

# Title: High spatial resolution global ocean metagenomes from Bio-GO-SHIP repeat hydrography transects

## Authors

Alyse A. Larkin<sup>1</sup>, Catherine A. Garcia<sup>1</sup>, Melissa L. Brock<sup>2</sup>, Jenna A. Lee<sup>1</sup>, Nathan Garcia<sup>1</sup>, Lucas J. Ustick<sup>2</sup>, Leticia Barbero<sup>3,4</sup>, Brendan R. Carter<sup>5,6</sup>, Rolf E. Sonnerup<sup>5,6</sup>, Lynne Talley<sup>7</sup>, Glen A. Tarran<sup>8</sup>, Denis L. Volkov<sup>3,4</sup>, and Adam C. Martiny<sup>1,2</sup>

## Affiliations

1. Department of Earth System Science, University of California at Irvine, Irvine, CA, USA
2. Department of Ecology and Evolution, University of California at Irvine, Irvine, CA, USA
3. NOAA Atlantic Oceanographic and Meteorological Laboratory, Miami, FL, USA
4. Cooperative Institute for Marine & Atmospheric Studies, University of Miami, Miami, FL, USA
5. NOAA Pacific Marine Environmental Laboratory, Seattle, WA, USA.
6. Cooperative Institute for Climate, Ocean, and Ecosystem Studies, University of Washington, Seattle, WA, USA
7. Climate, Atmospheric Sciences, and Physical Oceanography Division, Scripps Institution of Oceanography, University of California, San Diego, CA, USA
8. Plymouth Marine Laboratory, Plymouth, UK

corresponding author(s): Adam C. Martiny ([amartiny@uci.edu](mailto:amartiny@uci.edu))

## Abstract

Detailed descriptions of microbial communities have lagged far behind physical and chemical measurements in the marine environment. Here, we present 720 globally distributed surface ocean metagenomes collected at high spatio-temporal resolution. Our low-cost metagenomic sequencing protocol produced 2.75 terabases of data, where the median number of base pairs per sample was 3.48 billion. The median distance between sampling stations was 26 km. The metagenomic libraries described here were collected as a part of a biological initiative for the Global Ocean Ship-based Hydrographic Investigations Program, or “Bio-GO-SHIP.” One of the primary aims of GO-SHIP is to produce high spatial and vertical resolution measurements of key state variables to directly quantify climate change impacts on ocean environments. By similarly collecting marine metagenomes at high spatiotemporal resolution, we expect that this dataset will help answer questions about the link between microbial communities and biogeochemical fluxes in a changing ocean.

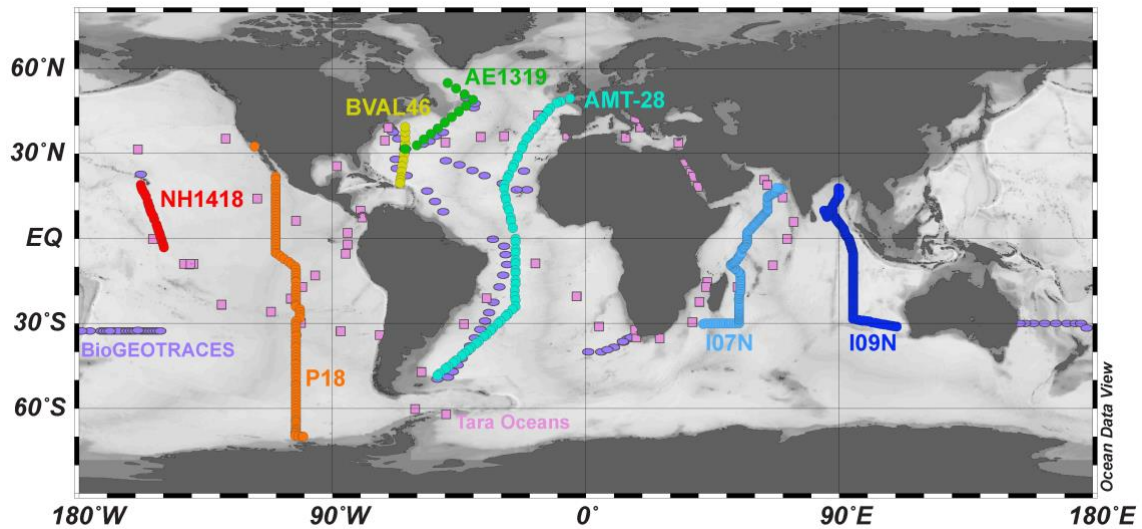
## 40 **Background & Summary**

41 A growing list of coordinated scientific efforts have produced deep metagenomic  
42 libraries of the surface ocean. Projects such as the Global Ocean Survey, Tara Oceans, and  
43 bioGEOTRACES<sup>1-3</sup> have significantly advanced our understanding of marine microbial  
44 biogeography and biodiversity. However, this ever-increasing abundance of metagenomic  
45 data raises the question of how do we move beyond analyses of biodiversity to linking  
46 microbial traits with ecosystem function and elemental fluxes<sup>4</sup>. In oceanography, it has been  
47 widely acknowledged that sparse sampling results in high noise and error rates that in turn  
48 prevent the characterization of dynamic chemical balances and limit biogeochemical models<sup>5</sup>.  
49 Thus, we propose that an increased emphasis on high resolution spatiotemporal sampling of  
50 marine microbial communities would allow for a more mechanistic understanding of the  
51 relationship between microbes and ocean biogeochemistry.

52 The Global Ocean Ship-based Hydrographic Investigations Program (GO-SHIP) seeks to  
53 produce high spatial and vertical resolution measurements of physical, chemical, and  
54 biological parameters over the full water column. This internationally-organized program  
55 coordinates a network of sustained hydrographic sections that are repeatedly measured on  
56 an approximately decadal time scale. Compared to autonomous programs such as Argo, which  
57 has significantly increased the spatial and temporal resolution of ocean observations<sup>6</sup>, ship-  
58 based programs have the advantage of a much broader range of biogeochemical  
59 measurement capabilities. To date, repeat hydrography programs have largely focused on  
60 physical (light, currents, water column thermohaline structure, etc.) and chemical (nutrients,  
61 oxygen, dissolved organic and inorganic carbon, pH, etc.) state variables. This work has  
62 significantly improved our understanding of the response of oxygen<sup>7</sup>, pH<sup>8</sup>, calcium carbonate  
63 saturation depth<sup>9</sup>, and sea level rise<sup>10</sup> to global warming and anthropogenic carbon  
64 accumulation<sup>11</sup>. By comparison, systematic and sustained biological measurements of the  
65 microbial component of ocean ecosystems has lagged far behind.

66 Here, we present a dataset of 720 ocean surface water metagenomes collected at high  
67 spatiotemporal resolution in an effort to more mechanistically link marine microbial traits and  
68 biodiversity to both chemical and hydrodynamic ecosystem fluxes as a part of a novel Bio-GO-  
69 SHIP sampling program. Samples were collected in the Atlantic, Pacific, and Indian Ocean  
70 basins (Fig 1, Table 1). This effort has been supported by GO-SHIP, the Plymouth Marine  
71 Laboratory Atlantic Meridional Transect (PML AMT), and three National Science Foundation  
72 (NSF) Dimensions of Biodiversity funded cruises (AE1319, BVAL46, and NH1418) (Table 2).  
73 Whereas the median distance between Tara Oceans sampling stations was 709 km and the  
74 median distance between bioGEOTRACES sampling stations was 191 km, the median distance  
75 between sampling stations in the current Bio-GO-SHIP dataset is 26 km (Fig 2). In addition, the  
76 majority of Bio-GO-SHIP samples were collected every 4-6 hours, allowing for analysis of diel  
77 fluctuations in microbial composition and gene content<sup>12</sup>. We anticipate that our high-  
78 resolution sampling scheme will allow for a more detailed examination of the relationship  
79 between the broad range of geochemical parameters measured across the various cruises  
80 (Table 2) and microbial diversity and traits.

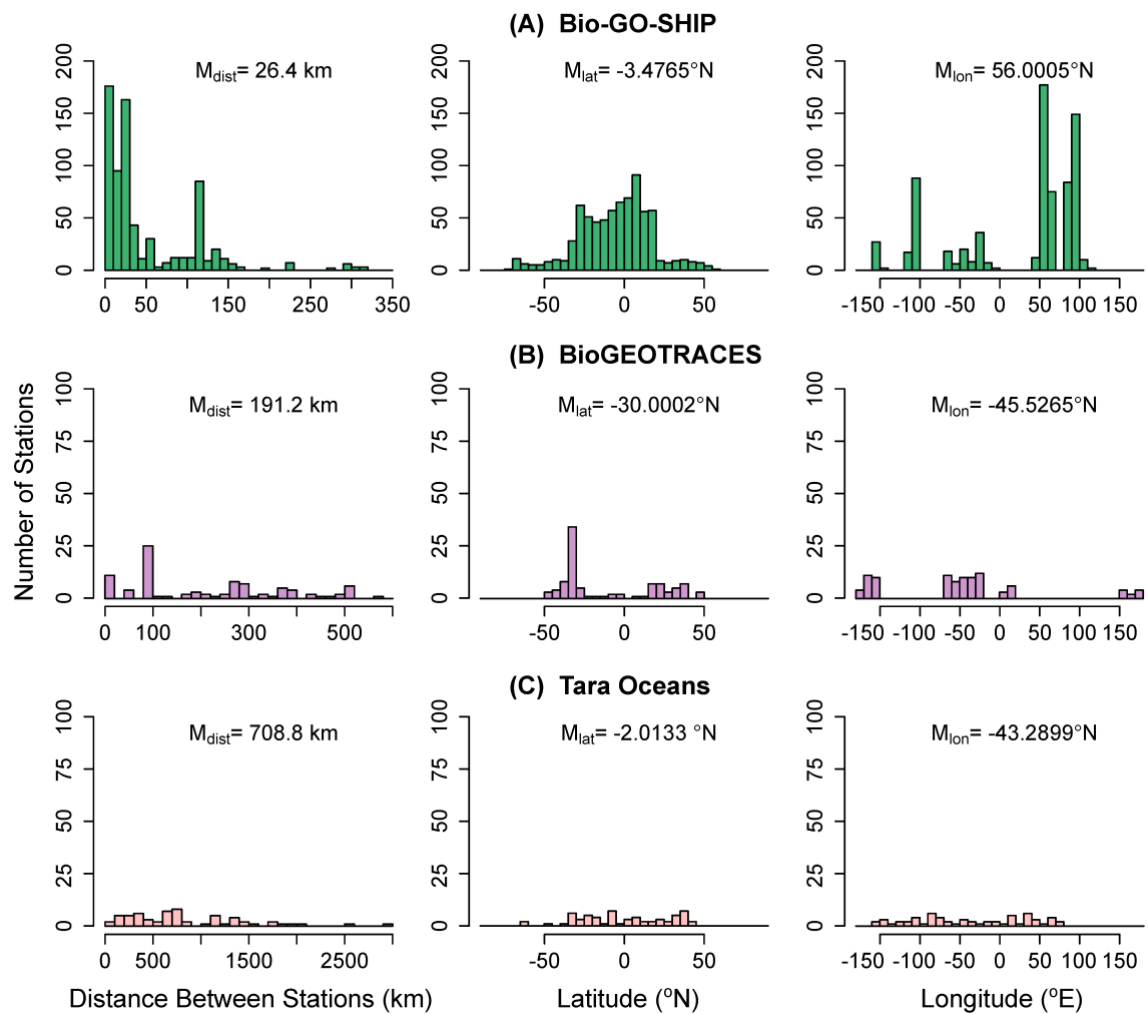
81 Due to their rapid generation times and high diversity, microbial genomes integrate  
82 the impact of environmental change<sup>13</sup> and can be used a 'biosensor' of subtle biogeochemical  
83 regimes that cannot be identified from physical parameters alone<sup>12, 14-16</sup>. Thus, the fields of  
84 microbial ecology and oceanography would benefit from coordinated, high resolution  
85 measurements of marine 'omics products (i.e., metagenomes, metatranscriptomes,  
86 metaproteomes, etc.). This dataset provides an important example of the benefits of a high  
87 spatial and temporal resolution sampling regime. Specifically, our data highlights the need for  
88 increased sampling of marine metagenomes in the Central and Western Pacific Ocean (Fig 1),  
89 areas above 50°N and 50°S (Fig 2), and below the euphotic zone. We hope and expect that  
90 these challenges will be addressed by the scientific community in the coming decade.



91

**Figure 1:** Distribution of global surface microbial metagenomes from Bio-GO-SHIP (circles) in comparison to Tara Oceans (squares) and bioGEOTRACES (ovals).

92



93

**Figure 2:** Comparison of the distance between stations, station latitudes, and station longitudes for global surface ocean metagenomes. Individual station locations from (A) Bio-GO-SHIP, (B) bioGEOTRACES and (C) Tara Oceans were examined. Plots are labelled with the median value,  $M$ . Station distance was calculated as the distance to the nearest station.

94

95 **Table 1:** Sampling protocols and read counts for global Bio-GO-SHIP surface ocean  
96 metagenomes.

Cruise	DNA Collection	DNA Volume	Station Count	Total Reads	Total Bases	Median Bases Per Sample	Range of Bases Per Sample
I09N	Underway	10 L	242	5.73x10 <sup>9</sup>	8.64x10 <sup>11</sup>	3.10x10 <sup>9</sup>	4.71x10 <sup>8</sup> – 1.22x10 <sup>10</sup>
I07N	Underway, CTD	4 L, 2-4 L	248	6.20x10 <sup>9</sup>	9.36x10 <sup>11</sup>	3.27x10 <sup>9</sup>	2.47x10 <sup>8</sup> – 1.42x10 <sup>10</sup>
P18	CTD	2 L	104	3.22x10 <sup>9</sup>	4.86x10 <sup>11</sup>	4.46x10 <sup>9</sup>	6.14x10 <sup>7</sup> – 1.77x10 <sup>10</sup>
AMT-28	CTD	2 L	63	2.18x10 <sup>9</sup>	3.29x10 <sup>11</sup>	4.95x10 <sup>9</sup>	1.62x10 <sup>9</sup> – 1.22x10 <sup>10</sup>
BVAL46	CTD	2 L	12	2.01x10 <sup>8</sup>	3.04x10 <sup>10</sup>	2.73x10 <sup>9</sup>	2.33x10 <sup>9</sup> – 4.88x10 <sup>9</sup>
AE1319	CTD	2 L	13	2.01x10 <sup>8</sup>	3.03x10 <sup>10</sup>	4.69x10 <sup>9</sup>	2.15x10 <sup>9</sup> – 7.43x10 <sup>9</sup>
NH1418	CTD	2 L	23	5.41x10 <sup>8</sup>	8.17x10 <sup>10</sup>	3.03x10 <sup>9</sup>	2.42x10 <sup>9</sup> – 1.08x10 <sup>10</sup>

97

98

99

## Methods

100 On all cruises, whole (i.e., no size fractionation) surface water was collected via either  
101 the Niskin rosette system (depth ~3-5m) or the ship's circulating seawater system (depth  
102 ~7m). Between 2-10 L of surface water (Table 1) was collected in triple-rinsed containers and  
103 gently filtered through a 0.22 µm pore size Sterivex filter (Millipore, Darmstadt, Germany)  
104 using sterilized tubing and a Masterflex peristaltic pump (Cole-Parmer, Vernon Hills, IL). DNA  
105 was preserved with 1620 µl of lysis buffer (4 mM NaCl, 750 µM sucrose, 50 mM Tris-HCl, 20  
106 mM EDTA) and stored at -20°C before extraction.

107 To extract DNA (modified from Bostrom et al. 2004)<sup>17</sup> Sterivex filters were incubated  
108 with 180 µl lysozyme (3.5 nM) at 37°C for 30 minutes followed by an overnight 55°C incubation  
109 with 180 µl Proteinase K (0.35 nM) and 100 µl 10% SDS buffer. DNA was extracted from the  
110 Sterivex with 1000 µl TE buffer (10 mM Tris-HCl, 1 mM EDTA), precipitated in an ice-cold  
111 solution of 500 µl isopropanol (100%) and 1980 µl sodium acetate (3 mM, pH 5.2), pelleted via  
112 centrifuge for 30 mins at 4°C, and resuspended in TE buffer in a 37°C water bath for 30 min.  
113 Next, DNA was purified using a genomic DNA Clean and Concentrator kit (Zymo Research  
114 Corp., Irvine, CA). Finally, DNA concentrations were quantified using a Qubit dsDNA HS Assay  
115 kit and Qubit fluorometer (ThermoFisher, Waltham, MA).

116 A total of 720 metagenomic libraries were prepared using Illumina-specific Nextera  
117 DNA transposase adapters and a Tagment DNA Enzyme and Buffer Kit (Illumina, San Diego,  
118 CA, cat. no. 20034197) (modified from Baym et al. 2015)<sup>18-20</sup>. Nextera adapter sequences to  
119 be used for bioinformatic quality trimming are: 5'-TCG TCG GCA GCG TCA GAT GTG TAT AAG  
120 AGA CAG-3' and 5'-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G-3'. Custom Nextera  
121 DNA-style 8bp unique dual index (UDI) barcodes I7 (5'-CAA GCA GAA GAC GGC ATA CGA GAT  
122 [NNN NNN NN]G TCT CGT GGG CTC GG-3') and I5 (5'-AAT GAT ACG GCG ACC ACC GAG ATC  
123 TAC AC[N NNN NNN N]TC GTC GGC AGC GTC-3') were used to multiplex the metagenomic  
124 libraries. A total of 1 µl of 2 ng/µl DNA was added to 1.5 µl tagmentation reactions (1.25 µl TD  
125 buffer, 0.25 µl TDE1) and incubated at 55°C for 10 minutes. After tagmentation, product (2.5  
126 µl) was immediately added to 22 µl reactions (1.02 µM per UDI barcode, 204 µM dNTPs,  
127 0.0204 U Phusion High Fidelity DNA polymerase and 1.02X Phusion HF Buffer [ThermoFisher,  
128 Waltham, MA] final concentration). Barcodes were annealed to tagmented products using the  
129 following polymerase chain reaction (PCR): 72°C for 2 min., 98°C for 30 s., followed by 13  
130 cycles of 98°C 10 s., 63°C 30 s., 72°C 30 s., and a final extension step of 72°C for 5 min.

131 To quality control tagmentation products, dimers that were less than 150 nucleotides  
132 long were removed using a buffered solution (1 M NaCl, 1 mM EDTA, 10 mM Tris-HCl, 44.4 M  
133 PEG-8000, 0.055% Tween-20 final concentration) of Sera-mag SpeedBeads (ThermoFisher,  
134 Waltham, MA). Metagenomic libraries were quantified using a Qubit dsDNA HS Assay kit

135 (ThermoFisher, Waltham, MA) and a Synergy 2 Microplate Reader (BioTek, Winooski, VT).  
136 Libraries were then pooled at equimolar concentrations. Pooled library concentration was  
137 verified using a KAPA qPCR platform (Roche, Basel, Switzerland). Finally, dimer removal as well  
138 as read size distribution were checked using a 2100 Bioanalyzer high sensitivity DNA trace  
139 (Agilent, Santa Clara, CA).

140 54 samples were sequenced on two Illumina HiSeq 4000 lanes using 150 bp paired-  
141 end chemistry with 300 cycles (Illumina, San Diego, CA). All remaining samples were  
142 sequenced on three Illumina NovaSeq lanes using S4 150 bp paired-end chemistry with 300  
143 cycles. The sequencing strategy produced a total of  $1.83 \times 10^{10}$  reads, or  $2.75 \times 10^{12}$  bp. The  
144 median number of bases per sample was 3.48 billion (range: 61,400,000 – 17.7 billion). The  
145 sequencing cost per bp in US dollars was  $\$8.2 \times 10^{-9}$ .

146

147

## 148 **Data Records**

149 The majority of the samples here were collected under the auspices of the international GO-  
150 SHIP program and the national programs that contribute to it<sup>21-24</sup>. A comprehensive data  
151 directory of metadata resources is available at <https://www.go-ship.org/>. Bottle data and  
152 cruise report links are provided in Table 1.

153

154 Metadata variables from the AMT-28 cruise (<https://www.amt-uk.org/>) are hosted by the  
155 British Oceanographic Data Centre, and may be requested through the following URL:  
156 <https://www.bodc.ac.uk/>. Select metadata are also available through GO-SHIP<sup>24</sup>.

157

158 The BVAL46, AE1319, and NH1418 cruises were collected as a part of the “Biological Controls  
159 on the Ocean C:N:P Ratios” project funded by the NSF Division of Ocean Sciences<sup>25-28</sup>. Data  
160 associated with these deployments are hosted by the NSF Biological and Chemical  
161 Oceanography Data Management Office (BCO-DMO). A comprehensive list of metadata  
162 resources is available at <https://www.bco-dmo.org/project/2178>.

163

164 All sequencing products associated with the Bio-GO-SHIP program can be found under  
165 BioProject ID PRJNA656268 hosted by the National Center for Biotechnology Information  
166 Sequence Read Archive (SRA)<sup>29</sup>. SRA accession numbers associated with each metagenome file  
167 are provided in Supplementary Table 1.



168 **Table 2:** Complete list of metadata variables collected on Bio-GO-SHIP cruises

Campaign	Metadata Access	Metadata Variables
I09N / GO-SHIP	<a href="https://cchdo.ucsd.edu/cruise/33RR20160321">https://cchdo.ucsd.edu/cruise/33RR20160321</a>	Temperature, Dissolved O <sub>2</sub> , Conductivity, Density, Salinity, Nutrients (NO <sub>3</sub> , NO <sub>2</sub> , NH <sub>4</sub> , PO <sub>4</sub> , SiO <sub>4</sub> ), Chlorofluorocarbons (CFCs) /SF <sub>6</sub> , Dissolved Inorganic Carbon (DIC), <sup>13</sup> C and <sup>14</sup> C of DIC, Total pH, Total Alkalinity, Stable gases (N <sub>2</sub> , N <sub>2</sub> O, Ar), <sup>18</sup> O, Chromophoric Dissolved Organic Matter (CDOM), Pigment HPLC, variable chlorophyll fluorescence, Dissolved Organic Carbon, underway Particulate Organic C N and P, underway pCO <sub>2</sub> , Lowered Acoustic Doppler Current Profiler, Chipods, Dissolved/ particulate/ cellular P and Fe, N P and Fe uptake rates, <i>Prochlorococcus</i> / <i>Synechococcus</i> / Picoeukaryotes/ Nanoeukaryotes cell counts
I07N / GO-SHIP	<a href="https://cchdo.ucsd.edu/cruise/33RO20180423">https://cchdo.ucsd.edu/cruise/33RO20180423</a>	Temperature, Dissolved O <sub>2</sub> , Conductivity, Density, Salinity, Chlorophyll, Nutrients (NO <sub>3</sub> , NO <sub>2</sub> , PO <sub>4</sub> , SiO <sub>4</sub> ), Dissolved Inorganic Carbon (DIC), Chlorofluorocarbons (CFCs) /SF <sub>6</sub> , <sup>14</sup> C of DIC, Dissolved Organic Carbon, Black Carbon, DO <sup>14</sup> C, Total pH, Total Alkalinity, Calcium, Dissolved Organics, Biomarkers, underway Particulate Organic C N and P, underway pCO <sub>2</sub>
P18 / GO-SHIP	<a href="https://cchdo.ucsd.edu/cruise/33RO20161119">https://cchdo.ucsd.edu/cruise/33RO20161119</a>	Temperature, Dissolved O <sub>2</sub> , Conductivity, Density, Salinity, Nutrients (NO <sub>3</sub> , NO <sub>2</sub> , PO <sub>4</sub> , SiO <sub>4</sub> ), Chlorofluorocarbons (CFCs) /N <sub>2</sub> O/SF <sub>6</sub> , Helium isotopes and noble gases (Ne, Ar, Kr, and Xe), Dissolved Inorganic Carbon (DIC), <sup>13</sup> C and <sup>14</sup> C of DIC, Total pH, Total Alkalinity, Stable gases (N <sub>2</sub> , O <sub>2</sub> , Ar), Dissolved Organic Carbon / Total Dissolved Nitrogen, Tritium, Black Carbon, DO <sup>14</sup> C/DO <sup>14</sup> C, Dissolved Organics, Biomarkers, underway Particulate Organic C N and P, underway pCO <sub>2</sub> , Wind speed, Wind direction, Air temperature
AMT-28 / PML AMT / GO-SHIP	<a href="https://www.bodc.ac.uk/data/hosted_data_systems/amt/">https://www.bodc.ac.uk/data/hosted_data_systems/amt/</a> <a href="https://amt-uk.org/Cruises/AMT28">https://amt-uk.org/Cruises/AMT28</a> <a href="https://cchdo.ucsd.edu/cruise/74JC20180923">https://cchdo.ucsd.edu/cruise/74JC20180923</a>	Temperature, Dissolved O <sub>2</sub> , Conductivity, Salinity, Nutrients (NO <sub>3</sub> , NO <sub>2</sub> , PO <sub>4</sub> , SiO <sub>4</sub> ), biogenic silica/silicon uptake, Total pH, Total Alkalinity, Pigment HPLC, Chlorophyll a, Dissolved Organic C N and P, <i>Prochlorococcus</i> / <i>Synechococcus</i> / Picoeukaryote/ Nanoeukaryote/ Heterotrophic bacteria cell counts, FlowCAM, 15N/13C, Respiration (total, bacterial, size-fractionated), Bacterial production, underway Particulate Organic C N and P, underway wave radar (Cband), Sky Infrared Brightness temperature, Hyperspectral radiance/irradiance, Wind speed, Wind direction, Aerosol size/composition
BVAL46 / NSF / BATS	<a href="https://www.bco-dmo.org/project/2178">https://www.bco-dmo.org/project/2178</a>	Temperature, Dissolved O <sub>2</sub> , Conductivity, Salinity, PAR irradiance, Chlorophyll a, NO <sub>3</sub> + NO <sub>2</sub> , NO <sub>2</sub> , PO <sub>4</sub> , SiO <sub>4</sub> , Soluble Reactive Phosphorus (SRP), Particulate Organic C N and P, P uptake (max. uptake, half saturation conc.), <i>Prochlorococcus</i> / <i>Synechococcus</i> / Picoeukaryote/ Nanoeukaryote cell counts
AE1319 / NSF	<a href="https://www.bco-dmo.org/project/2178">https://www.bco-dmo.org/project/2178</a>	Temperature, Dissolved O <sub>2</sub> , Conductivity, Salinity, PAR irradiance, Chlorophyll a, NO <sub>3</sub> + NO <sub>2</sub> , NO <sub>2</sub> , PO <sub>4</sub> , SiO <sub>4</sub> , Soluble Reactive Phosphorus (SRP), Particulate Organic C N and P, P uptake (max. uptake, half saturation conc.), <i>Prochlorococcus</i> / <i>Synechococcus</i> / Picoeukaryote/ Nanoeukaryote cell counts
NH1418 / NSF	<a href="https://www.bco-dmo.org/project/2178">https://www.bco-dmo.org/project/2178</a>	Temperature, Dissolved O <sub>2</sub> , Conductivity, Salinity, Density, Fluorescence, PAR irradiance, Chlorophyll a, NO <sub>3</sub> + NO <sub>2</sub> , NO <sub>2</sub> , Soluble Reactive Phosphorus (SRP), Particulate Organic C N and P, <i>Prochlorococcus</i> / <i>Synechococcus</i> / Picoeukaryote/ Nanoeukaryote / Crocospheera cell counts

## 169 **Technical Validation**

170 To ensure that no contamination of metagenomes occurred, negative controls were used. To  
171 ensure optimum paired-end short read sequencing, a 2100 Bioanalyzer high sensitivity DNA  
172 trace (Agilent, Santa Clara, CA) was used for each library to confirm that 90% of the sequence  
173 fragments were above 250 bp and below 600 bp in length. Qubit (ThermoFisher, Waltham,  
174 MA) and a KAPA qPCR platform (Roche, Basel, Switzerland) were used to ensure that all pooled  
175 libraries were submitted for sequencing at a concentration >15 nM.

176

177

## 178 **Usage Notes**

179 The genomic data described here have not been pre-screened or processed in any way. We  
180 recommend quality control parameters. Prior to our sequence analysis in subsequent projects,  
181 we removed adapter sequences, performed sequence quality control, and ensured there was  
182 no contamination from common genomic add-ins such as Phi-X using the following code  
183 parameters:

184

185 Trimmomatic (v0.35): PE ILLUMINACLIP:NexteraPE-PE.fa:2:30:10 SLIDINGWINDOW:4:15

186 MINLEN:36

187 BBMap (v37.50): bbduk.sh -Xmx1g ref=/BBMap/37.50/resources/phix174\_ill.ref.fa.gz k=31

188 hdist=1

189

190

## 191 **Code Availability**

192 Custom scripts were not used to generate or process this dataset. Software versions and non-  
193 default parameters used have been appropriately specified where required.

194

## 195 **Acknowledgements**

196 We would like to thank the captains and crew of the R/V Atlantic Explorer, R/V New Horizon,  
197 R/V Ronald H. Brown, R/V Roger Revelle, and the R.R.S. James Clark Ross. Financial support for  
198 this project was provided by the National Science Foundation (OCE-1046297, 1559002,  
199 1848576, and 1948842 to ACM). LB and DLV were supported in part under the auspices of the  
200 Cooperative Institute for Marine and Atmospheric Studies (CIMAS), a cooperative institute of  
201 the University of Miami and NOAA (cooperative agreement NA10OAR4320143). The PML AMT  
202 is funded by the UK Natural Environment Research Council through its National Capability  
203 Long-term Single Centre Science Programme, Climate Linked Atlantic Sector Science (grant  
204 number NE/R015953/1). This study contributes to the international IMBeR project and is  
205 contribution number ### of the AMT program. GO-SHIP is supported by NOAA Global Ocean  
206 Monitoring and Observation program (U8R1SE3-PRF) and by the National Science Foundation  
207 (OCE-1437015). P18 and I07N were NOAA-led cruises and I09N was an NSF-led cruise.  
208

## 209 **Author contributions**

210 A.A.L. wrote the manuscript, coordinated sample collection, collected/processed samples,  
211 designed protocols, performed metagenomic sequencing, and compiled metadata.

212 C.A.G. coordinated sample collection, collected/processed samples, performed metagenomic  
213 sequencing, and compiled metadata.

214 M.L.B performed metagenomic sequencing.

215 J.A.L. collected/processed samples and performed metagenomic sequencing.

216 N.G. coordinated sample collection and collected samples.

217 L.J.U. processed samples and compiled metadata.

218 L.B, B.G.C., R.E.S., L.T., and D.L.V. coordinated GO-SHIP collection and collaboration efforts.

219 G.T. coordinated PML AMT/GO-SHIP collection and collaboration efforts.

220 A.C.M. designed and supervised the study, secured funding, and coordinated GO-SHIP  
221 collection.

222 All authors contributed to manuscript editing and revision.  
223

## 224 **Competing interests**

225 The authors declare no competing interests.  
226



## 227 **References**

228

229 1. Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooseph, S., et  
230 al. The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern  
231 tropical Pacific. *PLoS Biol*, **5**, e77 (2007).

232 2. Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., Le Bescot, N., Gorsky, G., et al.  
233 Open science resources for the discovery and analysis of Tara Oceans data. *Scientific*  
234 *Data* **2**, 150023 (2015).

235 3. Biller, S. J., Berube, P. M., Dooley, K., Williams, M., Satinsky, B. M., Hackl, T., et al.  
236 Marine microbial metagenomes sampled across space and time. *Scientific Data* **6**,  
237 180176 (2019).

238 4. Hall, E. K., Bernhardt, E. S., Bier, R. L., Bradford, M. A., Boot, C. M., Cotner, J. B., et al.  
239 Understanding how microbiomes influence the systems they inhabit. *Nature*  
240 *Microbiology* **3**(9), 977-982 (2018).

241 5. Curtin, T. B., Bellingham, J. G., Catipovic, J., & Webb, D. Autonomous oceanographic  
242 sampling networks. *Oceanography* **6**(3), 86-94 (1993).

243 6. Riser, S. C., Freeland, H. J., Roemmich, D., Wijffels, S., Troisi, A., Belbeoch, M., et al.  
244 Fifteen years of ocean observations with the global Argo array. *Nature Climate Change* **6**  
245 (2), 145-153 (2016).

246 7. Schmidtko, S., Stramma, L., & Visbeck, M. Decline in global oceanic oxygen content  
247 during the past five decades. *Nature* **542**(7641), 335-339 (2017).

248 8. Doney, S. C., Fabry, V. J., Feely, R. A., & Kleypas, J. A. Ocean acidification: The other CO<sub>2</sub>  
249 problem. *Annual Review of Marine Science* **1**, 169-192 (2009).

250 9. Feely, R. A., Sabine, C. L., Lee, K., Berelson, W., Kleypas, J., Fabry, V. J., et al. Impact of  
251 anthropogenic CO<sub>2</sub> on the CaCO<sub>3</sub> system in the oceans. *Science* **305**(5682), 362-366  
252 (2004).

253 10. Purkey, S. G. & Johnson, G. C. Warming of global abyssal and deep Southern Ocean  
254 waters between the 1990s and 2000s: Contributions to global heat and sea level rise  
255 budgets. *Journal of Climate* **23**(23), 6336-6351 (2010).

256 11. Gruber, N., Clement, D., Carter, B. R., Feely, R. A., van Heuven, S., Hoppema, M., et  
257 al., The oceanic sink for anthropogenic CO<sub>2</sub> from 1994 to 2007. *Science* **363**(6432), 1193-  
258 1199 (2019).

259 12. Larkin, A. A., Garcia, C. A., Ingoglia, K. A., Garcia, N. S., Baer, S. E., Twining, B. S., et al.  
260 Subtle biogeochemical regimes in the Indian Ocean revealed by spatial and diel  
261 frequency of *Prochlorococcus* haplotypes. *Limnology and Oceanography* **65**, S220-S232  
262 (2020).

263 13. Larkin, A. A. & Martiny, A. C. Microdiversity shapes the traits, niche space, and  
264 biogeography of microbial taxa. *Environmental Microbiology Reports* **9**(2), 55-70 (2017).

265 14. Garcia, C. A., Hagstrom, G. I., Larkin, A. A., Ustick, L. J., Levin, S. A., Lomas, M. W., et al.  
266 Linking regional shifts in microbial genome adaptation with surface ocean

- 267 biogeochemistry. *Philosophical Transactions of the Royal Society B* **375**(1798), 20190254  
268 (2020).
- 269 15. Martiny, A. C., Ustick, L., Garcia, C. A., & Lomas, M. W. Genomic adaptation of marine  
270 phytoplankton populations regulates phosphate uptake. *Limnology and Oceanography*  
271 **65**, S340-S350 (2020).
- 272 16. Sonnewald, M., Dutkiewicz, S., Hill, C., & Forget, G. Elucidating ecological complexity:  
273 Unsupervised learning determines global marine eco-provinces. *Science Advances* **6**(22),  
274 eaay4740 (2020).
- 275 17. Bostrom, K. H., Simu, K., Hagstrom, A., & Riemann, L.. Optimization of DNA extraction  
276 for quantitative marine bacterioplankton community analysis. *Limnology and*  
277 *Oceanography-Methods* **2**, 365-373 (2004).
- 278 18. Baym, M., Kryazhimskiy, S., Lieberman, T. D., Chung, H., Desai, M. M., & Kishony, R.  
279 Inexpensive multiplexed library preparation for megabase-sized genomes. *Plos One*  
280 **10**(6), e0128036 (2015).
- 281 19. Wandro, S., Oliver, A., Gallagher, T., Weihe, C., England, W., Martiny, J. B. H., et al.  
282 Predictable molecular adaptation of coevolving *Enterococcus faecium* and lytic phage  
283 EfV12-phi1. *Frontiers in Microbiology* **9**, 3192 (2019).
- 284 20. Oliver, A., LaMere, B., Weihe, C., Wandro, S., Lindsay, K. L., Wadhwa, P. D., et al.  
285 Cervicovaginal microbiome composition is associated with metabolic profiles in healthy  
286 pregnancy. *mBio* **11**(4), e01851-20 (2020).
- 287 21. Barbero, L., Bottle data from Cruise 33RR20160321, exchange version. CCHDO cruise  
288 DOI: 10.7942/C2008W. CCHDO <https://cchdo.ucsd.edu/cruise/33RR20160321> (2016).
- 289 22. Sonnerup, R., Carter, B., Purkey, S., & Bourbonnais, A. Bottle data from  
290 Cruise 33RO20161119, exchange version. CCHDO cruise DOI: 10.7942/C21T0F. CCHDO  
291 <https://cchdo.ucsd.edu/cruise/33RO20161119> (2016).
- 292 23. Volkov, D. & Menezes, V. Bottle data from Cruise 33RO20180423, exchange version.  
293 CCHDO cruise DOI: 10.7942/C25H2B. CCHDO  
294 <https://cchdo.ucsd.edu/cruise/33RO20180423>, (2018).
- 295 24. Tarran, G. & Reed, A. Bottle data from Cruise 74JC20180923, exchange version. CCHDO  
296 <https://cchdo.ucsd.edu/cruise/74JC20180923> (2018).
- 297 25. Lomas, M. W. & Martiny, A. C., Bonachela, J. A., & Levin, S. Biogeochemistry of microbial  
298 phosphorus uptake from cruises in the Sargasso Sea Bermuda Atlantic Time-Series  
299 Station from 2011-2013 (Biological C:N:P ratios project). *Biological and Chemical*  
300 *Oceanography Data Management Office (BCO-DMO)* [http://lod.bco-](http://lod.bco-dmo.org/id/dataset/538091)  
301 [dmo.org/id/dataset/538091](http://lod.bco-dmo.org/id/dataset/538091) (2014).
- 302 26. Lomas, M. W. & Martiny, A. C. Depth profile data from Bermuda Atlantic Time-Series  
303 Validation cruise 46 (BVAL46) in the Sargasso Sea from Sept-Oct. 2011 (Biological C:N:P  
304 ratios project). *Biological and Chemical Oceanography Data Management Office (BCO-*  
305 *DMO)* <http://lod.bco-dmo.org/id/dataset/> (2020).

- 306 27. Lomas, M. W. & Martiny, A. C., Depth profile data from R/V Atlantic Explorer AE1319 in  
307 the NW Atlantic from Aug-Sept. 2013 (Biological C:N:P ratios project). *Biological and*  
308 *Chemical Oceanography Data Management Office (BCO-DMO)* [http://lod.bco-  
dmo.org/id/dataset/](http://lod.bco-<br/>309 <u>dmo.org/id/dataset/</u>) (2020).
- 310 28. Lomas, M. W., Martiny, A. C., Depth profile data from R/V New Horizons NH1418 in the  
311 tropical Pacific from Sept-Oct. 2014 (Biological C:N:P ratios project). *Biological and*  
312 *Chemical Oceanography Data Management Office (BCO-DMO)* [http://lod.bco-  
dmo.org/id/dataset/](http://lod.bco-<br/>313 <u>dmo.org/id/dataset/</u>) (2020).
- 314 29. NCBI Sequence Read Archive: <http://identifiers.org/ncbi/insdc.sra:SRP> (2020).  
315