coral-associated bacterial communities to heat stress differ with Symbiodiniumtype on the same coral host. *Mol. Ecol.* 19, 1978–1990. doi: 10.1111/j.1365-294x.2010.04620.x

Littman, R., Willis, B. L., and Bourne, D. G. (2011). Metagenomic analysis of the coral holobiont during a natural bleaching event on the great barrier reef. *Environ. Microbiol. Rep.* 3, 651–660. doi: 10.1111/j.1758-2229.2010.00234.x

Longo, M. C., Berninger, M. S., and Hartley, J. L. (1990). Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. *Gene* 93, 125–128. doi: 10.1016/0378-1119(90)90145-h

Lopez, I., Ruiz-Larrea, F., Cocolin, L., Orr, E., Phister, T., Marshall, M., et al. (2003). Design and evaluation of PCR primers for analysis of bacterial populations in wine by denaturing gradient gel electrophoresis. *Appl. Environ. Microbiol.* 69, 6801–6807. doi: 10.1128/AEM.69.11.6801-6807.2003

Lorenz, T. C. (2012). Polymerase chain reaction: basic protocol plus troubleshooting and optimization strategies. *J. Vis. Exp.* 63:e3998. doi: 10.3791/3998

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Maher, R. L., Epstein, H. E., and Vega Thurber, R. (2022). "Dynamics of bacterial communities on coral reefs: implications for conservation," in *Coral reef conservation and restoration in the omics age*, eds. OppenM. J. H. van and M. Aranda Lastra Cham: Springer International Publishing, 97–116.

Maher, R. L., Rice, M. M., McMinds, R., Burkepile, D. E., and Thurber, R. V. (2019). Multiple stressors interact primarily through antagonism to drive changes in the coral microbiome. *Sci. Rep.* 9:274. doi: 10.1038/s41598-019-43274-8

Mandal, S., Van Treuren, W., White, R. A., Eggesbø, M., Knight, R., and Peddada, S. D. (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microb. Ecol. Health Dis.* 26:27663. doi: 10.3402/mehd.v26.27663

Marchioro, G. M., Glasl, B., Engelen, A. H., Serrão, E. A., Bourne, D. G., Webster, N. S., et al. (2020). Microbiome dynamics in the tissue and mucus of acroporid corals differ in relation to host and environmental parameters. *PeerJ* 8:e9644. doi: 10.7717/peerj.9644

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380. doi: 10.1038/nature03959

Markoulatos, P., Siafakas, N., and Moncany, M. (2002). Multiplex polymerase chain reaction: a practical approach. *J. Clin. Lab. Anal.* 16, 47–51. doi: 10.1002/jcla.2058

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

McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., DeSantis, T. Z., Probst, A., et al. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* 6, 610–618. doi: 10.1038/ismej.2011.139

Mikheyev, A. S., and Tin, M. M. Y. (2014). A first look at the Oxford Nanopore min ION sequencer. *Mol. Ecol. Resour.* 14, 1097–1102. doi: 10.1111/1755-0998.12324

Morrow, K. M., Muller, E., and Lesser, M. P. (2018). "How does the coral microbiome cause, respond to, or modulate the bleaching process?," in *Coral bleaching: Patterns,*

processes, causes and consequences, eds. OppenM. J. H. van and J. M. Lough Cham: Springer International Publishing, 153–188.

Muscatine, L. (1969). New insights into reef-building corals. *Am. Biol. Teach.* 31, 367–371. doi: 10.2307/4442638

Neave, M. J., Rachmawati, R., Xun, L., Michell, C. T., Bourne, D. G., Apprill, A., et al. (2017). Differential specificity between closely related corals and abundant Endozoicomonas endosymbionts across global scales. *ISME J.* 11, 186–200. doi: 10.1038/ismej.2016.95

Nilakanta, H., Drews, K. L., Firrell, S., Foulkes, M. A., and Jablonski, K. A. (2014). A review of software for analyzing molecular sequences. *BMC. Res. Notes* 7:830. doi: 10.1186/1756-0500-7-830

Ogier, J.-C., Pagès, S., Galan, M., Barret, M., and Gaudriault, S. (2019). Rpo B, a promising marker for analyzing the diversity of bacterial communities by amplicon sequencing. *BMC Microbiol.* 19:171. doi: 10.1186/s12866-019-1546-z

Paul, N., and Yee, J. (2010). PCR incorporation of modified dNTPs: the substrate properties of biotinylated dNTPs. *BioTechniques* 48, 333–334. doi: 10.2144/000113405

Pei, A., Nossa, C. W., Chokshi, P., Blaser, M. J., Yang, L., Rosmarin, D. M., et al. (2009). Diversity of 23S rRNA genes within individual prokaryotic genomes. *PLoS One* 4:e5437. doi: 10.1371/journal.pone.0005437

Peixoto, R. S., Rosado, P. M., Leite, D. C. D. A., Rosado, A. S., and Bourne, D. G. (2017). Beneficial microorganisms for corals (BMC): proposed mechanisms for coral health and resilience. *Front. Microbiol.* 8:341. doi: 10.3389/fmicb.2017.00341

Plummer, E., Twin, J., Bulach, D. M., Garland, S. M., and Tabrizi, S. N. (2015). A comparison of three bioinformatics pipelines for the analysis of preterm gut microbiota using 16S rRNA gene sequencing data. *J. Proteomics & Bioinformatics* 8, 283–291. doi: 10.4172/jpb.1000381

Pollock, F. J., McMinds, R., Smith, S., Bourne, D. G., Willis, B. L., Medina, M., et al. (2018). Coral-associated bacteria demonstrate phylosymbiosis and cophylogeny. *Nat. Commun.* 9:4921. doi: 10.1038/s41467-018-07275-x

Pollock, F. J., Morris, P. J., Willis, B. L., and Bourne, D. G. (2011). The urgent need for robust coral disease diagnostics. *PLoS Pathog.* 7:e1002183. doi: 10.1371/journal.ppat.1002183

Pootakham, W., Mhuantong, W., Putchim, L., Yoocha, T., Sonthirod, C., Kongkachana, W., et al. (2018). Dynamics of coral-associated microbiomes during a thermal bleaching event. *Microbiology* 7:e00604. doi: 10.1002/mbo3.604

Pratte, Z. A., and Kellogg, C. A. (2021). Comparison of preservation and extraction methods on five taxonomically disparate coral microbiomes. *Front. Mar. Sci.* 8:161. doi: 10.3389/fmars.2021.684161

Prince, A. M., and Andrus, L. (1992). PCR: how to kill unwanted DNA. *BioTechniques* 12, 358–360.

Prodan, A., Tremaroli, V., Brolin, H., Zwinderman, A. H., Nieuwdorp, M., and Levin, E. (2020). Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. *PLoS One* 15:e0227434. doi: 10.1371/journal.pone.0227434

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., et al. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35, 7188–7196. doi: 10.1093/nar/gkm864

Rädecker, N., Pogoreutz, C., Voolstra, C. R., Wiedenmann, J., and Wild, C. (2015). Nitrogen cycling in corals: the key to understanding holobiont functioning? *Trends Microbiol.* 23, 490–497. doi: 10.1016/j.tim.2015.03.008

Reigel, A. M., Owens, S. M., and Hellberg, M. E. (2020). Reducing host DNA contamination in 16S rRNA gene surveys of anthozoan microbiomes using PNA clamps. *Coral Reefs* 39, 1817–1827. doi: 10.1007/s00338-020-02006-5

Reshef, L., Koren, O., Loya, Y., Zilber-Rosenberg, I., and Rosenberg, E. (2006). The coral probiotic hypothesis. *Environ. Microbiol.* 8, 2068–2073. doi: 10.1111/j.1462-2920.2006.01148.x

Ricci, F., Tandon, K., Black, J. R., Lê Cao, K.-A., Blackall, L. L., and Verbruggen, H. (2022). Host traits and phylogeny contribute to shaping coral-bacterial symbioses. *Msystems* 7:e0004422. doi: 10.1128/msystems.00044-22

Ritchie, K. B. (2006). Regulation of microbial populations by coral surface mucus and mucus-associated bacteria. *Mar. Ecol. Prog. Ser.* 322, 1–14.

Robbins, S. J., Singleton, C. M., Chan, C. X., Messer, L. F., Geers, A. U., Ying, H., et al. (2019). A genomic view of the reef-building coral Porites lutea and its microbial symbionts. *Nat. Microbiol.* 4, 2090–2100. doi: 10.1038/s41564-019-0532-4

Robinson, M. D., and Smyth, G. K. (2007). Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* 23, 2881–2887. doi: 10.1093/bioinformatics/btm453

Rodriguez-Lanetty, M., Granados-Cifuentes, C., Barberan, A., Bellantuono, A. J., and Bastidas, C. (2013). Ecological inferences from a deep screening of the complex bacterial consortia associated with the coral, Porites astreoides. *Mol. Ecol.* 22, 4349–4362. doi: 10.1111/mec.12392

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e 2584. doi: 10.7717/peerj.2584

Rohwer, F., Breitbart, M., Jara, J., Azam, F., and Knowlton, N. (2001). Diversity of bacteria associated with the Caribbean coral Montastraea franksi. *Coral Reefs* 20, 85–91. doi: 10.1007/s003380100138

Rohwer, F., Seguritan, V., Azam, F., and Knowlton, N. (2002). Diversity and distribution of coral-associated bacteria. *Mar. Ecol. Prog. Ser.* 243, 1–10. doi: 10.3354/meps243001

Roitman, S., López-Londoño, T., Joseph Pollock, F., Ritchie, K. B., Galindo-Martínez, C. T., Gómez-Campo, K., et al. (2020). Surviving marginalized reefs: assessing the implications of the microbiome on coral physiology and survivorship. *Coral Reefs* 39, 795–807. doi: 10.1007/s00338-020-01951-5

Rosado, P. M., Leite, D. C. A., Duarte, G. A. S., Chaloub, R. M., Jospin, G., Nunes da Rocha, U., et al. (2019). Marine probiotics: increasing coral resistance to bleaching through microbiome manipulation. *ISME J.* 13, 921–936. doi: 10.1038/s41396-018-0323-6

Rosales, S. M., Miller, M. W., Williams, D. E., Traylor-Knowles, N., Young, B., and Serrano, X. M. (2019). Microbiome differences in disease-resistant vs. susceptible Acropora corals subjected to disease challenge assays. *Sci. Rep.* 9:18279. doi: 10.1038/s41598-019-54855-y

Rosenberg, E. (2009). Coral microbiology. *Microb. Biotechnol.* 2:147. doi: 10.1111/j.1751-7915.2009.00090_15.x

Rosenberg, E., Koren, O., Reshef, L., Efrony, R., and Zilber-Rosenberg, I. (2007). The role of microorganisms in coral health, disease and evolution. *Nat. Rev. Microbiol.* 5, 355–362. doi: 10.1038/nrmicro1635

Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., et al. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475, 348–352. doi: 10.1038/nature10242

Roux, K. H. (1995). Optimization and troubleshooting in PCR. *PCR Methods Appl.* 4, S185–S194. doi: 10.1101/gr.4.5.s185

Salazar, O., and Asenjo, J. A. (2007). Enzymatic lysis of microbial cells. *Biotechnol. Lett.* 29, 985–994. doi: 10.1007/s10529-007-9345-2

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., et al. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* 12:87. doi: 10.1186/s12915-014-0087-z

Sambo, F., Finotello, F., Lavezzo, E., Baruzzo, G., Masi, G., Peta, E., et al. (2018). Optimizing PCR primers targeting the bacterial 16S ribosomal RNA gene. *BMC Bioinformatics* 19:343. doi: 10.1186/s12859-018-2360-6

Santoro, E. P., Borges, R. M., Espinoza, J. L., Freire, M., Messias, C. S. M. A., Villela, H. D. M., et al. (2021). Coral microbiome manipulation elicits metabolic and genetic restructuring to mitigate heat stress and evade mortality. *Sci. Adv.* 7:088. doi: 10.1126/sciadv.abg3088

Santos, H. F., Carmo, F. L., Leite, D. C. A., Jesus, H. E., Maalouf, P. D. C., Almeida, C., et al. (2012). Comparison of different protocols for the extraction of microbial DNA from reef corals. *Braz. J. Microbiol.* 43, 517–527. doi: 10.1590/S1517-83822012000200012

Sato, Y., Willis, B. L., and Bourne, D. G. (2013). Pyrosequencing-based profiling of archaeal and bacterial 16S rRNA genes identifies a novel archaeon associated with black band disease in corals. *Environ. Microbiol.* 15, 2994–3007. doi: 10.1111/1462-2920.12256

Scherczinger, C. A., Ladd, C., Bourke, M. T., Adamowicz, M. S., Johannes, P. M., Scherczinger, R., et al. (1999). A systematic analysis of PCR contamination. *J. Forensic Sci.* 44:12038J. doi: 10.1520/jfs12038j

Schloss, P. D. (2020). Reintroducing mothur: 10 years later. *Appl. Environ. Microbiol.* 86:e02343-19. doi: 10.1128/AEM.02343-19

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09

Schultz, J., Modolon, F., Rosado, A. S., Voolstra, C. R., Sweet, M., and Peixoto, R. S. (2022). Methods and strategies to uncover coral-associated microbial dark matter. *mSystems* 7:e0036722. doi: 10.1128/msystems.00367-22

Sekar, R., Kaczmarsky, L. T., and Richardson, L. L. (2009). Effect of freezing on PCR amplification of 16S rRNA genes from microbes associated with black band disease of corals. *Appl. Environ. Microbiol.* 75, 2581–2584. doi: 10.1128/AEM.01500-08

Shehadul Islam, M., Aryasomayajula, A., and Selvaganapathy, P. R. (2017). A review on macroscale and microscale cell lysis methods. *Micromachines* 8:83. doi: 10.3390/mi8030083

Shinzato, C., Shoguchi, E., Kawashima, T., Hamada, M., Hisata, K., Tanaka, M., et al. (2011). Using the Acropora digitifera genome to understand coral responses to environmental change. *Nature* 476, 320–323. doi: 10.1038/nature10249

Shnit-Orland, M., and Kushmaro, A. (2009). Coral mucus-associated bacteria: a possible first line of defense. *FEMS Microbiol. Ecol.* 67, 371–380. doi: 10.1111/j.1574-6941.2008.00644.x

Siboni, N., Ben-Dov, E., Sivan, A., and Kushmaro, A. (2008). Global distribution and diversity of coral-associated archaea and their possible role in the coral holobiont nitrogen cycle. *Environ. Microbiol.* 10, 2979–2990. doi: 10.1111/j.1462-2920.2008.01718.x

Sierra, M. A., Li, Q., Pushalkar, S., Paul, B., Sandoval, T. A., Kamer, A. R., et al. (2020). The influences of bioinformatics tools and reference databases in analyzing the human Oral microbial community. *Genes* 11:878. doi: 10.3390/genes11080878

Silva, D. P., Villela, H. D. M., Santos, H. F., Duarte, G. A. S., Ribeiro, J. R., Ghizelini, A. M., et al. (2021). Multi-domain probiotic consortium as an alternative to chemical remediation of oil spills at coral reefs and adjacent sites. *Microbiome* 9:118. doi: 10.1186/s40168-021-01041-w

Singer, G. A. C., Fahner, N. A., Barnes, J. G., McCarthy, A., and Hajibabaei, M. (2019). Comprehensive biodiversity analysis via ultra-deep patterned flow cell technology: a case study of eDNA metabarcoding seawater. *Sci. Rep.* 9:5991. doi: 10.1038/s41598-019-42455-9

Staley, C., Kaiser, T., Gidley, M. L., Enochs, I. C., Jones, P. R., Goodwin, K. D., et al. (2017). Differential impacts of land-based sources of pollution on the microbiota of Southeast Florida coral reefs. *Appl. Environ. Microbiol.* 83:e03378-16. doi: 10.1128/AEM.03378-16

Sudek, M., Work, T. M., Aeby, G. S., and Davy, S. K. (2012). Histological observations in the Hawaiian reef coral, Porites compressa, affected by Porites bleaching with tissue loss. *J. Invertebr. Pathol.* 111, 121–125. doi: 10.1016/j.jip.2012.07.004

Sunagawa, S., Woodley, C. M., and Medina, M. (2010). Threatened corals provide underexplored microbial habitats. *PLoS One* 5:e9554. doi: 10.1371/journal.pone.0009554

Sweet, M. J., Croquer, A., and Bythell, J. C. (2011). Bacterial assemblages differ between compartments within the coral holobiont. *Coral Reefs* 30, 39–52. doi: 10.1007/s00338-010-0695-1

Swift, D., Cresswell, K., Johnson, R., Stilianoudakis, S., and Wei, X. (2022). A review of normalization and differential abundance methods for microbiome counts data. *WIREs* e1586. doi: 10.1002/wics.1586

Team, R.C. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: RA Lang. Environ. Stat. Comput. R Found. Stat. Comput.

van Oppen, M. J. H., and Blackall, L. L. (2019). Coral microbiome dynamics, functions and design in a changing world. *Nat. Rev. Microbiol.* 17, 557–567. doi: 10.1038/s41579-019-0223-4

Vega Thurber, R., Mydlarz, L. D., Brandt, M., Harvell, D., Weil, E., Raymundo, L., et al. (2020). Deciphering Coral Disease Dynamics: Integrating Host, Microbiome, and the Changing Environment. *Front. Ecol. Evol.* 8:575927. doi: 10.3389/fevo.2020.575927

Vega Thurber, R., Schmeltzer, E. R., Grottoli, A. G., van Woesik, R., Toonen, R. J., Warner, M., et al. (2022). Unified methods in collecting, preserving, and archiving coral bleaching and restoration specimens to increase sample utility and interdisciplinary collaboration. *PeerJ* 10:e14176. doi: 10.7717/peerj.14176

Vega Thurber, R., Willner-Hall, D., Rodriguez-Mueller, B., Desnues, C., Edwards, R. A., Angly, F., et al. (2009). Metagenomic analysis of stressed coral holobionts. *Environ. Microbiol.* 11, 2148–2163. doi: 10.1111/j.1462-2920.2009.01935.x

Wada, N., Pollock, F. J., Willis, B. L., Ainsworth, T., Mano, N., and Bourne, D. G. (2016). In situ visualization of bacterial populations in coral tissues: pitfalls and solutions. *PeerJ* 4:e 2424. doi: 10.7717/peerj.2424

Walters, W., Hyde, E. R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., et al. (2016). Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. *mSystems* 1:009. doi: 10.1128/mSystems.00009-15

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi: 10.1128/AEM.00062-07

Watson, R. J., and Blackwell, B. (2000). Purification and characterization of a common soil component which inhibits the polymerase chain reaction. *Can. J. Microbiol.* 46, 633–642. doi: 10.1139/w00-043

Weber, L. G. (2020). Characterizing the ecology of coral reef microorganisms across different scales within the Caribbean. Available at: https://dspace.mit.edu/handle/1721.1/128592?show=full [Accessed June 13, 2022].

Weber, L., DeForce, E., and Apprill, A. (2017). Optimization of DNA extraction for advancing coral microbiota investigations. *Microbiome* 5:18. doi: 10.1186/s40168-017-0229-y

Webster, N. S., and Reusch, T. B. H. (2017). Microbial contributions to the persistence of coral reefs. *ISME J.* 11, 2167–2174. doi: 10.1038/ismej.2017.66

Wegley, L., Edwards, R., Rodriguez-Brito, B., Liu, H., and Rohwer, F. (2007). Metagenomic analysis of the microbial community associated with the coral Porites astreoides. *Environ. Microbiol.* 9, 2707–2719. doi: 10.1111/j.1462-2920.2007.01383.x

Weiler, B. A., Verhoeven, J. T. P., and Dufour, S. C. (2018). Bacterial communities in tissues and surficial mucus of the cold-water coral Paragorgia arborea. *Front. Mar. Sci.* 5:378. doi: 10.3389/fmars.2018.00378

Weiss, S., Amir, A., Hyde, E. R., Metcalf, J. L., Song, S. J., and Knight, R. (2014). Tracking down the sources of experimental contamination in microbiome studies. *Genome Biol.* 15:564. doi: 10.1186/s13059-014-0564-2

Willis, A. D. (2019). Rarefaction, alpha diversity, and statistics. *Front. Microbiol.* 10:2407. doi: 10.3389/fmicb.2019.02407

Wingett, S. W., and Andrews, S. (2018). Fast Q screen: a tool for multi-genome mapping and quality control. *F1000Res* 7:1338. doi: 10.12688/f1000research.15931.2

Wu, W. W., Phue, J.-N., Lee, C.-T., Lin, C., Xu, L., Wang, R., et al. (2018a). Robust sub-nanomolar library preparation for high throughput next generation sequencing. *BMC Genomics* 19:326. doi: 10.1186/s12864-018-4677-y

Wu, H., Wang, C., and Wu, Z. (2013). A new shrinkage estimator for dispersion improves differential expression detection in RNA-seq data. *Biostatistics* 14, 232–243. doi: 10.1093/biostatistics/kxs033

Wu, Y., Wu, J., Zhang, Z., and Cheng, C. (2018b). DNA decontamination methods for internal quality management in clinical PCR laboratories. *J. Clin. Lab. Anal.* 32:e22290. doi: 10.1002/jcla.22290

Yang, B., Wang, Y., and Qian, P.-Y. (2016). Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics* 17:135. doi: 10.1186/s12859-016-0992-y

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., et al. (2013). The SILVA and "all-species living tree project (LTP)" taxonomic frameworks. *Nucleic Acids Res.* 42, D643–D648. doi: 10.1093/nar/gkt1209

Zaneveld, J. R., Burkepile, D. E., Shantz, A. A., Pritchard, C. E., McMinds, R., Payet, J. P., et al. (2016). Overfishing and nutrient pollution interact with temperature to disrupt coral reefs down to microbial scales. *Nat. Commun.* 7, 833. doi: 10.1038/ncomms11833

Zaneveld, J. R., McMinds, R., and Thurber, R. V. (2017). Stress and stability: applying the Anna Karenina principle to animal microbiomes. *Nat. Microbiol.* 2:121. doi: 10.1038/nmicrobiol.2017.121

Zhang, Y.-Y., Ling, J., Yang, Q.-S., Wang, Y.-S., Sun, C.-C., Sun, H.-Y., et al. (2015). The diversity of coral associated bacteria and the environmental factors affect their community variation. *Ecotoxicology* 24, 1467–1477. doi: 10.1007/s10646-015-1454-4

Zhang, R.-Y., Zou, B., Yan, Y.-W., Jeon, C. O., Li, M., Cai, M., et al. (2020). Design of targeted primers based on 16S rRNA sequences in meta-transcriptomic datasets and identification of a novel taxonomic group in the Asgard archaea. *BMC Microbiol.* 20:25. doi: 10.1186/s12866-020-1707-0

Zhu, Y. S., Isaacs, S. T., Cimino, C. D., and Hearst, J. E. (1991). The use of exonuclease III for polymerase chain reaction sterilization. *Nucleic Acids Res.* 19:2511. doi: 10.1093/nar/19.9.2511

Ziegler, M., Seneca, F. O., Yum, L. K., Palumbi, S. R., and Voolstra, C. R. (2017). Bacterial community dynamics are linked to patterns of coral heat tolerance. *Nat. Commun.* 8:14213. doi: 10.1038/ncomms14213

Ziubrii, S. (2019). Comparative characteristic of effectiveness of some decontamination facilities used for carrying out of forensic molecular genetic examination. *Theory and Practice of Forensic Science and Criminalistics* 19, 489–502. doi: 10.32353/khrife.1.2019.39

Frontiers in Microbiology

Explores the habitable world and the potential of microbial life

The largest and most cited microbiology journal which advances our understanding of the role microbes play in addressing global challenges such as healthcare, food security, and climate change.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact