Sargasso Sea bacterioplankton community structure and drivers of variance as revealed by DNA metabarcoding analysis

J. Geoffrey Gill [1] , **Kristina M Hill-Spanik** [1] , **Kerry A Whittaker** [2, 3] , **Martin L Jones** [4] , **Craig Plante** [Corresp. 1]

[1] Grice Marine Laboratory, College of Charleston, Charleston, SC, United States

[2] Sea Education Association, Woods Hole, MA, United States

[3] Maine Maritime Academy, Castine, Maine, United States

[4] Department of Mathematics, College of Charleston, Charleston, SC, United States

Corresponding Author: Craig Plante
Email address: plantec@cofc.edu

Marine microbes provide the backbone for pelagic ecosystems by cycling and fixing nutrients and establishing the base of food webs. Microbial communities are often assumed to be highly connected and genetically mixed, with localized environmental filters driving minor changes in structure. Our study applied high-throughput Illumina 16S ribosomal RNA gene amplicon sequencing on whole-community bacterial samples to characterize geographic, environmental, and stochastic drivers of community diversity. DNA was extracted from seawater collected from the surface ($N = 18$) and at depth just below the deep chlorophyll-a maximum (DCM mean depth = 115.4 m; $N = 22$) in the Sargasso Sea and adjacent oceanographic regions. Discrete bacterioplankton assemblages were observed at varying depths in the North Sargasso Sea, with a signal for distance-decay of bacterioplankton community similarity found only in surface waters. Bacterial communities from different oceanic regions could be distinguished statistically but exhibited a low magnitude of divergence. Redundancy analysis identified temperature as the key environmental variable correlated with community structuring. The effect of dispersal limitation was weak, while variation partitioning and neutral community modeling demonstrated stochastic processes influencing the communities. This study advances understanding of microbial biogeography in the pelagic ocean and highlights the use of high-throughput sequencing methods in studying microbial community structure.

1 **Sargasso Sea bacterioplankton community structure and drivers of variance as revealed by**
2 **DNA metabarcoding analysis**

4 J. Geoffrey Gill[1], Kristina M. Hill-Spanik[1], Kerry A. Whittaker[2,3], Martin L. Jones[4], & Craig J.
5 Plante[1]
6
7 [1] Grice Marine Lab, College of Charleston; Charleston, SC
8 [2] Sea Education Association; Woods Hole, MA
9 [3] Maine Maritime Academy, Castine, ME
10 [4]Department of Mathematics, College of Charleston; Charleston, SC

13 **Abstract**

14   Marine microbes provide the backbone for pelagic ecosystems by cycling and fixing
15 nutrients and establishing the base of food webs. Microbial communities are often assumed to be
16 highly connected and genetically mixed, with localized environmental filters driving minor
17 changes in structure. Our study applied high-throughput Illumina 16S ribosomal RNA gene
18 amplicon sequencing on whole-community bacterial samples to characterize geographic,
19 environmental, and stochastic drivers of community diversity. DNA was extracted from seawater
20 collected from the surface ($N = 18$) and at depth just below the deep chlorophyll-a maximum
21 (DCM mean depth = 115.4 m; $N = 22$) in the Sargasso Sea and adjacent oceanographic regions.
22 Discrete bacterioplankton assemblages were observed at varying depths in the North Sargasso
23 Sea, with a signal for distance-decay of bacterioplankton community similarity found only in
24 surface waters. Bacterial communities from different oceanic regions could be distinguished
25 statistically but exhibited a low magnitude of divergence. Redundancy analysis identified
26 temperature as the key environmental variable correlated with community structuring. The effect
27 of dispersal limitation was weak, while variation partitioning and neutral community modeling
28 demonstrated stochastic processes influencing the communities. This study advances
29 understanding of microbial biogeography in the pelagic ocean and highlights the use of high-
30 throughput sequencing methods in studying microbial community structure.

32

34   **Introduction**

35   Marine bacteria are amongst the most abundant organisms in the global ocean and are

36   responsible for greater than half of primary productivity and nutrient fluxes in marine

37   environments (Azam et al. 1983, Azam 1998). The field of microbial biogeography has

38   expanded due to the advancement of genetic tools capable of classifying microbial taxa on a

39   whole-community scale (Giovannoni & Stingl 2005, Franzosa et al. 2015, Santoferrara et al.

40   2016). This has allowed investigation into the dispersal capacity of these planktonic microbes,

41   determining the extents to which dispersal barriers, environmental filters, and stochastic effects

42   structure their communities over space and time (Schwob et al. 2021, Vergin et al. 2017). By

43   analyzing the whole community at once, high-throughput genetic sequencing tools can help to

44   resolve the mechanisms driving these sweeping patterns in diversity.

45   Environmental filters influencing marine bacterial community structure may include

46   physical factors such as temperature and surface mixing depth, chemical variables including

47   salinity or nutrient concentrations, and biological parameters and interactions (Follows &

48   Dutkiewicz 2011, Friedline et al. 2012, Fuhrman et al. 2008). Marine bacteria are known for

49   their ability to rapidly respond to changing environmental conditions (Corno et al. 2007),

50   allowing for community structures to swiftly change to suit present conditions.

51   The hypothesis of cosmopolitan distributions, or that "everything is everywhere, but the

52   environment selects," has remained appealing for decades for its conciseness, but is impossible

53   to fully test given its near-infinite criteria for the influences of time and space with respect to

54   microbial diversity (Baas-Becking 1934). As genetic tools have advanced to directly probe this

55   theory, distance effects and geographic provincialism have been identified as contributing factors

56   to microbial community dissimilarities on global scales, suggesting that barriers to and means of

57   dispersal can have an impact on microbial community structuring (Hörtsmann et al. 2021,

58   Logares et al. 2018, Martiny et al. 2006). Physical features, such as defined pycnoclines or ocean

59   currents delineating water masses, may restrict the free, passive dispersal of microbes, bringing

60   about these community dissimilarities over certain time scales (Fuhrman et al. 2015, Hellweger

61   et al. 2014, Hörtsmann et al. 2021, Treusch et al. 2009). Many taxa of marine bacteria are found

62   globally, but local distributions of endemic microbes are frequently identified, particularly in

63   extreme environments (Giovannoni & Stingl 2005, Pedros-Alio 2006).

64   Stochastic influences may also play a significant role in structuring communities. A

65   neutral community model, based upon dynamics of birth, death, and immigration, has been

66   successfully applied to prokaryotic assemblages (Sloan et al. 2006), demonstrating that chance

67   events play a significant role despite microbes' rapid responses to deterministic environmental

68   factors and competition (Giovannoni & Stingl 2005). These tend to influence community

69   composition as a historical contingency, with random fluctuation in mortality or natality

70   (ecological drift), selection for past environmental conditions, or barriers to interbreeding

71   between distant or physically separated populations leading to dynamics of community

72   structuring in the present (Hanson et al. 2012, Sloan et al. 2006). Stochasticity and environmental

73   determinism are thought to co-influence microbial community structure (Caruso et al. 2011). For

74   instance, extremely low-resource environments have been shown to exert more severe limitations

75   on community competition, thus increasing the role of determinism (Stegen et al. 2012), but

76   neutral models have been successfully applied even in resource-poor cryptic desert bacterial

77   communities (Caruso et al. 2011).

78   The Sargasso Sea, a distinctly bounded open water mass just offshore of the eastern

79   United States, presents an appropriate ecosystem to study mechanisms structuring

80   bacterioplankton communities. High sea surface temperatures and a regional geostrophic

81   downwelling zone contribute to vertical stratification of the water column and oligotrophic

82   conditions, providing annually consistent physical structuring of the water column (Steinberg et

83   al. 2001). These conditions select for picoplankton such as marine bacteria as a productive base,

84   as these communities are able to more effectively absorb nutrients compared to the larger

85   eukaryotic phytoplankton dominant in higher-nutrient areas (Cotti-Rausch et al. 2016, Finkel et

86   al. 2010, Herlemann et al. 2011). Currents of the North Atlantic gyre, including the North

87   Equatorial Current, Gulf Stream, and Antilles Current, provide natural biogeographic boundaries

88   for the Sargasso ecosystem; these gyre currents also congregate the floating holopelagic

89   *Sargassum* that imparts ecological uniqueness to the Sea (Ardron et al. 2011). The eddies

90   distributed throughout the Sargasso provide additional biogeographic features, where a region of

91   higher eddy influence from the Gulf Stream may be distinguished from downwelling of the

92   central gyre, although the ecological implications of eddy influence are not fully understood

93   (Ardron et al. 2011). The physical dynamics of the Sargasso Sea have been largely explored with

94   regard to *Sargassum*-associated species (e.g., Govindarajan et al. 2019, Sehein et al. 2014, Siuda

95   2011), and the physical, chemical, and biological makeup of the area has been well documented

96   in the Bermuda Atlantic Time Series survey, which has surveyed the same region of the northern

97   Sargasso Sea monthly since 1988 (Lomas et al. 2013).

98   Surveys of bacterial communities in the Sargasso Sea and broader North Atlantic Ocean

99   have identified depth and chlorophyll-*a* (chl-*a)* concentration as significant determinants of

100   community structuring (Cotti-Rausch et al. 2016, Sjöstedt et al. 2014). Other studies focusing on

101   bacterial distribution have found distinct communities at the surface, the deep chlorophyll

102   maximum (DCM), and the upper mesopelagic, as well as distinct communities during periods of

103 winter convective mixing compared to other seasons (Treusch et al. 2009). Water bodies, both on

104 a meso-scale with eddies and surface mixing, as well as on a Longhurstian provincial scale, have

105 also been found to influence prokaryote β-diversity (Cotti-Rausch et al. 2016, Hahnke et al.

106 2013).

107   This study sought to further investigate the geographic, environmental, and stochastic

108 factors that influence bacterial communities in the Sargasso Sea and neighboring regions of the

109 North Atlantic using Illumina high-throughput amplicon DNA sequencing. We focused our

110 efforts on characterizing bacterial communities both vertically within the euphotic zone and

111 horizontally across environmentally-defined biogeographic regions. We hypothesized that there

112 would be no geographic dispersal limitation for marine bacteria within the euphotic zone,

113 extending from the surface to the DCM. Of the contemporary environmental filters examined,

114 we expected sampling depth to be the most significant influence on community structure, with

115 distinct communities at the surface and at the DCM driven by distinct nutrient and light

116 characteristics in these two regions of the water column. We further anticipated that stochastic

117 effects would significantly impact distribution and occurrence of these marine bacteria, despite

118 extreme environmental filtering, given their small size and the lack of dispersal barriers in the

119 open ocean.

120

121 **Materials and Methods**

122 *Sampling*

123   Water samples ($N = 40$) were collected during the daytime from the Sargasso Sea and

124 adjacent areas of the Atlantic Ocean during mornings from 21 April to 18 May 2018. The SSV

125 *Corwith Cramer* (Sea Education Association; (SEA)) followed a route travelling from Nassau,

126 Bahamas to Pennos Bay, Bermuda to New York City, New York, USA (Fig. 1) as a part of the

127 SEA's annual transect through the Sargasso Sea (Govindarajan et al. 2019, Law et al. 2010,

128 Sehein et al. 2014, Zettler et al. 2013). Cruise plans were filed with the US State Department,

129 who obtained the required collection permits. No permits were required for sampling in

130 international and US waters under federal jurisdiction. The cruise and permit numbers for the

131 samples collected in this study are as follows: C279, US State Department Cruise F2017-112,

132 Bermuda permit number SP171103 and Bahamas permit number MAMR/FIS/13.

133   A shipboard flow-through system continuously monitored sea surface temperature (SST),

134 salinity, and turbidity. SST measurements were used to define oceanographic regions observed

135 throughout the research cruise and refine the location of the subtropical convergence zone

136 (STCZ). Neuston tows conducted at each scientific station were used to collect surface

137 *Sargassum* and zooplankton, and surface *Sargassum* abundance was weighed manually. Water

138 samples were collected by bucket manually (surface only) or collected using a Seabird

139 Electronics SBE19PlusV2 Conductivity, Temperature, Depth (CTD) Carousel (Bellevue, WA)

140 equipped with a Model QSP-2300 photosynthetic active radiation (PAR) sensor, Sea Point *in-*

141 *vivo* chl-*a* fluorometer, and twelve 2.5-L Niskin bottles. Water samples were collected both from

142 surface waters and from the DCM as determined from the chl-*a* maximum detected by the CTD

143 fluorometer. Samples from six of 21 scientific stations were filtered *in situ* using a McLane

144 Research Laboratories Large Volume Water Transfer System (Falmouth, MA). Due to

145 equipment malfunctions, remaining samples for molecular analysis were collected by vacuum-

146 filtering 250-500 mL of water through a 0.22-μm Millipore mixed-cellulose membrane filter. All

147 filtered microbial samples were stored at -20 °C on board. Both filtration methodologies (*in-situ*

148 McLane filtration and Niskin bottle-collected shipboard filtration) involved vacuum filtration of

149   seawater onto mixed-cellulose membrane filters with identical pore sizes (0.22 μm), and

150   downstream sample processing, from filter handling to extraction to sequencing, was identical

151   across samples. *In-situ* sampling (from the McLane pump) may have been associated with a

152   lower probability for bacterial contamination, given the demand for less seawater handling.

153   However, we bioinformatically examined evidence for bacterial contamination across our

154   samples (detailed below), allowing us to confidently compare samples collected *in situ* with

155   those collected via Niskin bottle/CTD and filtered on board.

156   Nitrate concentrations corresponding with each sample were determined via cadmium

157   reduction and colorimetric methods, measured using an Ocean Optics Flame spectrophotometer

158   (Orange County, FL). To determine chl-*a* concentration, seawater was first filtered over 0.45-μm

159   GF/F filters. Chl *a* was then extracted by immersing filters in 5 ml of 90% acetone, and

160   extractions were performed in the dark at -20 ^{o}C. Chl-*a* concentrations were determined using a

161   Turner Designs Model 10-AU Benchtop Fluorometer (San Jose, CA) following Strickland and

162   Parsons (1972).

163
164   *Oceanographic regions*

165   Oceanographic regions used for statistical analysis and categorization of the data were

166   defined by geographic and oceanographic factors. The South Sargasso Sea (SSS) and North

167   Sargasso Sea (NSS) boundary, typically set in the literature at 29° N (Sjöstedt et al. 2014, Siuda

168   2011) was modified for this study to fall between samples taken at 27.2^o N and 28.5^o N (Fig. 1)

169   based on a decrease in SST at this latitude (SST < 24 ^{o}C), indicating the location of the STCZ

170   during the study. The Antilles Current (AC) and Sargasso Sea boundary was established using an

171   RDI Ocean Surveyor 75 kHz Acoustic Doppler Current Profiler (Teledyne RD Instruments;

172   Poway, CA) as well as at-depth CTD data indicating the presence of a warm (> 26 ^{o}C) water

173   mass near the DCM. The North Atlantic (NA) was defined as the region beyond the influence of

174   the Gulf Stream along the continental shelf of the northeastern US, as measured using the RDI

175   current profiler. Sampling was also conducted offshore of Bermuda during a port stop. The

176   Bermuda sample is included in descriptive data summaries for comparison to the open-ocean

177   sites, but was excluded from all statistical analyses.

178

179   *DNA extraction and library preparation*

180   DNA was extracted from the 0.22-µm water filters using a QIAGEN DNeasy Plant Mini

181   Kit (Germantown, MD) following the manufacturer's protocol. DNA was amplified following a

182   modified version of the 16S Metagenomic Sequencing Library Preparation protocol using S-D-

183   Bact-0341-b-S-17/S-D-Bact-0785-a-A-21 primers (Klindworth et al. 2013), targeting

184   hypervariable regions 3 and 4 of the 16S ribosomal RNA (rRNA) gene identified for use in next-

185   generation sequencing-based diversity studies, which included forward and reverse Illumina

186   overhang adapters, respectively. Triplicate 25-µl reactions were performed using 1x TaKaRa *Ex*

187   *Taq* Buffer (Clontech, Mountain View, CA), 0.4 µg ml^{-1} bovine serum albumin, 0.3 mM dNTPs,

188   0.2 µM of each primer, 0.025 U µl^{-1} TaKaRa *Ex Taq* DNA polymerase (Clontech), and 1 µl

189   template DNA. Cycling was as follows: 95 $^\circ$C for 3 min; 30 cycles at 95 $^\circ$C for 30 s, 50 $^\circ$C for 30

190   s, and 72 $^\circ$C for 30 s; a 5 min final extension at 72 $^\circ$C. Samples were electrophoresed on 1%

191   agarose gels stained with GelRed (Biotium, Hayward, CA), and triplicate products were pooled

192   based on band intensity to approximately equalize concentrations in the pool. All PCR negative

193   controls were also pooled and were processed along with samples to identify any potential

194   contamination during library preparation. All pooled products were indexed using a unique

195   combination of forward and reverse Illumina Nextera indices, and indexed PCR products were

196   cleaned using AMPureXP beads (Beckman Coulter, Brea, CA) following the manufacturer's

197   protocol. Successful indexing was confirmed using gel electrophoresis as above, with a subset

198   being analyzed on a Bioanalyzer 2100 using a DNA 1000 kit (Agilent Technologies, Santa Clara,

199   CA). Cleaned, indexed products were quantified using the dsDNA Broad Range assay for an

200   Invitrogen Qubit 1.0 fluorometer (Thermo Fisher, Waltham, MA) and then pooled in equimolar

201   ratios based on the sample with the lowest concentration. The pooled library was quantified

202   using a Qubit 1.0 (Thermo Fisher) and ran on the Bioanalyzer 2100 (Agilent) as above. The

203   library was sequenced on an Illumina MiSeq (San Diego, CA) using a 2 x 250 bp sequencing kit

204   at the Medical University of South Carolina Hollings Cancer Center Genomics Shared Resource

205   (Charleston, SC).

206

207   *Sequence data analysis*

208   Sequencing data were demultiplexed by the sequencing facility. Sequence quality was

209   examined using FastQC (www.bioinformatics.babraham.ac.uk/projects/), and primers and

210   adapter sequences were trimmed from each sequence using Cut-adapt (Martin 2011) in the

211   Galaxy platform (galaxyproject.org). Fastq-join (Aronesty 2013) in Quantitative Insights into

212   Microbial Ecology (QIIME) v. 1.9.1 (Caporaso et al. 2010) was used to join forward and reverse

213   reads with 8% maximum difference between matching segments and a 6-nucleotide minimum

214   length difference. QIIME was also used to remove sequences with quality scores < 24.

215   USEARCH V.5.2.236 (Edgar 2017) was then used to detect and remove chimeras using

216   UCHIME (Edgar et al. 2011) and cluster operational taxonomic unit (OTU) at a 97% sequence

217   similarity threshold. A representative set of sequences was selected and then compared to the

218   Greengenes 16S rRNA database (Desantis et al. 2011) using the RDP classifier to assign

219   taxonomy to each representative sequence (Wang et al. 2007). Any OTU found in the negative

220   control was considered a contaminant if the number of reads from any sample was < 10x the

221   number of reads from the negative control (as in Pagenkopp Lohan et al. 2017). Based on these

222   criteria, three contaminant OTUs identified as g. *Moraxella, Streptococcus*, and *Micrococcus*

223   were removed from the dataset.

224   A single rarefaction was performed to equalize the number of OTUs to the sample with

225   the fewest number of sequences (N = 15,468) prior to statistical analyses. Shannon diversity,

226   number of observed OTUs, Heip's evenness (E), and Chao1 α-diversity indices were calculated

227   for each sample using QIIME and used to determine species richness and evenness. For

228   phylogenetic-based β-diversity analysis (weighted UniFrac; Hamady et al. 2009), sequences

229   were aligned in QIIME using PyNAST (Python Nearest Neighbor Alignment Space Termination,

230   Caporaso et al. 2009). The alignment was filtered to remove positions where a gap was present in

231   80% of the sequences. The top 10% most entropic base positions were also removed. An

232   approximately maximum-likelihood phylogenetic tree was created using FastTree 2.1.3 (Price et

233   al. 2010) and used to create a weighted UniFrac dissimilarity matrix (Hamady et al. 2009).

234   α-diversity metrics between depths and among oceanographic regions were compared

235   using Kruskal-Wallis tests in JMP (SAS, Cary, NC). The effect of environmental and geographic

236   variables on α-diversity was explored using multiple linear regression. Standard least square

237   models with eight (date, latitude, longitude, temperature, salinity, nitrate, chl *a, Sargassum*

238   density) or nine environmental variables (adding depth) were run after a square-transformation

239   on α-diversity metrics (observed OTUs, Chao1, Shannon, and evenness) due to lack of normality

240   with raw data. β-diversity was compared across categories (oceanographic region or depth) using

241   PERMANOVA tests in QIIME, with PERMDISP tests conducted in R version 3.6.3 (R Core

242   Team, 2015) using the *betadisper* function in *vegan* (Oksanen et al. 2020) to assess sampling

243   dispersions. Pairwise PERMANOVA tests were conducted in QIIME on significant results to

244   assess the specific communities accounting for statistical separability. Principal coordinates

245   analysis (PCoA) was used to visualize community associations and results were plotted using

246   EMPeror (Vázquez-Baeza et al. 2013) in QIIME. Two-way SIMPER (similarity percentage)

247   analysis, performed in PRIMER-E v6 (Clarke & Gorley 2006), was used to identify taxa

248   responsible for differences between geographic regions and depths. Venn diagrams were

249   generated in VENN (http://bioinformatics.psb.ugent.be/webtools/Venn/) to examine the number

250   of shared and unique OTUs among sampling locations at the surface and at the DCM.

251   Partial distance-based redundancy analysis (db-RDA) was applied to examine

252   relationships between bacterioplankton community structure (weighted UniFrac distances) and

253   environmental variables (date, temperature, salinity, nitrate, chl *a*, depth, *Sargassum* density),

254   while controlling for geographic distance. Using the *capscale* function in *vegan* (Oksanen et al.

255   2020), forward selection was used to determine the subset of environmental factors that best

256   explained community variation, as determined by adjusted R^2 values. The principal coordinates

257   of neighbor matrix (PCNM; Borchard & Legendre 2002) eigenfunctions, which represent the

258   spectral decomposition of the spatial relationship across sampling locations, were considered as

259   the spatial variables in the ordination-based analyses. Variation partitioning was carried out to

260   partition community variation into environmental and spatial effects, enabling us to determine

261   the variation in the bacterial community explained by the various unique and combined fractions

262   of spatial and environmental data. Because the shared fraction of variation was negative, we used

263   proportional apportioning (Legendre et al. 2012) to correct the variation attributable to each

264   unique factor.

265   We also evaluated the fit of a neutral community model (NCM) to determine the potential

266   contributions of stochastic processes on bacterial community assembly. NCMs were fitted with

267   least squares to mean relative abundance and detection frequency data (Keymer et al. 2009) for

268   each sampling site, based on the equations originally derived by Sloan et al. (2006). Sloan's

269   model is an adaptation of the neutral theory (Hubbell 2001), specifically tailored to deal with

270   extremely large population sizes and the inability to completely describe microbial species

271   abundances. For these analyses, we separately employed the surface, deep, and combined

272   datasets. R^2 values represent goodness of fit to the NCM, and larger R^2 values indicate stochastic

273   dispersal and/or ecological drifts are more important than selection to community assembly

274   (Logares et al. 2013, Sloan et al. 2006). All computations were performed in R version 3.6.3 (R

275   Core Team, 2015).

276

277   **Results**

278   *Environmental parameters*

279   DCM depth varied throughout the sampling track from ~40 to 150 m (Table 1). At the

280   surface, temperature and salinity ranged from 14.8 to 25.7 °C and from 33.7 to 36.9 psu, resp.; at

281   depth, values ranged from 19.5 to 26.4 °C and from 36.4 to 36.9 psu, resp. (Table 1).

282   Temperature exhibited a significant negative correlation with latitude ($r = -0.88$, $P < 0.001$).

283   Nitrate concentrations were below detection level for all surface samples except the single

284   sample from Bermuda (0.69 µM), which was taken in an urbanized harbor, and one outlier in the

285   SSS (0.15 µM) (Table 1). Nitrate in the DCM ranged between 0 and 1.57 µM. *Sargassum*

286   density varied from 0.03 to 0.35 g m^{-1} (Table 1).

287

288   *α-level biodiversity*

289   An average of 1621 OTUs (± 65) per sample were found in the 39 open-ocean seawater

290   samples. Overall, the DCM samples demonstrated significantly greater species richness than

291   surface samples as indicated by the higher number of observed OTUs ($X^2 = 5.46$, $df = 1.00$, $P =$

292   0.020) and Chao1 index ($X^2 = 4.51$, $df = 1.00$, $P = 0.034$) (Table 2 and Table S1). This

293   distinction was attributable to significant differences in observed OTUs, Chao1 index, and

294   Shannon biodiversity between DCM and surface samples in the SSS (Table S1). Small sample

295   sizes prevented similar regionally-specific analysis of the two depth categories elsewhere.

296   α-diversity did not significantly differ among oceanographic regions whether sample sets

297   were considered as a whole or subdivided by comparable depths (Table 2, Table S1). No

298   environmental variables had significant predictive ability in regression models for any of the four

299   diversity metrics ($P > 0.05$).

300

301   *Community composition*

302   Proteobacteria was the most commonly observed phylum in all regions, both in surface

303   waters and at the DCM (~55-69%), followed by Cyanobacteria (9-17%) and Bacteroidetes (6-

304   11%). The Bermuda sample was most distinct, with notably higher contribution of the

305   Proteobacteria families Rhodospirillaceae and Rhodobacteraceae, and much higher dominance of

306   Synechococcaceae relative to Prochloraceae (Fig. 2). This single surface sample from Bermuda

307   coastal waters was more dominated by abundant OTUs than communities in pelagic

308   environments, as indicated by its low evenness (E = 0.10; Table 2).

309   Within the DCM, sequences from Pelagibacteraceae and Alteromonadaceae

310   (Proteobacteria), and Alphaproteobacteria OTU 1 were dominant in all regions, but sequences

311   from OTU 1 were relatively more dominant in the South Sargasso Sea (Fig. 2). As in the surface

312   communities, Synechococcaceae and Prochloraceae (Cyanobacteria) sequences were abundant in

313   the DCM. In the AC and both NSS and SSS samples, sequences from Prochloraceae were more

314   common than those from Synechococcaceae, whereas the reverse was true in the NA samples.

315   Community dissimilarity ranged from 62-74% based on two-way SIMPER analysis, with

316   the NSS and the AC being the least dissimilar (Table S2). No single OTU contributed more than

317   0.88% to the dissimilarity among communities among regions at each sampling depth, with 49 to

318   64 OTUs contributing to 10% dissimilarity between regions at each depth (Table S2).

319   The number of unique OTUs at a given sampling location at each depth exceeded that of

320   shared OTUs among all sampling locations (805 OTUs for surface, 737 for DCM samples), with

321   the exception of NA surface waters (Fig. 3). Overall, 71.5% of OTUs identified in surface

322   samples were unique to their region, while 52.7% of OTUs identified at the DCM were unique to

323   their respective regions. 37.3% of all OTUs observed appeared in only a single sample.

324

325   *Distance-decay of bacterial community similarity*

326   We observed no relationship between bacterial community similarity and geographic

327   distance among all samples analyzed together ($R^2 = 0.001$, $F_{1,1480} = 2.49$, $P = 0.115$; Fig. 4a).

328   When analyzed separately, the surface samples exhibited a significant but weak decay of

329   similarity with geographic distance ($R^2 = 0.037$, $F_{1,270} = 11.43$, $P < 0.001$). No distance-decay

330   relationship was observed among DCM samples ($R^2 = 0.001$, $F_{1,460} = 1.58$, $P = 0.210$; Fig. 4).

331

332   *β-diversity and community structure*

333  Community structures were significantly, but weakly, separable based on oceanographic

334 region ($N = 39$, pseudo-$F = 1.892$, $P = 0.028$; Table S3), but not by depth ($N = 39$, pseudo-$F =$

335 0.492, $P = 0.832$). When DCM and surface communities were considered independently and

336 compared by oceanographic region, greater separability was observed in each, with surface

337 communities being more separable by region ($N = 17$, pseudo-$F = 4.640$, $P = 0.001$) compared to

338 the DCM ($N = 22$, pseudo-$F = 2.765$, $P = 0.003$). Pairwise comparisons between oceanographic

339 regions indicated that differences between microbial communities in the SSS and NSS regions

340 primarily accounted for these results ($N = 15$, pseudo-$F = 8.252$, $P = 0.002$ at the surface; $N =$

341 17, pseudo-$F = 4.300$, $P = 0.004$ at the DCM) (Table S3). PERMDISP results revealed

342 significant differences between sample group dispersion based on oceanographic region (F-value

343 $= 5.65$, $P = 0.003$), but not based on sampling depth. Comparisons between depths at each

344 oceanographic region revealed a significant difference between NSS surface and DCM

345 communities ($N = 11$, pseudo-$F = 14.73$, $P = 0.004$) (Table S3).

346  PCoA analysis conducted on all collected samples revealed overlap among communities

347 across the geographic range of the study but demonstrated a clear separation of surface and DCM

348 communities collected in the NSS. Surprisingly, the most obvious overlap in communities was

349 NSS surface and SSS DCM bacterial communities (Fig. 5).

350

351 *Spatial and environmental determinants of microbial community structure*

352  Partial db-RDA performed with all environmental variables (temperature, salinity, chl *a*,

353 nitrate, depth, *Sargassum* density, and date) indicated that that a three-variable model including

354 temperature, nitrate concentration, and *Sargassum* density gave the best fit after partialling out

355 spatial effects ($F_{3,17} = 2.24$, $P = 0.014$), although only temperature and *Sargassum* density were

356   significant individual factors ($F_{1,17}$ = 2.38, P = 0.042 and $F_{1,17}$ = 2.75, P = 0.024, resp.). No

357   models were significant (P > 0.05) when redundancy analysis was performed separately for

358   surface or DCM samples. Variation partitioning resulted in a negative shared fraction of

359   variation (i.e., environmental *and* spatial influence), resulting from nonorthogonality of a key

360   environmental variable and the spatial factor – temperature and geographic distances were

361   significantly correlated (R^2 = 0.443, P = 0.001; Mantel test). After correcting with proportional

362   appropriation, environmental and spatial variables could together explain ~28% of the variation

363   in bacterial community structure. Pure spatial effects explained approximately ~11% of the

364   variation, while pure environmental factors explained 17%. Total residual variation was high at

365   72%.

366   The NCM explained a large fraction (R^2 = 0.86) of the variability in occurrence

367   frequency of the entire bacterioplankton metacommunity of the Sargasso Sea and surrounding

368   North Atlantic (Fig. 6). When surface and DCM samples were analyzed separately, the NCM

369   likewise explained a large proportion of variability (R^2 = 0.84 and 0.83, resp.; Fig. 6).

370

371   **Discussion**

372   Due to its clear biogeographic boundaries and internal patterns of environmental

373   structuring, the Sargasso Sea and adjacent waters provided an ideal setting to study the drivers of

374   bacterioplankton community structure. We employed high-throughput DNA sequencing to

375   compare communities from surface waters and the DCM, as well as among oceanographic

376   regions. Overall, we found a weak pattern of distance-decay and relatively small differences in

377   community structure among regions, although β-diversity was somewhat higher for surface

378   waters than the DCM. Environmental selection, especially temperature, was a stronger driver of

379 community variation than dispersal limitation. We also found that stochastic factors contributed

380 significantly to biogeographic patterns observed.

381   Bacterial species richness was greater in the DCM than at the surface, with this difference

382 being most pronounced in the South Sargasso Sea. Previous studies have found that bacterial

383 richness increases with depth through the water column up to the 0.1% light level (Treusch et al.

384 2009), peaking near the bottom of the epipelagic and gradually declining with increased depth

385 (Wang et al. 2020). The trend of increasing diversity with depth is often amplified in oceanic

386 regions such as the Sargasso Sea with significant stratification and oligotrophic surface

387 conditions (Gajigan et al. 2018, West et al. 2016), where surface waters are heavily impacted by

388 rainfall, sun exposure, flotsam, and other conditions that may permit more dynamic fluctuations

389 of opportunistic bacterial taxa.

390   Bacterioplankton composition was in general agreement with prior studies conducted in

391 epipelagic waters of the open ocean. We found that sequences from phyla Proteobacteria,

392 Cyanobacteria, and Bacteroidetes, in descending order of relative abundance, comprised the bulk

393 of bacterioplankton, which corroborates earlier studies from the epipelagic zone (e.g., Pommier

394 et al. 2007, Sjöstedt et al. 2014, Zheng et al. 2016). The predominance of sequences from the

395 SAR11 clade was unsurprising because it is widely thought to be the most abundant planktonic

396 microbe in the world's oceans (Giovannoni 2017, Morris et al. 2002). Observations of sequences

397 and relative abundances from the cyanobacteria *Prochlorococcus* and *Synechococcus* are also in

398 accord with prior research. For instance, Sjöstedt et al. (2014) observed these cyanobacteria in

399 the Sargasso Sea in high relative abundance and, as in our study, noted that *Prochlorococcus*

400 sequences were more prevalent than those from *Synechococcus*. In contrast, in the North

401 Atlantic, we observed a relative dominance of *Synechococcus* sequences over those from

402   *Prochlorococcus*, consistent with previous studies and likely due to cooler water temperatures

403   (Flombaum et al. 2013). Despite the broad similarity in dominant bacterial groups among all

404   regions and both depths, communities were distinct among the various sampling locations, either

405   due to differences in relative abundances or presence of less abundant OTUs.

406   Neither species richness nor biodiversity was found to be related to latitude or regional

407   water masses, contrasting with some studies that have suggested that these community

408   characteristics may be shaped by decreasing temperatures at higher latitudes or water-mass

409   specific dispersal barriers and oceanographic differences (Agogué et al. 2011, Costello &

410   Chaudhary 2017, Moran 2015). Although the question of a latitudinal diversity gradient for

411   bacterioplankton has been addressed by numerous investigators, no clear answer has emerged.

412   For instance, Sul et al. (2013) reported a negative correlation between latitude and

413   bacterioplankton diversity, whereas others have reported non-linear trends, with diversity

414   maxima at intermediate latitudes (Fuhrman et al. 2008, Raes et al. 2011, Sunagawa et al. 2015).

415   Trends with latitude have been attributed to temperature relations (Sunagawa et al. 2015), while

416   other studies have found stronger relationships with primary productivity or other environmental

417   factors (e.g., Raes et al. 2011).

418   Regional differences in β-diversity were statistically significant, but weak, and the PCoA

419   plot showed regional overlap, highlighting the relative influences of both environmental filters

420   and neutral processes in shaping bacterioplankton communities. When surface and DCM

421   communities were analyzed separately across all regions, surface communities demonstrated

422   greater separation than deep communities, particularly within the Sargasso Sea itself. This

423   disparity was corroborated by the slightly stronger distance-decay relationship for our surface

424   samples as compared to the DCM. Communities from all regions were not statistically separable

425   based on sampling depth. However, there was a clear separation of surface and DCM

426   communities when examining only those samples from the North Sargasso Sea. The distinct

427   structuring and biodiversity characteristics support our hypotheses that environmental pressures

428   differing at the surface and the DCM would lead to distinctions in these two microbial

429   communities. Given unbalanced sampling design and significant PERMDISP results, however,

430   the differences in these community structures (as per PERMANOVA) could be attributed to any

431   combination of differences in group location centroid or sampling dispersion. The PCoA plot

432   demonstrating general clustering of samples based on their region and depth (i.e., surface

433   Sargasso Sea samples were separated from their respective DCM samples) provide qualitative

434   evidence that differences in community structure were in fact due to differences in location

435   centroids but cannot conclusively rule out difference due to statistical dispersion.

436   Prior studies likewise observed provincialism among bacterial assemblages of the upper

437   pelagic (Friedline et al. 2012, Gómez-Pereira et al. 2010, Milici et al. 2016), and our results also

438   concur with several studies that found no distance-decay for bacterioplankton within the DCM

439   (Walsh et al. 2015, Milici et al. 2016). This pattern emphasizes the importance of contemporary

440   environmental factors in community assembly, as surface bacterioplankton are likely subjected

441   to more variable physical-chemical conditions than those in deeper waters (Ghiglione et al.

442   2012).

443   Variation partitioning revealed that combined environmental factors accounted for

444   approximately 17% of the variation among bacterial assemblages. Independently, the spatial

445   component explained 11% of the dissimilarity among communities. The distance effect indicates

446   that processes other than contemporary selective factors have left a legacy on current microbial

447   community compositions, e.g., through past selection or ecological drift. At least some level of

448   dispersal limitation is required for past events to leave a contemporary signature, otherwise

449   community changes due to past selection or drift would be counteracted. Therefore, a significant

450   distance effect can be interpreted as evidence of dispersal limitation over some time period

451   (Hellweger et al. 2014). Significant influence from both environmental and spatial factors has

452   been identified in 68% of studies reporting on the subject, according to a literature review by

453   Hanson et al. (2012).

454   Environmental selection, primarily temperature, was stronger than dispersal limitation in

455   accounting for variation among bacterial assemblages (17% vs. 11%); this finding agrees with

456   several prior studies of marine bacterioplankton (Hanson et al. 2012, Hu et al. 2020, Milici et al.

457   2016, Sunagawa et al. 2015). Relatively weaker spatial patterning in bacteria supports the "size-

458   dispersal" hypothesis, i.e., that smaller organisms have higher dispersal rates, especially when

459   dispersal is primarily passive (De Bie et al. 2012). In contrast, some investigators have noted a

460   stronger influence of dispersal limitation relative to environmental factors in bacterioplankton

461   (Wu et al. 2018), in support of the 'size-plasticity' hypothesis, i.e., that smaller organisms are

462   less environment-filtered compared to larger organisms because they are more likely to be plastic

463   in metabolic abilities and have greater environmental tolerances. Several recent studies have

464   likewise examined the biogeographic patterns of microeukaryotes, at times in direct comparison

465   to bacteria, to shed additional light on assembly processes of marine microbes. With respect to

466   the size-plasticity vs. size-dispersal hypotheses, findings of these comparative studies have been

467   mixed – some have supported the size-dispersal hypothesis (Logares et al. 2018, 2020), while

468   others supported the size-plasticity hypothesis (Wu et al. 2018, Allen et al. 2020), and still others

469   that showed little difference between bacteria and protists with respect to relative roles of spatial

470   and environmental effects (Bock et al. 2020, Milici et al. 2016).

471   Redundancy analysis in this study demonstrated the importance of temperature in

472   determining bacterioplankton community structure. In addition, the clustering of South Sargasso

473   DCM and North Sargasso surface samples in the PCoA analysis showed that although these two

474   water masses are separated by distance, as well as light level, salinity, and other abiotic factors,

475   their respective communities were more similar to one another than to those from the same depth

476   category in other regions, likely due to their temperature similarity (Table 1, Fig. 5). Further, the

477   only region with a significant difference in communities with respect to depth (the North

478   Sargasso) also had the greatest difference in temperature (2.8 °C) compared to differences

479   between depths at other regions. One study focusing on widely distributed marine microbes such

480   as SAR11 and Actinobacteria, which we found to be dominant in Sargasso Sea communities,

481   indicated that temperature significantly influences community dissimilarity, even when

482   environmental differences associated with distance are eliminated (West et al. 2016). Further,

483   seasonal mixing events and productivity cycles created by shifts in temperature regimes have

484   also been shown to impact bacterial diversity (Wang et al. 2020).

485   *Sargassum* density was also identified as an important environmental influence on

486   bacterial assemblages based on our spatially-controlled redundancy analysis. While *Sargassum*

487   and other flotsam and debris found in the Sargasso Sea have been shown to harbor distinct and

488   diverse microbial communities (Zettler et al. 2013), the buoyant seaweed's absence from DCM

489   environments equivocates its importance on correlated variables, including temperature or

490   unmeasured variables such as light availability, which may instead drive differentiation with

491   depth.

492   A large fraction (72%) of variation among bacterial assemblages in our variation

493   partitioning analysis was left unexplained. Hanson et al. (2012) found that on average 50% of

494 variation in microbial composition was unexplained, although that fraction ranged widely among

495 studies. Unexplained variation is typically attributed to unmeasured environmental factors or

496 biotic interactions (Vass et al. 2020), but may also be due to stochastic processes (Hanson et al.

497 2012; Plante et al. 2016).

498 The NCM explained a large portion (~86%) of the variation in occurrence frequency of

499 photic-zone bacteria in the Northwest Atlantic Ocean, including for separate analyses of surface

500 waters and the DCM. Our findings are in general accord with Mo et al. (2018) who likewise

501 found stochastic factors to be primary in bacterioplankton community assembly in open-ocean

502 habitat, albeit at a somewhat lower level of dominance (57%). However, several other studies

503 found deterministic factors to be more important relative to neutral processes in similar scenarios

504 (Allen et al. 2020, Hu et al. 2020, Vergin et al. 2017, Wu et al. 2018). Although spatial factors

505 comprise a portion of the neutral process (Hubbell 2001), the strength of the NCM in our study

506 does not contradict the relatively small spatial effect revealed by variation partitioning. Spatial

507 factors (i.e., dispersal limitation) comprise just one part of the neutral process, whereas other

508 stochastic processes such as ecological drift, also contribute to community dissimilarity (Sloan et

509 al. 2006, 2007). The significant but weak distance-decay pattern that we observed at the surface

510 also indicates a low level of dispersal limitation (Hanson et al. 2012). The large stochastic effect

511 is likely ascribed to high and random dispersal of bacteria from a common, regional source pool

512 (Chen et al. 2017).

513

514 **Conclusions**

515 In this study, we compared bacterial community structures at the surface and at the DCM

516 within the Sargasso Sea and adjacent oceanographic regions. Community structures were

517   separable on the basis of physically-defined oceanographic regions and separable by depth in the

518   North Sargasso Sea. While both environmental differences and physical distance were found to

519   be correlated with greater community dissimilarity, these dissimilarities were amplified in

520   surface communities, and stochastic effects explained a significant proportion of variability in

521   bacterial occurrence frequency, demonstrating the complexity of interactions governing this

522   community. Future research is needed to further investigate biological interactions or co-

523   occurrences, which could help drive community structuring, and to understand the role of a

524   wider range of physical and chemical variables, such as dissolved oxygen content or particulate

525   matter.

526

527   **Acknowledgements**

528   We would like to offer a special thanks to the crew, researchers, and students aboard SEA

529   Semester's C-279 program for making our work at sea possible. This is Grice Marine Laboratory

530   contribution number XXX.

531 **References**

532 1. Allen, R., Hoffmann, L.J., Larcombe, M.J., Louisson, Z., Summerfield, T.C., 2020.
533 Homogeneous environmental selection dominates microbial community assembly in the
534 oligotrophic South Pacific Gyre. Mol. Ecol. 29, 4680-4691.
535 2. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J., 1990. Basic local
536 alignment search tool. J. Mol. Biol. 215, 403-410.
537 3. Agogué, H., Lamy, D., Neal, P.R., Sogin, M.L., Herndl, G.J., 2011. Water mass-specificity
538 of bacterial communities in the North Atlantic revealed by massively parallel sequencing.
539 Mol. Ecol. 20(2), 258-274.
540 4. Ardron, J., Halpin, P., Roberts, J., Cleary, J., Moffitt, M., Donnelly, J. 2011. Where is the
541 Sargasso Sea? A Report Submitted to the Sargasso Sea Alliance. Duke University Marine
542 Geospatial Ecology Lab & Marine Conservation Institute. Sargasso Sea Alliance Science
543 Report Series, No 2, 24 pp. ISBN 978-0-9847520-3-4
544 5. Aronesty, E., 2013. Comparison of sequencing utility programs. Open Bioinform. J. 7, 1-8.
545 6. Azam, F., Fenchel, T., Field, J.G., Gray, J.S., Meyer-Reil, L.A., Thingstad, F., 1983. The
546 ecological role of water-column microbes in the sea. Mar. Ecol. Prog. Ser. 10, 257-263.
547 7. Azam, F., 1998. Microbial control of oceanic carbon flux: The plot thickens. Science.
548 280(5364), 694-696.
549 8. Baas Becking, L.G.M., 1934. Geobiologie of inleiding tot de milieukunde. Van Stockum &
550 Zoon, The Hague.
551 9. Bock, C., Jensen, M., Forster, D., Marks, S., Nuy, J., Psenner, R., Beisser, D., Boenigk, J.,
552 2020. Factors shaping community patterns of protists and bacteria on a European
553 scale. Environ. Microbiol. 22, 2243-2260.
554 10. Borcard, D., Legendre, P., 2002. All-scale spatial analysis of ecological data by means of
555 principal coordinates of neighbor matrices. Ecol. Modelling. 153, 51-68.
556 11. Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Anderson, G.L., Knight, R.,
557 2009. PyNAST: a flexible tool for aligning sequences to a template alignment.
558 Bioinformatics. 26(2), 266-267.
559 12. Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K.,
560 Fierer, N., Pea, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D.,
561 Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder,
562 J., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat.
563 Methods. 7(5), 335-336.
564 13. Caruso, T., Chan, Y., Lacap, D.C., Lau, M.C., McKay, C.P. and Pointing, S.B., 2011.
565 Stochastic and deterministic processes interact in the assembly of desert microbial
566 communities on a global scale. ISME J. 5(9), 1406-1413.
567 14. Chen, W., Pan, Y., Yu, L., Yang, J., Zhang, W., 2017. Patterns and processes in marine
568 microeukaryotic community biogeography from Xiamen coastal waters and intertidal
569 sediments, southeast China. Front. Microbiol. 8, 1912.
570 15. Clarke, K.R., Gorley, R.N., 2006. PRIMER v6: User manual/tutorial. PRIMER-E: Plymouth.
571 16. Corno, G., Karl, D.M., Church, M.J., Letelier, R.M., Lukas, R., Bidigare, R.R., Abbott, M.R.,
572 2007. Impact of climate forcing on ecosystem processes in the North Pacific Subtropical
573 Gyre. J. Geophys. Res. 112, C04021.
574 17. Costello, M.J., Chaudhary, C., 2017. Marine biodiversity, biogeography, deep-sea gradients,
575 and conservation. Curr. Biol. 27, R511–R527.

576   18. Cotti-Rausch, B.E., Lomas, M.W., Lachenmyer, E.M., Goldman, E.A., Bell, D.W.,
577   Goldberg, S.R., Richardson, T.L., 2016. Mesoscale and sub-mesoscale variability in
578   phytoplankton community composition in the Sargasso Sea. Deep Sea Res. I. 110, 106-122.
579   19. De Bie, T., De Meester, L., Brendonck, L., Martens, K., Goddeeris, B., Ercken, D., Hampel,
580   H., Denys, L., Vanhecke, L., Van der Gucht, K., Van Wichelen, J., 2012. Body size and
581   dispersal mode as key traits determining metacommunity structure of aquatic
582   organisms. Ecol. Lett. 15, 740-747.
583   20. DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T.,
584   Daveli, D., Hu, P., Anderson, G.L., 2011. Greengenes, a chimera-checked 16S rRNA gene
585   database and workbench compatible with ARB. Appl. Environ. Microbiol. 72(7), 5069–5072.
586   21. Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., Knight, R., 2011. UCHIME improves
587   sensitivity and speed of chimera detection. Bioinformatics. 27(16), 2194–2200.
588   22. Edgar, R.C., 2017. SEARCH_16S: A new algorithm for identifying 16S ribosomal RNA
589   genes in contigs and chromosomes. bioRxiv, 124131 DOI: https://doi.org/10.1101/124131.
590   23. Finkel, Z.V., Beardall, J., Flynn, K.J., Quigg, A., Van Rees, T.A., Raven, J.A., 2010.
591   Phytoplankton in a changing world: cell size and elemental stoichiometry. J. Plankton Res.
592   32(1), 119-137.
593   24. Flombaum, P., Gallegos, J.L., Gordillo, R.A., Rincón, J., Zabala, L.L., Jiao, N., Karl, D.M.,
594   Li, W.K., Lomas, M.W., Veneziano, D., Vera, C.S., Vrugt, J.A., Martiny, A.C., 2013.
595   Present and future global distributions of the marine Cyanobacteria Prochlorococcus and
596   Synechococcus. Proc. Natl. Acad. Sci. U S A. 110(24):9824-9.
597   25. Follows, M.J., Dutkiewicz, S., 2011. Modeling diverse communities of marine microbes.
598   Annu. Rev. Mar. Sci. 3, 427-451.
599   26. Friedline, C.J., Franklin, R.B., McCallister, S.L., Rivera, M.C., 2012. Bacterial assemblages
600   of the eastern Atlantic Ocean reveal both vertical and latitudinal biogeographic signatures.
601   Biogeosciences. 9, 2177-2193.
602   27. Franzosa, E.A., Hsu, T., Sirota-Madi, A., Shafquat, A., Abu-Ali, G., Morgan, X.C.,
603   Huttenhower, C., 2015. Sequencing and beyond: integrating molecular 'omics' for microbial
604   community profiling. Nat. Rev. Microbiol. 13, 360-372.
605   28. Fuhrman, J.A., Steele, J.A., Hewson, I., Schwalbach, M.S., Brown, M.V., Green, J.L.,
606   Brown, J.H., 2008. A latitudinal diversity gradient in planktonic marine bacteria. P. Natl.
607   Acad. Sci. 105(22), 7774–7778.
608   29. Fuhrman, J.A., Cram, J.A., Needham, D.M., 2015. Marine microbial community dynamics
609   and their ecological interpretation. Nat. Rev. Microbiol. 13, 133-146.
610   30. Gajigan, A.P., Yñiguez, A.T., Villanoy, C.L., San Diego-McGlone, M.L., Jacinto, G.S.,
611   Conaco, C., 2018. Diversity and community structure of marine microbes around the
612   Benham Rise underwater plateau, northeastern Phillipines. PeerJ. DOI 10.7717/peerj.4781
613   31. Ghiglione, J.F., Galand, P.E., Pommier, T., Pedrós-Alió, C., Maas, E.W., Bakker, K.,
614   Bertilson, S., Kirchman, D.L., Lovejoy, C., Yager, P.L. and Murray, A.E., 2012. Pole-to-pole
615   biogeography of surface and deep marine bacterial communities. Proc. Natl. Acad.
616   Sci. 109(43), 17633-17638.
617   32. Giovannoni, S.J., Stingl, U., 2005. Molecular diversity and ecology of microbial plankton.
618   Nature. 437, 343-348.
619   33. Giovannoni, S.J., 2017. SAR11 bacteria: the most abundant plankton in the ocean. Annu.
620   Rev. Mar. Sci. 9, 231-255.

621   34. Gómez-Pereira, P.R., Fuchs, B.M., Alonso, C., Oliver, M.J., van Beusekom, J.E., Amann, R.,
622   2010. Distinct flavobacterial communities in contrasting water masses of the north Atlantic
623   Ocean. ISME J. 4, 472–487.
624   35. Govindarajan, A.F., Cooney, L., Whittaker, K., Bloch, D., Burdorf, R.M., Canning, S.,
625   Carter, C., Cellan, S.M., Eriksson, F.A.A., Freyer, H., Huston, G., Hutchinson, S.,
626   McKeegan, K., Malpani, M., Merkle-Raymond, A., Ouellette, K., Petersen-Rockney, R.,
627   Schultz, M., Siuda, A.N.S., 2019. The distribution and mitochondrial genotype of the hydroid
628   *Aglaophenia latecarinata* is correlated with its pelagic *Sargassum* substrate type in the
629   tropical and subtropical western Atlantic Ocean. PeerJ 7 p.e7814. DOI: 10.7717/peerj.7814.
630   36. Hahnke, R.L., Probian, C., Fuchs, B.M., Harder, J., 2013. Variations in pelagic bacterial
631   communities in the North Atlantic Ocean coincide with water bodies. Aquat. Microb. Ecol.
632   71,131-140.
633   37. Hamady, M., Lozupone, C., Knight, R. 2009. Fast UniFrac: Facilitating high-throughput
634   phylogenetic analyses of microbial communities including analysis of pyrosequencing and
635   PhyloChip data. ISME J. 4, 17-27.
636   38. Hanson, C.A., Fuhrman, J. A., Horner-Devine, M.C., Martiny, J.B., 2012. Beyond
637   biogeographic patterns: processes shaping the microbial landscape. Nat. Rev. Microbiol. 10,
638   497-506.
639   39. Hellweger, F.L., van Sebille, E., Fredrick, N.D., 2014. Biogeographic patterns in ocean
640   microbes emerge in a neutral agent-based model. Science. 345 (6202), 1346-1349.
641   40. Herlemann, D.P.R., Labrenz, M., Jurgens, K., Bertilsson, S., Waniek, J.J., Andersson, A.F.,
642   2011. Transitions in bacterial communities along the 2000km salinity gradient of the Baltic
643   Sea. ISME J. 5, 1571-1579.
644   41. Hörstmann, C., Raes, E.J., Buttigieg, P.L., Lo Monaco, C., John, U., Waite, A.M., 2021.
645   Hydrographic fronts shape productivity, nitrogen fixation, and microbial community
646   composition in the southern Indian Ocean and the Southern Ocean. Biogeosciences. 18,
647   3733-3749.
648   42. Hu, H., He, J., Yan, H., Hou, D., Zhang, D., Liu, L., Wang, K., 2020. Seasonality in spatial
649   turnover of bacterioplankton along an ecological gradient in the East China Sea:
650   biogeographic patterns, processes and drivers. Microorganisms. 8(10), 1484.
651   43. Hubbell, S.P. 2001. The Unified Neutral Theory of Biodiversity and Biogeography.
652   Princeton University Press, Princeton, NJ.
653   44. Keymer, D.P., Lam, L.H., Boehm, A.B., 2009. Biogeographic patterns in genomic diversity
654   among a large collection of *Vibrio cholerae* isolates. Appl. Environ. Microbiol. 75(6), 1658–
655   1666.
656   45. Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., Glöckner, F.O.,
657   2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-
658   generation sequencing-based diversity studies. Nucleic Acids Res. 41(1), e1.
659   46. Law, K.L., Morét-Ferguson, S., Maximenko, N.A., Proskurowski, G., Peacock, E.E., Hafner,
660   J., Reddy, C.M., 2010. Plastic accumulation in the North Atlantic Subtropical Gyre. Science
661   329, 1185-1188.
662   47. Legendre, P., Borcard, D., Roberts, D.W., 2012. Variation partitioning involving orthogonal
663   spatial eigenfunction submodels. Ecology. 93, 1234-1240.
664   48. Logares, R., Deutschmann, I.M., Giner, C.R., Krabberød, A.K., Schmidt, T.S., Rubinat-
665   Ripoll, L., Mestre, M., Salazar, G., Ruiz-González, C., Sebastián, M., de Vargas, C., 2018.

666   Different processes shape prokaryotic and picoeukaryotic assemblages in the sunlit ocean
667   microbiome. bioRxiv DOI: 10.1101/374298
668   49. Logares, R., Deutschmann, I.M., Junger, P.C., Giner, C.R., Krabberød, A.K., Schmidt, T.S.,
669   Rubinat-Ripoll, L., Mestre, M., Salazar, G., Ruiz-González, C., Sebastian, M., 2020.
670   Disentangling the mechanisms shaping the surface ocean microbiota. Microbiome. 8, 1-17.
671   50. Logares, R., Lindström, E.S., Langenheder, S., Logue, J.B., Paterson, H., Laybourn-Parry, J.,
672   Rengefors, K., Tranvik, L., Bertilsson, S., 2013. Biogeography of bacterial communities
673   exposed to progressive long-term environmental change. ISME J. 7, 937-948.
674   51. Lomas, M.W., Bates, N.R., Johnson, R.J., Knap, A.H., Steinberg, D.K., Carlson, C.A., 2013.
675   Two decades and counting: 24-years of sustained open ocean biogeochemical measurements
676   in the Sargasso Sea. Deep Sea Res. II. 93, 16-32.
677   52. Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing
678   reads. EMBnet.Journal. 17.1, 10-12.
679   53. Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L.,
680   Horner-Devine, M.C., Kane, M., Krumins, J.A., Kuske, C.R., Morin, P.J., Naeem, S.,
681   Ovreas, L., Reysenbach, A-L., Smith, V.H., Staley, J.T., 2006. Microbial biogeography:
682   putting microorganisms on the map. Nat. Rev. Microbiol. 4(2), 102-112.
683   54. Milici, M., Tomasch, J., Wos-Oxley, M.L., Decelle, J., Jáuregui, R., Wang, H., Deng, Z.L.,
684   Plumeier, I., Giebel, H.A., Badewien, T.H., Wurst, M., Piper, D.H., Simon, M., Wagner-
685   Döbler, I., 2016. Bacterioplankton biogeography of the Atlantic Ocean: a case study of the
686   distance-decay relationship. Front. Microbiol. 7, 590.
687   55. Mo, Y., Zhang, W., Yang, J., Lin, Y., Yu, Z., Lin, S., 2018. Biogeographic patterns of
688   abundant and rare bacterioplankton in three subtropical bays resulting from selective and
689   neutral processes. ISME J. 12, 2198-2210.
690   56. Moran, M.A., 2015. The global ocean microbiome. Science. 350(6266), 1330-1336.
691   57. Morris, R.M., Rappé, M.S., Connon, S.A., Vergin, K.L., Siebold, W.A., Carlson, C.A.,
692   Giovannoni, S.J., 2002. SAR11 clade dominates ocean surface bacterioplankton
693   communities. Nature. 420(6917), 806-810.
694   58. Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin,
695   P.R., O'Hara, R.B., Simpson, G.L., Solymos, P. and Stevens, M.H.H., 2020. vegan:
696   Community Ecology Package. R package version 2.5-6. 2019.
697   59. Pagenkopp Lohan, K.M., Fleischer, R.C., Carney, K.J., Holzer, K.K., Ruiz, G.M., 2017
698   Molecular characterisation of protistan species and communities in ships' ballast water
699   across three U.S. coasts. Biodivers. Res. 23, 680-69.
700   60. Pedrós-Alió, C., 2006. Marine microbial biodiversity: can it be determined? Trends
701   Microbiol. 14(6), 257-263.
702   61. Plante, C., Fleer, V., Jones, M.L., 2016. Neutral processes and species sorting in benthic
703   microalgal community assembly: Effects of tidal resuspension. J. Phycol. 52, 827-839.
704   62. Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2 – approximately maximum-likelihood
705   trees for large alignments. PLos One. 5(3): e9490.
706   63. Pommier, T., Canbäck, B., Riemann, L., Boström, K.H., Simu, K., Lundberg, P., Tunlid, A.,
707   Hagström, Å., 2007. Global patterns of diversity and community structure in marine
708   bacterioplankton. Molec. Ecol. 16, 867-880.
709   64. Raes, E.J., Letunic, I., Yamada, T., Jensen, L.J., Bork, P., 2011. Toward molecular
710   trait-based ecology through integration of biogeochemical, geographical and metagenomic
711   data. Molec. Systems Biol. 7(1), 473.

712   65. R Core Team. 2015. R: A language and environment for statistical computing. Vienna: R
713   Foundation for Statistical Computing http://www.R-project.org/.
714   66. Santoferrara, L.F., Grattepanche, J.D., Katz, L.A. and McManus, G.B., 2016. Patterns and
715   processes in microbial biogeography: do molecules and morphologies give the same
716   answers? ISME J. 10, 1779-1790.
717   67. Schlitzer, R., Ocean Data View 4.7.8, https://odv.awi.de, 2016.
718   68. Schwob, G., Segovia, N., Gonzalez-Wevar, C., Cabrol, L., Orlando, J. and Poulin, E.A.,
719   2021. Exploring the microdiversity within marine bacterial taxa: Towards an integrated
720   biogeography in the Southern Ocean. bioRxiv DOI:
721   https://doi.org/10.1101/2021.03.03.433356
722   69. Sehein, T., Siuda, A.N.S., Shank, T.M., Govindarajan, A.F., 2014. Connectivity in the
723   slender *Sargassum* shrimp (*Latreutes fucorum*): implications for a Sargasso Sea protected
724   area. J. Plankton Res. 36, 1408-1412.
725   70. Sloan, W.T., Lunn, M., Woodcock, S., Head, I.M., Nee, S., Curtis, T.P., 2006. Quantifying
726   the roles of immigration and chance in shaping prokaryote community structure. Environ.
727   Microbiol. 8(4), 732-740.
728   71. Sloan, W.T., Woodcock, S., Lunn, M., Head, I.M., Curtis, T.P., 2007. Modeling taxa-
729   abundance distributions in microbial communities using environmental sequence data.
730   Microb. Ecol. 53, 443-455.
731   72. Sjöstedt, J., Martiny, J.B.H., Munk, P., Riemann, L., 2014. Abundance of broad bacterial
732   taxa in the Sargasso Sea explained by environmental conditions but not water mass. Appl.
733   Environ. Microbiol. 80(9), 2786-2795.
734   73. Steinberg, D.K., Carlson, C.A., Bates, N.R., Johnson, R.J., Michaels, A.F., and Knap, A.H.,
735   2001. Overview of the US JGOFS Bermuda Atlantic Time-series Study (BATS): A decade-
736   scale look at ocean biology and biochemistry. Deep Sea Res. Part II Top. Stud. Oceanogr.
737   48(8-9), 1405-1447.
738   74. Strickland, J.D.H. and Parsons, T.R., 1972. A practical handbook of seawater analysis. Bull.
739   Fish. Res. Bd. Can. 2(167), 310 pp.
740   75. Siuda, A.N.S., 2011. Summary of Sea Education Association long-term Sargasso Sea surface
741   net data. Sargasso Sea Alliance Report Series. 10.
742   76. Stegen, J.C., Lin, X., Konopka, A.E. and Fredrickson, J.K., 2012. Stochastic and
743   deterministic assembly processes in subsurface microbial communities. ISME J. 6(9), 1653-
744   1664.
745   77. Sul, W.J., Oliver, T.A., Ducklow, H.W., Amaral-Zetter, L.A., Sogin, M.L., 2013. Marine
746   bacteria exhibit a bipolar distribution. Proc. Natl. Acad. Sci. 110(6), 2342-2347.
747   78. Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G.,
748   Djahanschiri, B., Zeller, G., Mende, D.R., Alberti, A. and Cornejo-Castillo, F.M., 2015.
749   Structure and function of the global ocean microbiome. Science. 348(6237).
750   79. Treusch, A.H., Vergin, K.L., Finlay, L.A., Donatz, M.G., Burton, R.M., Carlson, C.A.,
751   Giovannoni, S.J., 2009. Seasonality and vertical structure of microbial communities in an
752   ocean gyre. ISME J. 3, 1148-1163.
753   80. Vázquez-Baeza, Y., Pirrung, M., Gonzalez, A., Knight, R. 2013. EMPeror: a tool for
754   visualizing high-throughput microbial community data. Gigascience. 2(1), 16.
755   81. Vass, M., Székely, A.J., Lindström, E.S. and Langenheder, S., 2020. Using null models to
756   compare bacterial and microeukaryotic metacommunity assembly under shifting
757   environmental conditions. Sci. Rep. 10: 1-13.

758 82. Vergin, K.L., Jhirad, N., Dodge, J., Carlson, C.A., Giovannoni, S.J. 2017. Marine
759   bacterioplankton consortia follow deterministic, non-neutral community assembly rules.
760   Aquat. Microb. Ecol. 79, 165-175.
761 83. Walsh, E.A., Smith, D.C., Sogin, M.L. and D'Hondt, S., 2015. Bacterial and archaeal
762   biogeography of the deep chlorophyll maximum in the South Pacific Gyre. Aquat. Microb.
763   Ecol. 75(1), 1-13.
764 84. Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naïve Bayesian classifier for rapid
765   assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol.
766   73(16), 5261–5267.
767 85. Wang, M., Ma, Y., Feng, C., Cai, L., & Li, W., 2020. Diversity of pelagic and benthic
768   bacterial assemblages in the western Pacific Ocean. Front. Microbiol. 11, 1730.
769 86. West, N.J., Lepère, C., de O. Manes, C-L., Catala, P., Scanlan, D.J., Lebaron, P., 2016.
770   Distinct spatial patterns of SAR11, SAR86, and Actinobacteria diversity along a transect in
771   the ultra-oligotrophic South Pacific Ocean. Front. Microbiol. 7, 234.
772 87. Wu, X., Wu, L., Liu, Y., Zhang, P., Li, Q., Zhou, J., Hess, N.J., Hazen, T.C., Yang, W.,
773   Chakraborty, R., 2018. Microbial interactions with dissolved organic matter drive carbon
774   dynamics and community succession. Front. Microbiol. 9, 1234.
775 88. Zettler, E.R., Mincer, T.J., Amaral-Zettler, L.A., 2013. Life in the "Plastisphere": Microbial
776   communities on plastic marine debris. Environ. Sci. Technol. 47, 7137–7146.
777 89. Zheng, X., Dai, X., Huang, L., 2016. Spatial variations of prokaryotic communities in surface
778   water from India Ocean to Chinese marginal seas and their underlining environmental
779   determinants. Front. Mar. Sci. 3, 17.

780

Mean environmental values by depth and region.

Columns describe: depth category (DCM = deep chlorophyll maximum), oceanographic region (AC = Antilles Current, NA = North Atlantic, NSS = North Sargasso Sea, SSS = South Sargasso Sea, B = Bermuda), number of samples taken (*N*), depth, temperature (Temp), salinity, nitrate concentration, chlorophyll-*a* (chl-*a*) concentration, *Sargassum* density (*Sarg.* dens.). No samples were taken at depth in Bermuda due to gear deployment restrictions at port, and no samples were taken at the surface in the Antilles Current due to modifications in sample collection. NA indicates samples for which contamination or equipment maintenance prevented collection of data, ND indicates samples for which particular concentrations were below detectable levels.

1   **Table 1:** Mean environmental values (± standard error) by depth and region.
2
3   Columns describe depth category (DCM = deep chlorophyll maximum), oceanographic
4   region (B = Bermuda, NA = North Atlantic, NSS = North Sargasso Sea, SSS = South
5   Sargasso Sea, AC = Antilles Current), number of samples taken (*N*), depth, temperature
6   (Temp), salinity, nitrate concentration, chlorophyll-*a* (chl-*a*) concentration, *Sargassum*
7   density (*Sarg.* dens.). Values in parentheses indicate standard error. No samples were
8   taken at depth in Bermuda due to gear deployment restrictions at port, and no samples
9   were taken at the surface in the Antilles Current due to modifications in sample
10   collection. ND indicates samples for which contamination or equipment maintenance
11   prevented collection of data, BD indicates samples for which particular concentrations
12   were below detectable levels.
13

Depth category	Region	*N*	Depth (m)	Temp (°C)	Salinity (psu)	Nitrate (µM)	Chl-*a* (µg g⁻¹)	*Sarg.* dens. (g m⁻¹)
Surface	B	1	-	ND	ND	0.69	0.283	0
	NA	2	-	18.5 (3.75)	35.1 (1.38)	BD	0.144 (0.09)	0.03 (0.03)
	NSS	6	-	22.9 (0.24)	36.1 (0.36)	BD	0.030 (0.01)	0.03 (0.02)
	SSS	9	-	25.3 (0.10)	36.7 (0.04)	0.69 (0.02)	0.023 (0.00)	0.12 (0.04)
DCM	AC	3	146.6 (10.2)	26.3 (0.10)	36.5 (0.06)	0.02 (0.02)	0.232 (0.08)	-
	NA	2	42.5 (22.5)	16.7 (2.85)	36.1 (0.59)	0.09 (ND)	0.126 (0.07)	-
	NSS	5	94.0 (8.3)	20.1 (0.24)	36.8 (0.03)	0.64 (0.11)	0.273 (0.07)	-
	SSS	12	87.8 (11.8)	23.3 (0.55)	36.7 (0.09)	0.36 (0.10)	0.214 (0.03)	-

16
17
18

Mean α-diversity values by depth and region.

Columns describe: depth category (DCM = deep chlorophyll maximum), oceanographic region (AC = Antilles Current, NA = North Atlantic, NSS = North Sargasso Sea, SSS = South Sargasso Sea, B = Bermuda), number of samples taken (*N*), Shannon Biodiversity Index (H'), number of observed OTUs (S), Heip's evenness (E), and Chao1 richness index.

1   **Table 2:** Mean α-diversity values (± standard error) by depth and region.
2
3   Columns describe: depth category (DCM = deep chlorophyll maximum), oceanographic
4   region (B = Bermuda, NA = North Atlantic, NSS = North Sargasso Sea, SSS = South
5   Sargasso Sea, AC = Antilles Current), number of samples taken (*N*), Shannon
6   Biodiversity Index (H'), number of observed operational taxonomic units (S), Heip's
7   evenness (E), and Chao1 richness index. Values in parentheses indicate standard error.

Depth category	Region	*N*	H'	S	E	Chao1
Surface	B	1	7.47	1787	0.10	3214
	NA	2	7.24 (0.14)	1157 (48.0)	0.13 (0.01)	2122 (62.8)
	NSS	6	8.13 (0.07)	2150 (13.9)	0.13 (0.01)	4940 (72.3)
	SSS	9	7.86 (0.13)	1710 (17.0)	0.14 (0.02)	3409 (80.7)
DCM	AC	3	8.21 (0.23)	1707 (47.9)	0.17 (0.02)	3878 (266.8)
	NA	2	7.70 (0.14)	1754 (42.5)	0.12 (0.00)	3708 (306.5)
	NSS	5	8.21 (0.11)	2183 (9.22)	0.14 (0.02)	4894 (81.5)
	SSS	12	7.54 (0.18)	1715 (21.6)	0.11 (0.02)	3732 (82.8)

Figure 1

Map of research cruise track and stations.

Oceanographic regions delineated by white bars and sampling stations indicated by white circles. Darkest gray = land masses. White to gray background gradient indicates relative depth from shallow to deep. NA = North Atlantic, NSS = North Sargasso Sea, SSS = South Sargasso Sea, B = Bermuda, NC = North Carolina, USA, SC = South Carolina, USA. Figure generated using Ocean Data View 4.7.8 (Schlitzer 2016).
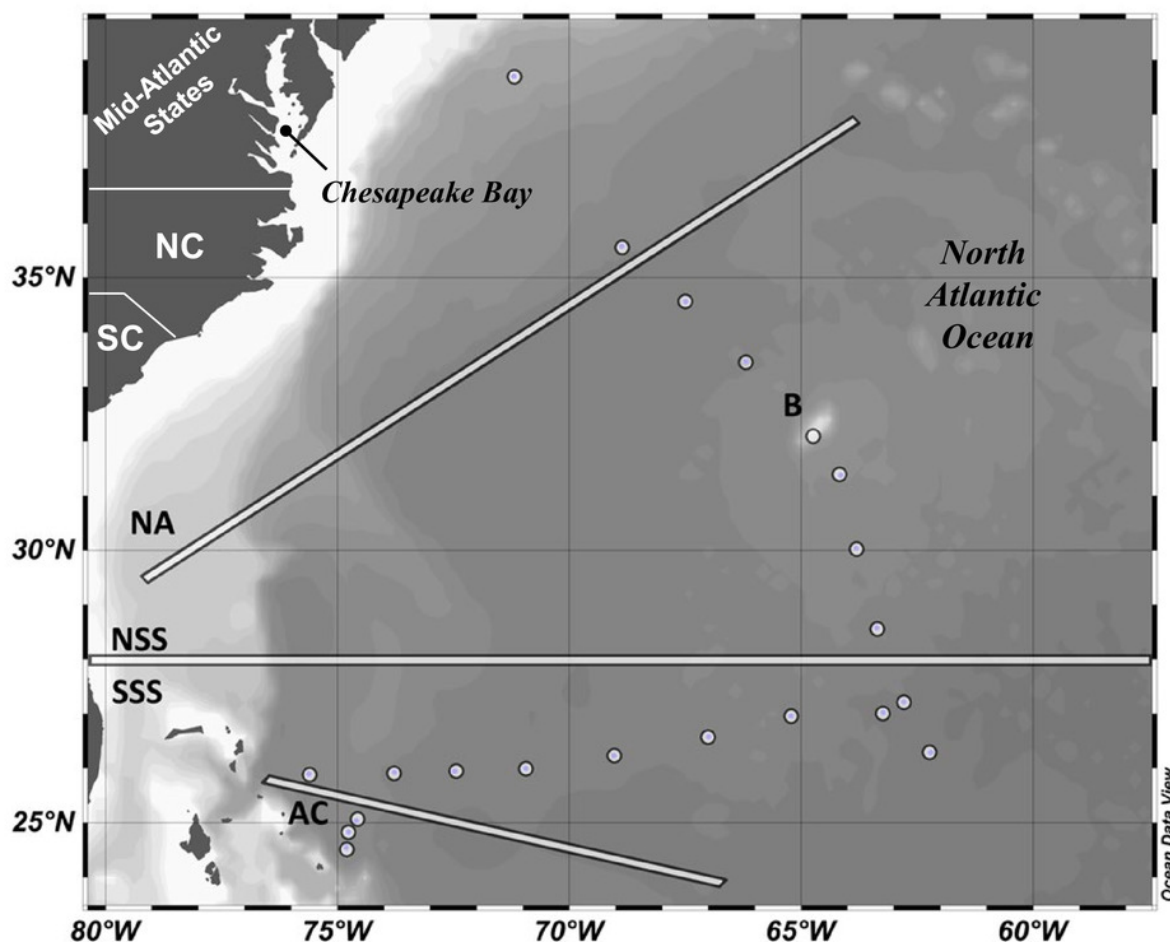
Figure 2

Community structure of all identified operational taxonomic units (OTUs) making up >3% of all samples taken from A) surface waters and B) the deep chlorophyll maximum (DCM).

Relative abundances do not sum to 100% as less abundant taxa were not included to allow for clear visualization of the data. Sequences of OTUs were compared to the National Center for Biotechnology Institute's GenBank database using BLAST (basic local alignment search tool; Altschul et al. 1990) to determine taxa identities to the lowest taxonomic level possible. All sequences from this dataset were 98-100% similar to sequences from GenBank, but most GenBank sequences were not identified to species level and only to genus, family (f), order (o), or class (c). Note that $N = 1$ for Bermuda and this sample was obtained from surface waters (A), and all Antilles Current samples ($N = 3$) were taken at the DCM. NA = North Atlantic, NSS = North Sargasso Sea, SSS = South Sargasso Sea, AC = Antilles Current, B = Bermuda.
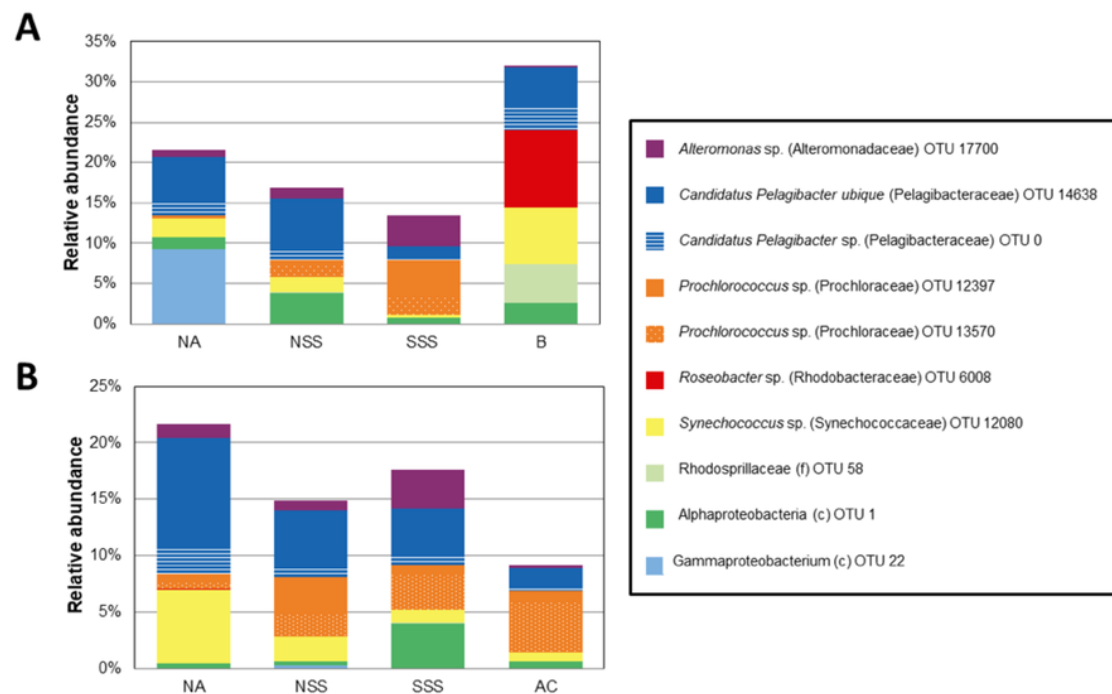
Figure 3

Relationship between number of unique and shared operational taxonomic units found among A) surface sampling locations and B) deep chlorophyll maximum sampling locations.

The diagram was generated using VENN (http://bioinformatics.psb.ugent.be/webtools/Venn/). NSS = North Sargasso Sea, SSS = South Sargasso Sea, NA = North Atlantic, AC = Antilles Current.

A

B
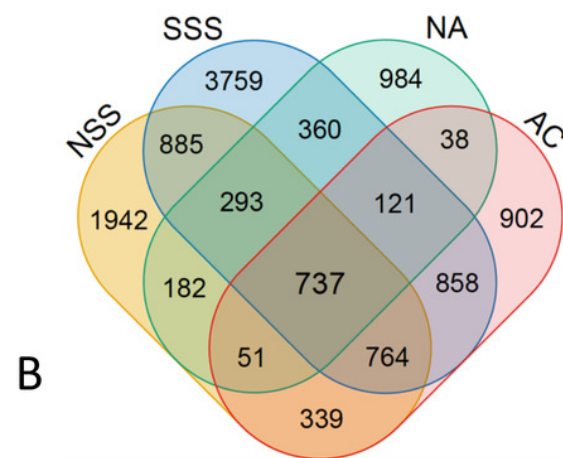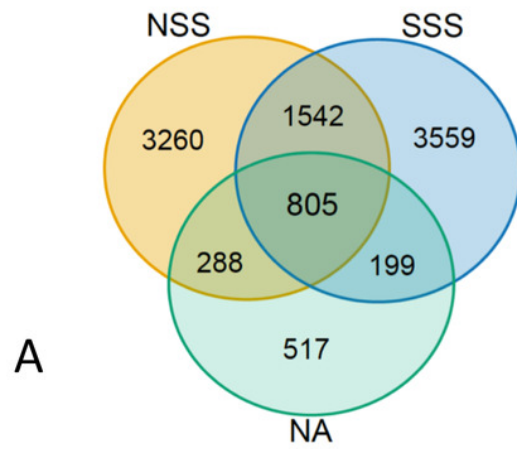
Figure 4

Distance-decay relationship between bacterial community similarity (weighted UniFrac) and geographic distance between sampling sites.

A) all data combined, B) surface samples only, and C) deep chlorophyll maximum samples only.

A) $R^2 = 0.001$

B) $R^2 = 0.037$

C) $R^2 = 0.001$
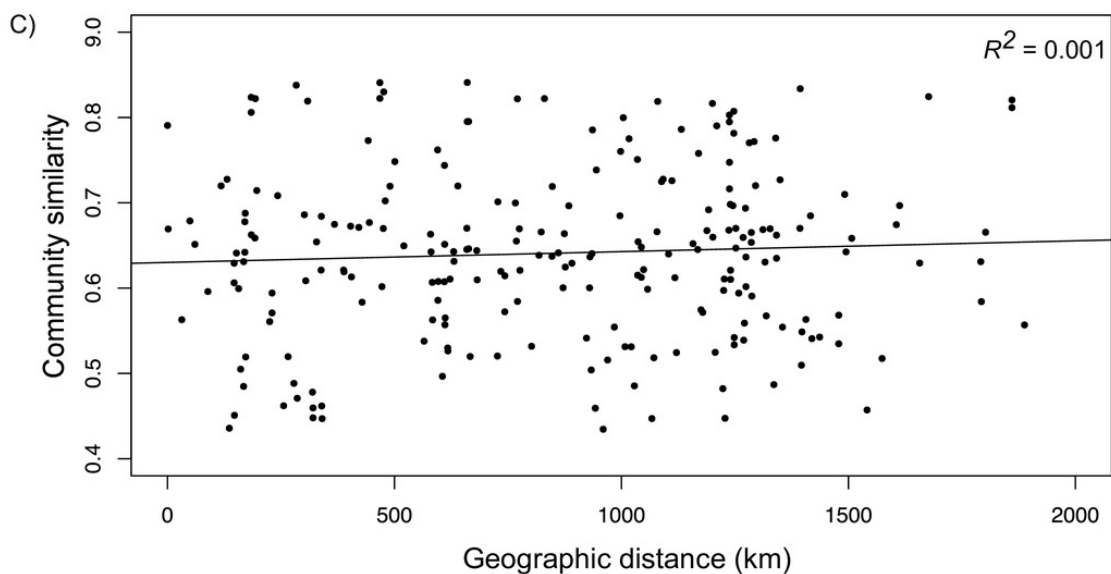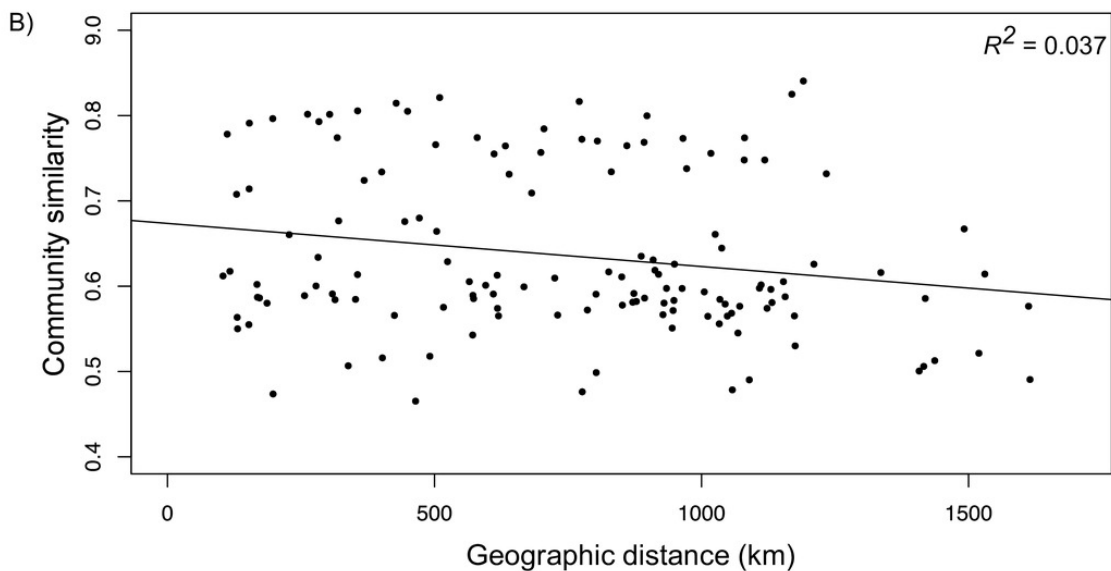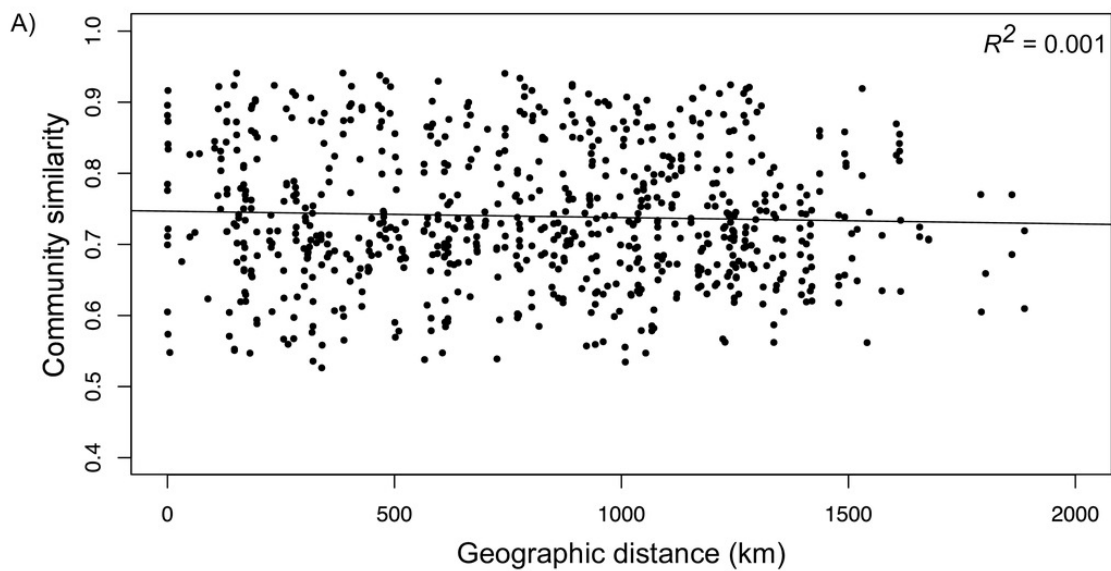
Figure 5

Principal coordinates analysis plot based on weighted UniFrac distance comparing bacterial communities from the oceanographic regions sampled.

AC = Antilles Current, SSS = South Sargasso Sea, NSS = North Sargasso Sea, NA = North Atlantic. Solid plot points represent communities from the deep chlorophyll maximum, outlined points represent surface communities.
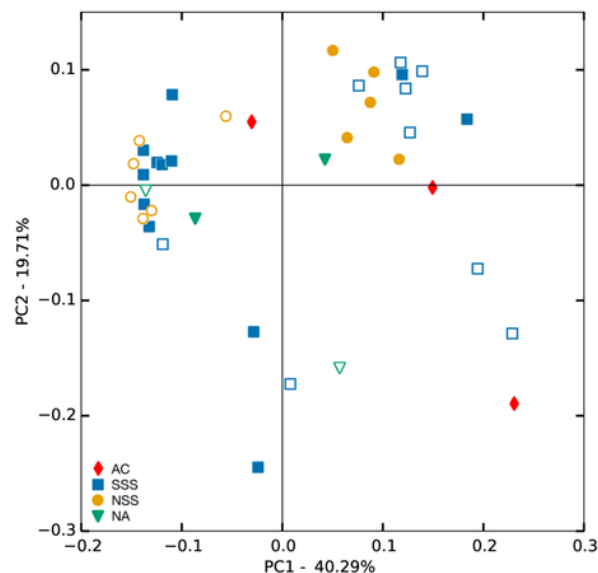
Figure 6

Fit of the neutral model of community assembly.

A) all, B) surface-only, and C) deep-only bacterioplankton. Red lines indicate the best fit to

the model as determined in Sloan et al. (2006). R^2 indicates the fit to the neutral model.