



OPEN ACCESS

EDITED BY

Jasna Vijayan,
Cochin University of Science and
Technology, India

REVIEWED BY

Xuhui Huang,
Fudan University, China
Joel Sanchez,
Estación Científica Coiba (Coiba AIP),
Panama

*CORRESPONDENCE

Scott J. Edmundson
✉ Scott.Edmundson@pnnl.gov
David D. Pollock
✉ David.Pollock@CUAnschutz.edu

RECEIVED 17 December 2025

REVISED 26 January 2026

ACCEPTED 30 January 2026

PUBLISHED 24 February 2026

CITATION

Boise NR, Leiser OP, Jones K, Peter-Frank M, Leach D, Crafton-Tempel S, Regier P, Wu R, Phillips C, Cheung MS, Johnson CGM, Edmundson SJ and Pollock DD (2026) Sampling microbial dynamics in the Salish Sea estuary: evaluating methods to capture cyanobacteria and cyanophage. *Front. Mar. Sci.* 13:1769457. doi: 10.3389/fmars.2026.1769457

COPYRIGHT

© 2026 Boise, Leiser, Jones, Peter-Frank, Leach, Crafton-Tempel, Regier, Wu, Phillips, Cheung, Johnson, Edmundson and Pollock. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Sampling microbial dynamics in the Salish Sea estuary: evaluating methods to capture cyanobacteria and cyanophage

Noelani R. Boise¹, Owen P. Leiser², Kristin Jones¹, Mahala Peter-Frank¹, Damon Leach³, Stephen Crafton-Tempel¹, Peter Regier¹, Ruonan Wu³, Conner Phillips⁴, Margaret S. Cheung^{3,5,6}, Connah G. M. Johnson⁷, Scott J. Edmundson^{1*} and David D. Pollock^{3,8*}

¹Coastal Sciences Division, Energy and Environment Directorate, Pacific Northwest National Laboratory, Sequim, WA, United States, ²Nuclear, Chemical, and Biological Technologies Division, National Security Directorate, Pacific Northwest National Laboratory, Seattle, WA, United States,

³Biological Sciences Division, Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA, United States, ⁴AI and Data Analytics Division, National Security Directorate, Pacific Northwest National Laboratory, Seattle, WA, United States, ⁵Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, United States, ⁶Physics Department, University of Washington, Seattle, WA, United States, ⁷Advanced Computing, Mathematics, and Data Division, Physical and Computational Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA, United States, ⁸Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, CO, United States

Introduction: Picocyanobacteria from the genera *Prochlorococcus* and *Synechococcus* thrive across the globe in aquatic environments, have relatively small genomes, and have growth dynamics regulated by both viral interactions and abiotic conditions, making them excellent model organisms for exploring host-pathogen coevolution.

Methods: We developed and refined methods to sample and sequence cyanobacteria, cyanophages, and measured features of their abiotic environment.

Results: The protocol described herein can successfully discriminate large-cell eukaryotic organisms, but size fractionation of picocyanobacteria appears to be affected by the presence of free DNA, multicellular structures, and abundant tycheoposons. Our preferred final protocol from this exploratory effort included a combination of in-line and single vacuum flask filtrations, which reduced filtration processing time by over threefold in some cases compared to other tested methods, such as a fully in-line sequence or in-site filtrations. We successfully extracted an average of approximately 400–1200 ng for all filter fractions, with some variations between kits.

Discussion: The protocol described herein can successfully discriminate large-cell eukaryotic organisms, but size fractionation of picocyanobacteria appears to be affected by the presence of free DNA, multicellular structures, and abundant tycheoposons.

KEYWORDS

phage host interactions, phytoplankton, picocyanobacteria, Salish Sea, metagenomics, viral ecology, sampling methods, sequence diversity

Introduction

The picocyanobacteria *Prochlorococcus* and *Synechococcus* are collectively a ubiquitous component of phytoplankton and play an outsized role in global primary production in both freshwater and marine systems (Visintini et al., 2021; Callieri et al., 2022). They have a large geographic distribution year-round (Flombaum et al., 2013), but are more abundant in warm summer months (Zinser et al., 2006). The picocyanobacteria *Prochlorococcus* and *Synechococcus* are collectively a ubiquitous component of phytoplankton and play an outsized role in global primary production in both freshwater and marine systems (Callieri et al., 2022; Visintini et al., 2021; Callieri et al., 2022). They have a large geographic distribution year-round (Flombaum et al., 2013; Flombaum et al., 2013), but are more abundant in warm summer months (Zinser et al., 2006; Zinser et al., 2006). Typically ranging in size from 0.5–0.8 and 0.8–1.2 μm diameter, respectively, (Mella-Flores et al. 2012) the picocyanobacteria *Prochlorococcus* and *Synechococcus* are collectively a ubiquitous component of phytoplankton and play an outsized role in global primary production in both freshwater and marine systems (Visintini et al. 2021; Callieri et al. 2022), *Prochlorococcus* and *Synechococcus* are among the most abundant photosynthetic organisms on Earth and are key indicators of ecosystem health (Sánchez-Gallego et al., 2025). Due to their prevalence in various marine and estuarine ecosystems, their relatively small genomes, and their anti-viral potential (e.g., cyanovirin-N, a potent antiviral lectin), these organisms are exceptional model organisms for studying viral-bacterial dynamics (Visintini et al., 2021; Scanlan et al., 2009; Reynolds et al., 2021; Edmundson et al., 2020). Well-adapted to oligotrophic conditions (Kent et al., 2019), they also thrive in the euphotic zone of tropical and subtropical waters and are expected to expand their distribution as ocean temperatures rise and impact oceanic communities (Flombaum et al., 2020; Flombaum et al., 2013). Given their ubiquity, these organisms are likely present in the coastal and estuarine waters of the northern Olympic Peninsula of Washington State, USA and surrounding Salish Sea waters, but few studies specifically focus on picocyanobacteria in this region (Del Bel Belluz et al., 2024). The Salish Sea lies along what was historically considered the northern boundary of the distribution of *Prochlorococcus* (defined by *Prochlorococcus* versus *Synechococcus* prevalence ratios), but this boundary is expected to have shifted and continue to change as ocean temperatures increase (Ribalet et al., 2025; Flombaum et al., 2013; Mioduchowska et al., 2023), making the changing interactions and relative success of *Prochlorococcus*, *Synechococcus*, and their viruses topics of considerable ecological and biopreparedness interest (Edmundson et al., 2020; Reynolds et al., 2021).

The Salish Sea is an expansive network of complex water bodies including the Strait of Juan de Fuca, the Strait of Georgia, and Puget Sound. Kathryn Sobocinski's *The State of the Salish Sea* report showcases the dynamic currents and biogeochemistry of these waters (Sobocinski, 2021). Strong upwelling of deep, nutrient-rich water due to outflow of freshwater from the Strait of Juan de Fuca produces dynamic seasonal fluctuations. This wind-driven upwelling of deep water can also decrease dissolved oxygen (DO) levels and increase nutrient-rich water from the coast which circulates through the strait. Wastewater from the heavily

populated coastal areas around the Salish Sea also has significant effects on nutrient levels and thus cyanobacterial populations in these water bodies; in the summer, reduced freshwater outflow can lead to stagnant wastewater effluent and agricultural runoff in small regions of the Salish Sea, increasing seasonal eutrophication (Sobocinski, 2021). Despite extensive research on impacts of environmental conditions on picocyanobacteria, there is far less work documenting the impact of environmental factors on picocyanobacteria-cyanophage dynamics. Some work shows examples of environmental impacts, such as diel cycles impacting cyanobacterial gene expression and phage-host interactions (Liu et al., 2019; Rozum et al., 2025) and infection rates increasing in euphotic zones (Fuchsman and Hays, 2023).

While there are many commonly-used methods to sample specific microorganisms from aqueous environments (Huang et al., 2023), especially for public health end-uses (Karamati et al., 2024), there are few methods that provide relatively rapid samples targeting multiple size fractions from the same source water while simultaneously producing enough sample biomass for sufficient downstream genomic sequencing to elucidate community dynamics (Karamati et al., 2024; Zehr, 2010; Hsieh and Maloney, 2025). Methods such as flow cytometry (Ruber et al., 2018), eDNA sampling (Sahu et al., 2025), filter plating (Kearney et al., 2022), dilution-to-extinction (Buchholz et al., 2021), and density gradient centrifugation (Rong et al., 2022; Lawrence and Steward, 2010) can successfully identify or isolate target organisms, but require expensive equipment and refined expertise, have more limitations on target organisms than sequential filtration, and/or lack sequencing depth potential (Adams et al., 2019; Sharifian Gh. and Norouzi, 2023). Our described sampling method enables the isolation of multiple size fractions—such as viruses and hosts—from a discrete timeframe without culturing, as well as the simultaneous collection of metadata to inform the environmental context in which these organisms were found. By establishing consistent methods to obtain quality genomic and environmental data across locations, seasons, tidal stages, and diel cycles, this work yields holistic information to assess changing dynamics between marine cyanophages and picocyanobacteria, which can inform models that predict the behavior of higher-risk species. In a time of increasing risks in viral-host adaptations and spillover due to rapid environmental change (Carlson et al., 2022), it is critical that the scientific community establish consolidated methods for pairing environmental data with molecular research at a large scale to understand past, present, and future adaptation and interactive responses between viruses and their hosts. In addition to elucidating genomic relationships between phages and hosts, understanding the broader ecological and biogeochemical processes of discrete sample collections is essential for increasing predictive accuracy and capacity in model development (Dorigatti et al., 2025).

In this study, we tested and improved methods to sample and filter raw environmental water samples to extract DNA, identify picocyanobacteria and their associated phages, and describe their abiotic environment, with the long-term aim to study host-pathogen interactions and coevolution. To achieve this, we collected samples from five locations at varying times of day and tidal conditions along the northern Olympic Peninsula over an initial year-long period while performing iterative method

TABLE 1 Sample location descriptions for marked sites in Figure 1.

site reference number	Sampling location	Site coordinates	# of sampling events
1	Cline Spit Loading Dock	48°09'06"N 123°09'09"W	n = 1
2	PNNL-Sequim Floating Dock	48°04'45"N 123°02'42"W	n = 26
3	Exterior southern side of John Wayne Marina	48°03'41"N 123°02'20"W	n = 1
4	Southern end of Discovery Bay	47°59'48"N 122°52'32"W	n = 1
5	Open water north of Port Townsend	48°09'56"N 122°48'24"W	n = 1

All sites were sampled only once due to logistical challenges except for the PNNL-Sequim dock site, which was sampled 26 times in year one.

development. Sampling involved collecting surface-level samples either via in-line *in situ* serial filtration or direct surface sampling. Due to the wide variety and size distribution of phytoplankton organisms and specifically cyanobacteria (Mehdizadeh Allaf and Peerhossaini, 2022), we performed serial filtrations to remove larger debris and to reduce the amount of larger-celled non-target organisms prior to sequencing. We focused on identifying relative read counts of target organisms in the genera *Prochlorococcus* and *Synechococcus*, along with previously identified cyanophage pathogens. Based on preliminary results showing that reads similar to eukaryotic amoeba were consistently observed in our samples, we used relative amoeba read counts as a marker for removal of large-celled organisms.

In summary, the present work describes the development and evaluation of methods for enriching and analyzing picocyanobacteria and associated cyanophages in the Salish Sea for metagenomic sequencing, along with preliminary observations of select lineages under various collection conditions over time. This work highlights the efficacy and replicability of sequential filtration

and DNA extraction protocols to support studies of host-pathogen dynamics and environmental impacts on microbial interactions.

Materials and methods

Sampling, filtering and processing

We performed initial targeted sampling in Salish Sea waters from November 2023 to February 2025, mostly from the PNNL-Sequim floating dock and pier, and less frequently at four nearby locations: Cline Spit, south of John Wayne Marina near an estuary outflow, open waters North of Port Townsend, and in Discovery Bay (Table 1; Figure 1). Sampling events, including filtering and collection (Figures 2, 3) occurred near-monthly at the PNNL-Sequim dock over the course of a year, with offsite sampling and multiple within-month sampling concentrated during peak summer months. A total of 30 primary sampling events were performed on 25 separate days at ebb, flood, and slack tides, with the majority occurring during flood tides. Water levels, temperature, salinity, and collection dates are shown in Figure 4.

Samples were collected using two pilot methods, either 1) time-discrete samples using a horizontal water sampler, Niskin bottle, or direct surface sampling using a sterilized carboy; or 2) time-extended *in situ* in-line filtrations at the sampling site. The *in situ* apparatus consisted of a custom-designed, weighted metal frame attached to flexible tubing that we tested using both peristaltic and vacuum pumps at approximately 550–680 kPa and 130 kPa, respectively (Supplementary Table 1). We configured the tubing as an *in-situ* mechanism for two- and three-step serial filtrations using in-line filters that could be deployed from height and immediately pump raw seawater directly through two or three sequentially smaller filter sizes to capture target species, with the final filtrate collected into a sterile vessel. Alternatively, we used a direct water sampler (a Niskin bottle, horizontal water sampler, and/or a sterilized carboy) to collect discrete samples followed by

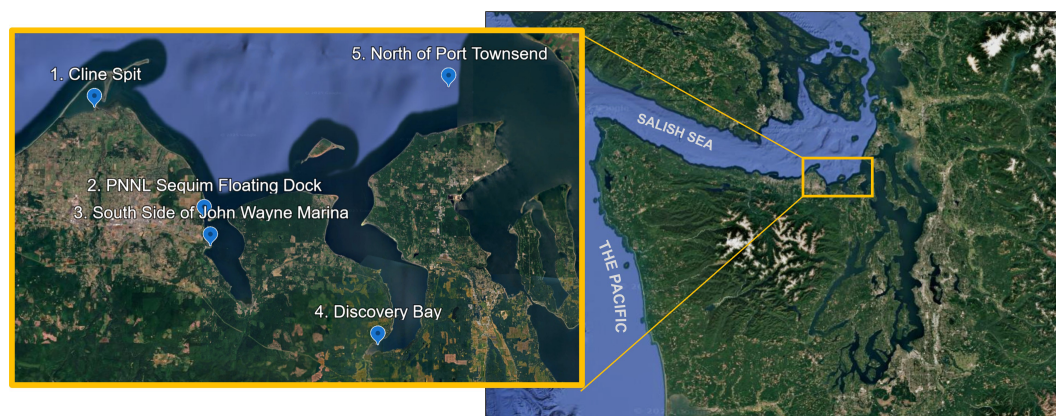


FIGURE 1

Sampling locations in the Salish Sea during the first year of sampling. 1) Cline Spit boat launch, 2) PNNL-Sequim floating dock, 3) Southern side of John Wayne Marina near estuary outflow, 4) Discovery Bay, 5) North of Port Townsend. All locations are located near the entrance of Admiralty Inlet from the Pacific Ocean. Satellite imagery generated using Google Earth. Inset (left) map data attribution: Google, Landsat/Copernicus Data SIO, NOAA, U.S. Navy, NGA, GEBCO Data LDEO-Columbia, NSF, NOAA Imagery from the dates: 12/14/2015–1/1/2021. Regional map (right) map data attribution: Google, Airbus Data, SIO, NOAA, U.S. Navy, NGA, GEBCO Imagery from the dates: 9/27/2004–10/14/2025.

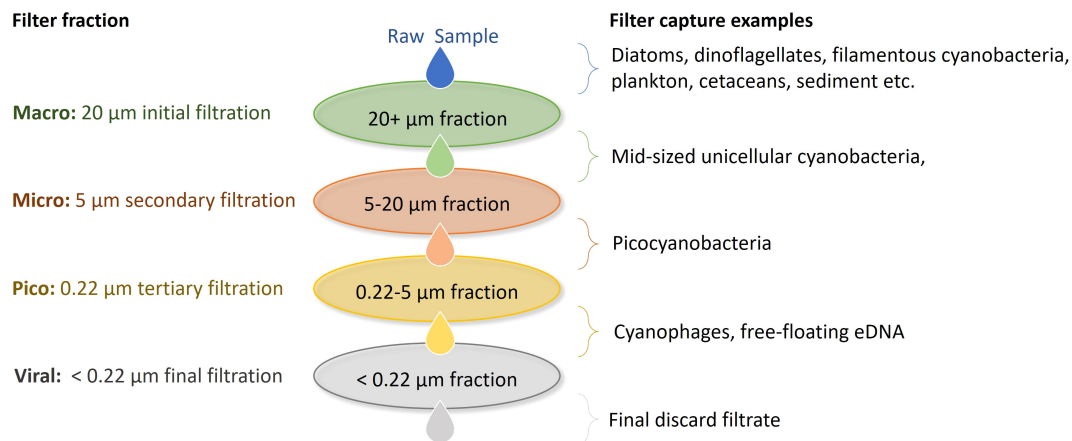


FIGURE 2

Serial filtration fractions. The projected target species enrichment goals for size fractionation are shown. The four fractions are: macro ($> 20 \mu\text{m}$); micro ($5\text{--}20 \mu\text{m}$); pico (and small micro, $0.22\text{--}5 \mu\text{m}$); and viral (filtered through $0.22 \mu\text{m}$, then precipitated, then collected on $0.22 \mu\text{m}$ filters). Note that the labels for these fractions are rough approximations, and the “pico” fraction in particular is generous at the upper end ($5 \mu\text{m}$) to avoid filtering out organisms with irregularly grouped cells at the larger end of the picocyanobacterial range (*Prochlorococcus* $0.5\text{--}0.7 \mu\text{m}$, *Synechococcus* $0.8\text{--}1.5 \mu\text{m}$).

filtrations in a laboratory. The in-line system collected samples over extended periods of time (up to 3 hours with especially dense water samples through a three-step filtration), compared to sampling durations of less than one minute using a direct water sampler.

Prior to sampling events, all sample collection vessels (carboys, Niskin bottles, and horizontal water samplers) were sterilized by autoclave or disinfected using a 10% bleach solution. We primed the collection vessel, tubing, and priming filters prior to every sample collection with water from the site, then discarded the priming water away from the collection point to avoid contamination. After collecting a pre-determined target volume, we immediately recorded metadata and sampling event context and assigned a single parent sampling ID to all associated data. Sampling conditions were noted, and additional weather data from the day of sampling were downloaded from the PNNL-Sequim Marine and Coastal Research Laboratory (MCRL) data site (MCRL Data, 2025).

Depending on which fractions were already processed, we then transported samples to our laboratory in Sequim WA to process the remaining filter fractions. All tubing and vessels were rinsed with 70% ethanol and/or autoclaved prior to priming with the appropriate matrix (we used each sequential filtrate to prime the subsequent collection vessels). For most samples, this involved a combination of in-line filtrations ($20\mu\text{m}$ 47 mm nylon filter and a $5 \mu\text{m}$ 47 mm mixed-cellulose ester (MCE) membrane secured in either a Millipore 47 mm polypropylene nylon filter holder or a Cytiva 47mm polycarbonate in-line filter holder using a peristaltic pump, Supplementary Tables 2, 3) and single vacuum flask filtrations. In-line filtration was followed by mixing in a single-flask and filtration onto a sterile $0.22 \mu\text{m}$ polyethersulfone (PES) or MCE filter in a vacuum flask to capture the pico fraction. Flow-throughs from the $0.22 \mu\text{m}$ filter were precipitated using a method adapted from John et al. (2011); iron chloride hexahydrate was added to the filtrate at 1 mL of 10g/L $\text{FeCl} \cdot 6\text{H}_2\text{O}$ solution in deionized water per 20 L of seawater and incubated for 1 hour, and precipitates were captured on a sterile $0.22 \mu\text{m}$ PES filter via vacuum filtration.

Additional filtration setups we tested included a $0.05 \mu\text{m}$ filter after $0.22 \mu\text{m}$ filtration, and various alternate size combinations with the three filter types: nylon, MCE membranes, and PES. Specifically, we tested 47 mm and 90 mm 0.22 or $0.2 \mu\text{m}$ MCE or PES filters, either as individual, sterile membranes or sterile vacuum units. We tested the $5 \mu\text{m}$ and $0.22 \mu\text{m}$ MCE membranes in an in-line filtration setup using either a polycarbonate filter-holder purchased from Avantor Sciences or a polypropylene filter-holder purchased from Millipore Sigma. The long-term use of 47 and 90 mm filters for differing filter fractions was driven by limited commercial availability of each material-pore size-diameter combination and for time-efficiency. A list of the tested filters, in-line filter holders, and other product information can be found in Supplementary Tables 1–4.

In the final sampling scheme, the sampling process included separating environmental samples by serial filtration into four fractions (Figure 2), labelled by their target organism size and the nominal pore sizes of the filters which they passed through and on which they were collected: macro ($> 20 \mu\text{m}$), micro (large micro, $5\text{--}20 \mu\text{m}$), pico (small micro, $0.22\text{--}5 \mu\text{m}$), and viral ($< 0.22 \mu\text{m}$). Further details are provided in Figure 3. Note that filter subsamples for the $20 \mu\text{m}$ and $5 \mu\text{m}$ filters were obtained by changing filters as they became clogged (as estimated qualitatively by reduced flow rate), whereas most filter subsamples for the $0.22 \mu\text{m}$ filters (both the pico and the viral fractions) were obtained by physically cutting the filters into thirds or quarters after completion of filtration (Figure 3).

DNA extraction

DNA was extracted from selected filter subsamples using Zymo’s Quick DNA Plant/Seed Miniprep Kit, and three Qiagen kits: DNeasy PowerWater Kit, PowerSoil Pro Kit, or Blood and Tissue Kit (Supplementary Table 4). Following extraction, DNA quality and quantity were assessed using a NanoDrop One spectrophotometer (ThermoFisher Scientific, Waltham, MA,

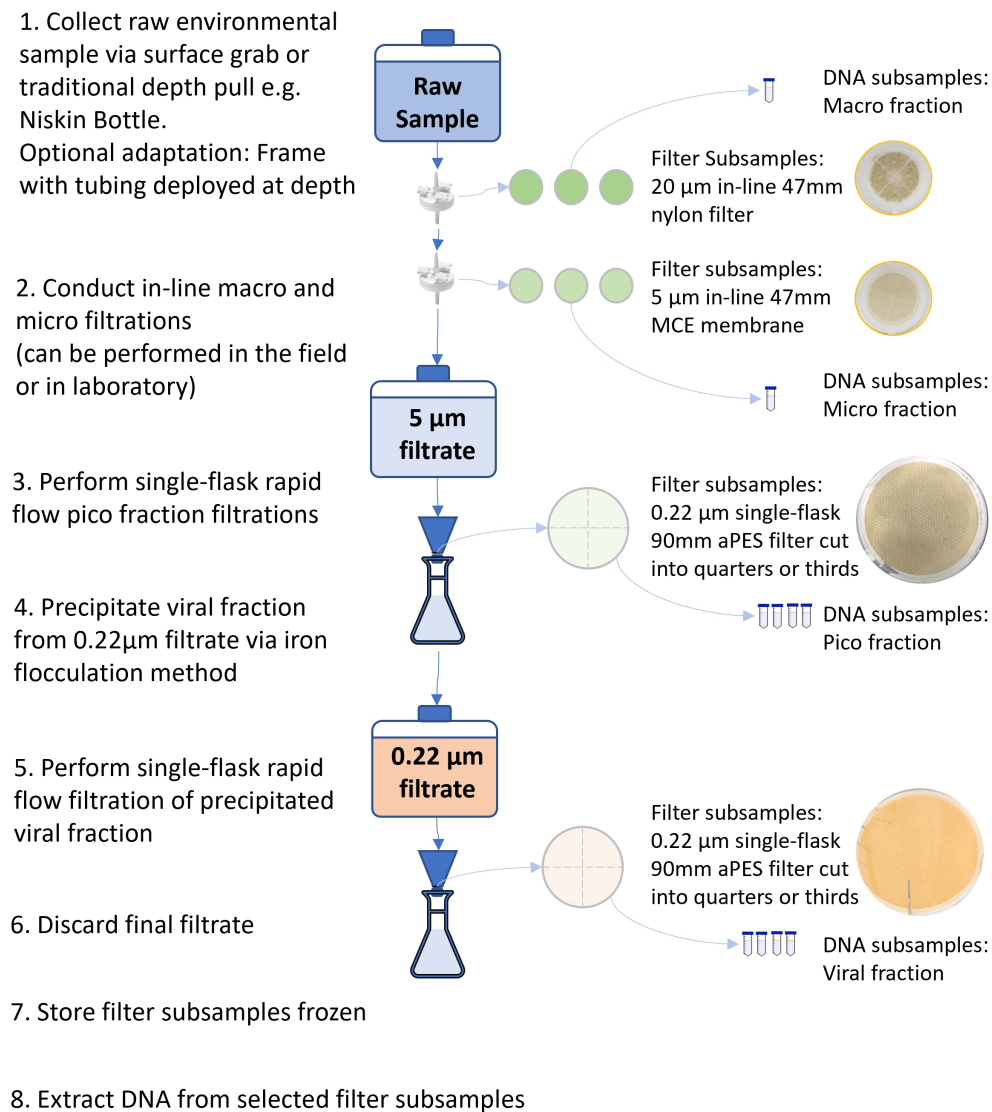


FIGURE 3

Final sampling scheme. Scheme incorporates a combination of in-line filtrations and single-flask filtrations to increase time-efficiency, minimize material waste, and improve yield of pico and viral fractions. Filters were replicated either as subsequent replacement filters in the in-line or flask filtration process, or as cut fractions through the midpoint of the filter, obtained by physically cutting individual filters after sample collection.

USA) and Qubit4 fluorometer system (ThermoFisher Scientific, Waltham, MA, USA). Samples were then frozen, and a select few were later analyzed on an Agilent 4150 TapeStation for quality assessment (Supplementary Data Figure 1). Samples were opportunistically selected for comparative analyses and were chosen based on sample factors such as number of technical replicates, the filter size, date of sampling and extractions, and tidal stage. While we did not have sufficient sample numbers for controlled comparisons in many cases, we chose the most comparable samples from the larger exploratory set and excluded samples that were too variable.

For direct comparisons of the kit effect on DNA recovery, DNA from select sampling events were extracted using two separate kits (Blood & Tissue and PowerWater) and analyzed on an Agilent 4150 TapeStation using a D5000 ScreenTape and the relevant kit reagents (D5000 buffer and D5000 ladder) following the kit protocol with no modifications (Agilent Technologies, Palo Alto, CA, USA). To

understand whether levels of variability due to kits are large compared to other sources of variation, we also make comparisons of DNA recovery without controlling for other known variables, which we refer to as 'kitchen sink' comparisons (a mix of conditions and replicates) to clarify the context in which they should be interpreted.

Metadata collection and processing

PNNL-Sequim is uniquely positioned for marine environmental research, leveraging its bench-to-bay capabilities through direct access to Sequim Bay, raw seawater access in the facilities, and a unique array of permanently deployed environmental sensors (MCRL Data, 2025). Most sampling events took place from the PNNL-Sequim floating dock at the mouth of Sequim Bay with an extensive range of environmental parameters documented from the MCRL Data, while other locations relied solely on parameters captured from a YSI ProDSS multiparameter

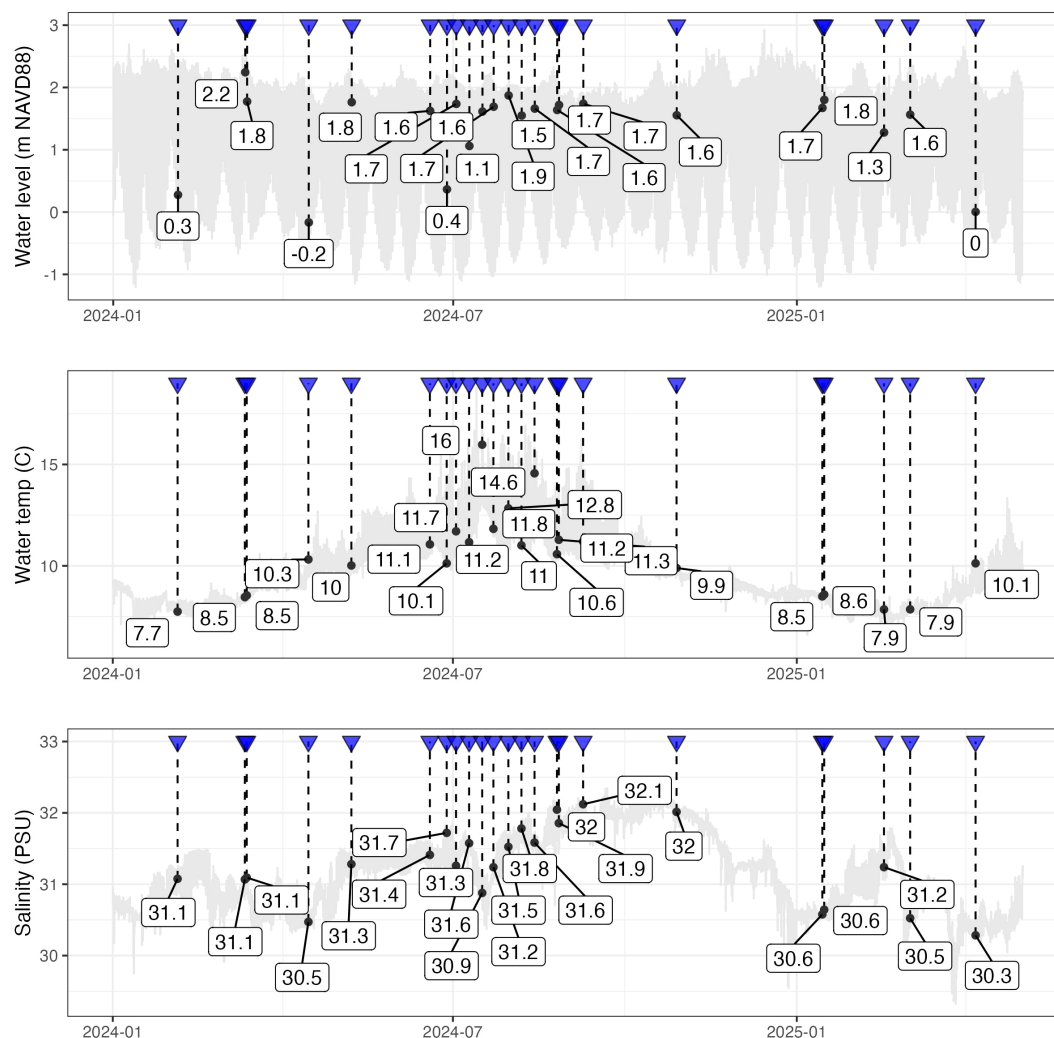


FIGURE 4
Seasonal data for water levels, temperature, and salinity. Values were measured using the MCRL Data sensor suite with denoted sampling events from 2024–2025.

water quality meter (YSI Xylem, Yellow Springs OH) (Supplementary Data Table 5). For PNNL-Sequim sampling events, we acquired data post-sampling using targeted scripts (MCRLdata Sandbox, 2025a; MCRLdata Sandbox, 2025b) to process environmental data from an online data repository (MCRL Data, 2025). One sampling event used for a controlled membrane test was executed in PNNL's marine mesocosm system, which circulates unfiltered seawater through a large (ca. 62,000 L) outdoor tank (Marine and Coastal Research Laboratory, 2025). This mesocosm system was used for a single sampling event due to restricted access to the PNNL-Sequim floating dock during the scheduled sampling time. While the samples from the mesocosm and dock are not directly comparable, the mesocosm system is constantly circulated with raw seawater with a submerged intake a few meters away the floating dock.

We used data collected for tidal stage, water temperature, and salinity by the sensors described above to provide environmental context for sampling events off the PNNL-Sequim dock (Figure 4). Data collected by sensors are sent to the cloud where they undergo

automated initial quality control and displayed in near real-time (MCRL Data, 2025) and are provided as part of the joint metadata summary (MCRLdata Sandbox, 2025a; MCRLdata Sandbox, 2025b). Additional quality control combining automated (e.g. de-spiking) and manual visual assessment were then applied, resulting in several data gaps originating from sensor maintenance, sensor errors, or unreliable data (identified during quality control). To fill any gaps, we used the *ranger* R package to construct Random Forest models (Wright and Ziegler, 2017). First, we gap-filled water levels by building a Random Forest model using water levels (NOAA Tides and Currents, 2025) pulled using the *VulnToolkit* R package along with hour of day, and day of year as predictors (Hill et al., 2021). The resulting model ($R^2 = 0.96$) was then used to gap-fill water levels. Those gap-filled water levels were used (along with hour of day and day of year) to predict both water temperature and salinity measured off the MCRL dock by CTD (conductivity, temperature, and depth sensor), and those models were used to gap-fill both temperature and salinity time-series. Model goodness-of-fit was high for both temperature and salinity models ($R^2 = 0.93$ and 0.95 , respectively).

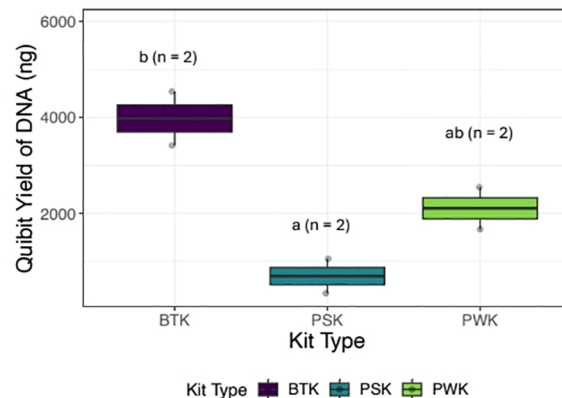


FIGURE 5

Comparative DNA yields from different extraction kits: Blood & Tissue Kit (BTK, purple), PowerSoil Kit (PSK, blue), and PowerWater Kit (PWK, green). DNA recovered (measured using a Qubit fluorometer) is shown for pico fractions (small micro) of a single sampling event on 26 August, 2024 during a high slack tide. Letters indicate significance groups as denoted by a Tukey's *post hoc* test ($p < 0.05$) after performing an analysis of variance (ANOVA), while $n=2$ indicates sample size in each plot.

Sequencing and preliminary data processing

DNA subsamples were shipped on ice to Azenta Life Sciences (South Plainfield, NJ, USA) for sequencing. Sequencing libraries were prepared using NEBNext Ultra DNA Library Prep Kit (New England Biolabs, Ipswich, MA, USA) according to manufacturer's instructions. Briefly, genomic DNA was acoustically fragmented using a Covaris S220 instrument (Covaris, Woburn, MA, USA) and fragmented DNA was cleaned and end repaired. Sequencing adapters were ligated to 3' ends of fragments after adenylation and enriched by limited-cycle PCR. Sequence libraries were validated using a High Sensitivity D1000 ScreenTape on an Agilent TapeStation (Agilent Technologies, Palo Alto, CA, USA) and quantified using a Qubit 3.0 fluorometer (ThermoFisher Scientific, Waltham, MA, USA). Sequence libraries were also quantified using real-time PCR (Applied Biosystems, Carlsbad, CA, USA) libraries from samples that passed all quality control steps and were multiplexed and clustered onto an Illumina NovaSeq flowcell (Illumina, San Diego, CA, USA) and sequenced using 2x150 paired-end configuration. Base calling and image analysis were conducted using the NovaSeqControl Software. Raw sequence data was converted to fastq format and de-multiplexed using Illumina's bcl2fastq software v2.20, allowing one mismatch for index sequence identification. Sequencing datasets were processed through the sequence processing pipelines described below for evaluation of the efficacy of filter fractionation methods and to refine strategies for further sampling. Raw paired-end reads were sequentially pre-processed by first interleaving paired reads into a single sequence read file preserving read pair information using the reformat functionality of BBTtools v37.62 (Bushnell et al., 2017). Interleaved reads were clustered, with optical and PCR duplicate reads removed using clumpify within BBTtools. Reads derived from low-quality regions of flowcells were removed using the filterbytile functionality from BBTtools. BBTtools bbduk was used to sequentially remove Illumina sequencing adapters, bases from the ends of reads below a phred33 quality score of ten, any reads with

an overall phred33 quality score lower than ten, and any reads mapping to PhiX spike-in. Read quality was assessed before and after preprocessing using FastQC v0.12.1 (Babraham Institute, 2023).

We generated metagenome contig assemblies using MEGAHIT v1.2.9 (Li et al., 2015) under default settings in paired-end mode. Output contigs were filtered by length as required for downstream analyses using the reformat functionality of BBTtools. Assembly qualities were assessed using QUAST v5.3.0 (Gurevich et al., 2013). Putative coding regions and putative protein sequences were predicted within metagenome assemblies filtered to contigs larger than 2500 bases using Metaprokka v1.15.0 (Seemann, 2014).

Raw short read pairs, interleaved pre-processed reads, contig assemblies, and predicted genes and proteins are found in [Supplementary Data 3](#) and were deposited to BioProject PRJNA1336290.

FAIR data management

We assigned each sampling event a unique parent ID that was tagged to the relevant environmental data for every filter and DNA subsample, which also had its own unique subsample ID. To track sampling events, we created a standardized field form with nomenclature and categorization based on the National Microbiome Data Collaborative (NMDC) water sample template for data documentation compatibility with the other public repositories (National Microbiome Data Collaborative, 2026). The unique sample identifiers for all parent and subsamples were used to further facilitate consistent data integration. Sequences, processed sequences, analytical results and associated metadata were captured and a) stored for sharing among the team using DataFed federated data storage and Globus-enabled federated IDs to ensure that this data is searchable and well-structured (George et al., 2025; EcoKMER GitHub Software Repository, 2025) and b) harmonized and visualized using the concurrently-developed EcoKMER program (George et al., 2025) with final output stored as a CSV file (Supplementary Data 1). This integrated metadata file combined

sequencing summary information with environmental and contextual data, adhering to FAIR (Findable, Accessible, Interoperable, and Reusable) principles (Wilkinson et al., 2016) and ensures that this data is searchable and well-structured.

Microbial sequence diversity estimation

To assess sequence diversity across environmental metagenomes, we applied Nonpareil (v3.5.5) to interleaved pre-processed reads (Rodriguez-R et al., 2018). Because the default kmer size for Nonpareil is 24, reads shorter than 24 base pairs (bp) were excluded. Nonpareil estimates average community coverage and diversity using kmer redundancy, modeling the rate at which new 24 bp sequences are encountered with increasing sequencing effort. The resulting Nonpareil diversity index (Nd) used does not cleanly reflect microbial diversity, as it reports diversity inferred from identical 24mers sampled, and does not account for variable rates along genomes, or variable genome size, but it has nevertheless been demonstrated to be useful in detecting differences in community composition (Rodriguez-R et al., 2018).

Relative read coverage of target taxa

61 DNA subsamples produced in year one were sequenced, filtered, assembled into contigs, and 24mer sequence diversities were calculated. To quantify taxon-specific enrichment of organisms with varying sizes, we mapped metagenomic reads to curated reference genome sets using BBDMap (v39.13) (Bushnell, 2014). Five reference genome sets were constructed from high-quality, complete NCBI genomes representing *Prochlorococcus* (26 reference genomes), *Synechococcus* (78 reference genomes), their phages (341 *Synechococcus* phage genomes; 37 *Prochlorococcus* phage genomes), and *Acanthamoeba* spp (5 reference genomes; accession numbers of each reference set in Supplementary Data 2). Our main reason to focus on these taxon groups was to identify the relative occurrence of the two picocyanobacterium and cyanophages that might be associated with them. In addition, we separately mapped reads to a mycetezoan amoeba to determine the effectiveness of filtering out large-celled organisms. Each set was indexed using the BBDMap indexing utility. Quality-filtered reads from each sample were aligned to each reference genome set using default parameters, and coverage statistics were calculated with the BBDMap pileup utility (Supplementary Data 1), which attaches each read to the reference genome in a run that matches it best, and leaves unattributed any reads that do not pass the minimum matching cutoff to any included genome(s). We repeated this analysis using the curated CyanoRak database of *Prochlorococcus* and *Synechococcus* reference samples (Supplementary Figure 3), with highly correlated results between our analysis and that using CyanoRak (Garczarek et al., 2021).

Each of the six genome groups listed above were analyzed separately. During preliminary analysis, we observed that the bulk of *Acanthamoeba* reads were consistently attributed to the five major chromosomes of *Dictyostelium discoideum*, so we opted to present the sum of plus and minus reads (relative to genome orientation) most preferentially attributed to these five chromosomes as a marker

for large-celled amoeboid organisms. *Dictyostelium discoideum* is nominally a soil organism, but the representation of diverse amoeboid organisms in the genome references was not strong, and these results should not be construed to positively indicate an undue degree of soil contamination in estuarine waters or to identify the actual genus or species source in these aquatic samples—relative amoeba read counts were only used as a marker for removal of large-celled organisms. For other runs, we report the sum of the plus and minus reads for all genomes in the reference set. To normalize differences in sequencing depth, total read counts per sample were used to compute reads per million (RPM). For comparison, we ran Kraken/Bracken (Lu et al., 2022) using version 2.0.7-beta of the Kraken database using default parameters, outputting the total reads that were attributable to the groups *Prochlorococcus*, *Synechococcus*, and *Dictyostelium discoideum*. The output is reported in Supplementary Data 1. Read counts using Bracken are generally lower than for BBDMap, likely due to use of a smaller reference database and differences in default parameters.

Results

Sample collection and processing

Available pumps (see Methods and Supplementary Table 1), including eDNA and peristaltic models, did not reliably generate enough power to overcome sampling challenges at height from piers or vessels, so direct surface sampling was preferred over *in-situ* filtration. Biomass accumulation quickly clogged the 0.22 μm inline filter, which inhibited consistent flow and led to eventual removal of 0.22 μm inline filters from our protocol in favor of flask filtration of the mixed 5 μm filtrate in the tertiary filtration step (Figures 2, 3).

Millipore's 47 mm polypropylene nylon in-line filter holders improved flow rates over Cytiva's 47 mm polycarbonate filter holders, reducing macro and micro filtration time by over 50% and reducing sample loss due to inhibitive filtration times and insufficient gasket seals in alternative holders. The 30 primary sampling events were ultimately segregated by filtration and technical replicates into 294 filter subsamples and 424 DNA subsamples, with filters cut and used for replicate DNA extractions. With the final protocol detailed in Figure 3, we were able to produce all four filter fractions in as little as three hours for water with relatively low biomass, though this could be greatly reduced if processing smaller volumes. Filtrations could extend up to 10 hours in some cases when the collection volume was high (up to 60 liters) and/or the sampled substrate contained higher levels of biomass, such as during an algal bloom. Compared to similar methods used for the isolation of picocyanobacteria and cyanophages, such as flow cytometry, eDNA, density-gradient centrifugations, culturing, dilution to extinction, and titer plating, this protocol can be completed the same day as sampling with relatively inexpensive equipment, moderate processing volumes, and can be easily adapted for varying target organisms based on size. Additionally, this protocol can be used with sufficient preparation *in-situ* as well as in a laboratory setting for filtration

processing, and produced multiple filter fraction samples that reduce sequencing noise in downstream applications. The 30 primary sampling events were ultimately segregated by filtration and technical replicates into 294 filter subsamples and 424 DNA subsamples, with filters cut and used for replicate DNA extractions.

Environmental metadata

As expected, we found that water temperatures increased at the entrance of Sequim Bay through summer months, peaking from late July to early August 2024, with lows in February 2025 (Figure 4). The lowest recorded water temperature of the 26 sampling events from PNNL Sequim was 7.7 °C on 2 January, 2024, and the highest recorded temperature was 14.7 °C on 17 July, 2024. The sampling event at Cline Spit had the highest recorded temperature of 19.8 °C on 7 August, 2024. The highest practical salinity value derived from the MCRL dock data was 32.1 on 18 August, 2024 and the lowest was 30.3 on 7 April, 2025. Surface sampling events occurred across a range of tidal events, ranging from -0.2 to 2.2 meters in water depth. Of all 30 sampling events, two were sampled during a slack low tide, three during a slack high tide, 17 during a flood tide, seven during an ebb tide, and one event from the Sequim outdoor mesocosm tank did not have an assigned tidal stage.

DNA yields under various conditions

DNA yields for various method and environmental context comparisons were statistically analyzed by using a Tukey's *post hoc* test ($p < 0.05$) after performing an analysis of variance (ANOVA). Due to the highly limited sample sizes in some comparisons (as little as two in each group), these results can only be considered preliminary. Conclusions should be validated with further studies and greater sample sizes. The results for low sample size comparisons must be interpreted with caution, acknowledging that the impact of outliers or deviations may be amplified, and group variance and differences may not be accurately depicted.

In an exploratory controlled comparison of DNA yields, the Blood & Tissue kit outperformed the PowerWater and PowerSoil kits, yielding approximately 3900, 690, and 2100 ng DNA, respectively (Figure 5). For this controlled comparison, two pico filters (90mm PES) from a single sampling event were cut into thirds, and DNA was extracted from each third using a different kit ($n =$ two extractions per kit). Samples labeled "ab" in Figure 5 denote an overlap between groups a and b, indicating that DNA yields in these samples are not significantly different from either group a or group b, while groups a and b are significantly different from each other, as determined by a Tukey's *post hoc* test ($p < 0.05$) after performing an analysis of variance (ANOVA). However, a 'kitchen sink' comparison (a mix of conditions and replicates, controlling only for filter pore size and location) of each filter fraction across all DNA subsamples collected at PNNL-Sequim (macro, micro, pico, and viral) that also included the Zymo Plant/Seed kit (Figure 6) resulted in highly variable yields under each condition such that there were no significant differences in the micro, pico, and viral fractions, and overlapping significance groups in the macro fraction. Yields in the three direct filter fractions

(macro, micro, and pico) were on average relatively high (~500–1000 ng or more), with a large proportion having acceptable yields for downstream sequencing (50 ng or greater; Figure 6). This supports our approach of running filters to saturation to obtain sufficient material for DNA extraction. Perhaps unsurprisingly, the viral fractions had generally lower DNA yields, averaging 450 ng compared to an average of 1210, 1272, and 924 ng for the pico, micro, and macro fractions, respectively. However, viral DNA yields were often still high enough for sequencing.

A similar kitchen sink comparison (controlling only for filter pore size and location) of MCE versus PES membranes also had high variation in yield. While the MCE membranes yielded slightly more DNA than the PES membranes for the pico fraction (1260 ng average for pico MCE and 1143 ng average for pico PES), the difference was insubstantial compared to the variation among samples (Figure 7a). In a more controlled, exploratory comparison of DNA yields from samples, the MCE membranes in the pico fraction were significantly greater than the PES membranes ($p < 0.05$, 1818 and 1673 ng DNA, respectively), but the differences were small relative to the DNA yield (Figure 7b). Differences between PES and MCE in the viral fraction were also relatively small (324 ng difference in broad comparison, and 54 ng difference in focused comparison) and either were not significantly different or had overlapping significance groups based on a Tukey's *post hoc* test ($p < 0.05$) after performing an ANOVA. Losses of yield when extracting DNA from filters after a long freezer storage period (over 100 days) were small and not significant (Figure 8).

In-line serial filtration of the macro and micro fractions onto nylon and MCE membranes, respectively, using the Millipore Sigma polypropylene in-line filter holders (see Methods and Supplementary Data Tables) was observed to be the fastest method to remove larger-sized organisms, and facilitated large-volume processing in a laboratory while monitoring and changing out 20 μ m or 5 μ m filters if flow lessened as they became clogged. We note that for this reason the replicates for the 5 μ m filters are taken sequentially from the same sample flow through a single 20 μ m filter, and may differ based on differential coating, clogging, or disruption of the 20 μ m filter over that time series (see relative read counts section below). In contrast, the single-flask sterile PES filters were observed to be more time-efficient for pico and viral fractions than in-line filtration on an MCE membrane, reducing processing time per liter from over an hour to a matter of minutes in some cases. The larger (90 mm) filter size also allowed collection of quality technical replicates of DNA extraction and sequencing by cutting PES filters into equal fractions prior to freezing. A list of all yield data for samples used in these analyses can be found in Supplementary Data 3.

Shotgun sequencing and relative read counts assigned to taxon groups

Read mapping results for four sites (PNNL-Sequim, John Wayne Marina, Cline Spit, and Discovery Bay) show that amoeba read counts were dramatically reduced between 3- and 30-fold after passage through 5 μ m to 0.22 μ m filters (Figure 9), with the greatest fold reduction from the highest count 5 μ m filter (John Wayne

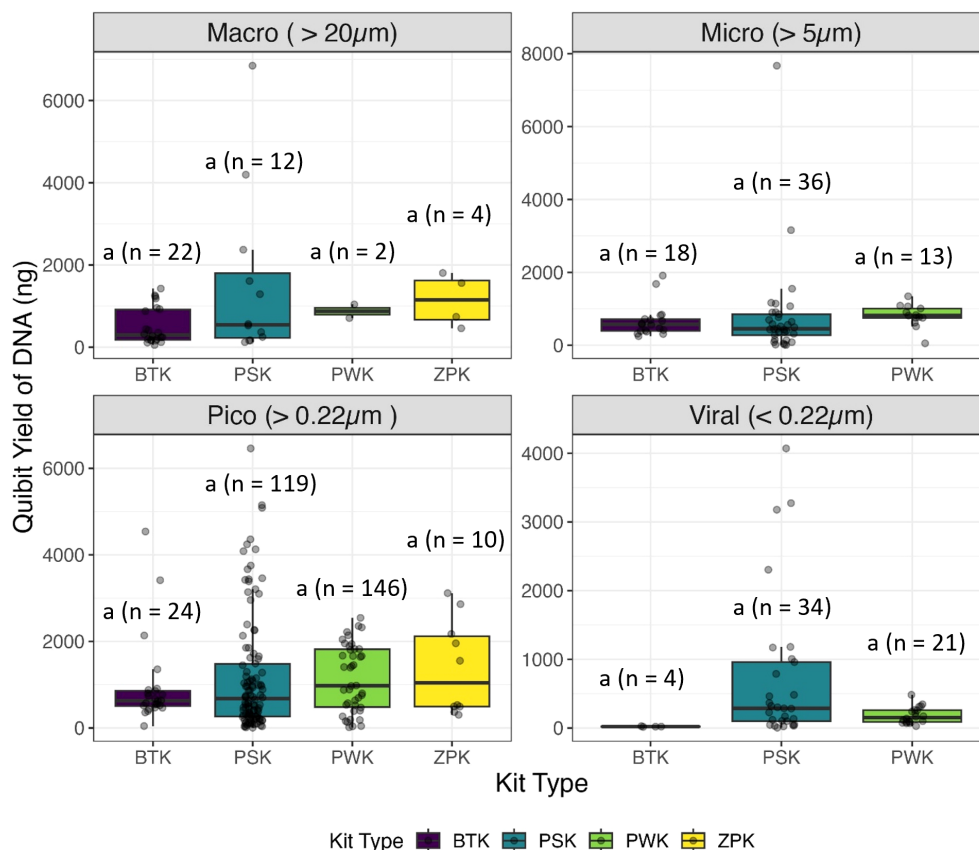


FIGURE 6

Kitchen sink comparison of DNA yields from different extraction kits for all filter fractions. DNA recovered (measured using a Qubit fluorometer) using the Blood & Tissue Kit (BTK, purple), PowerSoil Kit (PSK, blue), PowerWater Kit (PWK, green), or Zymo Plant/Seed (ZPK, yellow) across all PNNL-Sequim sampling events. Filter fractions included in this figure are: macro (top left), micro (top right), pico (bottom left) and viral (bottom right) fractions. All conditions are labelled "a" indicating that there was only one significance group as denoted by a Tukey's *post hoc* test ($p < 0.05$) after performing an ANOVA to determine the extraction kit had a significant impact on yield, and n indicates sample size in each plot. These results include all samples from all PNNL-Sequim sampling events, filter types, and dates over the year. Variable conditions are not taken into account in each test other than extraction method.

Marina) and the least fold reduction from the lowest count 5 µm filter (Cline Spit). In the Cline Spit sequences, there were only 287 reads from the 5 µm filter that matched amoeba sequences. Although this indicates that large-bodied cells are effectively

screened out, relative *Prochlorococcus* reads were not substantially increased or decreased in any of the 0.22 µm filter samples, while relative *Synechococcus* reads were substantially reduced in some 0.22 µm filter samples by up to a third (Figure 9; Supplementary

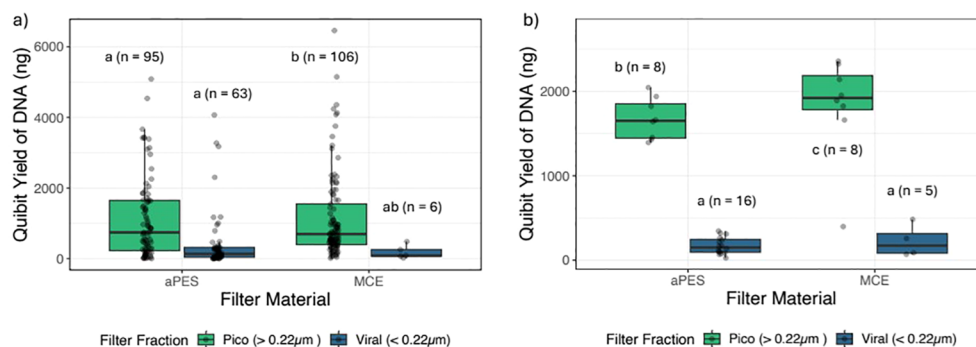


FIGURE 7

DNA yields for different filter types. DNA recovered for pico (green) and viral (blue) filter fractions obtained using PES filters (left in each subfigure), and MCE membrane filters (right in each subfigure). (A) Kitchen sink comparison of all filters collected from PNNL-Sequim or (B) focused comparison of filters collected from a seawater mesocosm tank on 3 December, 2024. Lowercase letters inside graph area indicate significance groups as denoted by a Tukey's *post hoc* test ($p > 0.05$) after performing an ANOVA to determine if material and fraction had a significant impact on yield. Samples labeled "ab" denote an overlap between groups a and b, indicating that DNA yields in these samples are not significantly different from either group a or group b, while groups a and b are significantly different from each other.

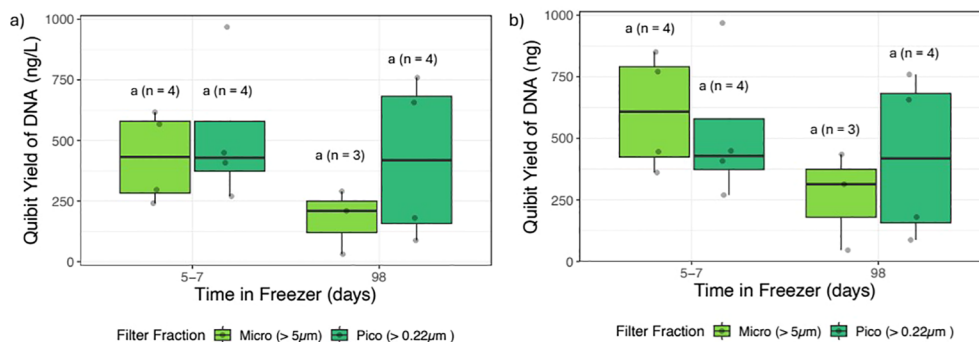


FIGURE 8

DNA yield before and after extended freezer storage. DNA recovered (measured using a Qubit fluorometer), presented as ng/L (a), and ng (b), with extractions taking place either 5–7 days after sampling (left plots within each figure) or 98 days after sampling (right plots within each figure). All data in this plot reflect DNA extracted from filters produced from a single sampling event on 31 July, 2024 extracted using the PowerSoil Kit and otherwise with the same processing methods. All conditions are labelled “a” indicating that there was only one significance group as denoted by a Tukey’s *post hoc* test ($p < 0.05$) after performing an ANOVA to determine if material and fraction had a significant impact on yield, and n indicates sample size in each plot.

Data 1). *Prochlorococcus* normalized read counts were similar across sites and filters (range 718–1300 per 10 M reads), indicating that substantial numbers are trapped by the larger filters and possibly that the differential filtering of larger cells is not enough to increase the relative enrichment of small cells on the subsequent filter; other possibilities are discussed below. The coefficient of variation among *Prochlorococcus* mean-normalized read counts for the Figure 9 technical replicates (replicates of all steps from filtration onwards) on the 0.22 µm filters was also small at 6.8% (7.5%, 8.8%, and 5.0% for data in Figures 9a–c, respectively). The *Synechococcus* read counts were also similar across samples and showed slightly higher reads than those of *Prochlorococcus* with the

exception of Discovery Bay, in which the *Synechococcus* read counts were approximately 10-fold higher than at Cline Spit (Figures 9c, d). The amoeba reads from the 5 µm filters at Cline Spit and Discovery Bay, were about 50%-100% greater than *Synechococcus* reads at PNNL-Sequim and John Wayne Marina. The number of reads mapping to amoeba and *Synechococcus* fluctuate much more than reads mapping to *Prochlorococcus*, but they are not in sync with each other across sites. It is also notable that the 24-mer nonpareil sequence diversity measures across all sites (Figure 9e) were consistently higher and more variable in the 5 µm filter fractions (range ~20.5–22.5) than the 0.22 µm filter fractions (range ~19.5–20). The 5 µm filters are thus at levels that are comparable to marine

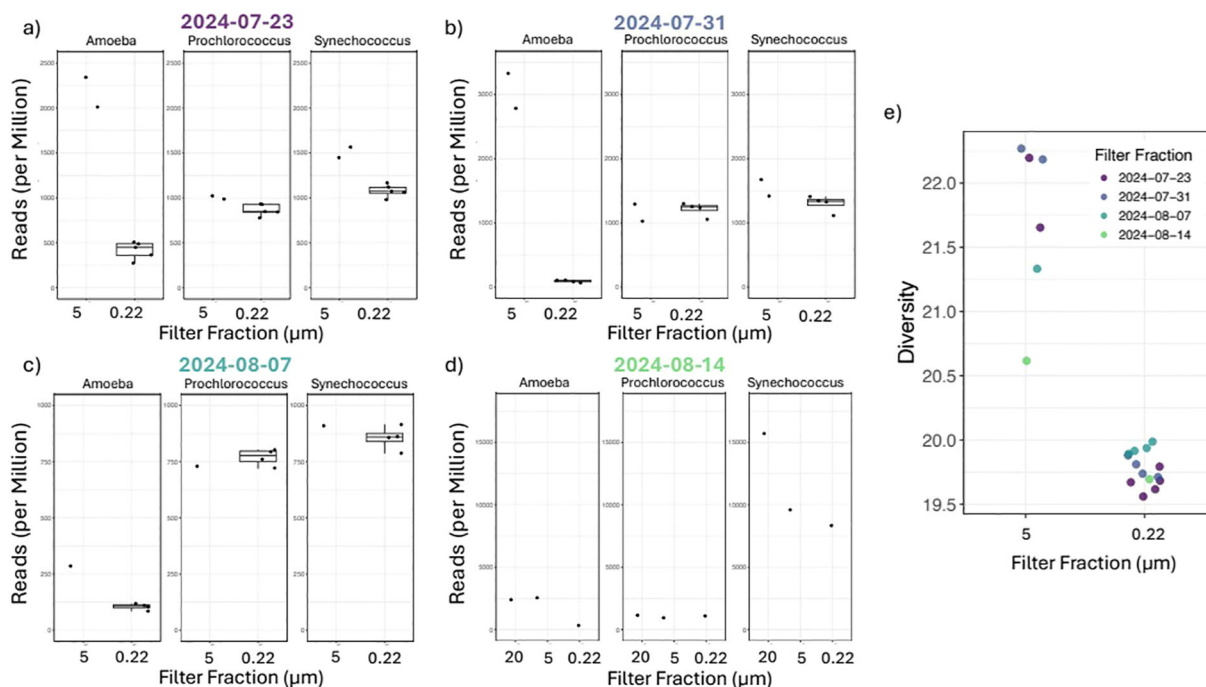


FIGURE 9

Taxon assignments and sequence diversity measures across sampling locations. Sampling occurred in mid-late Summer 2024 at four sites (a) PNNL-Sequim 23 July 2024; (b) John Wayne Marina 31 July 2024; (c) Cline Spit 7 August 2024; and (d) Discovery Bay 14 August 2024. (e) Nonpareil 24-mer diversity measures (log scale) of the same samples shown in a–d, color-coded by site and sampling date as indicated in legend.

and soil samples (21-25) or freshwater and sandy soils (20-22) (Rodriguez-R et al., 2018), while the 0.22 μm diversity is comparatively reduced.

Synechococcus read counts were remarkably similar at ebb and flood tides for both the 5 μm and 0.22 μm filters (Figure 10). In contrast, amoeba reads were about 3x less frequent at flood versus ebb tide on the 5 μm filter and greatly reduced in frequency on the 0.22 μm filters (Figure 10). *Prochlorococcus* reads were also reduced in flood versus ebb tides and were reduced by 50%-60% on the 0.22 μm filter compared to the 5 μm filter. We note that the flood tide samples were extracted considerably later than the ebb stage samples, but time stored in freezer prior to extraction was not expected to change relative frequencies based on our exploratory analysis showing no significant impact of freeze-time on yield (Figure 8).

Relative frequencies of microbial taxa on the 0.22 μm filters over time at PNNL-Sequim, which lasted from mid-April to late October 2024, are shown in Figure 11. These comparisons unavoidably included variation in tidal stage, and also in extraction method and time between collection and extraction, and are across a single year (see Figure 11 legend), so must be interpreted in that context. As in previous figures, the amoeba reads are generally low, indicating relative exclusion by the 5 μm filter. Amoeba and *Prochlorococcus* reads had high relative frequencies in early spring and late fall, indicating that they were relatively more frequent than other organisms in the filter sample). In contrast, *Synechococcus* counts were higher from April to October, with the exception of counts in July. Counts on July 17, 2024 were clustered by processing conditions for all three taxon groups, indicating that the combination of kit and time in freezer prior to extraction can affect results at nearly the same level of variation as change over time (PSK kit and short time in freezer had much higher counts). September 9, 2024 amoeba samples were strongly clustered by condition, with high relative counts for all samples with a short freezer storage duration and extracted with the PWK kit (the *Prochlorococcus* samples were also clustered in the same fashion).

Across all three plots, DNA samples extracted within a few days of filtration had consistently clustered higher counts than samples extracted greater than 50 days after filtration (stored frozen and thawed). This suggests that extended freezer time may contribute to decreased identifiable sequence reads, yet our focused comparisons to assess the impact of freezer storage duration on DNA yield found no significant effect (Figure 8).

Discussion

In this paper, we present and test the efficacy of a sampling approach to enrich taxonomic fractions by size exclusion filtration prior to targeted metagenomic shotgun sequencing. Our prior expectation was that organisms with cell sizes in the given size range would be enriched in the corresponding filter fraction, but that results would vary due to variable filter performance, variable cell-cell and cell-particle attachment frequencies among organisms, cell-free DNA, and horizontal gene transfer, especially involving viral particles. We define our sample fractions as macro (greater than 20 μm), micro (5 to 20 μm), pico (aka small micro, 0.2 to 5 μm), and viral (less than 0.2 μm), and sampling occurred in conjunction with abiotic metadata collection from the environment. Our final methodology balances material recovery, filter type and size, and collection speed. It incorporates a multi-step process: inline serial filtration through a combination of filters with sequentially smaller pore sizes, followed by single-flask filtration to capture various target organism size classes, followed by precipitation of viral particles in the filtrate from the smallest 0.22 μm filters, which are expected to screen out nearly all intact cells. The final filter types were 20 μm Millipore nylon filters, 5 μm Millipore MCE membrane filters, and 0.22 μm PES filters. By increasing the filter volume processed when there was little visible biological material, we successfully produced ≥ 50 ng of DNA in the majority of extractions, our minimal cutoff for sequencing

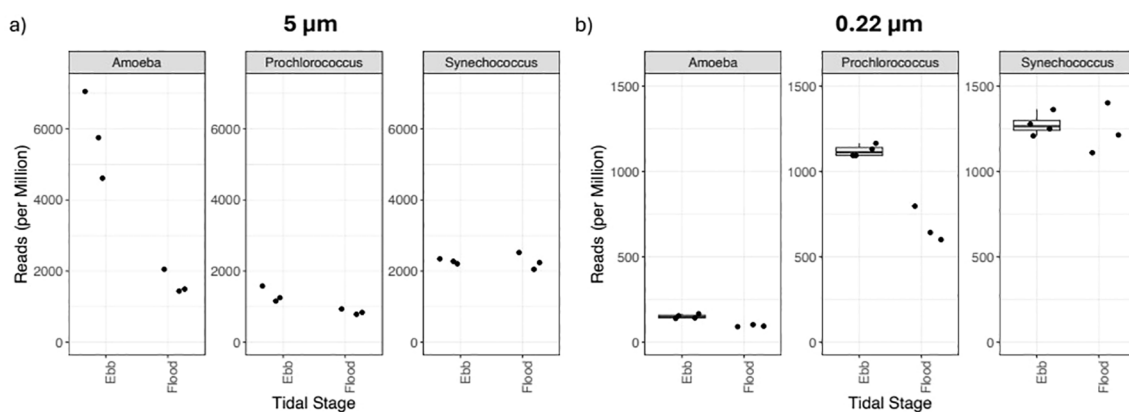


FIGURE 10

Taxon assignments of same-day samples collected at PNNL-Sequim at Ebb and Flood tides. All samples were collected on 31 July 2024 at dawn or midday (ebb and flood tides, respectively), filtered with 5 and 0.22 μm MCE membrane filters and extracted with Qiagen PowerSoil kits. For readability, the magnitudes of the scales differ between the two filter pore sizes, with 5.0 μm scale max 8000 reads per million and 0.22 μm scale max 1500 reads per million. All conditions with multiple points are technical replicates of the process from sample collection and filtration through DNA extraction and sequencing process.

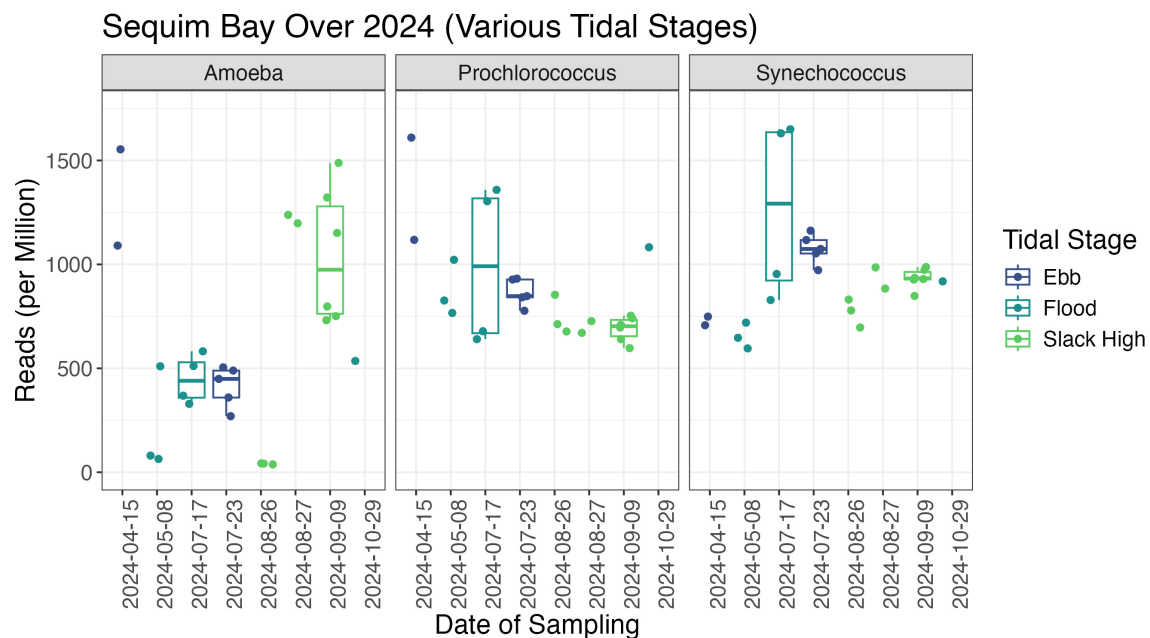


FIGURE 11

Diversity reads (per million) over six months in 2024 for all pico fraction samples from the PNNL-Sequim floating dock. PNNL-Sequim dock samples included samples from multiple sampling events, methods, kits, and dates over the year. 8 distinct sampling events on 8 different days are captured in this figure: April 15 (n = 2 DNA samples extracted with PSK 3 days after filtration), May 8 (n = 3 DNA samples extracted using PSK 5 days after filtration), July 17 (n = 4 DNA samples; 2 extracted using PSK 1 day after filtration, and 2 extracted using PWK 124 days after filtration), July 23 (n = 5 samples extracted using PSK 13 days after filtration), July 31 (n = 4 DNA samples extracted using PSK 5 days after filtration), August 26 (n = 1 DNA sample extracted using PWK 9 days after filtration), September 9 (n = 6 DNA samples; 3 extracted using PWK 2 days after sampling, and 3 extracted using PSK 58 days after filtration), and October 29, 2024 (n = 1 DNA sample extracted using PSK 8 days after sampling).

submission. This required extensive filtration processing time of up to 3 hours for a single filter in some cases with higher biomass levels. All kit and filter types tested yielded sufficient DNA quantity and quality, and the more critical factors were processing time, cost, and material availability. While this protocol is advantageous for time-efficiency, material and equipment cost, and versatility in target organisms, other established methods such as eDNA sampling, culturing, and dilution to extinction may be better suited or used in conjunction with this protocol for applications that require axenic isolates, small samples, long-term cultivated strains, and/or in-depth sequencing is not required.

The sequencing results presented in this paper were focused on exploratory (many conditions had small numbers or no replicates) statistical analysis to demonstrate detection of the picocyanobacteria genera *Prochlorococcus* and *Synechococcus*, the times of year they were detectable, whether we could enrich the number of reads for these genera in the pico fractions, the technical repeatability of the results, and whether we could detect enrichment for sequences homologous to viruses that might be associated with picocyanobacteria in viral precipitate fractions. The clearest and perhaps most important sequencing result was that both *Prochlorococcus* and *Synechococcus* were detected in the Salish Sea samples, usually but not always with similar counts between them. *Synechococcus* was generally more common, especially in Discovery Bay which had dramatically higher *Synechococcus* counts per million. Detection of picocyanobacteria sequences is important because extraction of high-quality DNA from picocyanobacteria remains challenging primarily due to relatively low biomass collected in competitive and low-nutrient (oligotrophic) marine water (Wagner et al., 2024; Hept and Greene, 2023).

Complicated by their potential inability to contribute proportionally as much to biomass in the more competitive and nutrient-rich estuarine waters of the Salish Sea. The relatively constant ratio of sequences from the two genera across sites and over the year is also important because of possible temperature, nutrient level, and virus interaction effects that at this latitude could have strongly affected their relative distributions (Carlson et al., 2022; Coello-Camba and Agustí, 2024; Zhang et al., 2022; Zufia et al., 2022). The validation of picocyanobacterial relative type abundance was essential because while some studies determined that picocyanobacteria are found in all regions of the euphotic zone, there is little research to validate these findings in the northern Pacific Ocean and specifically in the Salish Sea (Babić et al., 2017). It is also notable, and important to establish, that the technical replicability (from filtering onwards) of the relative *Prochlorococcus* counts on the 0.22 μm filters was high (Figure 9 and throughout), allowing preliminary confidence in results differing by 10% or higher.

We note that although one viral sample had anomalously low counts, the top 5 viral samples have only 3.5x higher relative read counts (average 3417) than the remainder of the samples (average 882), indicating that a substantial amount of viral DNA is retained with the large-sized fractions. This may be due to particles or free DNA sticking to the filters, or to incorporation/infection or attachment to the cells that are stuck on those filters. We also note that phages derived from *Synechococcus* were the best matches to slightly over 50% of the cyanophage reads, although we caution against over-interpreting this result in terms of what types of organisms are infected with cyanophage, both because phages can often infect multiple species, and because the reference dataset

contains many more *Synechococcus*-derived phages than cyanophages isolated from other organisms. Nevertheless, it is interesting that some of the highest read counts observed for *Synechococcus*-derived phages (2698, 2696, and 1240 respectively on the 20, 5 and 0.22 μm filters from August 14, 2024 at Discovery Bay) were from samples with some of the highest *Synechococcus* genome read counts (respectively 15,251, 9653, and 7989).

Salinity and water temperature followed similar trends to each other, with slight increases during summer months and lows during cooler seasons. The slight changes in salinity across seasons are likely due to the heavy influence of rivers and precipitation in the estuary system, with increased outflow and precipitation in colder months that decrease salinity (Walker et al., 2022). Cline Spit's increased water temperature compared to PNNL-Sequim may be influenced by its location within Dungeness Bay, which is protected by a long spit that reduces water volume exchange and creates shallower sites. There are also large variances in the distance from the shore between sampling sites, which may inform differences between samples (e.g. sampling less than one meter from shore at Cline Spit, and greater than 50 meters from shore near Port Townsend).

The trends in relative frequencies of *Prochlorococcus*, *Synechococcus*, and amoeba-like sequences over the course of the year at PNNL Sequim (Figure 11) were somewhat surprising. Although the numbers were small, relative numbers of *Prochlorococcus* were highest towards the beginning of the year and decreased throughout the year, while *Synechococcus* numbers were lowest early in the year and then relatively constant. One might have thought that *Prochlorococcus*, which dominates at tropical and subtropical altitudes, would do comparatively better in the warmer months. Indeed, the highest counts were expected during the late spring and summer months (Zinser et al., 2006), so this may indicate that the estuarine dynamics of an inland sea are more unpredictable than expected. This trend was visible despite differences in tidal stage at time of sampling (which also impact relative counts, Figure 10). The trends in relative amoeba-like read counts (we remind readers that with only one amoeba sequence included, the phylogenetic definition of these sequences is poor) did not correspond to either of the picocyanobacteria, being most common in the earliest and latest months of the year sampled. Further definition of these trends will await results from planned sampling and sequencing across 2025 and 2026, and will particularly benefit from more detailed analysis of the relative frequency of sub-lineages within these broad taxonomic groups (i.e. at the species or ecotype level or below).

The relative read counts of various taxa under different filtration conditions were also unexpected. The amoeba-like sequences were more frequent on the 5 μm filters than the 0.22 μm filters, and combined with the relatively high nonpareil diversity on the 5 μm filters, this indicated that the size exclusion filtration was effective. However, *Prochlorococcus* and *Synechococcus* relative read counts were not comparatively higher on the 0.22 μm filters, and were fairly frequent in the viral fraction, which passed through the 0.22 μm filters. Given these results, we may have been overly conservative in using 5 μm filters (rather than, say, 2 μm filters) to screen out large-

bodied cells, but it also appears that considerable amounts of picocyanobacterial DNA both accumulated on the larger filters and passed through the smaller filters. Although free DNA and inadequacies of the filtration process need to be considered (e.g., high variability in pore size, micro breaks in filters, DNA or cell binding to filters as they pass through), it is possible that there are alternative explanations. Tycheponsons, which are numerically common DNA-containing particles from *Prochlorococcus*, are emitted into the environment as particles similar to viruses (Hackl et al., 2023). These and other undiscovered elements in the taxa under study may strongly impact the filtration properties of DNA-containing objects. Higher-level conglomerates of cells or cell-particle mixes are now also recognized as important features of both *Prochlorococcus* and *Synechococcus*, which would contribute to their retention on larger pore-size filters. Furthermore, viruses are commonly retained with infected cells, and due to gene transfer mechanisms both viruses and host targets can contain each other's genes. These sequencing features will also strongly benefit from more detailed analysis of the relative frequency of sub-lineages within these broad taxonomic groups, as well as detailed analysis of genomic regions that may be under- or over-sampled in different filter partitions.

Technical challenges

The extraction of high-quality DNA from picocyanobacteria remains challenging, likely due to relatively low biomass collected in oligotrophic marine water, and potentially due to environmental contaminants, cell lysis inefficiencies due to cyanobacterial extracellular polymeric substances, and polysaccharide production (Wagner et al., 2024). Additionally, the consistently low biomass found at sampling sites—excluding Cline Spit, which had much higher biological material than the other sampling sites—necessitated high filtration volumes and extensive processing time.

The low sample size for many compared groups was a challenge to use for analyses and must be interpreted with caution. While ANOVA requires a minimum of one sample above the number of groups and can technically be performed under these conditions, it is not considered a sufficient sample size for statistical conclusions. We recognize that these results are only exploratory analyses to guide further research with larger sample sizes.

For our preliminary estimates, we used limited high-level estimates of sequence composition, and small numbers of technical replicates from identical samples (same time, location, and water sample with same filtration protocol). We also had few replicates of biological/environmental effects such as tidal state, time of year, or sampling location. Due to continuous sampling methodology adaptations throughout year one, there was high variation in sample production, which significantly decreased the number of quality technical replicates for focused comparisons. Especially for the non-controlled comparisons, results should be validated with larger sample sets and using consistent methods; however, the results presented here provide a valuable understanding of the impact that methodological variation can have.

Conclusions

To address a gap of relatively rapid methods for isolating multiple target microorganisms from aqueous environments, we explored methods for consistent size-fractionated sampling and metadata collection. While this work was completed to support host-virus interaction studies influenced by environmental factors and to provide shareable, high-quality metadata for diverse applications, this work may also be useful in environmental DNA pre-filtration, ecosystem health monitoring, phytoplankton carbon cycling, harmful algal bloom detection, water quality assessments, metagenomics, single-cell genomics, aquaculture, algal anti-viral research, and fisheries management. By analyzing community composition, species, and ecotypes at the community level, as well as variations in host and phage genes, protein structures, and interactomes at the molecular level, this approach may inform phage-host co-evolution and viral-host interactions while reducing sequencing noise. This refined sampling and metadata production method for downstream analyses may also be useful in environmental DNA pre-filtration, ecosystem health monitoring, phytoplankton carbon cycling, harmful algal bloom detection, water quality assessments, metagenomics, single-cell genomics, aquaculture, algal anti-viral research, and fisheries management.

Using this refined sampling and metadata production method for downstream analyses, we were able to successfully produce sufficient DNA from multiple filter fractions to conduct preliminary metagenomic analyses in conjunction with environmental context over approximately a year. The processing time to produce four filter fractions for up to 60 liters of raw sample was reduced to as little as three hours in some cases, though processing time did increase with samples containing relatively high levels of biomass. While purifying enough DNA from marine aqueous environments for metagenomic applications is challenging, we were able to produce sufficient yields in all filter fractions, averaging 450, 1210, 1272, and 924 ng for the viral, pico, micro, and macro fractions. These results allowed us to conduct preliminary metagenomic analyses to investigate picocyanobacterial and viral dynamics in the Salish Sea.

In future work, more data should be collected using the final sampling scheme to generate larger sample sizes for more robust comparisons. This will produce valuable, standardized biological and physiochemical datasets to feed in-depth research applications such as evolutionary modeling of virus-host interactions. In addition, varying filter sizes may be helpful in refining the isolation and retention of target species in each fraction.

Finally, we considered additional experimental approaches to obtain more data on genetic aspects of the metagenomic samples, including RNA extraction and sequencing, individual cell and/or phage isolation and amplification in culture, and single cell sequencing, but concluded that these were not within our current budgetary constraints. These approaches would likely require further testing and modification of the protocol used to collect samples. Single cell sequencing or cell monoculture sequencing would likely have a tradeoff of fewer organisms sampled (leading to lower genetic diversity in sequencing results) for much better linkage information along the genome, while RNA sequencing in

particular would reflect functional activity rather than just the presence of organisms measured in DNA sequencing. In future experiments with sufficient funding, adding these or similar types of approaches would be highly beneficial to better understand host-pathogen relationships, the ultimate aim of our research project.

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

NB: Validation, Data curation, Resources, Visualization, Conceptualization, Methodology, Writing – review & editing, Supervision, Investigation, Formal analysis, Writing – original draft. OL: Conceptualization, Writing – review & editing, Methodology, Data curation, Supervision. KJ: Data curation, Methodology, Investigation, Writing – review & editing. MP-F: Investigation, Methodology, Writing – original draft, Data curation. DL: Software, Writing – review & editing, Visualization, Formal analysis. SC-T: Data curation, Methodology, Visualization, Formal analysis, Investigation, Writing – original draft. PR: Formal analysis, Methodology, Visualization, Data curation, Writing – review & editing, Writing – original draft, Software. RW: Data curation, Formal analysis, Writing – original draft. CP: Writing – review & editing, Software, Visualization. MC: Project administration, Funding acquisition, Writing – review & editing, Supervision, Resources. CJ: Visualization, Writing – review & editing, Investigation, Project administration, Supervision, Formal analysis, Writing – original draft, Data curation, Software. SE: Writing – original draft, Resources, Writing – review & editing, Supervision, Methodology, Conceptualization. DP: Project administration, Conceptualization, Resources, Writing – original draft, Formal analysis, Funding acquisition, Supervision, Writing – review & editing, Investigation.

Funding

The author(s) declared that financial support was received for this work and/or its publication. The research described in this paper is supported by the NW-BRaVE for Biopreparedness project funded by the U. S. Department of Energy (DOE), Office of Science, Office of Biological and Environmental Research, under FWP 81832. A portion of this research was performed on a project award (Enhancing biopreparedness through a model system to understand the molecular mechanisms that lead to pathogenesis and disease transmission) from the Environmental Molecular Sciences Laboratory, a DOE Office of Science User Facility sponsored by the Biological and Environmental Research

program. Pacific Northwest National Laboratory is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RL01830. This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Advanced Scientific Computing Research programs in the DOE Office of Science under Contract No. DE-AC05-00OR22725.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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.

References

- Adams, C. I. M., Knapp, M., Gemmell, N. J., Jeunen, G. J., Bunce, M., Lamare, M. D., et al. (2019). Beyond biodiversity: can environmental DNA (eDNA) cut it as a population genetics tool? *Genes* 10, 192. doi: 10.3390/genes10030192
- Babić, I., Petrić, I., Bosak, S., Sunčica, M., Mišanović, H., Radić, I., Dupčić, and Ljubešić, Z. (2017). "Distribution and Diversity of Marine Picocyanobacteria Community: Targeting of *Prochlorococcus* Ecotypes in Winter Conditions (Southern Adriatic Sea)," in *Marine Genomics, Navigating the Future: Cross-sector Marine Genomics*, (Amsterdam, Netherlands: Elsevier), vol. 36, 3–11. doi: 10.1016/j.margen.2017.05.014
- Babraham Institute (2023). *Babraham Bioinformatics - FastQC A Quality Control Tool for High Throughput Sequence Data*. Available online at: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (Accessed November 5, 2025).
- Buchholz, H. H., Michelsen, M. L., Bolaños, L. M., Browne, E., Allen, M. J., and Temperton, B. (2021). Efficient Dilution-to-extinction isolation of novel virus–host model systems for fastidious heterotrophic Bacteria. *ISME J.* 15, 1585–1598. doi: 10.1038/s41396-020-00872-z
- Bushnell, B. (2014). "BBMap: A Fast, Accurate, Splice-Aware Aligner," in *LBNL-7065E*. (Berkeley, California, United States: LBL Publications). Available online at: <https://escholarship.org/uc/item/1h3515gn> (Accessed November 20, 2025).
- Bushnell, B., Rood, J., and Singer, E. (2017). BBMerge - accurate paired shotgun read merging via overlap. *PloS One* 12, e0185056. doi: 10.1371/journal.pone.0185056
- Callieri, C., Cabello-Yeves, P. J., and Bertoni, F. (2022). The 'Dark side' of picocyanobacteria: life as we do not know it (Yet). *Microorganisms* 10, 546. doi: 10.3390/microorganisms10030546
- Carlson, C. J., Albery, G. F., Merow, C., Trisos, C. H., Zipfel, C. M., Eskew, E. A., et al. (2022). Climate change increases cross-species viral transmission risk. *Nature* 607, 555–562. doi: 10.1038/s41586-022-04788-w
- Coello-Camba, A., and Agustí, S. (2024). Seawater temperature drives the diversity of key cyanobacteria (*Synechococcus* and *Prochlorococcus*) in a warming sea. *Front. Mar. Sci.* 11. doi: 10.3389/fmars.2024.1456799
- Del Bel Belluz, J., Jackson, J. M., Kellogg, C. T.E., Peña, M.A., Giesbrecht, I. J.W., and Hobson, L. A. (2024). Phytoplankton community composition links to environmental drivers across a fjord to shelf gradient on the central coast of british columbia. *Front. Mar. Sci.* 11. doi: 10.3389/fmars.2024.1458677
- Dorigatti, I., Gaythorpe, K. A.M., Cox, V. M., Windram, F. A., and Cator, L. (2025). Priorities for modelling arbovirus transmission under climate change. *Trends Mol. Med.* 31, 885–894. doi: 10.1016/j.molmed.2025.05.010

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmars.2026.1769457/full#supplementary-material>

SUPPLEMENTARY TABLE 1

Peristaltic and Vacuum Pumps used in sampling method iteration.

SUPPLEMENTARY TABLE 2

Filter types tested for serial filtration optimization.

SUPPLEMENTARY TABLE 3

In-line filter holders tested for serial filtration optimization.

SUPPLEMENTARY TABLE 4

DNA Extraction kits used to test filter subsample DNA extraction optimization.

SUPPLEMENTARY TABLE 5

Environmental metadata: acquisition methods distinguished by data acquired from MCRL Data at PNNL-Sequim or other locations.

SUPPLEMENTARY FIGURE 1

TapeStation Data: DNA fragment analysis of 6 pico fraction samples and 1 viral fraction sample from parent sample NW_FE_024, sampled from Sequim Bay during a slack tide in September of 2024.

SUPPLEMENTARY FIGURE 2

Read alignment comparison for different cyanobacteria databases. The plots show good agreement between reads mapped using the NCBI-derived references used in the main manuscript and *Prochlorococcus* and *Synechococcus* references derived from the Cyanorak database. Each data point represents all sequences from a metagenomic sample. Red dashed line is the line $y=x$.

SUPPLEMENTARY FIGURE 3

Reads aligned to macroscopic reference strains across sample sequences. A range of reference strains demonstrate similar qualitative filtering behaviors across the sampling events, where the x-axis represents the sample sequences and y-axis the map read counts (per million). Amoeba (Chromosomes) and several marine diatoms common in the Salish Sea were chosen as a macroscopic reference for filter comparisons. Specific diatom strains considered were *Skeletonema costatum* strain CNS00243 (GCA_046619385.1), *Chaetoceros tenuissimus* strain NIES-3715 (GCA_021927905.1), *Chaetoceros muelleri* strain CCMP1316 (GCA_019693545.1), *Thalassiosira oceanica* strain CCMP1005 (GCA_019693575.1), *Thalassiosira pacifica* strain AJA261-08 (GCA_036939875.1).

- EcoKMER GitHub Software Repository (2025). Available online at: <https://github.com/NWBRaVE/EcoKMER> (Accessed October 21, 2025).
- Edmundson, S., Huesemann, M., Cady, S., Kuo, L.-J., Anderson, B., and Reynolds, D. (2020). *VITAL-viral inhibitors from ALgae: generating extracts for antiviral activity assays*. (Alexandria, Virginia, United States: National Technical Information Service), 1776864. doi: 10.2172/1776864
- Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincon, J., Zabala, L. L., Jiao, N., et al. (2013). Present and future global distributions of the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc. Natl. Acad. Sci. Unit States America* 110, 9824–9829. doi: 10.1073/pnas.1307701110
- Flombaum, P., Wang, W.-L., Primeau, F. W., and Martiny, A. C. (2020). Global picophytoplankton niche partitioning predicts overall positive response to ocean warming. *Nat. Geosci.* 13, 116–120. doi: 10.1038/s41561-019-0524-2
- Fuchsman, C. A., and Hays, M. D. (2023). Increased cyanophage infection at the bottom of the euphotic zone, especially in the fall. *Environ. Microbiol.* 25, 3349–3363. doi: 10.1111/1462-2920.16525
- Garczarek, L., Guyet, U., Doré, H., Farrant, G. K., Hoebeke, M., Brillet-Guéguen, L., et al. (2021). Cyanorak v2.1: A scalable information system dedicated to the visualization and expert curation of marine and brackish picocyanobacteria genomes. *Nucleic Acids Res.* 49, D667–D676. doi: 10.1093/nar/gkaa958
- George, A., Leach, D., Philips, C., Brown, J., Duncan, J. M., Nedved, B., et al. (2025). EcoKMER: one-stop shop for spatio-temporal metagenomic exploration using dataFed. *bioRxiv*. doi: 10.64898/2025.12.15.694490
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUASt: quality assessment tool for genome assemblies. *Bioinf. (Oxf Engl)* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086
- Hackl, T., Laurenceau, R., Ankenbrand, M. J., Bliem, C., Cariani, Z., Thomas, E., et al. (2023). Novel integrative elements and genomic plasticity in ocean ecosystems. *Cell* 186, 47–62.e16. doi: 10.1016/j.cell.2022.12.006
- Hept, M. A., and Greene, L. H. (2023). Extracting high-molecular weight DNA from cyanobacteria using promega's wizard® HMW DNA extraction kit with a modified protocol, METIS. *MethodsX* 11, 102341. doi: 10.1016/j.mex.2023.102341
- Hill, T., cre., and Anisfeld, S. (2021). *VulnToolkit: analysis of tidal datasets*. Available online at: <https://cran.r-project.org/web/packages/VulnToolkit/index.html> (Accessed November 1, 2025).
- Hshieh, K., and Maloney, J. G. (2025). Sequencing on the edge: tools and techniques for sequencing low biomass isolates from unicellular eukaryotic parasites and other challenging samples. *J. Microbiol. Methods* 237, 107233. doi: 10.1016/j.jmimet.2025.107233
- Huang, Z., Fang, F., Ding, L., Yu, K., Zhang, L., and Lu, H. (2023). Technological advancements in field investigations of marine microorganisms: from sampling strategies to molecular analyses. *J. Mar. Sci. Eng.* 11, 1981. doi: 10.3390/jmse11101981
- 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
- Karamati, N., Elnaz, L., Weese, J.S., McCarthy, D. T., and Murphy, H. M. (2024). Passive sampling of microbes in various water sources: A systematic review. *Water Res.* 266, 122284. doi: 10.1016/j.watres.2024.122284
- Kearney, S. M., Coe, A., Castro, K. G., and Chisholm, S. W. (2022). Filter plating method for rendering picocyanobacteria cultures free of heterotrophic bacterial contaminants and clonal. *Front. Microbiol.* 13. doi: 10.3389/fmicb.2022.821803
- Kent, A. G., Baer, S. E., Mouginot, C., Huang, J., Larkin, A., Lomas, M. W., et al. (2019). Parallel phylogeography of *Prochlorococcus* and *Synechococcus*. *ISME J.* 13, 430–441. doi: 10.1038/s41396-018-0287-6
- Lawrence, J. E., and Steward, G. F. (2010). "Purification of viruses by centrifugation," in *Manual of aquatic viral ecology*. Eds. S. Wilhelm, M. Weinbauer and C. Suttle (Waco, Texas, United States: American Society of Limnology and Oceanography). doi: 10.4319/mave.2010.978-0-9845591-0-7.166
- 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. *Bioinf. (Oxf Engl)* 31, 1674–1676. doi: 10.1093/bioinformatics/btv033
- Liu, R., Liu, Y., Chen, Y., Zhan, Y., and Zeng, Q. (2019). Cyanobacterial viruses exhibit diurnal rhythms during infection. *Proc. Natl. Acad. Sci. Unit States America* 116, 14077–14082. doi: 10.1073/pnas.1819689116
- Lu, J., Rincon, N., Wood, D. E., Breitwieser, F. P., Pockrandt, C., Langmead, B., et al. (2022). Metagenome analysis using the kraken software suite. *Nat. Protoc.* 17, 2815–2839. doi: 10.1038/s41596-022-00738-y
- Marine and Coastal Research Laboratory (2025). "Outdoor Experimental Tanks at PNNL-Sequim," in *Pacific northwest national laboratory*. (Richland Washington, United States: Pacific Northwest National Laboratory). Available online at: <https://www.pnnl.gov/outdoor-experimental-tanks-pnnl-sequim> (Accessed December 1, 2025).
- MCRL Data (2025). *Marine and coastal research laboratory data*. Available online at: <https://mcrldata.pnnl.gov/> (Accessed December 1, 2025).
- MCRLdata Sandbox (2025a). *MCRLdata-sandbox CTD*. Available online at: https://github.com/MCRLdata-Sandbox/data_prep/blob/main/data/outputs/L1/250815_ctd_water_temp_salinity_L1.csv (Accessed December 1, 2025).
- MCRLdata Sandbox (2025b). *MCRLdata-sandbox tide gauge*. Available online at: https://github.com/MCRLdata-Sandbox/data_prep/blob/main/data/outputs/L1/250515_tidegauge_L1.csv (Accessed December 1, 2025).
- Mehdizadeh Allaf, M., and Peerhossaini, H. (2022). Cyanobacteria: model microorganisms and beyond. *Microorganisms* 10, 696. doi: 10.3390/microorganisms10040696
- Mella-Flores, D. M., Six, C., Ratin, M., Partensky, F., Boutte, C., Le Corguillé, G., et al. (2012). *Prochlorococcus* and *Synechococcus* have evolved different adaptive mechanisms to cope with light and UV stress. *Front. Microbiol.* 3. doi: 10.3389/fmicb.2012.00285
- Mioduchowska, M., Pawłowska, J., Mazanowski, K., and Weydmann-Zwolicka, A. (2023). Contrasting marine microbial communities of the fram strait with the first confirmed record of cyanobacteria *Prochlorococcus marinus* in the arctic region. *Biology* 12, 1246. doi: 10.3390/biology12091246
- National Microbiome Data Collaborative (2026). *The NMDC metadata standards documentation* (Github. Inc: National Microbiome Data Collaborative). Available online at: https://microbiomedata.github.io/nmdc-schema/Metadata_Documentation_Overview/ (Accessed October 28, 2025).
- NOAA Tides and Currents (2025). *Port Angeles, WA - Station ID: 9444090, National Oceanic and Atmospheric Administration*. Available online at: <https://tidesandcurrents.noaa.gov/stationhome.html?id=9444090> (Accessed August 2025).
- Reynolds, D., Huesemann, M., Edmundson, S., Sims, A., Hurst, B., Cady, S., et al. (2021). Viral inhibitors derived from macroalgae, microalgae, and cyanobacteria: A review of antiviral potential throughout pathogenesis. *Algal Res.* 57, 102331. doi: 10.1016/j.algal.2021.102331
- Ribalet, François, Dutkiewicz, S., Monier, E., and Armbrust, E.V. (2025). Future ocean warming may cause large reductions in *Prochlorococcus* biomass and productivity. *Nat. Microbiol.* 10, 2441–2453. doi: 10.1038/s41564-025-02106-4
- Rodriguez-R, L. M., Gunturu, S., Tiedje, J. M., Cole, J. R., and Konstantinidis, K. T. (2018). Nonpareil 3: fast estimation of metagenomic coverage and sequence diversity. *mSystems* 3, e00039–e00018. doi: 10.1128/mSystems.00039-18
- Rong, C., Zhou, K., Li, S., Xiao, K., Xu, Y., Zhang, R., et al. (2022). Isolation and characterization of a novel cyanophage encoding multiple auxiliary metabolic genes. *Viruses* 14, 887. doi: 10.3390/v14050887
- Rozum, J. C., Sineath, W., Bohutskyi, P., Quenneville, J., Kim, D. N., Johnson, C., et al. (2025). Synergy and antagonism in a genome-scale model of metabolic hijacking by bacteriophages. *bioRxiv*. doi: 10.1101/2024.12.11.628001
- Ruber, J., Geist, J., Hartmann, M., Millard, A. D., Raeder, U., Zubkov, M. V., et al. (2018). Spatio-temporal distribution pattern of the picocyanobacterium *Synechococcus* in lakes of different trophic states: A comparison of flow cytometry and sequencing approaches. *Hydrobiologia* 811, 77–92. doi: 10.1007/s10750-017-3368-z
- Sahu, A., Singh, M., Amin, A., Malik, M. M., Qadiri, S. S. N., Abubakr, A., et al. (2025). A systematic review on environmental DNA (eDNA) science: an eco-friendly survey method for conservation and restoration of fragile ecosystems. *Ecol. Indic.* 173, 113441. doi: 10.1016/j.ecolind.2025.113441
- Sánchez-Gallego, J., Curtis, N. P., Paerl, H. W., and Paerl, R. W. (2025). New perspectives on picocyanobacteria and understudied cyanobacterial diversity in the alabarle pamlico sound system, north carolina, USA. *Front. Microbiol.* 16. doi: 10.3389/fmicb.2025.1539050
- Scanlan, D. J., Ostrowski, M., Mazard, S., Dufresne, A., Garczarek, L., Hess, W. R., et al. (2009). Ecological genomics of marine picocyanobacteria. *Microbiol. Mol. Biol. Rev.* 73, 249–299. doi: 10.1128/MMBR.00035-08
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Sharifian Gh., M., and Norouzi, F. (2023). Guidelines for an optimized differential centrifugation of cells. *Biochem. Biophys. Rep.* 36, 101585. doi: 10.1016/j.jbbrep.2023.101585
- Sobocinski, K. (2021). *State of the salish sea*. (Bellingham, Washington. United States: Salish Sea Institute at Western CEDAR, Western Washington University). doi: 10.25710/vfbb-3a69
- Visintini, N., Martiny, A. C., and Flombaum, P. (2021). *Prochlorococcus*, *Synechococcus*, and picocaryotic phytoplankton abundances in the global ocean. *Limnol. Oceanogr. Lett.* 6, 207–215. doi: 10.1002/lol2.10188
- Wagner, A. D., Ahmed, M. M.A., Starks, V., and Boudreau, P. D. (2024). Using repeated lysis steps fractionates between heterotrophic and cyanobacterial DNA extracted from xenic cyanobacterial cultures. *bioRxiv*. doi: 10.1101/2024.08.22.609136
- Walker, S., Mozaria-Luna, H. N., Kaplan, I., and Petatán-Ramírez, D. (2022). Future temperature and salinity in puget sound, washington state, under CMIP6

- climate change scenarios. *J. Water Climate Change* 13, 4255–4272. doi: 10.2166/wcc.2022.282
- 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
- Wright, M. N., and Ziegler, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Softw* 77, 1–17. doi: 10.18637/jss.v077.i01
- Zehr, J. P. (2010). Microbes in earth's aqueous environments. *Front. Microbiol.* 1. doi: 10.3389/fmicb.2010.00004
- Zhang, J., Li, F., Long, L., and Huang, S. (2022). Quantification of marine picocyanobacteria on water column particles and in sediments using real-time PCR reveals their role in carbon export. *mSphere* 7, e00499–e00422. doi: 10.1128/msphere.00499-22
- Zinser, E. R., Coe, A., Johnson, Z. I., Martiny, A. C., Fuller, N. J., Scanlan, D. J., et al. (2006). *Prochlorococcus* ecotype abundances in the north atlantic ocean as revealed by an improved quantitative PCR method. *Appl. Environ. Microbiol.* 72, 723–732. doi: 10.1128/AEM.72.1.723-732.2006
- Zufia, J. A., Legrand, C., and Farnelid, H. (2022). Seasonal dynamics in picocyanobacterial abundance and clade composition at coastal and offshore stations in the baltic sea. *Sci. Rep.* 12, 14330. doi: 10.1038/s41598-022-18454-8.