



OPEN

DATA DESCRIPTOR

A high-resolution diel survey of surface ocean metagenomes, metatranscriptomes, and transfer RNA transcripts

Sarah J. Tucker^{1,2,3,17} , Jessika Füssel^{4,17}, Kelle C. Freel^{5,17}, Evan Kiefl^{6,7}, Evan B. Freel⁵, Oscar Ramfelt⁵, Clarisse E. S. Sullivan^{8,9}, Andrian P. Gajigan¹⁰, Hanako Mochimaru^{5,11}, Mariana Rocha de Souza⁵, Mariko Quinn⁵, Ciara Ratum¹², Leon L. Tran⁵, Marek Sobczyk¹³, Samuel E. Miller¹, Florian Trigodet^{2,3}, Karen Lolans⁶, Hilary G. Morrison¹, Bailey Fallon¹, Bruno Huettel¹⁴ , Tao Pan¹³, Michael S. Rappé^{5,18} & A. Murat Eren^{1,2,3,15,16,18}

The roles of marine microbes in ecosystem processes are inherently linked to their ability to sense, respond, and ultimately adapt to environmental change. Capturing the nuances of this perpetual dialogue and its long-term implications requires insight into the subtle drivers of microbial responses to environmental change that are most accessible at the shortest scales of time. Here, we present a multi-omics dataset comprising surface ocean metagenomes, metatranscriptomes, tRNA transcripts, and biogeochemical measurements, collected every 1.5 hours for 48 hours at two stations within coastal and adjacent offshore waters of the tropical Pacific Ocean. We expect that this integrated dataset of multiple sequence types and environmental parameters will facilitate novel insights into microbial ecology, microbial physiology, and ocean biogeochemistry and help investigate the different mechanisms of adaptation that drive microbial responses to environmental change.

Background & Summary

Microbial communities experience environmental fluctuations over a broad range of time-scales, and their responses are evident at multiple levels – from changes in community composition to the physiological reactions of individual cells¹. Time-series studies of marine microbes have largely focused on time-scales of weeks, months, and seasons^{2–8}, with a particular emphasis on changes to microbial community composition. While

¹Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA, 02543, USA. ²Helmholtz Institute for Functional Marine Biodiversity, 26129, Oldenburg, Germany. ³Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, 27570, Bremerhaven, Germany. ⁴Marine Geochemistry, Institute for Chemistry and Biology of the Marine Environment (ICBM), Carl von Ossietzky University, Oldenburg, 26129, Germany. ⁵Hawai'i Institute of Marine Biology, School of Ocean and Earth Science and Technology, University of Hawai'i at Mānoa, Kāne'ohe, HI, 96744, USA. ⁶Department of Medicine, The University of Chicago, Chicago, IL, 60637, USA. ⁷Graduate Program in Biophysical Sciences, University of Chicago, Chicago, IL, 60637, USA. ⁸University of Hawai'i Foundation, Honolulu, HI, 96822, USA. ⁹School of Ocean and Earth Science and Technology (SOEST), University of Hawai'i at Mānoa, Honolulu, HI, 96822, USA. ¹⁰Department of Oceanography, University of Hawai'i at Mānoa, Honolulu, HI, 96822, USA. ¹¹National Institute of Advanced Industrial Science and Technology (AIST), Ibaraki, Japan. ¹²Hawai'i Pacific University, Kāne'ohe, HI, 96744, USA. ¹³Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL, 60637, USA. ¹⁴Max Planck Genome-centre Cologne, Max Planck Institute for Plant Breeding Research, Carl-von-Linne-Weg 10, 50829, Cologne, Germany. ¹⁵Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, 26129, Oldenburg, Germany. ¹⁶Max Planck Institute for Marine Microbiology, 28359, Bremen, Germany. ¹⁷These authors contributed equally: Sarah J. Tucker, Jessika Füssel, Kelle C. Freel. ¹⁸These authors jointly supervised this work: Michael S. Rappé, A. Murat Eren. ✉e-mail: rappe@hawaii.edu; meren@hifmb.de

these studies have provided critical insights into marine microbial biogeography, diversity, and ecology, the physiological shifts of individual microbial populations, especially at short timescales, are less well understood.

Microbes can rapidly respond to environmental change via transcriptional and translational regulation, often without any immediate impact on community composition^{9,10}. These changes enable rapid functional adaptations that modulate microbial metabolism¹¹, influence microbial contributions to ocean biogeochemical cycles¹², shape biogeographic distributions¹³, and affect long-term ecological and evolutionary trajectories¹⁴. Yet, relatively few studies have sampled gene expression and community composition with high enough temporal resolution to track the reactions of microbial populations to environmental fluctuations and the activities of other community members^{9,15–17}.

Here, we present a multi-omics dataset from our recent Hawai'i Diel Sampling (HaDS) survey, which represents a high-resolution sampling of surface ocean microbial communities every 1.5 hours for a two-day period across two environments: Kāne'ohe Bay on the windward coast of O'ahu, Hawai'i, and the adjacent offshore (Fig. 1a,b). The goal of HaDS is to characterize microbial responses to diel changes in ocean biogeochemistry across physically connected, yet environmentally distinct habitats. We designed HaDS to focus on sampling stations previously established within the Kāne'ohe Bay Time-series (KByT)^{18,19}, so that the diel data can be further contextualized with insights from monthly observations.

Altogether, HaDS generated 202 'omics sequencing products totalling 6.62 billion paired-end short reads and 27.63 million single-end long reads that capture multi-dimensional, high-resolution observations of near-shore and offshore microbial communities (Table S1). At both the coastal Kāne'ohe Bay station (HP1) and the adjacent offshore station (STO1), we sampled at 33 time points across 48 hours, and subsequently produced 59 metatranscriptomes, 65 short-read metagenomes, 8 long-read metagenomes, and 66 transfer RNA (tRNA) transcript libraries (Table 1), as well as biogeochemical measures. To the best of our knowledge, HaDS includes the first high-throughput sequencing of transfer RNA transcripts from marine microbial communities, offering a unique dataset for analysis. We also generated four deeply-sequenced short-read metagenomes from samples collected in the late fall and spring prior to HaDS through routine KByT sampling (Fig. 1c, Table 1).

In addition to the high-quality, raw sequencing data and contextual measures of the environment, we provide station-specific co-assemblies based on both the short-read and long-read metagenomes that were collected during the diel sampling. We reconstructed a total of 160 metagenome-assembled genomes (MAGs) from the short-read metagenomes, including 63 high-quality (>90% completeness, <5% contamination) and 97 medium-quality (>70% completeness, <10% contamination) MAGs (Table S1). Co-assemblies of long-read metagenomes per-station yielded 73 contigs over 1 Mbp, 14 high-quality MAGs, and 16 medium-quality MAGs. By providing the reconstructed MAGs, co-assemblies for short- and long-read metagenomes, and associated workflows in addition to the raw data, we aim to reduce the computational burden associated with this analysis step and ultimately increase the accessibility of these datasets for the broader research community. However, future analyses should consider multiple assembly approaches including hybrid assembly²⁰ and eukaryotic-specific assemblers²¹, as we expect that these analyses may capture additional populations potentially missed in our co-assemblies.

In summary, the Hawai'i Diel Sampling effort provides novel and cutting-edge multi-omics data alongside complementary environmental parameters, metagenomic assembly data products, and reproducible bioinformatic workflows that enhance the utility and value of these unique resources.

Methods

Hawai'i Diel Sampling (HaDS) collection and processing. Seawater collection took place between August 18, 2021 9:00 AM and August 20, 2021 9:00 AM Hawai'i Standard Time (HST), from a depth of ~0.25 m at station HP1 within coastal Kāne'ohe Bay, O'ahu, Hawai'i, and station STO1 in the adjacent offshore waters. Our periodic sampling every 1.5 hours over a 48 hour period yielded 33 sampling events at each site. For each time point, we collected seawater samples for flow cytometry, fluorometric chlorophyll *a* concentrations, and nucleic acids. Additionally, we recorded *in situ* measurements of seawater temperature, pH, and salinity with a YSI 6600 or YSI EcoSense EC300 (YSI Incorporated, Yellow Springs, OH, USA) and *in situ* measurements of Photosynthetic Photon Flux Fluence Rate (PPFRR) with a LI-193 Underwater Spherical Quantum Sensor and LI-250A Light Meter (LI-COR Environmental, Lincoln, NE, USA).

Following prefiltration of ~10–20 L of seawater through 85-μm Nitex mesh, we filtered ~4 (HP1) or ~9 (STO1) L of seawater through 0.2-μm pore-sized polyethersulfone (PES) Sterivex cartridge filters (MilliporeSigma, Burlington, MA, USA) to collect microbial cells for nucleic acids isolation. We immediately added RNAlater (Invitrogen, Waltham, MA, United States) to the Sterivex filters and stored the samples at –80 °C until further processing.

We collected seawater subsamples (125 mL) on 25-mm diameter GF/F glass microfiber filters (Whatman, GE Healthcare Life Sciences, Chicago, IL, USA) and stored the filters in aluminum foil at –80 °C until fluorometric analysis of chlorophyll *a* concentrations. We extracted chlorophyll *a* in 100% acetone and determined the concentration with a Turner 10-AU fluorometer (Turner Designs, Sunnyvale, CA, USA) according to established techniques²². We preserved seawater for cellular enumeration in 2 mL aliquots with a final concentration of 0.95% (v:v) paraformaldehyde (Electron Microscopy Services, Hatfield, PA, USA) and stored the samples at –80 °C until analysis. The SOEST Flow Cytometry Facility performed flow cytometric cellular enumeration of non-cyanobacterial (presumably heterotrophic) bacteria and archaea (reported here as heterotrophic bacteria), photosynthetic picoeukaryotes, *Synechococcus*, and *Prochlorococcus* on a Beckman Coulter CytoFLEX S, following the method of Monger & Landry²³. The SOEST Laboratory for Analytical Biogeochemistry (S-LAB) measured concentrations of dissolved inorganic nutrients on a Seal Analytical AA3 HR Nutrient Autoanalyzer (detection limits: NO₂– + NO₃–, 0.009 μM; SiO₄ 0.09 μM; PO₄^{3–} 0.009 μM; NH₄ 0.03 μM) and dissolved organic carbon and total nitrogen on a Shimadzu High-Temperature TOC-L Combustion Analyzer.

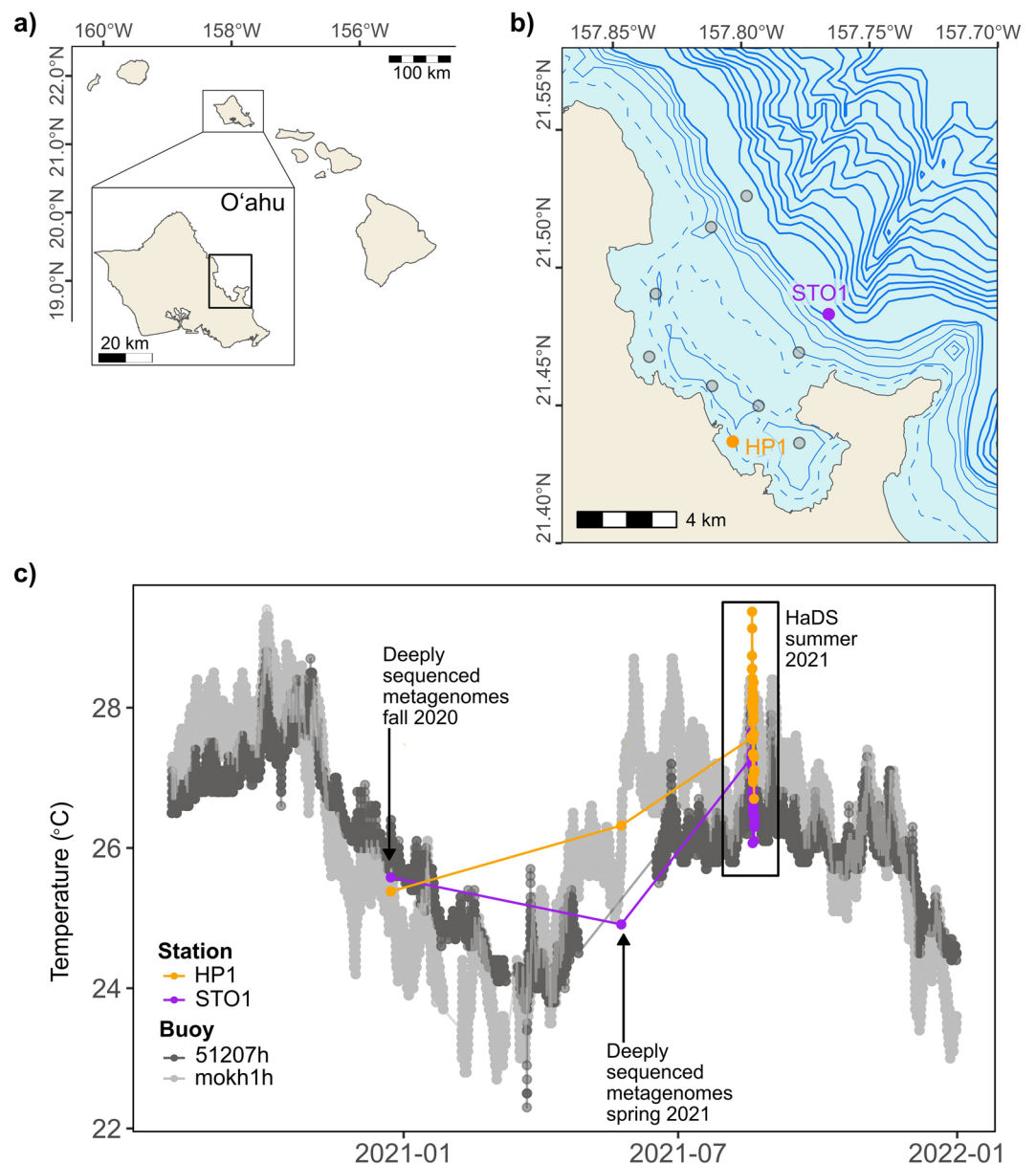


Fig. 1 Sampling regime of HaDS. (a) Map showing the location of Kāne'ohe Bay on the windward side of the island of O'ahu, Hawai'i. (b) We sampled stations HP1 (coastal Kāne'ohe Bay) and STO1 (adjacent offshore) during HaDS. These stations are part of the monthly sampling of the Kāne'ohe Bay Time-series (KByT), which also includes 8 additional KByT sampling stations shown in gray. Dashed bathymetric contours mark a depth of 5 m. Solid bathymetric contours indicate depth increases of 10 m up to 50 m, with thicker lines marking 50 m depth intervals. (c) We conducted HaDS sampling during a period of elevated summer temperatures. Two additional sampling events contributed four deeply sequenced metagenomes that contextualize the summer diel sampling. Temperature data collected during the sampling events and from buoys located in coastal Kāne'ohe Bay (mokh1h; National Buoy Data Center) and the adjacent offshore (51207h; National Buoy Data Center) show differences in surface water temperature between coastal and offshore stations as well as seasonal variation throughout the data collection period.

We used various monitoring platforms to collect additional environmental data. We accessed surface ocean temperature data covering the period from 2020–2021 from the NOAA station mokh1 in Kāne'ohe Bay and surface ocean temperatures and wave height from the NOAA station 51207 in the adjacent offshore waters. We accessed rainfall and wind speed from the NOAA Moku o Lo'e Weather station and tidal height from the NOAA/NOS/CO-OPS station 1612480, both of which are positioned on the island of Moku o Lo'e in southern Kāne'ohe Bay, along with the Hawai'i Institute of Marine Biology. We report tidal height as the mean lower low water, or the average of the lower low water height of each tidal day examined over the National Tidal Datum Epoch. We downloaded these data from their respective websites between November 20, 2024 and February 26, 2025.

Project	Station	Sampling dates	Short-read metagenomes	Long-read metagenomes	Metatranscriptomes	tRNA transcripts
HaDS	HP1	8/18/2021-8/20/2021	32	4	29	33
HaDS	STO1	8/18/2021-8/20/2021	33	4	30	33
KByT	HP1	5/24/2021	1	—	—	—
KByT	STO1	5/24/2021	1	—	—	—
KByT	HP1	12/23/2020	1	—	—	—
KByT	STO1	12/23/2020	1	—	—	—

Table 1. Summary of multi-omics datasets collected as part of the Hawai'i Diel Sampling (HaDS) and the Kāne'ohe Bay Time-series (KByT).

DNA extraction, preparation, and sequencing for short-read metagenomics. We adapted a previously published protocol²⁴ to extract DNA from Sterivex cartridge filters for short-read metagenomic sequencing. First, we collected any cells suspended in the RNAlater contained in the Sterivex filter cartridges by attaching a 10 mL air-filled clean syringe to the Sterivex cartridge and pushing the liquid into a nuclease-free 1 mL centrifuge vial. After centrifugation for 15 min at 4 °C at 13,000 rpm, we discarded the supernatant and kept the cell pellet on ice until further processing. We then opened the cartridge holding the filter with a sterilized PVC pipe cutter and removed the cylinder with the PES membrane. Using a sterilized scalpel, we cut along the barrel at the top and bottom of the filter membrane and down one side to remove the filter from the barrel. We cut the filter into ~6–8 small pieces in a sterile petri dish and transferred the pieces to a Lysing Matrix E vial (MP Biomedicals, Irvine, CA, USA) and added 450 µL Buffer A, 210 µL 20% SDS and 500 µL phenol:chloroform:IAA 125:24:1 pH 8 (50 mL Buffer A: 2 mL 5 M NaCl, 2 mL 0.5 M EDTA, 46 mL dH₂O, all solutions were molecular grade and nuclease-free). We resuspended the cell pellet obtained from RNAlater in 50 µL Buffer A by vortexing vigorously for 30 sec followed by a quick centrifugation step in a microcentrifuge and transferred the solution to the Lysing Matrix E vial. We lysed the cells with a Bead Ruptor Elite (OMNI International, Kennesaw, GA, USA) on setting 6 for 30 sec followed by two min on ice to avoid overheating of the samples and another lysing cycle with the same settings. We centrifuged the vials for 3 min at 4 °C at 13,000 rpm and removed the upper aqueous phase to a fresh, nuclease-free 1.5 mL centrifuge vial. Following the addition of 500 µL phenol:chloroform:IAA, we repeated the centrifugation step. Following transfer of the aqueous phase to a new centrifuge vial, we added 3 M sodium acetate at 1/10th of the transferred volume and mixed the solution by inversion. We added 500 µL ice-cold pure ethanol and placed the samples in the freezer at –20 °C overnight.

The next day, we thawed the samples on ice, centrifuged them at 13,000 rpm at 4 °C for 20 min, and carefully decanted the supernatant to retain the nucleic acid pellet. We washed the pellet with 500 µL 70% ethanol at 4 °C and vortexed shortly followed by centrifugation at 13,000 rpm at 4 °C for 10 min. We decanted the liquid carefully, dried the lip of the tube on a clean tissue, and air-dried the pellet with the centrifuge tubes open and upside down on a clean tissue for 15 min. We added 100 µL of nuclease-free dH₂O to the pellet and vortexed the pellet for 20 min with a Vortex-Genie 2 (Scientific Industries, Bohemia, NY, USA) and a Vortex adapter 24 (Qiagen, Hilden, Germany). To remove any impurities and chemical contaminations, we purified the DNA with the DNA Clean & Concentrator-100 kit (Zymo Research, Irvine, CA, USA) and eluted in a final volume of 30 µL.

Following shearing of DNA to 400 bp on a Covaris S220 focused-ultrasonicator system, we cleaned and concentrated the DNA with AMPure XP beads (Beckman Coulter, Brea, CA, USA) and prepared metagenomic sequencing libraries using the Ovation Ultralow System V2 with indexed sequencing adapters (Tecan Genomics, Männedorf, Switzerland). We pooled final libraries at equimolar amounts and size-selected the library using a BluePippin 1.5% agarose gel cassette (Sage Science, Beverly, MA, USA) targeting the 468–572 bp region. We paired-end sequenced the final pool on a NextSeq 500/550 High Output v2.5 using the 300 cycle kit (Illumina, San Diego, CA, USA). The Keck Sequencing Facility at the Marine Biological Laboratory in Woods Hole, MA, USA carried out the short-read metagenomic library preparation and sequencing.

Nucleic acid extraction, preparation, and sequencing for long-read metagenomics and metatranscriptomics. To extract nucleic acids from the Sterivex cartridge filters, we first removed the RNAlater by attaching a 10 mL air-filled clean syringe to the Sterivex cartridge to remove and discard the liquid. We then opened the cartridge holding the filter with a sterilized PVC pipe cutter and removed the cylinder with the PES membrane. Using a sterilized scalpel, we cut along the barrel at the top and bottom of the filter membrane and down one side to remove the filter from the barrel. We cut the filter into ~6–8 small pieces in a sterile petri dish and transferred them to 2 mL BashingBead lysis tubes with 0.1 & 0.5 mm beads (Zymo Research Europe, Germany) for subsequent cell disruption with a Qiagen TissueLyser (3 min. at 25 Hz). Next, we isolated DNA and total RNA from the same starting material using the ZymoBIOMICS DNA/RNA Miniprep kit (Zymo Research Europe, Germany), with an additional DNase I treatment of the extracted RNA as recommended by the vendor.

DNA directly served as input for a barcoded DNA library preparation without prior fragmentation. We prepared HiFi SMRTbell libraries (PacBio, Menlo Park, CA, USA) for long-read DNA sequencing following manufacturer protocols using the Ultra Low DNA Input approach and sequenced the libraries on a PacBio Revio device. For demultiplexing, adapter trimming, and deduplication of the sequencing data we relied on the PacBio SMRTlink software suite.

We prepared RNAseq libraries with the Universal Prokaryotic RNA-Seq Prokaryotic Any Deplete kit by Tecan Genomics, including the proprietary adapter addition step of the kit and dual 8-mer barcode indexing. We sequenced the libraries on an Illumina NextSeq2000 platform in 150 bp paired-end read mode. The MPI

Cologne Genomics Center conducted nucleic acid extraction, library preparation, and sequencing for long-read metagenomics and metatranscriptomics.

RNA extraction and preparation for transfer RNA sequencing. To maximize recovery of total RNA from the Sterivex filter cartridges, we performed pre-processing steps that enhance bacterial cell lysis. For enzymatic cell lysis, we prepared the following solutions: 1) marine microbe lysis solution: 40 mM EDTA, 50 mM TRIS and 0.75 M sucrose final concentration, 2) lysozyme stock solution: 100 mg/mL in TE buffer pH 8, RNase free, and 3) a final lysis solution of 670 μ L per sample, consisting of 600 μ L marine microbe lysis solution, 67 μ L lysozyme stock solution, and 0.2 μ L SUPERase (Invitrogen™, Thermo Fisher Scientific, Waltham, MA, USA, 20 U/ μ L).

For each Sterivex cartridge, we collected and pelleted cells that were suspended in RNAlater following the same steps described above. We resuspended each pellet in 70 μ L of lysis solution. We removed the filter from the Sterivex cartridges using the same protocol described above, cut the filter into small pieces using a clean scalpel, and transferred the pieces to a 14 mL culture tube containing 600 μ L of pre-prepared lysis solution. We added the resuspended cell lysate to each 14 mL tube that contained the corresponding filter pieces and vortexed vigorously for 2 min. We then incubated the tubes in a 37 °C water bath for 45–60 min and gently agitated the samples every 10 min.

After the initial incubation, we added 7.4 μ L of proteinase K (800 U/mL, New England Biolabs, Ipswich, MA, USA) and 74 μ L of 10% SDS for a final concentration of 1% to each sample and incubated at 37 °C for an additional 2 h, gently agitating every 10 min. We stored the lysed samples at –80 °C until further processing.

After thawing the pre-processed RNA samples on ice, we transferred the lysate to Lysing Matrix C tubes (MP Biomedicals, Santa Ana, CA, USA) that contained 1 mm silica beads. For a subset of samples (tRNA sample IDs: 1–12), we used Biosphere® SC Micro Tubes (Oakland, CA, USA) with 100 mg of 1 mm silica lysing beads for bead beating. We added 400 μ L of 0.3 M NaOAc/HOAc, 10 mM EDTA, pH 4.8 and an equal volume of acetate-saturated phenol/chloroform and vortexed briefly. We then placed the samples in a reciprocating bead beater for 2 min at maximum intensity, pausing after 1 min to avoid overheating the samples, and centrifuged at 13,000 \times g for 15 min at 4 °C.

We transferred the aqueous phase to a fresh Lysing Matrix C tube and repeated the phenol/chloroform extraction steps. To precipitate the RNA, we added 1 volume of 100% isopropanol, vortexed each sample thoroughly, and incubated them overnight at –20 °C. After incubation we centrifuged the sample at 13,000 \times g for 15 min at 4 °C, discarded the supernatant, and washed each pellet with 1 mL of chilled 75% ethanol. Following centrifugation for 5 min at 13,000 \times g at 4 °C and removal of the ethanol, we let the RNA pellet air dry for 10 min at room temperature to remove residual ethanol. We resuspended each pellet in 30 μ L of an acid-elution buffer (10 mM NaOAc/1 mM EDTA, pH 4.8).

We cleaned the RNA using the Zymo Research Oligo Clean & Concentrate and eluted the RNA in RNase-free water. Finally, we added an equivalent amount of acid-elution buffer to each extract to reach a final concentration of 5 mM NaOAc/0.05 mM EDTA, pH 4.8. All samples yielded at least 200 ng of RNA except for tRNA sample 11.

Sequencing of tRNA transcripts. We used approximately 200 ng of total RNA from each sample to build tRNA sequencing libraries following a previously published multiplex small RNA-seq (MSR-seq) protocol²⁵. Following incubation at 37 °C in a 33 mM sodium tetraborate buffer, pH 9.5 for 30 min to deacetylate RNAs, we added 5 μ L of a polynucleotide kinase (PNK, New England Biolaboratories (NEB), USA) reaction stock (4 U/ μ L T4 PNK, 40 mM MgCl₂, 200 mM Tris-HCl, pH 6.8) and incubated at 37 °C for 20 min to repair the 3' end. We inactivated the PNK by incubating the samples at 65 °C for 10 min. We then added 30 μ L of an RNA ligation reaction mix at a final concentration of 15% polyethylene glycol (PEG) 8000 (NEB, USA), 1 \times T4 RNA ligase I buffer (NEB, USA), 50 μ M ATP, 5% DMSO, 1 mM hexaammine cobalt (III) chloride, and 1 U/ μ L T4 RNA ligase I (NEB, USA) and incubated the samples overnight at 16 °C. This mix also contained the barcoded RNA ligation linker/RT primer oligo at a 1.2:1 molar ratio to the input RNA to allow multiplexing of samples. After ligating the RNA to the barcoded and biotinylated hairpin oligonucleotide overnight, we added streptavidin-coated MyOne C1 dynabeads (ThermoFisher Scientific, USA) and incubated at room temperature on rotation for 15 min. This immobilized RNA molecules on the dynabeads in order to minimize sample loss and enabled rapid washes and buffer exchanges between reactions. To liberate the 3' end of the reverse transcriptase (RT) primer by dephosphorylation, we added a phosphatase reaction mix (final concentration: 0.2 U/ μ L Calf Intestinal Phosphatase (CIP, NEB, USA), 10 mM MgCl₂, 0.5 mM ZnCl₂, 20 mM HEPES, pH 7.5) and incubated the samples for 30 min at 37 °C. To facilitate reverse transcription of the RNA molecules, we resuspended the samples in 25 μ L of 1 \times SuperScript IV VILLO (ThermoFisher Scientific, USA) mix and incubated at 55 °C for 10 min and then at 37 °C overnight. The following day, we added an RNase H reaction mix at a final concentration of 0.4 U/ μ L RNase H in 1 \times RNase H buffer (NEB, USA) for 15 min at 37 °C to digest the RNA molecules. To oxidize any non-extended reverse-transcriptase primer, we added a periodate solution at a final concentration of 50 mM sodium periodate in 15 mM sodium acetate, pH 5.0, and incubated the samples for 30 min at room temperature. We then ligated the barcoded cDNA ligation oligo to each cDNA molecule by incubating the samples overnight at room temperature in a 50 μ L reaction with 2 U/ μ L T4 RNA ligase I (NEB), 25% PEG 8000, 7.5% DMSO, 50 μ M ATP, 1 mM hexaammine cobalt (III) chloride, 2 mM barcoded cDNA ligation oligo and 1 \times T4 RNA ligase I buffer to enable polymerase chain reaction (PCR) amplification of the cDNA. Finally, we amplified the libraries via PCR with Illumina primers and performed multiplexed small RNA sequencing on an Illumina NovaSeq 6000 with 100 bp paired-end sequencing. We used custom scripts to demultiplex the sequence data.

Kāneʻohe Bay Time-series (KByT) collection and processing. To contextualize the diel data, we deeply sequenced four metagenomes collected at stations HP1 and STO1 on December 23, 2020 and May 24, 2021. These samples were part of the monthly sampling of the Kāneʻohe Bay Time-series and we processed the nucleic acids and other associated biogeochemical parameters following previously published protocols^{18,19}. We followed the same short-read metagenomic library preparation and sequencing protocols as described above.

Sequence data statistics and quality-control metrics. We exported read count and length statistics from the raw 'omics data using 'seqkit stats' command in Seqkit v2.8.1²⁶ with the '-all' flag. To assess the quality of short-read metagenome and metatranscriptome libraries, we first examined the output of FastQC v0.12.1 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). FastQC identified adapter sequences within the metatranscriptomic libraries, thus we used cutadapt v4.9²⁷, with the following arguments to further trim adapter sequences and remove any subsequent reads with a length of less than 100 bp: 'cutadapt -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCA -A AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT -n 5 -o trimmed-R1.fastq -p trimmed-R2.fastq input-R1.fastq input-R2.fastq -m 100'. We then used sortMeRNA v4.3.7²⁸ with the 'paired_out', 'fastx', and 'out2' flags to assess the percentage of reads aligning to ribosomal RNA databases, specifically the SILVA v119²⁹ silva-arc-16s-id95, silva-arc-23s-id98, silva-bac-16s-id90, silva-bac-23s-id98, silva-euk-18s-id95, and silva-euk-28s-id98 databases and Rfam 11.0³⁰ rfam-5.8s-database-id98 and rfam-5s-database-id98. Next, we filtered the reads that did not match to rRNA databases using the illumina-utils library v1.4.1 called 'iu-filter-quality-minoche'³¹, which uses quality filtering parameters previously described³². We then evaluated the proportion of tRNA-sequencing transcripts that had features of tRNAs using 'anvi-trnaseq' from anvi'o v8.0³³. Briefly, sequences were profiled from the 3' end for base-paired stems of the proper length and a group of canonically conserved nucleotides, requiring the presence of a characteristic set of nucleotides that form and extend through the anticodon loop for a read to be identified as having a tRNA origin.

Metagenomic co-assembly reconstruction and profiling. For both short- and long-read metagenomic data from HaDS, we co-assembled the data by grouping samples into those collected at coastal station HP1 and those collected at offshore station STO1 during the diel study. For the short-read metagenomic data, we used a Snakemake³⁴ metagenomic workflow in anvi'o v8.0³³ with the 'anvi-run-workflow' command and the 'metagenomics' flag. Briefly, the workflow quality controlled metagenomic short reads with illumina-utils³¹ and co-assembled metagenomic samples with IDBA-UD v1.1.3³⁵. Subsequently, we created anvi'o contig databases (<https://anvi.org/m/contigs-db>) for each co-assembly with 'anvi-gen-contigs-database', recruited metagenomic reads from all diel short-read metagenomes to the co-assemblies with bowtie2 v2.5.4³⁶, and linked the read recruitment results back to the contigs database using 'anvi-profile'. We further binned the co-assembled data using Metabat2 v2.12.1³⁷, and imported these binned data into anvi'o as a collection with the command 'anvi-import-collection'. To identify the taxonomic identity of the metagenomic bins, we ran 'gtdbtk classify' using the Genome Taxonomy Database Toolkit (GTDB-Tk) v2.4.0³⁸. We used the 'lineage_wf' in CheckM v1.2.3³⁹ to assess the contamination and completion of the binned data and the 'seqkit stats' command in Seqkit v2.8.1²⁶ with the '-all' flag to examine total genome length, %GC, and N50 scores.

We co-assembled long-read metagenomes with Hifiasm v. 0.13-r308⁴⁰ using default parameters. We split the largest contigs (>1 Mbp) from the co-assembly fasta file to individual fasta files to examine their quality and taxonomic identity using the 'seqkit split' command in Seqkit v2.8.1²⁶. For each large contig, we built contig databases in anvi'o and assigned taxonomic identities and examined quality metrics as described above for the short-read metagenomes.

To evaluate the percent of reads mapped to each of the four coassemblies, we used samtools v1.21⁴¹ to extract information from mapping files. We mapped the long-read metagenomes back to the co-assemblies with minimap2 v2.28 (r1209)⁴² using the '-split-prefix' and '-ax map-hifi' flags. We mapped all quality-controlled short-read metagenomes and metatranscriptomes to the co-assemblies with bowtie2 v2.5.4³⁶.

Taxonomic profiling of short-read metagenomes, metatranscriptomes, and tRNA transcripts. Using quality-controlled short-read metatranscriptomes and metagenomes, we profiled the taxonomic composition of the reads using kaiju v1.10.1⁴³ and the nr_euk database, which includes a subset of the NCBI BLAST nr database of proteins belonging to archaea, bacteria, and viruses, in addition to proteins from fungi and microbial eukaryotes. We summarized taxonomic profiles at the class level using the 'kaiju2table' command with the '-r class' flag. For the tRNA-sequencing reads that were identified as tRNA transcripts, we dereplicated profiled tRNA sequences across samples using 'anvi-merge-trnaseq', yielding a set of tRNA seed sequences. Seeds were taxonomically classified by 'anvi-run-trna-taxonomy' via alignment to tRNAs identified by tRNAscan-SE v2.0⁴⁴ in the prokaryotic genomes of the GTDB database (release 95)⁴⁵. Finally, we used 'anvi-tabulate-trnaseq' to output summary tables of tRNA abundances and taxonomy.

Table and figure production. We plotted figures using R v4.4.1⁴⁶ and ggplot2 v3.5.1⁴⁷. We mapped sampling locations of HaDS and KByT using R with 'geom_sf' from ggplot2 based on geospatial data of the main Hawaiian Islands (USGS Digital Line Graphs). To draw contours, we accessed bathymetry data hosted by NOAA using marmap v1.0.10⁴⁸. Figures were edited in Inkscape v1.3.2 (<https://inkscape.org/>).

Data Records

Metadata are available on the Simons Foundation Collaborative Marine Atlas Project (CMAP)⁴⁹ and at the NSF Biological and Chemical Oceanography Data Management Office (BCO-DMO) under Project 947647⁵⁰, in addition to Table S1. All sequencing products can be assessed through BioProject ID PRJNA1201851⁵¹ hosted by NCBI with SRA IDs available in Table S1. High-quality short-read metagenome-assembled genomes (MAGs) are available under BioProject ID PRJNA1235278⁵² with accessions JBMHBP000000000-JBMHBDZ000000000.

The remaining long-read and medium-quality short-read MAGs are available on FigShare⁵³. Short-read and long-read metagenomic coassemblies are also available on FigShare⁵⁴.

Technical Validation

For short-read metagenomic libraries, we evaluated the DNA concentration using the Qubit dsDNA high-sensitivity (HS) assays (Thermo Fisher Scientific) and estimated the purity with a NanoDrop Microvolume Spectrophotometer (Thermo Fisher Scientific) before library preparation. We quantified the short-read metagenomic libraries using real-time PCR using KAPA library quantification reagents (Roche 07960204001) prior to sequencing. For DNA used in long-read metagenome sequencing, we quality-controlled DNA using HS Qubit assays (Thermo, Waltham, USA) and pulse-field capillary electrophoresis (Agilent FEMTOpulse). We evaluated the quality of RNA that served as input for metatranscriptome libraries with HS Qubit assays (Thermo, Waltham, USA) and a PicoChip (Agilent Bioanalyzer) for total RNA. Prior to transfer RNA sequencing library preparation we evaluated the quality and quantity of the RNA with the Qubit RNA HS assay (ThermoFisher Scientific, USA) and a Nanodrop Microvolume Spectrophotometer (ThermoFisher Scientific, USA).

Initial processing of HaDS raw sequences further confirmed the quality of the data. We filtered short-read metagenomes using *illumina-utils*³¹, and retained $94.3 \pm 0.6\%$ of reads ($n = 69$; mean \pm sd; Table S1). We used *cutadapt* v4.9²⁷ and *sortMeRNA* v4.3.7²⁸ to remove reads matching to adapter sequences and ribosomal RNA (rRNA) databases, and then filtered low-quality reads from paired-end sequences³¹. The libraries retained $74.8 \pm 6.7\%$ of the original paired reads and only a small fraction matched to rRNA databases (9.7 ± 7.8 ; $n = 59$; Table S1). These proportions demonstrate effective rRNA removal during library preparation and suggest that a large fraction of reads represent messenger RNA, underscoring their utility for gene expression analyses. We used 'anvi-trna-seq' within *anvi'o* v8.0³³ to identify the proportion of reads that had characteristics of tRNA sequences. On average, tRNA transcripts accounted for $24.1 \pm 14.8\%$ of the merged tRNA-sequencing reads, equating to an average of $1,871,536 \pm 2,931,481$ tRNA transcripts per library ($n = 66$; Table S1). These proportions are in line with previous tRNA-sequencing experiments of the gut microbiome of mice ($22.9 \pm 13.7\%$; $n = 6$)⁵⁵.

Taxonomic profiles of the quality-controlled short-read metagenomic, metatranscriptomic, and tRNA sequences showed communities consistent with the expected composition of coastal and offshore marine microbial communities from the oligotrophic tropical Pacific (Fig. 2; Fig. S1). Reads unclassified at the phylum-level or above were common among all three sequencing types (tRNA libraries: $33.14 \pm 7.26\%$, $n = 66$; metagenomes: $36.61 \pm 3.89\%$, $n = 65$; metatranscriptomes: $66.02 \pm 8.50\%$, $n = 59$). These proportions reflect known limitations in the sensitivity and specificity of short-read taxonomic profiling⁵⁶, particularly in complex and under-characterized microbial communities. *Alphaproteobacteria* and *Cyanophyceae* dominated short-reads classified at the phylum-level or below (Fig. 2, Table S1), as expected for surface ocean marine communities⁵⁷. All three sequencing types captured differences in community composition between the two stations, with *Mamiellophyceae* and *Flavobacteriia* elevated in coastal samples (Fig. 2, Table S1), as previously described^{18,19}. Shifts in community composition between day and night samples were evident in coastal metagenomes, metatranscriptomes, and tRNA transcripts, and discernible, although more subtle, in the offshore sequences (Fig. 2, Fig. S1). The relative abundances of the dominant taxa broadly aligns across sequencing types, suggesting that future integrated data analysis has the potential to reveal drivers of microbial physiologies across time and space.

A high proportion of short-reads mapped to the respective station-specific short-read and long-read co-assemblies, with 67% to 82% of short reads mapped to assemblies on average (Table S1). Similarly, a high proportion of long-read libraries mapped to the respective station-specific short-read and long-read co-assemblies, where 61% to 75% of the total length of the long-read library mapped to the station-specific assembly on average (Table S1). The station-specific short-read and long-read co-assemblies recruited a smaller yet still substantial percentage of the metatranscriptomic reads (33–60% of reads mapped on average; Table S1). To ensure the quality of the MAGs, we examined N50 metrics and completeness and contamination scores (Table S1), and only reported those of high and medium quality, >90% completion and <5% contamination or >70% completion and <10% contamination, respectively. The high-quality long-read MAGs included a circularized contig belonging to *Pelagibacterales* (SAR11), a prevalent marine microbial clade with genomes that are notoriously difficult to reconstruct from short-read metagenomes⁵⁸. While recent work highlighted pervasive instances of misreporting of circular contigs from long-read assemblers⁵⁹, our manual inspection of this genome suggests that it does not suffer from the kinds of issues (such as extensive read-clipping and excessive repeats) previously reported⁵⁹.

Finally, we cross-referenced and contextualized the biogeochemical observations from the HaDS dataset with long-term environmental data from nearby buoys and weather stations (Fig. 1; Fig. S2), as well as published values from the Kāne'ohe Bay Time-series^{18,19}.

Usage Notes

Biogeochemical context of HaDS. The environmental variation sampled across space and time offers valuable opportunities to explore how microorganisms respond to biotic and abiotic factors. Stations HP1 and STO1 differ in their physical characteristics and biogeochemistry^{18,19}. During HaDS, station HP1 was characterized by warmer waters, lower salinity, and elevated phytoplankton biomass and inorganic nutrient concentrations relative to STO1, as expected during the summer months (Fig. 3a; Table S1). These environmental differences align with shifts in the microbial community, such as higher abundances of photosynthetic picoeukaryotes, *Synechococcus*, and heterotrophic bacteria at station HP1 compared to the offshore environment. In contrast, offshore station STO1 is enriched in *Prochlorococcus* cells relative to coastal Kāne'ohe Bay (Fig. 3a).

Over the 48 h sampling period, HP1 experienced temporal patterns that likely reflect a diel signal (Fig. 3b, c). Although station STO1 exhibited subtle diel patterns, the dominant temporal trend emerged over several days where, throughout the 48-hour diel sampling, chlorophyll *a* concentrations declined while *Prochlorococcus* cell numbers increased (Fig. 3d,e). The dominant temporal pattern at STO1 likely reflects

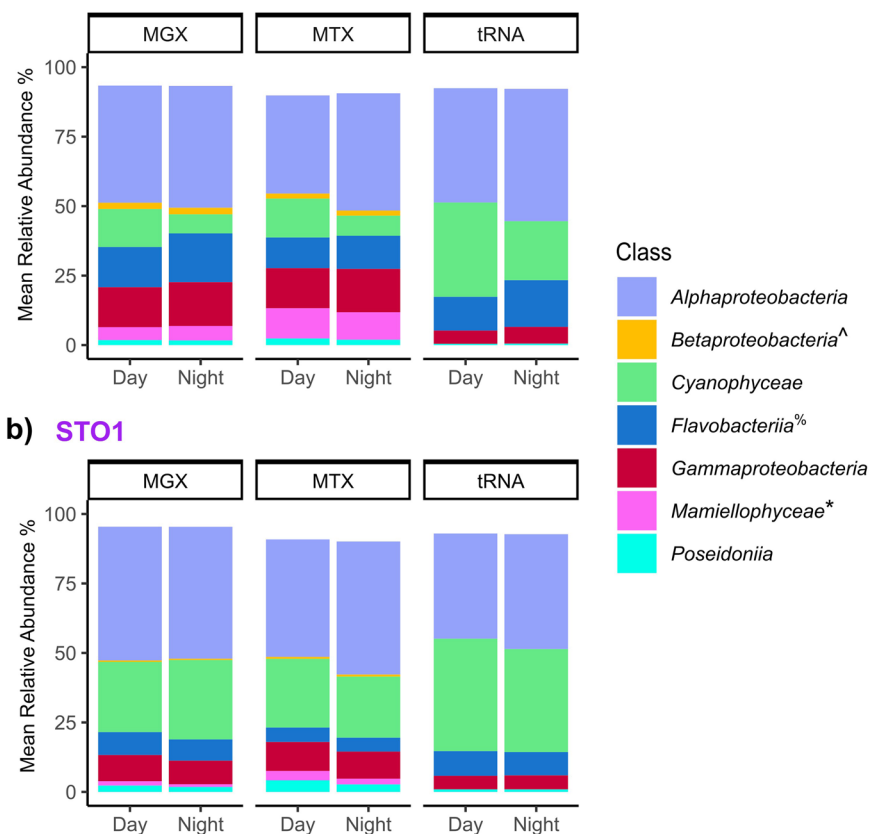
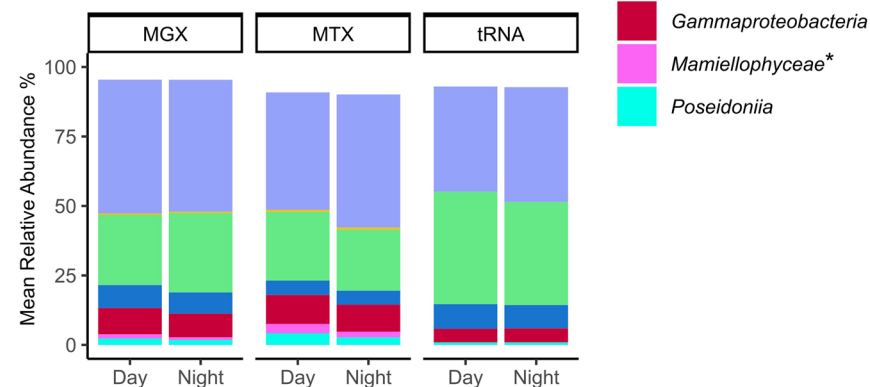
a) HP1**b) STO1**

Fig. 2 Taxonomic profiling across sample types and stations. We profiled the taxonomy of quality-controlled, paired-end metagenomic (MGX) and metatranscriptomic (MTX) reads and tRNA transcripts for coastal HP1 (a) and offshore STO1 (b). We averaged the relative abundances for classes that represent > 1% mean abundance of the total community across samples collected during sunlit hours (day) and those collected after sunset (night). The bar chart shows the fraction of reads with assignments resolving to at least phylum-level taxonomy. The identification of tRNA sequences was based on the GTDB database, and the identification of metatranscriptomic and metagenomic reads was based on the nr_euk database that utilizes taxonomy from NCBI. *Mamiellophyceae: GTDB does not identify eukaryotic organisms. %Flavobacteriia: Flavobacteriia is an NCBI class and an order in GTDB (c_Bacteroidia; o_Flavobacteriales). For tRNA sequences, the GTDB Bacteroidia class was dominated by reads matching to the order Flavobacteriales. ^Betaproteobacteria: Betaproteobacteria is not its own class within GTDB, and is part of the class Gammaproteobacteria.

broader climatic conditions that developed over the days leading up to the diel sampling (Fig. S2). Wind and wave-driven currents typically transport offshore waters into the bay across the centrally-located barrier reef. Variations in wind speeds and wave energy can strongly impact the circulation patterns that influence nearshore to offshore gradients both within and adjacent to Kāneʻohe Bay⁶⁰. Additionally, nutrient-enriched stream discharge following heavy rain events can trigger rapid phytoplankton responses in the bay^{19,61–63}. The relatively subdued onshore wind and wave height prior to the diel sampling, in combination with a brief but intense rainfall event during an outgoing tide shortly before sampling (Fig. S2), may have facilitated the export of nearshore waters from coastal Kāneʻohe Bay to offshore station STO1 and temporarily altered the biogeochemical conditions at STO1.

Label convention. Sample labels follow a standardized naming convention to enable cross-referencing with relevant environmental data and between ‘omic data types. For example, in the following sample name,

HADS_20210818_H1030_MGX_STO1_003

Each of the following fields are separated by an underscore:

- Project name: HADS stands for Hawaiʻi Diel Sampling.
- Sampling date: Four digit year, two digit month, and two digit day, where day restarts at midnight.
- Sampling hour: Starts with letter H, military time for hours and minutes with no other characters in between. Times given are in Hawaiʻi Standard Time.

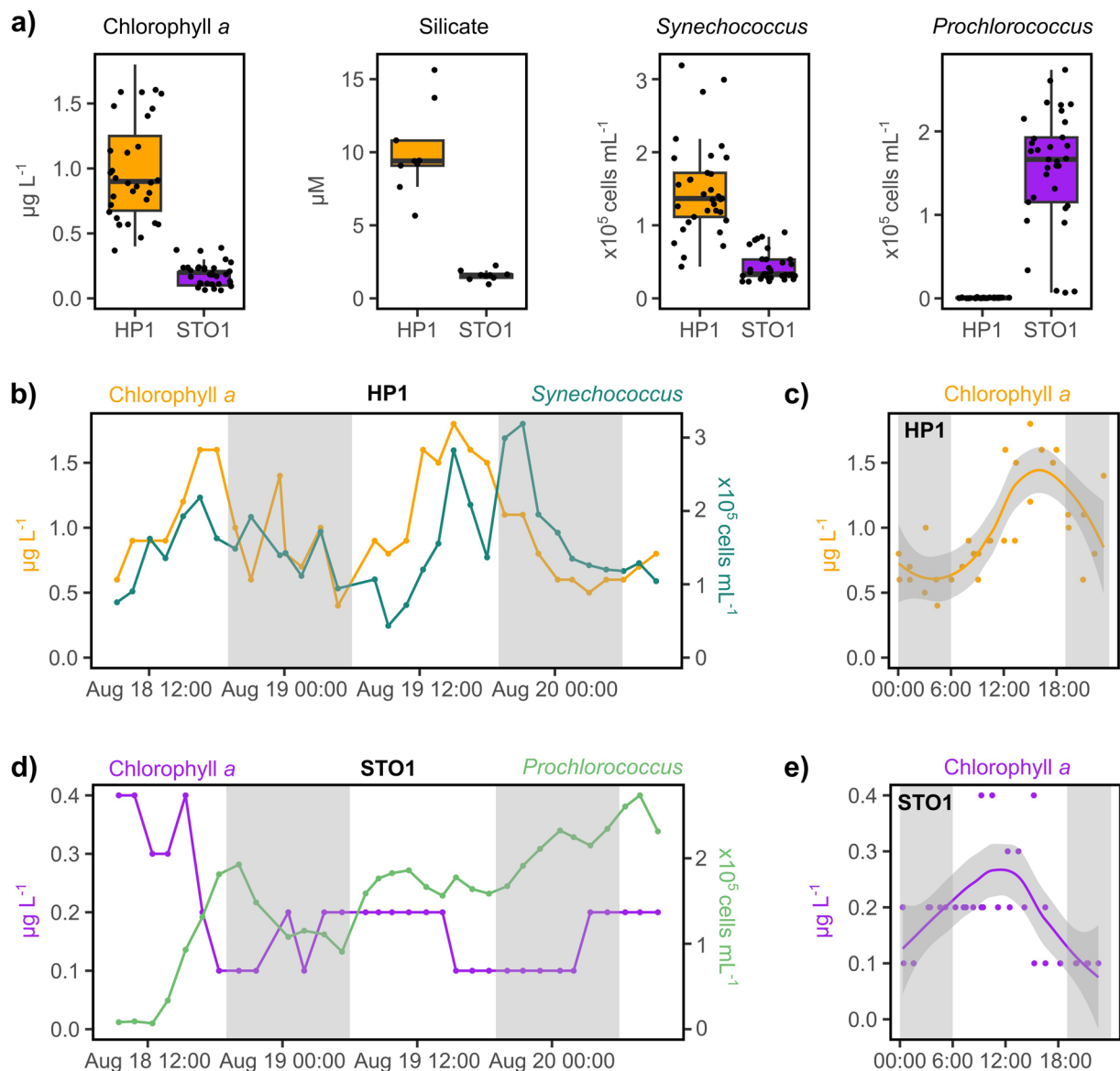


Fig. 3 Variation in biogeochemistry and phytoplankton communities across space and time during HaDS. (a) Chlorophyll *a* and silicate concentrations and *Synechococcus* cellular abundances are higher at coastal station HP1 compared to offshore station STO1, while *Prochlorococcus* cells are more abundant at offshore station STO1. (b) At station HP1, chlorophyll *a* concentrations and *Synechococcus* cellular abundances follow diel oscillations, with early morning (~03:00 HST) minimums and late afternoon (~15:00–21:00 HST) maximums. (c) Chlorophyll *a* concentrations at station HP1 rise during daylight hours and decline after sunset. Samples are plotted by time of day over a 24 h cycle. (d) At station STO1, *Prochlorococcus* cellular abundances show a clear trend that emerges on the order of days while chlorophyll *a* concentrations decline briefly at the beginning of the sampling period. (e) Chlorophyll *a* concentrations at station STO1 reach a minimum in the late evening. Like panel c, this plot shows samples by time of day over a 24-hour cycle. The shaded areas in panels c and e depict the standard error around a local polynomial regression fit line for chlorophyll *a* concentrations per hour of the day. Gray bars in panels b–e indicate nighttime.

- Sample type: Three letter code that can either be MGX (short-read metagenomes), MTX (short-read metatranscriptomes), TRN (short-read tRNA transcripts), or HMW (long-read metagenomes from high-molecular weight DNA).
- Sampling station: Can be STO1 or xHP1. Note that although we use xHP1 in the sample name to represent the station HP1 the formatting is solely to create equal-length names, and reference to the station should continue to explicitly use HP1.
- Sample ID. This number is unique within each sample type.

Metatranscriptomes. We recommend additional removal of adapter sequences and the removal of short reads (<100 bp) prior to use of the metatranscriptomic libraries, as outlined in the methods section.

Code availability

The URL <https://merenlab.org/data/HaDS> gives access the bioinformatics workflow we implemented to perform quality control, metagenomic co-assembly, read mapping, and recovery of metagenome-assembled genomes, as well as the reproducible anvi'o data artifacts^{53,54}.

Received: 2 May 2025; Accepted: 13 October 2025;

Published online: 04 December 2025

References

1. Fuhrman, J. A., Cram, J. A. & Needham, D. M. Marine microbial community dynamics and their ecological interpretation. *Nat. Rev. Microbiol.* **13**, 133–146 (2015).
2. Lambert, S. *et al.* Rhythmicity of coastal marine picoeukaryotes, bacteria and archaea despite irregular environmental perturbations. *ISME J.* **13**, 388–401 (2019).
3. Teeling, H. *et al.* Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* **336**, 608–611 (2012).
4. Needham, D. M., Sachdeva, R. & Fuhrman, J. A. Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *ISME J.* **11**, 1614–1629 (2017).
5. Cram, J. A. *et al.* Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. *ISME J.* **9**, 563–580 (2015).
6. Raes, E. J. *et al.* Seasonal patterns of microbial diversity across the world oceans. *Limnol. Oceanogr. Lett.* **9**, 512–523 (2024).
7. Yeh, Y.-C. & Fuhrman, J. A. Contrasting diversity patterns of prokaryotes and protists over time and depth at the San-Pedro Ocean Time series. *ISME Commun.* **2**, 36 (2022).
8. Giovannoni, S. J. & Vergin, K. L. Seasonality in ocean microbial communities. *Science* **335**, 671–676 (2012).
9. Satinsky, B. M. *et al.* Microspatial gene expression patterns in the Amazon River Plume. *Proc. Natl. Acad. Sci. USA*. **111**, 11085–11090 (2014).
10. Salazar, G. *et al.* Gene expression changes and community turnover differentially shape the global ocean metatranscriptome. *Cell* **179**, 1068–1083.e21 (2019).
11. Manck, L. E. *et al.* Iron limitation of heterotrophic bacteria in the California Current System tracks relative availability of organic carbon and iron. *ISME J.* **18**, wræ061 (2024).
12. Landa, M., Burns, A. S., Roth, S. J. & Moran, M. A. Bacterial transcriptome remodeling during sequential co-culture with a marine dinoflagellate and diatom. *ISME J.* **11**, 2677–2690 (2017).
13. Martiny, A. C. *et al.* Marine phytoplankton resilience may moderate oligotrophic ecosystem responses and biogeochemical feedbacks to climate change. *Limnol. Oceanogr.* **67**, S378–S379 (2022).
14. Becker, J. W., Hogle, S. L., Rosendo, K. & Chisholm, S. W. Co-culture and biogeography of *Prochlorococcus* and SAR11. *ISME J.* **13**, 1506–1519 (2019).
15. Ottesen, E. A. *et al.* Pattern and synchrony of gene expression among sympatric marine microbial populations. *Proc. Natl. Acad. Sci. USA*. **110**, E488–97 (2013).
16. Aylward, F. O. *et al.* Microbial community transcriptional networks are conserved in three domains at ocean basin scales. *Proc. Natl. Acad. Sci. USA*. **112**, 5443–5448 (2015).
17. Coesel, S. N. *et al.* Diel transcriptional oscillations of light-sensitive regulatory elements in open-ocean eukaryotic plankton communities. *Proc. Natl. Acad. Sci. USA*. **118**, e2011038118 (2021).
18. Tucker, S. J. *et al.* Spatial and temporal dynamics of SAR11 marine bacteria across a nearshore to offshore transect in the tropical Pacific Ocean. *PeerJ* **9**, e12274 (2021).
19. Tucker, S. J. *et al.* Seasonal and spatial transitions in phytoplankton assemblages spanning estuarine to open ocean waters of the tropical Pacific. *Limnol. Oceanogr.* **70**, 1693–1708 (2025).
20. Brown, C. L. *et al.* Critical evaluation of short, long, and hybrid assembly for contextual analysis of antibiotic resistance genes in complex environmental metagenomes. *Sci. Rep.* **11**, 3753 (2021).
21. Zhao, D. *et al.* Eukfinder: a pipeline to retrieve microbial eukaryote genome sequences from metagenomic data. *MBio* **16**, e0069925 (2025).
22. Welschmeyer, N. A. Fluorometric analysis of chlorophyll a in the presence of chlorophyll b and pheopigments. *Limnol. Oceanogr.* **39**, 1985–1992 (1994).
23. Monger, B. C. & Landry, M. R. Flow cytometric analysis of marine bacteria with hoechst 33342. *Appl. Environ. Microbiol.* **59**, 905–911 (1993).
24. Donaldson, G. P. *et al.* Spatially distinct physiology of *Bacteroides fragilis* within the proximal colon of gnotobiotic mice. *Nat. Microbiol.* **5**, 746–756 (2020).
25. Watkins, C. P., Zhang, W., Wylder, A. C., Katanski, C. D. & Pan, T. A multiplex platform for small RNA sequencing elucidates multifaceted tRNA stress response and translational regulation. *Nat. Commun.* **13**, 2491 (2022).
26. Shen, W., Sipos, B. & Zhao, L. SeqKit2: A Swiss army knife for sequence and alignment processing. *Imeta* **3**, e191 (2024).
27. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10 (2011).
28. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).
29. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–6 (2013).
30. Burge, S. W. *et al.* Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.* **41**, D226–32 (2013).
31. Eren, A. M., Vineis, J. H., Morrison, H. G. & Sogin, M. L. A filtering method to generate high quality short reads using illumina paired-end technology. *PLoS One* **8**, e66643 (2013).
32. Minoche, A. E., Dohm, J. C. & Himmelbauer, H. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol.* **12**, R112 (2011).
33. Eren, A. M. *et al.* Community-led, integrated, reproducible multi-omics with anvi'o. *Nat. Microbiol.* **6**, 3–6 (2021).
34. Mölder, F. *et al.* Sustainable data analysis with Snakemake. *F1000Res.* **10**, 33 (2021).
35. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: *ade novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–1428 (2012).
36. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
37. Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).

38. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2019).
39. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
40. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
41. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
42. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
43. Menzel, P., Ng, K. L. & Krogh, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* **7**, 11257 (2016).
44. Chan, P. P., Lin, B. Y., Mak, A. J. & Lowe, T. M. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **49**, 9077–9096 (2021).
45. Parks, D. H. *et al.* GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* **50**, D785–D794 (2022).
46. R Core Team. *R: A Language and Environment for Statistical Computing*. (2023).
47. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York (2016).
48. Pante, E. & Simon-Bouhet, B. marmap: A package for importing, plotting and analyzing bathymetric and topographic data in R. *PLoS One* **8**, e73051 (2013).
49. Meren Lab. Diel Multi Omics from the Hawaii Diel Sampling (HaDS). *Zenodo* <https://doi.org/10.5281/ZENODO.15643373> (2025).
50. Tucker, S. J., Fuessel, J., Freil, K. C., Rappé, M. S. & Eren, A. M. Diel multi-omics data from surface ocean microbial community samples collected during Hawai'i Diel Sampling (HaDS) in Kāne'ohe Bay and adjacent offshore waters of O'ahu, Hawai'i from December 2020 to August 2021. Biological and Chemical Oceanography Data Management Office (BCO-DMO) <https://doi.org/10.26008/1912/BCO-DMO.963210.1> (2025).
51. Marine Biological Laboratory. Hawaii Diel Sampling for Marine Microbial Multi Omics. <https://identifiers.org/ncbi/bioproject:PRJNA1201851> (2024).
52. Marine Biological Laboratory. Hawaii Diel Sampling Metagenome-Assembled Genomes. <https://identifiers.org/ncbi/bioproject:PRJNA1235278> (2025).
53. Eren, A. M. HADS MAGs from long-read and short-read metagenomes. *Figshare* <https://doi.org/10.6084/M9.FIGSHARE.28784762.V2> (2025).
54. Eren, A. M. HADS co-assemblies of Long-read and Short-read Metagenomes. *Figshare*. <https://doi.org/10.6084/m9.figshare.28784717.v1> (2025).
55. Schwartz, M. H. *et al.* Microbiome characterization by high-throughput transfer RNA sequencing and modification analysis. *Nat. Commun.* **9**, 5353 (2018).
56. Breitwieser, F. P., Lu, J. & Salzberg, S. L. A review of methods and databases for metagenomic classification and assembly. *Brief. Bioinform.* **20**, 1125–1136 (2019).
57. Sunagawa, S. *et al.* Ocean plankton. Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).
58. Chang, T., Gavelis, G. S., Brown, J. M. & Stepanauskas, R. Genomic representativeness and chimerism in large collections of SAGs and MAGs of marine prokaryoplankton. *Microbiome* **12**, 126 (2024).
59. Trigodet, F., Sachdeva, R., Banfield, J. F. & Eren, A. M. Assemblies of long-read metagenomes suffer from diverse errors. *bioRxiv*, <https://doi.org/10.1101/2025.04.22.649783> (2025).
60. Lowe, R. J., Falter, J. L., Monismith, S. G. & Atkinson, M. J. A numerical study of circulation in a coastal reef-lagoon system. *J. Geophys. Res.* **114**, C06022 (2009).
61. De Carlo, E. H., Hoover, D. J., Young, C. W., Hoover, R. S. & Mackenzie, F. T. Impact of storm runoff from tropical watersheds on coastal water quality and productivity. *Appl. Geochem.* **22**, 1777–1797 (2007).
62. Hoover, R. S. *et al.* Zooplankton response to storm runoff in a tropical estuary: bottom-up and top-down controls. *Marine Ecology Progress Series* **318**, 187–201 (2006).
63. Yeo, S. K., Huggett, M. J., Eiler, A. & Rappé, M. S. Coastal bacterioplankton community dynamics in response to a natural disturbance. *PLoS One* **8**, e56207 (2013).

Acknowledgements

We thank Jason Jones and Andrew Brown for their assistance with field operations. Data collection and curation efforts were supported by the W. M. Keck Foundation, the Simons Foundation (through grants #687269 to AME and #989028 to SJT), and the Center for Chemical Currencies of a Microbial Planet (C-CoMP, a U.S. National Science Foundation Science and Technology Center, grant #OCE-2019589). This is C-CoMP publication #80, SOEST contribution #12012, and HIMB contribution #2015. Our study makes use of data provided by PacIOOS (<https://www.pacioos.org>), part of the U.S. Integrated Ocean Observing System (IOOS), and funded in part by National Oceanic and Atmospheric Administration (NOAA) Awards #NA16NOS0120024 and #NA21NOS0120091.

Author contributions

Conceptualization: S.J.T., J.F., K.C.F., T.P., M.S.R., A.M.E. Methodology: S.J.T., J.F., K.C.F., M.S., S.E.M., F.T., K.L., H.G.M., B.F., B.H., T.P., M.S.R., A.M.E. Investigation: S.J.T., J.F., K.C.F., E.K., E.B.F., O.R., C.E.S.S., A.P.G., H.M., M.R.S., M.Q., C.R., L.L.T., M.S., S.E.M., F.T., K.L., H.G.M., B.F., B.H., T.P., M.S.R., A.M.E. Visualization: S.J.T., A.M.E. Supervision: M.S.R., A.M.E. Writing—original draft: S.J.T., J.F., K.C.F. Writing—review & editing: S.J.T., J.F., K.C.F., M.S.R., A.M.E.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-025-06166-3>.

Correspondence and requests for materials should be addressed to M.S.R. or A.E.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025