Spatial and temporal dynamics of SAR11 marine bacteria across a nearshore to offshore transect in the tropical Pacific Ocean

Sarah J Tucker [1, 2], **Kelle C Freel** [1], **Elizabeth A Monaghan** [1, 2], **Clarisse ES Sullivan** [1, 3], **Oscar Ramfelt** [1, 3], **Yoshimi M Rii** [1, 4], **Michael S Rappe** [Corresp. 1]

1 Hawaiʻi Institute of Marine Biology, School of Ocean and Earth Science and Technology, University of Hawaiʻi at Mānoa, Kāneʻohe, Hawaiʻi, United States

2 Marine Biology Graduate Program, University of Hawaiʻi at Mānoa, Honolulu, Hawaiʻi, United States

3 Department of Oceanography, School of Ocean and Earth Science and Technology, University of Hawaii at Mānoa, Honolulu, Hawaiʻi, United States

4 Heʻeia National Estuarine Research Reserve, Kāneʻohe, Hawaiʻi, United States

Corresponding Author: Michael S Rappe
Email address: rappe@hawaii.edu

Surveys of microbial communities across transitions coupled with contextual measures of the environment provide a useful approach to dissect the factors determining distributions of microorganisms across ecological niches. Here, monthly time-series samples of surface seawater along a transect spanning the nearshore coastal environment within Kāneʻohe Bay on the island of Oʻahu, Hawaiʻi, and the adjacent offshore environment were collected to investigate the diversity and abundance of SAR11 marine bacteria (order Pelagibacterales) over a two-year time period. Using 16S ribosomal RNA gene amplicon sequencing, the spatiotemporal distributions of major SAR11 subclades and exact amplicon sequence variants (ASVs) were evaluated. Seven of eight SAR11 subclades detected in this study showed distinct subclade distributions across the coastal to offshore environments. The SAR11 community was dominated by 7 (of 106 total) SAR11 ASVs that made up an average of 77% of total SAR11. These seven ASVs spanned five different SAR11 subclades (Ia, Ib, IIa, IV, and Va), and were recovered from all samples collected from either the coastal environment, the offshore, or both. SAR11 ASVs were more often restricted spatially to coastal or offshore environments (64 of 106 ASVs) than they were shared among coastal, transition, and offshore environments (39 of 106 ASVs). Overall, offshore SAR11 communities contained a higher diversity of SAR11 ASVs than their nearshore counterparts, with the highest diversity within the little-studied subclade IIa. This study reveals ecological differentiation of SAR11 marine bacteria across a short physiochemical gradient, further increasing our understanding of how SAR11 genetic diversity partitions into distinct ecological units

1 **Spatial and temporal dynamics of SAR11 marine bacteria across a nearshore to offshore**

2 **transect in the tropical Pacific Ocean**

3 Sarah J. Tucker [1,2], Kelle C. Freel [1], Elizabeth A. Monaghan [1,2], Clarisse E.S. Sullivan [1,3], Oscar

4 Ramfelt [1,3], Yoshimi M. Rii [1,4] & Michael S. Rappé [1]

6 [1]Hawaiʻi Institute of Marine Biology, School of Ocean and Earth Science and Technology,

7 University of Hawaiʻi at Mānoa, Kāneʻohe, Hawaiʻi, USA

8 [2]Marine Biology Graduate Program, University of Hawaiʻi at Mānoa, Honolulu, Hawaiʻi, USA

9 [3]Department of Oceanography, School of Ocean and Earth Science and Technology, University

10 of Hawaiʻi at Mānoa, Honolulu, Hawaiʻi, USA

11 [4]Heʻeia National Estuarine Research Reserve, Kāneʻohe, Hawaiʻi, USA

13 Corresponding Author:

14 Michael S. Rappé[1]

15 46-007 Lilipuna Road, Kāneʻohe, HI, 96744, USA

16 Email address: rappe@hawaii.edu

20 **Abstract**

21 Surveys of microbial communities across transitions coupled with contextual measures of the

22 environment provide a useful approach to dissect the factors determining distributions of

23 microorganisms across ecological niches. Here, monthly time-series samples of surface seawater

24 along a transect spanning the nearshore coastal environment within Kāneʻohe Bay on the island

25 of Oʻahu, Hawaiʻi, and the adjacent offshore environment were collected to investigate the

26 diversity and abundance of SAR11 marine bacteria (order Pelagibacterales) over a two-year time

27 period. Using 16S ribosomal RNA gene amplicon sequencing, the spatiotemporal distributions of

28 major SAR11 subclades and exact amplicon sequence variants (ASVs) were evaluated. Seven of

29 eight SAR11 subclades detected in this study showed distinct subclade distributions across the

30 coastal to offshore environments. The SAR11 community was dominated by 7 (of 106 total)

31 SAR11 ASVs that made up an average of 77% of total SAR11. These seven ASVs spanned five

32 different SAR11 subclades (Ia, Ib, IIa, IV, and Va), and were recovered from all samples

33 collected from either the coastal environment, the offshore, or both. SAR11 ASVs were more

34 often restricted spatially to coastal or offshore environments (64 of 106 ASVs) than they were

35 shared among coastal, transition, and offshore environments (39 of 106 ASVs). Overall, offshore

36 SAR11 communities contained a higher diversity of SAR11 ASVs than their nearshore

37 counterparts, with the highest diversity within the little-studied subclade IIa. This study reveals

38 ecological differentiation of SAR11 marine bacteria across a short physiochemical gradient,

39 further increasing our understanding of how SAR11 genetic diversity partitions into distinct

40 ecological units.

Introduction

43   The SAR11 order Pelagibacterales of the class Alphaproteobacteria is one of the most

44   abundant and ubiquitous bacterial lineages on Earth (Morris et al., 2002). While found

45   throughout the global ocean and freshwater environments, SAR11 bacteria are particularly

46   abundant in stratified, oligotrophic surface oceans, often making up 25% or more of all

47   bacterioplankton cells (Morris et al., 2002). SAR11 are chemoheterotrophic, free-living

48   microorganisms that are uniquely adapted to nutrient-poor environments through small cell sizes

49   and streamlined genomes (Rappé et al., 2002; Grote et al., 2012; Noell & Giovannoni, 2019).

50   Similar to other abundant marine bacteria and archaea, most SAR11 lineages are difficult to

51   culture in the laboratory (Swan et al., 2013), and the current capacity to interrogate SAR11

52   strains in a controlled laboratory setting remains limited to a small portion of the total SAR11

53   phylogenetic breadth (e.g. Jimenez-Infante et al., 2017; Henson et al., 2018; Monaghan, Freel &

54   Rappé, 2020).

55   The SAR11 clade is genetically diverse, with up to 18% small subunit (SSU) rRNA gene

56   sequence divergence distributed among at least five major subgroups (Brown et al., 2012) and

57   ten subclades including Ia, Ib, Ic, IIa, IIb, IIIa, IIIb, IV, Va, and Vb (Giovannoni, 2017). It

58   remains uncertain, however, if all five major subgroups derive from a single monophyletic

59   lineage within the Alphaproteobacteria as compositional biases and long evolutionary branches

60   have even made it a difficult phylogeny to resolve, particularly with respect to subgroup V

61   (Muñoz-Gómez et al. 2019). Some SAR11 subclades exhibit spatiotemporal distributions that are

62   delineated by depth, season, or geographical location (Field et al., 1997; Carlson et al., 2008;

63   Brown et al., 2012), revealing ecological differentiation at a broad phylogenetic level. The

64   ecotype concept describes an ecologically homogeneous group of closely related bacteria whose

65 genetic diversity is guided by cohesive forces such as periodic selection, recombination, and

66 genetic drift (Cohan, 2006; Koeppel et al., 2008). This concept has been helpful in discerning

67 populations among highly diverse and widely-distributed bacterial groups such as

68 *Prochlorococcus* (Biller et al., 2014) and *Bacillus* (Kopac et al., 2014). Previous studies of

69 SAR11 have shown evidence for ecotypic differentiation (Brown et al., 2012; Vergin et al.,

70 2013; Tsementzi et al., 2019; Kraemer et al., 2019). Yet, other research has attributed at least a

71 portion of the genetic diversity harbored by the SAR11 lineage to neutral processes (Hellweger,

72 van Sebille & Fredrick, 2014; Manrique & Jones, 2017).

73   The genetic diversity within SAR11 has presented a challenge in understanding the

74 ecological roles and evolutionary origins of the lineage. For example, rather than gene content,

75 measures of SAR11 genomic microdiversity including single-nucleotide polymorphisms and

76 single-amino acid variants were necessary to characterize two ecological niches among closely

77 related populations within the global Ia.3.V subgroup (Delmont et al., 2019). High intra-

78 population genomic variation has made it difficult to reconstruct SAR11 genomes from

79 metagenomic data (Tully, Graham & Heidelberg, 2018; Delmont et al., 2019). Even when

80 SAR11 genomes have been reconstructed from environmental samples, understanding the

81 boundaries between the sympatric populations that they represent has proven challenging

82 (Delmont et al., 2019).

83   16S rRNA gene sequencing surveys have been central to defining SAR11 subclades and

84 examining their spatiotemporal distributions in the environment (e.g. Field et al., 1997; Carlson

85 et al., 2008; Brown et al., 2012; Vergin et al., 2013; Salter et al., 2014; Herlemann et al., 2014;

86 West et al., 2016). For example, a multi-year study from the Bermuda Atlantic Time-Series

87 (BATS) in the Sargasso Sea has revealed that the relative abundance of some SAR11 subclades

88   changes with depth and seasonal regimes (Carlson et al., 2008; Vergin et al., 2013). In other

89   work, a combination of 16S rRNA gene and internal genomic spacer (ITS) analyses were able to

90   discern cold-water (Ia.1) and warm-water (Ia.3) variants of SAR11 subclade Ia (Brown et al.,

91   2012). Despite continued efforts to investigate SAR11 bacteria through 16S rRNA gene and ITS

92   surveys (Brown & Fuhrman, 2005; Ngugi & Stingl, 2012; Needham, Sachdeva & Fuhrman,

93   2017), fluorescence *in situ* hybridization (Alonso-Sáez et al., 2007; Salcher, Pernthaler & Posch,

94   2011), the cultivation of new strains (Jimenez-Infante et al., 2017; Henson et al., 2018;

95   Monaghan, Freel & Rappé, 2020), metagenomics (Tsementzi et al., 2016; Tsementzi et al., 2019;

96   Delmont et al., 2019; Haro-Moreno et al., 2019), and single-cell sequencing efforts (Thrash et

97   al., 2014; Thompson et al., 2019; Pachiadaki et al., 2019), the spatiotemporal characterization of

98   the majority of SAR11 diversity, particularly outside of subclade Ia, remains limited.

99   One useful approach to investigate the interplay between ecology and evolution within

100   natural microbial communities is to sample across environmental gradients. In the marine

101   environment, light, temperature, pressure, nutrients and other variables form steep gradients with

102   depth in the water column and have been used to investigate the partitioning of microorganisms

103   within different depth horizons of stratified offshore systems (Morris et al., 2005; Cram et al.,

104   2015; Haro-Moreno et al., 2018; Mende, Boeuf & DeLong, 2019). The nearshore to offshore

105   transition can also offer steep physiochemical gradients in nutrients, salinity, biomass, and

106   productivity which are likely to impact the structure of microbial communities (Herlemann et al.,

107   2014; Wang et al., 2019b). In tropical coastal environments like those found on islands and

108   atolls, this gradient can be acute; near-island biological, anthropogenic, and physical

109   oceanographic processes provide a substantial source of nutrients for increased biological

110   productivity in otherwise oligotrophic oceanic waters (Gove et al., 2016).

111   Kāneʻohe Bay is a well-studied, semi-enclosed embayment on the windward side of the

112   island of Oʻahu, Hawaiʻi (Bahr, Jokiel & Toonen, 2015). In this study, we used the steep

113   physiochemical gradient provided by the transition from the nearshore environment within

114   Kāneʻohe Bay to the adjacent offshore environment as a natural laboratory to investigate SAR11

115   marine bacteria across space and time. This system was sampled monthly over two years to

116   capture the microbial community from ten sites spanning the interior of the bay and the

117   surrounding offshore. We used 16S rRNA gene amplicon sequencing to assess the distribution of

118   SAR11 across a phylogenetic resolution that spanned subclades to individual SAR11 amplicon

119   sequence variants (ASVs). Our findings provide new insight into the ecological differentiation of

120   SAR11 subclades and the distribution of these subclades across the nearshore to offshore

121   continuum.

Materials and Methods

Sample collection and environmental parameters

Between August 2017 and June 2019, seawater was collected from a depth of 2 m at ten

sites in and around Kāneʻohe Bay, Oʻahu, Hawaiʻi, on a near-monthly basis (20 sampling events

over 23 months; **Fig. 1**). At each station, seawater samples for biogeochemical analyses and

nucleic acids were collected, and *in situ* measurements of seawater temperature, pH, and salinity

were made with a YSI 6600 sonde (YSI Incorporated, Yellow Springs, OH). Approximately 1 L

of seawater was prefiltered using an 85 μm Nitex mesh and subsequently collected on a 25-mm

diameter 0.1-μm pore-sized polyethersulfone (PES) membrane for nucleic acids (Supor-100, Pall

Gelman Inc., Ann Arbor, MI). The filters were submerged in DNA lysis buffer (Suzuki et al.,

2001; Yeo et al., 2013) and stored at -80°C until extraction.

Subsamples for chlorophyll *a* were collected by filtering 125 mL of seawater onto 25-mm

diameter GF/F glass microfiber filters (Whatman, GE Healthcare Life Sciences, Chicago, IL,

USA), and stored in aluminum foil at -80°C until extraction in 100% acetone and subsequent

measurement of fluorescence with a Turner 10AU fluorometer (Turner Designs, Sunnyvale, CA)

followed standard techniques (Welschmeyer, 1994). Seawater for cellular enumeration was

preserved in 2-mL aliquots in a final concentration of 0.95% (v:v) paraformaldehyde (Electron

Microscopy Services, Hatfield, PA) at -80°C until analyzed via flow cytometry. Cellular

enumeration of cyanobacterial picophytoplankton (marine *Synechococcus* and *Prochlorococcus*),

eukaryotic picophytoplankton, and non-cyanobacterial (presumably heterotrophic) bacteria and

archaea (hereafter referred to as heterotrophic bacteria) was performed on an EPICS ALTRA

flow cytometer (Beckman Coulter Inc., Brea, CA), following the method of Monger and Landry

(Monger & Landry, 1993).

146   Seasons were delineated by fitting a harmonic function to surface seawater temperature

147   collected hourly between 2010-2019 at NOAA station MOKH1 in Kāneʻohe Bay

148   (https://www.ndbc.noaa.gov/station_page.php?station=mokh1; **Fig. S1A**). Transition seasons

149   (spring and fall) typically experience the greatest amount of change in seawater temperature.

150   Thus, using the time-derivative of the function, thresholds were defined where the derivative was

151   at its greatest absolute value: $\geq$0.0255 (spring; 30 March through 27 June 2017-19) and $\leq$ -

152   0.0255 (fall; 29 September through 26 December 2017-19; **Fig. S1B**). Summer and winter were

153   defined as the periods between the thresholds (i.e. when the derivative was <0.0255 to >-0.0255;

154   summer: 28 June through 28 September 2017-19 and winter: 27 December through 29 March

155   2018-19).

156   In RStudio (version 1.1.456) (R Core Team, 2018) with the Vegan package (Oksanen et

157   al., 2019), stations were grouped into environment types based on non-metric rank-based

158   analysis using the 'metaMDS' function with k=2 and a Bray-Curtis transformed distance matrix

159   of environmental parameters (surface seawater temperature, pH, salinity), chlorophyll *a*

160   concentrations, and cellular abundances of *Prochlorococcus, Synechococcus*, eukaryotic

161   picophytoplankton, and heterotrophic bacteria. The pairwiseAdonis package (Martinez Arbizu,

162   2020) and the 'betdisper' function in the Vegan package were used to evaluate the statistical

163   significance and dispersion of these groupings. Comparisons of environmental variables, cellular

164   abundances, and sequencing depth across the spatial and temporal groupings were conducted

165   using the package multcomp (Hothorn, Bretz & Westfall, 2008) with one-way ANOVAs testing

166   for multiple comparisons of means with Holm correction and Tukey contrasts.

167

168   **DNA extraction and sequencing**

169  Genomic DNA was extracted using a Qiagen Blood and Tissue Kit (Qiagen Inc.,

170 Valencia, CA) with modifications (Becker, Brandon & Rappé, 2007). For each sample, 16S

171 rRNA gene fragments were amplified by polymerase chain reaction using a dual-index

172 sequencing strategy where barcoded universal primers 515-Y-F and 926R (Parada, Needham &

173 Fuhrman, 2016) are complete with Illumina sequencing adapters, barcode, and index. The 25 µL

174 reactions included 13 µL H$_2$O, 0.5 µL each of forward and reverse primer at 0.2 µM final

175 concentration, 1 µL gDNA (0.5 ng), and 10 µL 1x 5PRIME Hot Master Mix (0.5 U *Taq* DNA

176 polymerase, 45 mM KCl, 2.5 mM Mg^{2+}, 200 µM dNTPs) (Quantabio, Beverly, MA, USA). PCR

177 conditions included an initial denaturation at 95 °C for 2 min followed by 30 cycles of 95 °C for

178 45 s, 50 °C for 45 s, and 68 °C for 90 s, and a final 5 min extension at 68 °C. PCR products were

179 inspected on a 1.5% agarose gel and quantified using the Qubit dsDNA HS Kit (Qubit 2.0, Life

180 Technologies, Foster City, CA, USA). PCR products were normalized to 240 ng each, pooled,

181 and purified using the QIAquick PCR Purification Kit (Qiagen). Pooled libraries were then

182 sequenced on an Illumina MiSeq v2 250 bp paired-end run at the Oregon State University Center

183 for Genome Research & Biocomputing.

184

185 **Sequence analysis**

186  The sequence data were demultiplexed and assessed for quality in Qiime2 v 2019.4

187 (Bolyen et al., 2019). Forward reads were truncated to 245 bp using the --p-trunc-len command

188 in Qiime2 to improve their quality. Reverse reads were not used in these analyses due to poor

189 quality. Sequences were denoised using DADA2 (Callahan et al., 2016) and delineated into

190 exact amplicon sequence variants (ASVs). Denoised sequence data were then assigned taxonomy

191 using SILVA v132 as a reference database (Quast et al., 2012). Mock communities were

192 assessed to establish that the denoising of sequences using single-end reads was representative of

193 taxa and abundances expected in mock community samples.

194   ASVs that contained a minimum of 10 reads in at least two independent samples were

195 retained for subsequent analyses. ASVs undefined at the level of domain and uncharacterized at

196 the level of phylum were excluded from subsequent analyses. Sequences that matched to

197 chloroplasts or mitochondria were included in analyses of sequencing depth (**Table S2, S3, S4**),

198 ASV occurrence (**Table S5**), and community structure (**Figure 2A, Figure S4B, Table S6**), but

199 removed in all subsequent analyses of SAR11 including calculations of the relative proportion of

200 SAR11 within the microbial community and DESeq2 analysis of differences in SAR11,

201 *Prochlorococcus*, and *Synechococcus* abundance across environments and seasons.

202   ASVs classified by the SILVA v132 database as the bacterial order SAR11 and the

203 family AEGEAN-169 within the order Rhodospirillales (for SAR11 subgroup V) were assigned

204 to SAR11 subclades using phylogenetic placement methods. First, a reference alignment was

205 created in ARB (Ludwig, 2004) from 16S rRNA gene sequences downloaded from NCBI and

206 SAR11 ASVs identified in this study. Reference sequences were selected based on similarity to

207 ASVs, and also included all known cultivated SAR11 strains. Two outgroup representatives were

208 included: *Escherichia coli* (NR_024570.1) and *Magnetococcus marinus* (NR_074371.1). We

209 applied a mask to the alignment using filter by frequency in ARB in order to remove columns

210 containing gaps or no data, and exported the resulting 1224 bp alignment. The mask did not

211 remove any columns in the 245 bp region where the ASVs aligned. Using only the aligned

212 reference sequences, the best reference tree and model parameters were selected by performing

213 tree inference with a random starting tree, maximum-likelihood estimate of substitution rates and

214   nucleotide frequencies, a general time-reversible + gamma model, and 100 non-parametric

215   bootstrap replicates via RAxML- NG (Kozlov et al., 2019).

216   Next, the SAR11 ASVs aligned to the reference sequences were phylogenetically placed

217   using the EPA-NG algorithm (Barbera et al., 2018), specifying the model parameters

218   [GTR(.960433/4.369000/2.298538/0.794304/6.601978/1.000000)+FU(0.271858/0.192063/0.300

219   579/0.235499)+G4m(0.276190)], which were previously determined when conducting tree

220   inference with RAxML-NG. Placement results of the ASVs on the reference tree were visualized

221   in GAPPA (Czech & Stamatakis, 2019) using the function 'gappa examine graft' **(Fig. S2)** and

222   likelihood weights of each ASV placement to SAR11 subclades were evaluated using 'gappa

223   examine assign'. The tree was manipulated in FigTree v1.4.3. (available from

224   http://tree.bio.ed.ac.uk/software/figtree/) to improve visualization.

225   Statistical analyses and visualizations were conducted in R packages phyloseq

226   (McMurdie & Holmes, 2013), plotly (Sievert, 2013), ggplot2 (Wickham, 2016), and pheatmap

227   (Kolde, 2019), as well as the online tool Venny 2.1 (Oliveros, 2017). Generalized linear models

228   (GLMs) were built in mvabund (Wang et al., 2019a) using the 'manyglm' function to test for

229   differences in the number of ASVs belonging to SAR11 subclades across seasons and

230   environments. DESeq2 (Love, Huber & Anders, 2014) was used to normalize ASV abundance

231   and to evaluate whether individual ASVs exhibited significantly different distributions across

232   seasons or environments using Wald Tests, Bonferroni correction for multiple comparisons, and

233   a corrected p-value of 0.05. The same approach was used to normalize abundance and test for

234   spatiotemporal differences in the top 14 most abundant bacterial orders, SAR11 subclades, and

235   total SAR11, *Synechococcus*, *Prochlorococcus*, and chloroplast abundance.

236   The frequency of SAR11 ASVs was assessed by segregating the ASVs into rare (0-5%),

237 low-frequency (>5-25%), mid-frequency (25-75%), and high-frequency (75-100%) categories.

238 Frequency was calculated as the number of samples an ASV was detected in per environment,

239 divided by the number of total samples for a given environment (coastal n=120; transition n=40;

240 offshore n=40), multiplied by 100.

241   Variance stabilized transformed DESeq2 normalized counts of SAR11 subclades and

242 dominant ASVs and picocyanobacteria along with environmental parameters were used to

243 conduct Spearman's Rank correlation analysis using the 'cor' function in the stats package (R

244 Core Team, 2018). Using a matrix of the Spearman's correlation coefficients, between-group

245 average (UPGMA) linkage hierarchical clustering was conducted using the function 'hclust' in

246 the stats package (R Core Team, 2018). This dendrogram was then used to order the correlation

247 coefficients matrix for visualization.

248   SAR11 ASVs were compared to SAR11 isolates by aligning the SAR11 ASVs to all

249 publicly available 16S rRNA gene sequences from SAR11 isolates using the MUSCLE

250 algorithm in MEGAX (Kumar et al., 2018).

251

252 **Data availability**

253   Amplicon sequencing data are available in the National Center for Biotechnology

254 Information (NCBI) Sequencing Read Archive (SRA) under BioProject number PRJNA706753.

255

256 **Results**

257 **Environmental conditions**

258   Non-DNA sequence-based parameters partitioned the Kāneʻohe Bay Time-series (KByT)

259   sampling sites into three spatial groups, hereafter referred to as coastal (to define the nearshore

260   grouping), transition, and offshore (**Fig. S3**). A pairwise-adonis analysis showed statistically

261   significant differences among the three groups (coastal vs. transition: p=0.001, R^2=0.31; coastal

262   vs. offshore: p=0.001, R^2=0.43; transition vs. offshore: p=0.012, R^2=0.071). While significant,

263   we note that the difference between the transition and offshore sampling sites is small.

264   Dispersion tests found non-significant dispersion (p=0.115, F-value=2.19).

265   Compared to the transition and offshore stations, coastal stations were defined by lower

266   salinity, higher chlorophyll *a* concentrations, higher cellular abundances of heterotrophic

267   bacteria, *Synechococcus*, and eukaryotic picophytoplankton, and lower cellular abundances of

268   *Prochlorococcus* (**Fig. 1, Tables 1 & S1**). The cellular abundance of *Prochlorococcus* was

269   significantly different across all three environments (one-way ANOVA with Tukey's multiple

270   comparisons test, p<0.001), ranging from $0.2 \pm 0.7 \times 10^4$ cells mL^{-1} at coastal stations [mean ±

271   standard deviation (s.d.), n=120] to $2.2 \pm 2.1 \times 10^4$ cells mL^{-1} at transition stations (n=40) and 6.2

272    $\pm 2.3 \times 10^4$ cells mL^{-1} at offshore stations (n=40). pH was significantly different between coastal

273   (7.9 ± 0.2, n=120) and offshore stations (8.0 ± 0.2, n=40, p=0.002) and transition (7.9 ± 0.2,

274   n=40) and offshore stations (p=0.015); no significant difference in pH was detected between

275   coastal and transition stations (p=0.841). While not statistically significant, transition stations

276   had lower salinity, higher chlorophyll *a* concentrations, and higher cellular abundances of

277   heterotrophic bacteria, *Synechococcus*, and eukaryotic picophytoplankton in comparison to

278   offshore stations (**Table 1**). Sea surface water temperature did not statistically differ across the

279   three environments.

280   An increase in salinity and cellular abundances of heterotrophic bacteria and

281   *Synechococcus* were observed during spring (**Tables 1 & S1**). At coastal stations, salinity ranged

282   from 34.0 ± 1.2 (mean $\pm$ s.d., n=24) during summer to 35.1 ± 1.3 (n=36) during spring, with

283   significant differences between spring vs. summer (one-way ANOVA with Tukey's multiple

284   comparisons test, p<0.001), fall (p=0.037), and winter (p=0.002). Salinity also reached its

285   highest during spring at transition (35.5 ± 1.2, n=12) and offshore (35.6 ± 1.2, n=12) stations.

286   Increases in the cellular abundance of heterotrophic bacteria between spring and fall were

287   observed in the coastal (p=0.033) and in the offshore environments (p=0.037). The abundance of

288   *Synechococcus* cells increased during spring ($25.6 \pm 9.7 \times 10^4$ cells mL^{-1}, n=36) and summer

289   ($26.6 \pm 12.7 \times 10^4$ cells mL^{-1}, n=24) in the coastal environment in comparison to fall (16.8 ± 13.5

290    $\times 10^4$ cells mL^{-1}, n=24) and winter ($16.5 \pm 12.\times 10^4$ cells mL^{-1}, n=36; p=0.012). In contrast, the

291   abundance of *Prochlorococcus* cells increased in the coastal environment during fall (0.7 ± 0.1

292    $\times 10^4$ cells mL^{-1}, n=24). pH, chlorophyll *a* concentration, and the abundance of eukaryotic

293   picophytoplankton cells showed no significant changes over seasons. Surface seawater

294   temperatures were significantly different across all but two seasonal comparisons (spring vs. fall

295   in the coastal environment (p=0.876) and summer vs. fall in the transition and offshore

296   environments (p=0.093 and p=0.157, respectively)).

297

298   **Community composition**

299   Samples were sequenced to a depth of $25,684 \pm 11,226$ quality-controlled reads (mean $\pm$

300   s.d., n=200), with a range of 9,393 to 100,278 (**Table S2**). Read depth did not differ across

301   environment type or station (one-way ANOVA with Tukey's multiple comparisons test; p=1 all

302   comparisons, **Table S3 & S4**), although significant variation was detected among some seasonal

303   comparisons (spring vs fall, winter vs fall, summer vs spring, and winter vs summer, p<0.001;

304   **Table S3**). A total of 2,280 unique ASVs were detected across the 200 samples, including 2,241

305   bacteria and 39 archaea (**Table S5**). The majority of ASVs were distributed across the bacterial

306   phyla Proteobacteria (n=1,052), Bacteriodetes (n=325), Verrucomicrobia (n=60), Marinimicrobia

307   (n=54), and Cyanobacteria (n=39). A total of 350 ASVs belonged to chloroplasts. A large

308   portion (22.3%, n=509) of ASVs were detected only in the coastal environment, compared to

309   7.4% unique to the offshore (n=169) and 2.8% unique to transition (n=63) stations.

310   Community composition was further assessed by examining the top 14 most abundant

311   orders measured as a proportion of the entire community. Synechococcales was the most

312   abundant order detected in the coastal environment (26.3 ± 10.6%; mean ± s.d., n=120), while

313   Flavobacteriales (24..7 ± 8.1%, n=40) and Pelagibacterales (31.6 ± 7.6%; n=40) were the most

314   abundant orders detected in the transition and offshore environments, respectively (**Fig. 2A**). 8 of

315   the 14 orders examined showed significant differences between all comparisons of environment

316   types (coastal vs transition, transition vs offshore and coastal vs offshore, p<0.001) when

317   evaluated with DESeq2 normalization (Pelagibacterales, Rhodobacterales, Puniceispirillales,

318   SAR86, Actinomarinales, Betaproteobacteriales, Euryarchaeota Marine Group II, and

319   Parvibaculales; **Table S5)**. The remaining orders showed significant differences across two

320   comparisons of environment types (coastal vs transition and transition vs offshore:

321   Synechococcales, Rickettsiales, Flavobacteriales; coastal vs offshore and transition vs offshore:

322   chloroplast and Cellvibrionales; coastal vs transition and coastal vs offshore: Oceanospirillales;

323   p<0.001, **Table S6)**.

324   Within the Synechococcales, the combination of *Synechococcus* and *Prochlorococcus*

325   compromised between 89.0 ± 4.4% and 98.0 ± 2.5% of the total Synechococcales relative

326   abundance at a given station (**Fig. 2B**). *Synechococcus* dominated the microbial community in

327   the coastal stations (27.4 ± 10.5%, mean ± s.d., n=120), declining in the transition (12.5 ± 8.3%,

328   n=40) and offshore environments (10.7 ± 8.0%, n=40) (**Fig. 2B**). In contrast, *Prochlorococcus*

329   was most abundant in the offshore environment (13.3 ± 0.1%), declining in the transition (3.3 ±

330   0.0%) and coastal environments (0.3 ± 0.0%). The relative abundance of *Synechococcus* was

331   roughly two-fold higher in sequence data than in flow cytometry data (**Fig. 2B-C**). In contrast,

332   *Prochlorococcus* had similar relative abundances between sequence data and flow cytometry

333   data.

334   The seasonality of phytoplankton taxa was evaluated using DESeq2 normalized

335   comparisons of chloroplast, *Synechococcus*, and *Prochlorococcus* total abundance as defined by

336   sequence data. No significant seasonal patterns in chloroplast abundance were detected across

337   the three environments (Wald test, p>0.05 all comparisons, **Fig. S4)**. In the coastal environment,

338   *Synechoccocus* increased in spring and summer (winter vs spring, p<0.001; summer vs winter,

339   p=0.015; fall vs spring, p=0.042). *Synechococcus* decreased in the winter and increased in the

340   summer in both the transition (winter vs summer, p=0.049; winter vs spring= 0.009) and

341   offshore environments (winter vs summer, p=0.002; winter vs spring, p=0.014).

342   *Prochlorococcus* showed an increase in fall in the coastal environment (fall vs spring, p=0.006;

343   fall vs summer, p<0.001; winter vs summer, p<0.001 **Fig. S4)**.

344

345   **Spatial and temporal distributions of SAR11 subclades**

346   SAR11 accounted for 106 ASVs distributed across eight subclades: Ia, Ib, IIa, IIb, IIIa,

347   IV, Va, and Vb (**Table S7**). Individual samples harbored between 7 and 55 SAR11 ASVs,

348   averaging 14 ± 4 SAR11 ASVs in the coastal environment (mean ± s.d., n=120), 17 ± 6 in the

349   transition environment (n=40), and 28 ± 9 in the offshore environment (n=40). Generalized

350   linear models with Poisson distributions revealed that the number of SAR11 ASVs recovered per

351   sample varied significantly across environments [Likelihood Ratio Test (LRT)=1168, p=0.001],

352   but did not vary significantly across seasons (LRT=73.49, p=0.114). While sequencing depth did

353   not vary over environment or station (**Table S4**), we do note that significant seasonal differences

354   in sampling depth could obscure these results (**Table S3**). Of the eight SAR11 subclades

355   identified across KByT, subclade IIa harbored the highest number of unique ASVs (n=41),

356   followed by Ia (n=22) and Ib (n=19) (**Table S7**). Within individual subclades, the average

357   number of ASVs detected in a sample significantly differed between environments for subclades

358   Ib (LRT=641.52, p=0.001), IIa (LRT= 194.92, p=0.001), IIb (LRT= 19.31, p=0.001), and Vb

359   (LRT= 13.78 p=0.001) (**Table S7**).

360   Collectively, the relative abundance of SAR11 increased from coastal stations (21.6 ±

361   4.7%, mean ± s.d., n=120), to transition (23.2 ± 6.4%, n=40) and offshore stations (34.2 ± 7.2%,

362   n=40). DESeq2 normalization revealed significant differences in total SAR11 abundance

363   between the three environments (Wald test, coastal vs transition, p<0.001; transition vs offshore,

364   p<0.001; coastal vs offshore, p<0.001, **Fig. S5**). As a whole, the abundance of SAR11 did not

365   significantly differ across seasons within the transition or offshore environments, but did in the

366   coastal environment where it reached is maximum in spring (winter and spring p=0.002; fall and

367   spring, p<0.001; winter and summer, p=0.050; fall and summer, p=0.035; **Fig. S5).**

368   SAR11 subclade Ia was the most abundant subclade detected across all stations, with a

369   relative abundance that was fairly consistent across coastal (14.6 ± 3.9%, mean ± s.d., n=120),

370   transition (14.2 ± 3.9%, n=40), and offshore (14.3 ± 2.3%, n=40) environments (**Fig. 3**). In

371   contrast, subclade IIb was the least abundant of all SAR11 subclades detected throughout KByT,

372   reaching a maximum relative abundance of $0.01 \pm 0.04\%$ (mean $\pm$ s.d., n=20) at offshore station

373   NTO1. Both subclades Ib and IIa increased in relative abundance in offshore stations compared

374   to their coastal counterparts (**Fig. 3**). While spatial differences were most pronounced for SAR11

375   subclades Ib and IIa, significant differences in relative abundance across the coastal to offshore

376   transect were detected for seven of the eight subclades recovered in this study (**Fig. S7**). DESeq2

377   normalization revealed that subclades Ib, IIa, IV, and Vb were more abundant in offshore

378   compared to coastal stations (Wald test; $p<0.001$), while subclades Ia and IIIa were more

379   abundant at coastal stations compared to the offshore ($p<0.001$). Subclades Ib, IIa, IV, and Va

380   also differed between the coastal and transition environments ($p<0.001$, except for IV where

381    $p=0.002$), while subclades Ia, Ib, IIa, IIIa, IV, Va, and Vb differed between transition and

382   offshore environments ($p<0.001$, except for Ib where $p=0.017$).

383   Significant seasonal differences in the abundance of individual subclades were observed

384   in both coastal and offshore environments, but not in the transition environment (**Fig. S8**). These

385   include (i) subclade IIIa, which decreased during winter in both the coastal (Wald test, winter vs

386   fall, $p=0.003$; winter vs spring and winter vs summer, $p<0.001$) and offshore environments

387   (winter vs fall, $p<0.001$; winter vs spring, $p=0.002$; winter vs summer, $p=0.005$); (ii) subclade

388   Va, which increased in spring in the coastal environment (winter vs spring, $p<0.001$; spring vs

389   fall, $p=0.029$); (iii) subclade Ia, which increased during winter in the coastal environment (winter

390   vs fall, $p=0.002$; spring vs fall, $p=0.041$); (iv) subclade IV, which increased in spring in the

391   coastal environment (winter vs spring and spring vs summer, $p<0.001$); and (v) subclade Vb,

392   which decreased during winter in the coastal environment (winter vs summer, $p=0.018$; winter vs

393   spring, $p=0.011$).

394

Spatial and temporal distributions of SAR11 ASVs

395

396   Of 106 SAR11 ASVs in total, 39 were found at least once in each of the coastal,

397   transition, and offshore environments (**Fig. 4A**). 20 were unique to the coastal/coastal+transition

398   stations and 44 were unique to the offshore/offshore+transition stations (**Fig. 4A**). All subclades

399   except IIb, which was exceedingly rare overall, contained differentially distributed ASVs across

400   the three environments when evaluated by DESeq2 normalization (**Fig. 5; Table S8**).

401   Over half (55) of the 106 SAR11 ASVs appeared across all four seasons (**Fig. 4B**). The

402   fewest number of SAR11 ASVs were detected in the fall and summer (73 ASVs and n=40

403   samples each), while a higher number of SAR11 ASVs were detected during spring (85 ASVs in

404   n=60 samples) and winter (94 ASVs in n=60 samples). Some ASVs (n=14) were restricted to a

405   single season: two in each of spring, summer, and fall, and eight in winter. All of the season-

406   specific ASVs also had restricted spatial distributions, were infrequently recovered (n=2-5

407   samples), and were relatively low in abundance (<1.5% of the SAR11 community). Using

408   DESeq2 normalization, 28 ASVs showed seasonal differences in at least one of the three

409   environments (**Fig. 5; Table S8**). While six of these seasonal ASVs had significant seasonal

410   differences across two (n=5) or all three environment types (i.e. coastal, transition, offshore;

411   n=1), the majority of seasonal ASVs were detected in the coastal environment (n=15). Three

412   seasonal ASVs were detected in the transition environment and four were recovered in the

413   offshore environment.

414

Frequency of SAR11 ASVs across KByT

415

416   SAR11 ASVs were grouped into four categories based on the frequency they were

417   detected in samples within each of the three environments: rare (<5%), low-frequency (5-25%);

418   mid-frequency (25-75%); and high-frequency (>75%). In general, most SAR11 ASV diversity

419   was comprised of either rare or low-frequency ASVs, including 45 of 62 ASVs in the coastal

420   environment (73%), 47 of 67 ASVs from the transition environment (70%), and 52 of 86 ASVs

421   from the offshore environment (60%) (**Fig. S9**). The offshore environment harbored 16 ASVs in

422   the high-frequency category, followed by nine in the transition environment and seven in the

423   coastal environment. Of these, only three were high-frequency across all three environments and

424   none were unique to the transition zone.

425   Rare ASVs were most commonly members of SAR11 subclade IIa (coastal n=14;

426   transition n=16; offshore n=8). All subclades had an ASV that was high-frequency, with the

427   exception of the exceedingly rare subclade IIb. Two high-frequency ASVs were ubiquitous

428   across all 200 samples: ASV47 from subclade Ia and ASV78 from subclade IIa. Two high-

429   frequency ASVs were ubiquitous across all coastal samples: ASV16 from subclade Ia and

430   ASV96 from subclade Va. Three high-frequency ASVs were ubiquitous across all offshore

431   samples: ASV18 from subclade Ia, ASV34 from Ib, and ASV4 from subclade IV.

432   In general, ubiquitous ASVs were also typically the most abundant. As a proportion of

433   the total SAR11 fraction, ASV47 (28.5 ± 10.9%), ASV16 (24.2 ± 15.9%), ASV78 (8.5 ± 3.0%),

434   ASV18 (6.9 ± 6.9%), ASV34 (3.9 ± 4.8%), and ASV96 (5.1 ± 4.2%) were the most abundant

435   SAR11 ASVs across all samples (mean ± s.d., n=200) (**Fig. 3c**). One exception was ASV4; at 0.8

436   ± 0.9% of the SAR11 fraction, it was only the 14th most abundant SAR11 ASV yet ubiquitous in

437   the offshore. The seven ubiquitous ASVs made up 83.0 ± 5.9% (mean ± s.d., n=120) of the total

438   SAR11 community within the coastal environment, t, 77.9 ± 12.8% within the transition

439   environment (n=40), and 62.8 ± 7.9% within the offshore (n=40) (**Fig. 3c**).

440

Correlations between SAR11, picocyanobacteria, and environmental parameters

441

442   Hierarchal clustering of environmental parameters, the abundance of dominant SAR11

443   ASVs, SAR11 subclades, and *Synechococcus* and *Prochlorococcus* picocyanobacteria, by the

444   magnitude and direction of spearman rank correlations(r_s) grouped these measures into two main

445   clusters (**Fig. S9**). The first cluster contained dominant SAR11 ASVs ASV4, ASV18, ASV34,

446   ASV47, and ASV78, SAR11 subclades Ia, Ib, IIa, IIb, IV, and Vb, *Prochlorococcus*, and

447   environmental parameters (salinity and pH) that increased in the offshore environment, while the

448   second cluster included dominant SAR11 ASVs ASV16 and ASV96, SAR11 subclades IIIa and

449   Va, *Synechococcus*, and environmental parameters including the abundance of heterotrophic

450   bacteria, the concentration of chlorophyll *a*, temperature, and the abundance of eukaryotic

451   picophytoplankton that increased in the coastal environment (**Table 1, Figure S9**). Three of

452   seven dominant SAR11 ASVs (ASV78, ASV34, ASV18) had strong positive correlations with

453   the subclades that they belonged to ($r_s > 0.7$) (**Figure S9**). Subclade Ib, ASV34 (subclade Ib), and

454   ASV18 (subclade Ia) had strong positive correlations with *Prochlorococcus* sequence abundance

455   ($r_s > 0.7$), while more moderate positive correlations were detected between most other offshore

456   SAR11 subclades and dominant ASVs (i.e. subclades IIa, IV, and Vb; ASV4, ASV78) and

457   *Prochlorococcus* ($r_s = 0.4$-0.7). Subclades Ia and IIb and ASV47 showed weak correlations with

458   *Prochlorococcus* ($r_s < 0.4$) as well as weak correlations across most comparisons. Moderate

459   positive correlations were found between SAR11 subclades IIIa and Va and ASV16 and ASV96

460   and *Synechococcus* abundance ($r_s = 0.4$-0.7).

461

Comparison of KByT SAR11 ASVs and isolated strains

462

463   Of the 106 SAR11 ASVs detected in KByT, 11 were 100% identical to cultured strains of

464   SAR11, including seven from subclade Ia (**Table S9**). This included the most dominant and

465   ubiquitous ASV in this data set, ASV47 within subclade Ia, which was identical to the 16S rRNA

466   gene of fifteen SAR11 strains including several isolated from Kāneʻohe Bay (**Table S9**). ASV16

467   and ASV18 from SAR11 subclade Ia each matched 100% to cultivated strains and possessed

468   contrasting patterns of distribution, with ASV16 significantly higher in relative abundance in the

469   coastal environment and ASV18 significantly higher in relative abundance in the transition and

470   offshore environments (**Fig. 2B; Table S7**). ASV78, the other ubiquitously-distributed ASV and

471   the second-most abundant overall, was a 100% match to strain HIMB58 in subclade IIa. One

472   subclade IIIa ASV (ASV11) with ubiquitous distribution in coastal samples was identical to

473   strain HIMB114. Finally, ASV96 in subclade Va exactly matched strain HIMB59 and was

474   ubiquitously distributed in the coastal environment.

475   The majority (7 of 11) of ASVs with matches to previously isolated SAR11 strains

476   occurred in high frequency and abundance. The remaining ASVs that matched isolated strains

477   were all from subclade Ia (ASV50, ASV20, ASV22, ASV12), and were less frequent and

478   abundant (**Table S9**).

479

480   **Discussion**

481   The Hawaiian island landmasses provide a useful and convenient platform to investigate

482   the ecotypic differentiation of planktonic marine bacteria over an abrupt environmental gradient.

483   Across two years of monthly sampling, the coastal environment within Kāneʻohe Bay, Oʻahu

484   resolved itself as a marine *Synechococcus*-dominated system that contains elevated inorganic

485   nutrients, chlorophyll *a*, and cellular abundances of planktonic marine bacteria. These features

486 are consistent with an "Island Mass Effect", where an increase in phytoplankton biomass

487 proximate to near-island and atoll-reef ecosystems is caused by localized increases in nutrient

488 delivery through physical oceanographic, biological, land-based, and anthropogenic processes

489 (Doty & Oguri, 1956). A short, three nautical mile transit from the interior of the bay across a

490 transition zone leads to a *Prochlorococcus*-dominated system characterized by depressed

491 inorganic nutrients, low chlorophyll *a* concentrations, and lower abundances of planktonic

492 marine bacteria - characteristics typical of offshore waters (Partensky, Blanchot & Vaulot, 1999).

493   The persistent differences between coastal and offshore environments across the KByT

494 transect also manifested in distinct distributions of SAR11 marine bacteria. While SAR11

495 accounted for roughly 20% of the microbial community within the coastal environment of

496 Kāneʻohe Bay, their relative abundance sharply increased to 30-35% in the offshore waters

497 surrounding the bay. In fact, seven of eight SAR11 subclades detected throughout KByT

498 displayed distinct patterns of distribution across coastal Kāneʻohe Bay and the adjacent offshore,

499 the most dramatic of which were an increase in abundance of subclades Ib and IIa in the

500 offshore. Increases in these two subclades have been associated with oligotrophic ocean gyres

501 (Morris, Frazar & Carlson, 2012) and increased *Prochlorococcus* cellular abundances (Salter et

502 al., 2014). Our results show that the environmental differences between coastal Kāneʻohe Bay

503 and the neighboring offshore is a strong determinant of SAR11 subclade distribution, providing

504 further support that adaptations to environmental niches may shape the evolution of a majority of

505 SAR11 marine bacteria (Vergin et al., 2013).

506   In previous studies, abiotic parameters including depth, salinity, temperature, and latitude

507 have been identified to co-vary with the distribution of SAR11 subclades (Field et al., 1997;

508 Carlson et al., 2008; Brown et al., 2012; Vergin et al., 2013; Salter et al., 2014; Herlemann et al.,

509   2014; West et al., 2016). We observed a fundamental change in the picocyanobacteria that

510   dominate coastal versus offshore surface seawater across the KByT system, and hypothesize that

511   this is a major driver determining the distinct distributions of SAR11 subclades and ASVs across

512   KByT. Different SAR11 subclades have shown to preferentially feed on exudates from

513   *Synechococcus* (Nelson & Carlson, 2012), and similar patterns may be possible between SAR11

514   subclades and *Prochlorococcus* (Becker et al., 2019). Some evidence of how SAR11 populations

515   may specialize in coastal versus oceanic environments comes from analysis of genomes

516   sequenced from isolated SAR11 strains belonging to subclade Ia (Schwalbach et al., 2010;

517   Giovannoni et al., 2019). For example, genomic analyses of coastal, high-latitude subclade Ia

518   strain HTCC1062 and ocean gyre, low-latitude subclade Ia strain HTCC7211 have shown that

519   the coastal strain is capable of utilizing glucose for chemoheterotrophic growth, while the ocean

520   gyre strain could not (Schwalbach et al., 2010). This appears to be a variable metabolic property

521   within SAR11 subclade Ia, and is more commonly detected in productive environments that offer

522   higher concentrations of labile carbon sources (Schwalbach et al., 2010). Recent genomic

523   investigations and co-culture studies are expanding our understanding of the diverse

524   phytoplankton byproducts SAR11 cells are able to metabolize, such as volatile organic

525   compounds (Halsey et al., 2017; Moore et al., 2020) and dimethyl arsenate (Giovannoni et al.,

526   2019), emphasizing the importance of co-evolutionary relationships between SAR11 and

527   phytoplankton (Braakman, Follows & Chisholm, 2017; Becker et al., 2019), and further

528   suggesting that primary productivity plays a significant role in structuring SAR11 genetic and

529   ecological diversity.

530   Across surface seawater within the three environments of KByT, the relative abundance

531   of SAR11 was dominated by seven ASVs. This observation was similar to that of Ortmann &

532   Santos, who found that ten abundant SAR11 OTUs represented >80% of all of the SAR11

533   sequences collected from surface seawaters along a transect from coastal Mobile Bay, Alabama,

534   to the offshore Gulf of Mexico (Ortmann & Santos, 2016). Across KByT, part of this dominance

535   may arise from the relatively conserved nature of the 16S rRNA gene fragment, which can mask

536   genomic heterogeneity and associated ecological differentiation (Chase et al., 2018; Chevrette et

537   al., 2019). High similarity of the 16S rRNA gene despite large genomic variation has been

538   previously reported in the SAR11 subclade, with members of subclade Ia sharing 16S rRNA

539   gene identities of 98% or greater occurring within genomes that share average amino acid

540   identities as low as 71% (Grote et al., 2012). In addition, comparisons among intergenic spacer

541   (ITS) and 16S rRNA gene ASVs have shown that SAR11 ITS ASVs provided more and higher

542   correlations between SAR11 and viral communities, suggesting that important strain-specific

543   interactions remain unresolved by the 16S rRNA gene (Needham, Sachdeva & Fuhrman, 2017).

544   Despite its potential limitations, our analysis of 16S rRNA gene ASVs did show

545   ecologically relevant patterns below the subclade-level. Subclade Ia contained three ASVs

546   (ASV16, 18, 47) within the seven most abundant, which is not surprising considering that it is

547   the most abundant SAR11 subclade in surface oceans overall (Delmont et al., 2019). Yet, our

548   observation that ASV16 is more abundant nearshore and ASV18 is more abundant offshore, and

549   their segregation into two different clusters based on correlation analyses, provides evidence that

550   their coexistence may derive from functional differences between closely related SAR11 ASVs

551   that result in their differential distribution across end member environments of this study system.

552   In our study, spring and summer coincide with the dry season in Hawaiʻi; a period of

553   increased salinity, increased solar irradiance, and diminished dissolved organic nutrients in the

554   coastal environment due in part to decreased nutrient delivery from freshwater streams (Cox,

555   Ribes & Kinzie RA, 2006; Bryant et al., 2016). In contrast to offshore KByT stations that

556   underwent little seasonal change, coastal KByT stations within Kāneʻohe Bay experienced

557   notable seasonal changes during spring including increased salinity, *Synechococcus* and

558   heterotrophic bacteria cellular abundance, and an increase in the total relative abundance of

559   SAR11. Based on our results and those of past studies (Cox, Ribes & Kinzie RA, 2006; Yeo et

560   al., 2013), it appears that SAR11 and *Synechococcus* within Kāneʻohe Bay tend to increase in

561   abundance during dry periods and decrease seasonally and periodically (i.e. immediately

562   following storm events) with increased rainfall and nutrient concentrations.

563   Seasonal changes in the diversity and abundance of individual SAR11 subclades have

564   been observed in temperate (Salter et al., 2014; Meziti et al., 2015), tropical (Vergin et al., 2013),

565   freshwater (Heinrich, Eiler & Bertilsson, 2013), and subtropical environments (Chow et al.,

566   2013). While significant seasonal changes in the number of SAR11 ASVs were not observed in

567   KByT, we did discern strong seasonal patterns in the relative abundance of SAR11 subclades

568   and within individual ASVs. At the offshore stations of KByT, we found that SAR11 subclade

569   IIIa peaked during fall in a similar fashion to observations in surface waters of the Sargasso Sea

570   (Vergin et al., 2013). At the coastal stations of KByT, subclades IIIa, IV, and Va peaked in

571   relative abundance during spring, while subclade Ia peaked during the winter, and subclade Vb

572   peaked during spring. Several factors may contribute to differences in subclade seasonality

573   between the coastal and offshore environments. First, different ASVs that comprise these

574   subclades in each environment may display heterogeneous responses to seasonal changes.

575   Second, seasonal fluctuations in the structure and activity of the phytoplankton and

576   cyanobacterial communities that make up the base of the food web that are specific to the coastal

577   or offshore environment may elicit changes in the growth of SAR11 ASVs. In the coastal

578   environment, the average relative abundances of SAR11 and *Synechococcus* reached their

579   maximum during spring, with highly significant changes in abundance across multiple seasonal

580   comparisons. Similarly, in the offshore environment SAR11 and *Prochlorococcus* both reached

581   their maximum average relative abundances during the summer, however these changes in

582   abundance were less dynamic, with non-significant comparisons across seasons. It is plausible

583   that the abundance of the major consumer of dissolved organic matter, SAR11, is linked to the

584   abundance of dominant photoautotrophs *Synechococcus* and *Prochlorococcus* in the coastal and

585   offshore environments, respectively (Malmstrom et al., 2010). We note, however, that it is also

586   plausible that seasonal changes in terrestrial and eukaryotic phytoplankton-derived dissolved

587   organic carbon might differentially impact the structure and abundance of SAR11 assemblages

588   across the nearshore to offshore continuum investigated here.

589   Relatively little is known about the genetic structure of natural populations of marine

590   microbes in coastal tropical environments, as most studies have come from time-series in

591   temperate systems (Gilbert et al., 2009; Lindh et al., 2015) and oligotrophic ocean gyres (Eiler,

592   Hayakawa & Rappé, 2011; Vergin et al., 2013; Bryant et al., 2016). Our time-series analyses

593   from tropical coastal Kāneʻohe Bay to its adjacent offshore system show distinct SAR11

594   communities across environments, that are dominated by a few highly abundant ASVs. This

595   study contributes to a growing knowledge of how coexisting, closely-related populations of

596   marine bacteria are distributed across environmental gradients.

597

598   **Conclusions**

599   SAR11 is the most abundant organism inhabiting seawater of the global ocean. Yet, due to its

600   immense genetic diversity, an understanding of the determinants underlying its spatial and

601   temporal distributions in natural systems remains limited to specific subclades and environments.

602   Here we show sharp changes in the relative distribution of 16S rRNA gene ASVs, subclades, and

603   total relative abundance of SAR11 bacteria across a coastal to offshore transition in the tropical

604   Pacific Ocean. Nearly all of the SAR11 subclades detected throughout the Kāneʻohe Bay Time-

605   series displayed distinct patterns of distribution across coastal, transition, and adjacent offshore

606   environments. While seasonal patterns of distribution for SAR11 subclades occupying offshore

607   stations were few, the seasonality of SAR11 subclades in the coastal environment were unique

608   and more dynamic. Despite their close proximity and a constant exchange of seawater between

609   stations, our findings suggest that environmental selection shapes the stark spatial and seasonal

610   patterns of distribution in SAR11 lineages we observed. We anticipate that a genomic approach

611   will help to elucidate the underlying mechanisms responsible for this ecotypic differentiation.

612

613 **Acknowledgements**

614 We thank Catherine M. Foley for her generous help with the creation of maps, Evelyn Hoffman,

615 Helen Li, Rachel Ouye, and Hanako Mochimaru for their laboratory and field assistance, Daniel

616 Schar for assistance with fluorometric measurements, Karen Selph for flow cytometry

617 measurements, Jason Jones, Rebecca Weible, and Evan Barba for assistance with sample

618 collection, Jed Fuhrman for providing mock microbial communities to use in the 16S rRNA gene

619 analyses, and Brian Powell for his advice regarding the delineation of seasons. Any opinions,

620 findings, and conclusions or recommendations expressed in this material are those of the

621 author(s) and do not necessarily reflect the views of the National Science Foundation. This is

622 SOEST contribution xxx and HIMB contribution xxx.

623

References

Alonso-Sáez L, Balagué V, Sà EL, Sánchez O, González JM, Pinhassi J, Massana R, Pernthaler J, Pedrós-Alió C, Gasol JM 2007. Seasonality in bacterial diversity in north-west Mediterranean coastal waters: assessment through clone libraries, fingerprinting and FISH. *FEMS Microbiology Ecology* 60:98–112. DOI: 10.1111/j.1574-6941.2006.00276.x.

Bahr KD, Jokiel PL, Toonen RJ 2015. The unnatural history of Kāneʻohe Bay: coral reef resilience in the face of centuries of anthropogenic impacts. *PeerJ* 3:e950–26. DOI: 10.7717/peerj.950.

Barbera P, Kozlov AM, Czech L, Morel B, Darriba D, Flouri T, Stamatakis A 2018. EPA-ng: Massively Parallel Evolutionary Placement of Genetic Sequences. *Bioinformatics* 68:365–369. DOI: 10.1093/sysbio/syy054.

Becker JW, Brandon ML, Rappé MS 2007. Cultivating microorganisms from dilute aquatic environments: melding traditional methodology with new cultivation techniques and molecular methods. In: Hurst CJ, Crawford RL, Garland JL, Lipson DA, Mills AL, Stetzenbach LD eds. *Manual of Environmental Microbiology*. Washington DC, USA: American Society of Microbiology, 399–406. DOI: 10.1128/9781555815882.ch32.

Becker JW, Hogle SL, Rosendo K, Chisholm SW 2019. Co-culture and biogeography of *Prochlorococcus* and SAR11. *The ISME Journal* 13:1506–1519. DOI: 10.1038/s41396-019-0365-4.

Biller SJ, Berube PM, Berta-Thompson JW, Kelly L, Roggensack SE, Awad L, Roache-Johnson KH, Ding H, Giovannoni SJ, Rocap G, Moore LR, Chisholm SW 2014. Genomes of diverse isolates of the marine cyanobacterium *Prochlorococcus*. *Scientific Data* 1:140034–11. DOI: 10.1038/sdata.2014.34.

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu Y-X, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, Hippel von M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology* 37:852–857. DOI: 10.1038/s41587-019-0209-9.

Braakman R, Follows MJ, Chisholm SW 2017. Metabolic evolution and the self-organization of ecosystems. *Proceedings of the National Academy of Sciences of the United States of America* 114:E3091–E3100. DOI: 10.1073/pnas.1619573114.

669 Brown MV, Fuhrman JA 2005. Marine bacterial microdiversity as revealed by internal
670   transcribed spacer analysis. *Aquatic Microbial Ecology* 41:15–23. DOI:
671   10.3354/ame041015.
672 Brown MV, Lauro FM, DeMaere MZ, Muir Les, Wilkins D, Thomas T, Riddle MJ, Fuhrman JA,
673   Andrews-Pfannkoch C, Hoffman JM, McQuaid JB, Allen A, Rintoul SR, Cavicchioli R
674   2012. Global biogeography of SAR11 marine bacteria. *Molecular Systems Biology* 8:1–13.
675   DOI: 10.1038/msb.2012.28.
676 Bryant JA, Aylward FO, Eppley JM, Karl DM, Church MJ, DeLong EF 2016. Wind and sunlight
677   shape microbial diversity in surface waters of the North Pacific Subtropical Gyre. *The ISME
678   Journal* 10:1308–1322. DOI: 10.1038/ismej.2015.221.
679 Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP 2016. DADA2:
680   High-resolution sample inference from Illumina amplicon data. *Nature Methods* 13:581–
681   583. DOI: 10.1038/nmeth.3869.
682 Carlson CA, Morris R, Parsons R, Treusch AH, Giovannoni SJ, Vergin K 2008. Seasonal
683   dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern
684   Sargasso Sea. *The ISME Journal* 3:283–295. DOI: 10.1038/ismej.2008.117.
685 Chase AB, Gomez-Lunar Z, Lopez AE, Li J, Allison SD, Martiny AC, Martiny JBH 2018.
686   Emergence of soil bacterial ecotypes along a climate gradient. *Environmental Microbiology*
687   20:4112–4126. DOI: 10.1111/1462-2920.14405.
688 Chevrette MG, Carlos-Shanley C, Louie KB, Bowen BP, Northen TR, Currie CR 2019.
689   Taxonomic and Metabolic Incongruence in the Ancient Genus Streptomyces. *Frontiers in
690   Microbiology* 10:2170. DOI: 10.3389/fmicb.2019.02170.
691 Chow C-ET, Sachdeva R, Cram JA, Steele JA, Needham DM, Patel A, Parada AE, Fuhrman JA
692   2013. Temporal variability and coherence of euphotic zone bacterial communities over a
693   decade in the Southern California Bight. *The ISME Journal* 7:2259–2273. DOI:
694   10.1038/ismej.2013.122.
695 Cohan FM 2006. Towards a conceptual and operational union of bacterial systematics, ecology,
696   and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*
697   361:1985–1996. DOI: 10.1098/rstb.2006.1918.
698 Cox EF, Ribes M, Kinzie RA I 2006. Temporal and spatial scaling of planktonic responses to
699   nutrient inputs into a subtropical embayment. *Marine Ecology Progress Series* 324:19–35.
700   DOI: 10.3354/meps324019.
701 Cram JA, Chow C-ET, Sachdeva R, Needham DM, Parada AE, Steele JA, Fuhrman JA 2015.
702   Seasonal and interannual variability of the marine bacterioplankton community throughout
703   the water column over ten years. *The ISME Journal* 9:563–580. DOI:
704   10.1038/ismej.2014.153.
705 Czech L, Stamatakis A 2019. Scalable methods for analyzing and visualizing phylogenetic
706   placement of metagenomic samples. *PLoS ONE* 14:e0217050–50. DOI:
707   10.1371/journal.pone.0217050.
708 Delmont TO, Kiefl E, Kilinc O, Esen OC, Uysal I, Rappé MS, Giovannoni SJ, Eren AM 2019.
709   Single-amino acid variants reveal evolutionary processes that shape the biogeography of a
710   global SAR11 subclade. *eLife*:1–26. DOI: 10.7554/eLife.46497.001.
711 Doty MS, Oguri M 1956. The island mass effect. *J. Cons. perm. int. Explor. Mer* 22:33–37.
712 Eiler A, Hayakawa DH, Rappé MS 2011. Non-Random assembly of bacterioplankton
713   communities in the Subtropical North Pacific Ocean. *Frontiers in Microbiology* 2. DOI:
714   10.3389/fmicb.2011.00140.

715 Field KG, Gordon D, Wright T, Rappé M, Urback E, Vergin K, Giovannoni SJ 1997. Diversity
716   and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic
717   bacteria. *Applied and Environmental Microbiology* 63:63–70. DOI: 10.1128/AEM.63.1.63-
718   70.1997.
719 Gilbert JA, Field D, Swift P, Newbold L, Oliver A, Smyth T, Somerfield PJ, Huse S, Joint I
720   2009. The seasonal structure of microbial communities in the Western English Channel.
721   *Environmental Microbiology* 11:3132–3139. DOI: 10.1111/j.1462-2920.2009.02017.x.
722 Giovannoni SJ 2017. SAR11 bacteria: the most abundant plankton in the oceans. *Annual Review*
723   *of Marine Science* 9:231–255. DOI: 10.1146/annurev-marine-010814-015934.
724 Giovannoni SJ, Halsey KH, Saw J, Muslin O, Suffridge CP, Sun J, Lee C-P, Moore ER,
725   Temperton B, Noell SE 2019. A parasitic arsenic cycle that shuttles energy from
726   phytoplankton to heterotrophic bacterioplankton. *mBio* 10. DOI: 10.1128/mBio.00246-19.
727 Gove JM, McManus MA, Neuheimer AB, Polovina JJ, Drazen JC, Smith CR, Merrifield MA,
728   Friedlander AM, Ehses JS, Young CW, Dillon AK, Williams GJ 2016. Near-island
729   biological hotspots in barren ocean basins. *Nature Communications* 7:10581–8. DOI:
730   10.1038/ncomms10581.
731 Grote J, Thrash JC, Huggett MJ, Landry ZC, Carini P, Giovannoni SJ, Rappé MS 2012.
732   Streamlining and core genome conservation among highly divergent members of the SAR11
733   clade. *mBio* 3:806. DOI: 10.1128/mBio.00252-12.
734 Halsey KH, Giovannoni SJ, Graus M, Zhao Y, Landry Z, Thrash JC, Vergin KL, de Gouw J
735   2017. Biological cycling of volatile organic carbon by phytoplankton and bacterioplankton.
736   *Limnology and Oceanography* 62:2650–2661. DOI: 10.1002/lno.10596.
737 Haro-Moreno JM, López Pérez M, la Torre de JR, Picazo A, Camacho A, Rodriguez-Valera F
738   2018. Fine metagenomic profile of the Mediterranean stratified and mixed water columns
739   revealed by assembly and recruitment. *Microbiome* 6:128–19. DOI: 10.1186/s40168-018-
740   0513-5.
741 Haro-Moreno JM, Rodriguez-Valera F, Rosselli R, Martinez Hernandez F, Roda Garcia JJ,
742   Gomez ML, Fornas O, Martinez-Garcia M, López Pérez M 2019. Ecogenomics of the
743   SAR11 clade. *Environmental Microbiology* 22:1748–1763. DOI: 10.1111/1462-2920.14896.
744 Heinrich F, Eiler A, Bertilsson S 2013. Seasonality and environmental control of freshwater
745   SAR11 (LD12) in a temperate lake (Lake Erken, Sweden). *Aquatic Microbial Ecology*
746   70:33–44. DOI: 10.3354/ame01637.
747 Hellweger FL, van Sebille E, Fredrick ND 2014. Biogeographic patterns in ocean microbes
748   emerge in a neutral agent-based model. *Science* 345:1346–1349. DOI:
749   10.1126/science.1254421.
750 Henson MW, Lanclos VC, Faircloth BC, Thrash JC 2018. Cultivation and genomics of the first
751   freshwater SAR11 (LD12) isolate. *The ISME Journal* 12:1846–1860. DOI: 10.1038/s41396-
752   018-0092-2.
753 Herlemann DPR, Woelk J, Labrenz M, Jürgens K 2014. Diversity and abundance of
754   "Pelagibacterales" (SAR11) in the Baltic Sea salinity gradient. *Systematic and Applied*
755   *Microbiology* 37:601–604. DOI: 10.1016/j.syapm.2014.09.002.
756 Hothorn T, Bretz F, Westfall P 2008. Simultaneous Inference in General Parametric Models.
757   *Biometrical Journal* 50:346–363. DOI: 10.1002/bimj.200810425.
758 Jimenez-Infante F, Ngugi DK, Vinu M, Blom J, Alam I, Bajic VB, Stingl U 2017. Genomic
759   characterization of two novel SAR11 isolates from the Red Sea, including the first strain of
760   the SAR11 Ib clade. *FEMS Microbiology Ecology* 93:e82210. DOI: 10.1093/femsec/fix083.

761 Koeppel A, Perry EB, Sikorski J, Krizanc D, Warner A, Ward DM, Rooney AP, Brambilla E,
762   Connor N, Ratcliff RM, Nevo E, Cohan FM 2008. Identifying the fundamental units of
763   bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics.
764   *Proceedings of the National Academy of Sciences of the United States of America* 105:2504–
765   2509. DOI: 10.1073/pnas.0712205105.
766 Kolde R 2019. *pheatmap: Pretty Heatmaps. R package version 1.0.12.*
767 Kopac S, Wang Z, Wiedenbeck J, Sherry J, Wu M, Cohan FM 2014. Genomic heterogeneity and
768   ecological speciation within one subspecies of *Bacillus subtilis. Applied and Environmental*
769   *Microbiology* 80:4842–4853. DOI: 10.1128/AEM.00576-14.
770 Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A 2019. RAxML-NG: a fast, scalable and
771   user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35:4453–
772   4455. DOI: 10.1093/bioinformatics/btz305.
773 Kraemer S, Ramachandran A, Colatriano D, Lovejoy C, Walsh DA 2019. Diversity and
774   biogeography of SAR11 bacteria from the Arctic Ocean. *The ISME Journal*:1–12. DOI:
775   10.1038/s41396-019-0499-4.
776 Kumar S, Stecher G, Li M, Knyaz C, Tamura K 2018. MEGA X: Molecular evolutionary
777   genetics analysis across computing platforms. *Molecular biology and evolution* 35:1547–
778   1549. DOI: 10.1093/molbev/msy096.
779 Lindh MV, Sjöstedt J, Andersson AF, Baltar F, Hugerth LW, Lundin D, Muthusamy S, Legrand
780   C, Pinhassi J 2015. Disentangling seasonal bacterioplankton population dynamics by high-
781   frequency sampling. *Environmental Microbiology* 17:2459–2476. DOI: 10.1111/1462-
782   2920.12720.
783 Love MI, Huber W, Anders S 2014. Moderated estimation of fold change and dispersion for
784   RNA-seq data with DESeq2. *Genome biology* 15:31–21. DOI: 10.1186/s13059-014-0550-8.
785 Ludwig W 2004. ARB: a software environment for sequence data. *Nucleic Acids Research*
786   32:1363–1371. DOI: 10.1093/nar/gkh293.
787 Malmstrom RR, Coe A, Kettler GC, Martiny AC, Frias-Lopez J, Zinser ER, Chisholm SW 2010.
788   Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific oceans. *The*
789   *ISME Journal* 4:1252–1264. DOI: 10.1038/ismej.2010.60.
790 Manrique JM, Jones LR 2017. Are ocean currents too slow to counteract SAR11 evolution? A
791   next-generation sequencing, phylogeographic analysis. *Molecular Phylogenetics and*
792   *Evolution* 107:324–337. DOI: 10.1016/j.ympev.2016.11.015.
793 Martinez Arbizu P 2020. pairwiseAdonis: Pairwise multilevel comparison using adonis.
794 McMurdie PJ, Holmes S 2013. phyloseq: an R package for reproducible interactive analysis and
795   graphics of microbiome census data. *PLoS ONE* 8:e61217. DOI:
796   10.1371/journal.pone.0061217.
797 Mende DR, Boeuf D, DeLong EF 2019. Persistent core populations shape the microbiome
798   throughout the water column in the North Pacific Subtropical Gyre. *Frontiers in*
799   *Microbiology* 10:2273. DOI: 10.3389/fmicb.2019.02273.
800 Meziti A, Kormas KA, Moustaka-Gouni M, Karayanni H 2015. Spatially uniform but temporally
801   variable bacterioplankton in a semi-enclosed coastal area. *Systematic and Applied*
802   *Microbiology* 38:358–367. DOI: 10.1016/j.syapm.2015.04.003.
803 Monaghan EA, Freel KC, Rappé MS 2020. Isolation of SAR11 marine bacteria from
804   cryopreserved seawater. *mSystems* 5:e00954–20. DOI: 10.1128/mSystems.00954-20.
805 Monger BC, Landry MR 1993. Flow cytometric analysis of marine bacteria with Hoechst 33342.
806   *Applied and Environmental Microbiology* 59:905–911.

807 Moore ER, Davie Martin CL, Giovannoni SJ, Halsey KH 2020. Pelagibacter metabolism of
808   diatom-derived volatile organic compounds imposes an energetic tax on photosynthetic
809   carbon fixation. *Environmental Microbiology* 22:1720–1733. DOI: 10.1111/1462-
810   2920.14861.
811 Morris RM, Frazar CD, Carlson CA 2012. Basin-scale patterns in the abundance of SAR11
812   subclades, marine *Actinobacteria* (OM1), members of the *Roseobacter* clade and OCS116 in
813   the South Atlantic. *Environmental Microbiology* 14:1133–1144. DOI: 10.1111/j.1462-
814   2920.2011.02694.x.
815 Morris RM, Rappé MS, Connon SA, Vergin KL, Siebold WA, Carlson CA, Giovannoni SJ 2002.
816   SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* 420:806–810.
817   DOI: 10.1038/nature01240.
818 Morris RM, Vergin KL, Cho J-C, Rappé MS, Carlson CA, Giovannoni SJ 2005. Temporal and
819   spatial response of bacterioplankton lineages to annual convective overturn at the Bermuda
820   Atlantic Time-series Study site. *Limnology and Oceanography* 50:1687–1696. DOI:
821   10.4319/lo.2005.50.5.1687.
822 Muñoz-Gómez SA, Hess S, Burger G, Lang BF, Susko E, Slamovits CH, Roger AJ 2019. An
823   updated phylogeny of the *Alphaproteobacteria* reveals that the parasitic *Rickettsiales* and
824   *Holosporales* have independent origins. Elife 8:e42535. DOI: 10.7554/eLife.42535.
825 Needham DM, Sachdeva R, Fuhrman JA 2017. Ecological dynamics and co-occurrence among
826   marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *Nature*
827   *Publishing Group*:1–16. DOI: 10.1038/ismej.2017.29.
828 Nelson CE, Carlson CA 2012. Tracking differential incorporation of dissolved organic carbon
829   types among diverse lineages of Sargasso Sea bacterioplankton. *Environmental*
830   *Microbiology* 14:1500–1516. DOI: 10.1111/j.1462-2920.2012.02738.x.
831 Ngugi DK, Stingl U 2012. Combined Analyses of the ITS Loci and the Corresponding 16S
832   rRNA Genes Reveal High Micro- and Macrodiversity of SAR11 Populations in the Red Sea.
833   *PLoS ONE* 7:e50274–13. DOI: 10.1371/journal.pone.0050274.
834 Noell SE, Giovannoni SJ 2019. SAR11 bacteria have a high affinity and multifunctional glycine
835   betaine transporter. *Environmental Microbiology* 21:2559–2575. DOI: 10.1111/1462-
836   2920.14649.
837 Oksanen J, Blanchet F, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB,
838   Simpson GL, Solymos P, Stevens M, Szoecs E, Wagner H 2019. Vegan: community ecology
839   package. R package version 2.5-6. :1–292.
840 Oliveros JC 2017. *2007–2015. Venny. An interactive tool for comparing lists with Venn's*
841   *diagrams*.
842 Ortmann AC, Santos TTL 2016. Spatial and temporal patterns in the Pelagibacteraceae across an
843   estuarine gradient. *FEMS Microbiology Ecology* 92:fiw133. DOI: fiw133.
844 Pachiadaki MG, Brown JM, Brown J, Bezuidt O, Berube PM, Biller SJ, Poulton NJ, Burkart
845   MD, La Clair JJ, Chisholm SW, Stepanauskas R 2019. Charting the complexity of the
846   marine microbiome through single-cell genomics. *Cell* 179:1623–1635.e11. DOI:
847   10.1016/j.cell.2019.11.017.
848 Parada AE, Needham DM, Fuhrman JA 2016. Every base matters: assessing small subunit rRNA
849   primers for marine microbiomes with mock communities, time series and global field
850   samples. *Environmental Microbiology* 18:1403–1414. DOI: 10.1111/1462-2920.13023.

851 Partensky F, Blanchot J, Vaulot D 1999. Differential distribution and ecology of
852   *Prochlorococcus* and *Synechococcus* in oceanic waters: a review. *Bulletin de l'Institut*
853   *Océanographique - Special issue: Marine cyanobacteria* 19:457–476.
854 Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO 2012. The
855   SILVA ribosomal RNA gene database project: improved data processing and web-based
856   tools. *Nucleic Acids Research* 41:D590–D596. DOI: 10.1093/nar/gks1219.
857 R Core Team 2018. R: A language and environment for statistical computing.
858 Rappé MS, Connon SA, Vergin KL, Giovannoni SJ 2002. Cultivation of the ubiquitous SAR11
859   marine bacterioplankton clade. *Nature* 418:630–633. DOI: 10.1038/nature00917.
860 Salcher MM, Pernthaler J, Posch T 2011. Seasonal bloom dynamics and ecophysiology of the
861   freshwater sister clade of SAR11 bacteria "that rule the waves" (LD12). *The ISME Journal*
862   5:1242–1252. DOI: 10.1038/ismej.2011.8.
863 Salter I, Galand PE, Fagervold SK, Lebaron P, Obernosterer I, Oliver MJ, Suzuki MT, Tricoire C
864   2014. Seasonal dynamics of active SAR11 ecotypes in the oligotrophic Northwest
865   Mediterranean Sea. *The ISME Journal* 9:347–360. DOI: 10.1038/ismej.2014.129.
866 Schwalbach MS, Tripp HJ, Steindler L, Smith DP, Giovannoni SJ 2010. The presence of the
867   glycolysis operon in SAR11 genomes is positively correlated with ocean productivity.
868   *Environmental Microbiology* 12:490–500. DOI: 10.1111/j.1462-2920.2009.02092.x.
869 Sievert C 2013. *plotly for R. https://plotly-r.com*.
870 Suzuki MT, Béjà O, Taylor LT, DeLong EF 2001. Phylogenetic analysis of ribosomal RNA
871   operons from uncultivated coastal marine bacterioplankton. *Environmental Microbiology*
872   3:323–331. DOI: 10.1046/j.1462-2920.2001.00198.x.
873 Swan BK, Tupper B, Sczyrba A, Lauro FM, Martinez-Garcia M, González JM, Luo H, Wright
874   JJ, Landry ZC, Hanson NW, Thompson BP, Poulton NJ, Schwientek P, Acinas SG,
875   Giovannoni SJ, Moran MA, Hallam SJ, Cavicchioli R, Woyke T, Stepanauskas R 2013.
876   Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the
877   surface ocean. *Proceedings of the National Academy of Sciences of the United States of*
878   *America* 110:11463–11468. DOI: 10.1073/pnas.1304246110.
879 Thompson LR, Haroon MF, Shibl AA, Cahill MJ, Ngugi DK, Williams GJ, Morton JT, Knight
880   R, Goodwin KD, Stingl U 2019. Red Sea SAR11 and *Prochlorococcus* single-cell genomes
881   reflect globally distributed pangenomes. *Applied and Environmental Microbiology* 85:3805–
882   18. DOI: 10.1128/AEM.00369-19.
883 Thrash JC, Temperton B, Swan BK, Landry ZC, Woyke T, DeLong EF, Stepanauskas R,
884   Giovannoni SJ 2014. Single-cell enabled comparative genomics of a deep ocean SAR11
885   bathytype. *The ISME Journal* 8:1440–1451. DOI: 10.1038/ismej.2013.243.
886 Tsementzi D, Rodriguez-R LM, Ruiz-Perez CA, Meziti A, Hatt JK, Konstantinidis KT 2019.
887   Ecogenomic characterization of widespread, closely-related SAR11 clades of the freshwater
888   genus "*Candidatus* Fonsibacter" and proposal of *Ca.* Fonsibacter lacus sp. nov. *Systematic*
889   *and Applied Microbiology* 42:495–505. DOI: 10.1016/j.syapm.2019.03.007.
890 Tsementzi D, Wu J, Deutsch S, Nath S, Rodriguez-R LM, Burns AS, Ranjan P, Sarode N,
891   Malmstrom RR, Padilla CC, Stone BK, Bristow LA, Larsen M, Glass JB, Thamdrup B,
892   Woyke T, Konstantinidis KT, Stewart FJ 2016. SAR11 bacteria linked to ocean anoxia and
893   nitrogen loss. *Nature* 536:179–183. DOI: 10.1038/nature19068.
894 Tully BJ, Graham ED, Heidelberg JF 2018. The reconstruction of 2,631 draft metagenome-
895   assembled genomes from the global oceans. *Scientific Data* 5:170203. DOI:
896   10.1038/sdata.2017.203.

897   Vergin KL, Beszteri B, Monier A, Cameron Thrash J, Temperton B, Treusch AH, Kilpert F,
898   Worden AZ, Giovannoni SJ 2013. High-resolution SAR11 ecotype dynamics at the Bermuda
899   Atlantic Time-series Study site by phylogenetic placement of pyrosequences. *The ISME*
900   *Journal* 7:1322–1332. DOI: 10.1038/ismej.2013.32.
901   Wang Y, Naumann U, Eddelbuettel D, Wright ST, Warton DI 2019a. Statistical Methods for
902   Analysing Multivariate Abundance Data. R package version 4.0.1.
903   Wang Z, Juarez DL, Pan JF, Blinebry SK, Gronniger J, Clark JS, Johnson ZI, Hunt DE 2019b.
904   Microbial communities across nearshore to offshore coastal transects are primarily shaped
905   by distance and temperature. *Environmental Microbiology* 21:3862–3872. DOI:
906   10.1111/1462-2920.14734.
907   Welschmeyer NA 1994. Fluorometric analysis of chlorophyll a in the presence of chlorophyll b
908   and pheopigments. *Limnology and Oceanography* 39:1985–1992. DOI:
909   10.4319/lo.1994.39.8.1985.
910   West NJ, Lepère C, Manes C-L de O, Catala P, Scanlan DJ, Lebaron P 2016. Distinct Spatial
911   Patterns of SAR11, SAR86, and Actinobacteria Diversity along a Transect in the Ultra-
912   oligotrophic South Pacific Ocean. *Frontiers in Microbiology* 7:551–16. DOI:
913   10.3389/fmicb.2016.00234.
914   Wickham H 2016. ggplot2: elegant graphics for data analysis. Springer-Verlag New York.
915   Yeo SK, Huggett MJ, Eiler A, Rappé MS 2013. Coastal bacterioplankton community dynamics
916   in response to a natural disturbance. *PLoS ONE* 8:e56207–14. DOI:
917   10.1371/journal.pone.0056207.
918
919

920 **Figure Legends**

921 **Figure 1**. **Location and characteristics of the Kāneʻohe Bay Time-series (KByT) sampling**

922 **stations. (A)** Map of sampling stations within and immediately adjacent to Kāneʻohe Bay on the

923 island of Oʻahu, Hawaiʻi. **(B-G)** Distribution of environmental parameters over two years of

924 sampling across the coastal, transition, and offshore regions of KByT, including **(B)** temperature,

925 **(C)** chlorophyll *a*, **(D)** cellular abundance of heterotrophic bacteria, **(E)** cellular abundance of

926 *Synechococcus*, **(F)** cellular abundance of eukaryotic picophytoplankton, and **(G)** cellular

927 abundance of *Prochlorococcus*. Box plots show the mean and first and third quartile. Map

928 plotted in ArcGIS Pro 2.7.

929

930 **Figure 2. Community structure of dominant groups across environment types in KByT. (A)**

931 Average relative abundance of the fourteen most abundant order-level 16S rRNA gene-based

932 groups across coastal (n=120), transition (n=40), and offshore (n=40) environments sampled

933 over two years of KByT. Taxonomy based on Silva v132. **(B)** Distribution of *Synechococcus* and

934 *Prochlorococcus* ASVs across KByT. N=20 samples per station. **(C)** Distribution of

935 *Synechococcus* and *Prochlorococus* cellular abundances relative to the total prokaryotic

936 community. N=20 samples per station.

937

938 **Figure 3**. **Spatial distribution of SAR11 subclades. (A)** Average number of SAR11 ASVs per

939 subclade, and **(B)** subclade relative abundance across KByT stations. n=20 samples per station.

940 **(C)** Average proportion of the total SAR11 relative abundance for the ubiquitous SAR11 ASVs.

941 Seven ASVs are either ubiquitous within coastal (ASV16, ASV96), offshore (ASV34, ASV18,

942 ASV4), or both (ASV47, ASV78) environments. Circle size indicates the relative proportion of

943 SAR11 in the total microbial community. Map plotted in ArcGIS Pro 2.7.

944

945 **Figure 4. Environmental and seasonal distributions of SAR11 ASVs.** Venn diagrams

946 comparing the distribution of total number of ASVs detected across **(A)** environments and **(B)**

947 seasons.

948

949 **Figure 5. SAR11 ASV distributions.** Heatmap indicating the relative abundance of SAR11

950 ASVs (rows) per sample (columns). ASVs are ordered vertically by subclade, and samples are

951 ordered horizontally by environment, season, date sampled, and site. "E" next to ASVs denotes a

952 significant difference between environments, while "S" denotes a significant differences among

953 seasons within a given environment. Wald tests to assess the significance of differences among

954 environment and season for each SAR11 ASV were conducted using DESeq2 normalized

955 counts.

956

957 **Table Legends**

958 **Table 1. Environmental parameters and cellular abundances from KByT.** Samples are

959 averaged (mean ± s.d.) over environmental category and season.

960

961 **Supplementary Figure Legends**

962 **Figure S1. Seasonal delineations across KByT. (A)** Harmonic mean function fit to sea surface

963 water temperature from Kāneʻohe Bay, Hawaiʻi. **(B)** Annual time-derivate for all 10 years of

964 data.

965

966   **Figure S2. Phylogenetic analysis of SAR11 16S rRNA gene ASVs recovered through KByT.**

967   Cultivated isolates are indicated by "str." (strain). Genbank accession numbers for reference

968   sequences obtained from Genbank or Silva v132 are included in parentheses.

969

970   **Figure S3. Delineation of environments across KByT.** nMDS (k=2) of a Bray-Curtis distance

971   matrix calculated from temperature, salinity, pH, cellular abundances of heterotrophic bacteria,

972   cyanobacteria, and eukaryotic picophytoplankton, and chlorophyll-a concentrations for the entire

973   KByT dataset (n=200 samples). nMDS stress was 0.03.

974

975   **Figure S4. Seasonal and environmental distributions of dominant photoautotrophs across**

976   **KByT.** Average abundance and standard deviation of **(A)** *Synechococcus*, **B)** chloroplast, and **C)**

977   *Prochlorococcus* sequences relative to the total community in each environment and season.

978   Seasonal comparisons with significant seasonal differences as defined by DESeq2 normalization

979   are noted: * if $p<0.05$, ** if $p \le 0.01$, and *** if $p \le 0.001$.

980

981   **Figure S5. Distribution of total SAR11 relative abundance across KByT. (A)** Average

982   abundance and standard deviation of SAR11 relative to the total community in each

983   environment. **(B)** Average abundance and standard deviation of SAR11 relative to the total

984   community in each season. Comparisons with significant seasonal or environmental differences

985   as defined by DESeq2 normalization are noted: * if $p<0.05$, ** if $p \le 0.01$, and *** if $p \le 0.001$.

986

987

988   **Figure S6. Environmental distributions of SAR11 subclades across KByT.** Average

989   abundance and standard deviation of SAR11 subclades relative to total SAR11 abundance in

990   each environment. Comparisons with significant environmental differences as defined by

991   DESeq2 normalization are noted: * if p<0.05, ** if p ≤ 0.01, and *** if p ≤ 0.001.

992

993   **Figure S7. Seasonal distributions of SAR11 subclades across KByT.** Average abundance and

994   standard deviation of SAR11 subclades relative to total SAR11 abundance in each environment

995   (coastal, transition, offshore) per season. Seasonal comparisons with significant seasonal

996   differences as defined by DESeq2 normalization are noted: * if p<0.05, ** if p ≤ 0.01, and *** if

997   p ≤ 0.001.

998

999   **Figure S8. Frequency of detection of SAR11 ASVs across subclades and environments.**

1000   Sample size per environment of KByT included coastal (n=120), transition (n=40), and offshore

1001   (n=40).

1002

1003   **Figure S9. Correlation analysis of SAR11 subclades, dominant SAR11 ASVs,**

1004   **picocyanobacteria, and environmental parameters.** Spearman's Rank correlation coefficients

1005   between variance stabilized transformed DESeq2 normalized abundances of SAR11 subclades,

1006   dominant SAR11 ASVs, and picocyanobacterial and environmental parameters are shown. The

1007   dendrogram was built from a matrix of the correlation coefficients that clustered these correlated

1008   variables into two main groups- those associated to offshore conditions and those associated to

1009   coastal conditions.

1010

1011

1012   **Supplementary Tables Legends**

1013   **Table S1. Comparisons of water column parameters and cellular counts over time and**

1014   **space using one-way ANOVA with Tukey's multiple comparisons test.** Significance (Holm

1015   corrected p-values) are shown.

1016

1017   **Table S2. Read depth per sample.**

1018

1019   **Table S3. Comparisons of read depth over time and space using one-way ANOVA with**

1020   **Tukey's multiple comparisons test.** Significance (Holm corrected p-values) are shown.

1021

1022   **Table S4. Comparisons of sequencing depth across stations using one-way ANOVA with**

1023   **Tukey's multiple comparisons test.** Significance (Holm corrected p-values) are shown.

1024

1025   **Table S5. Detection of SAR11 ASVs across the three KByT environment types: coastal,**

1026   **transition, and offshore.** Taxonomy was assigned in SILVA v132. SAR11 ASVs were assigned

1027   to subclades through phylogenetic placement as described in the methods.

1028

1029   **Table S6. Comparisons of the relative abundance of the top 14 most abundant orders.**

1030   Significant DESeq2 normalized differences across environments are shown.

1031

1032   **Table S7**. **Number of SAR11 ASVs detected across KByT.**

1033

1034   **Table S8: Summary of SAR11 ASVs recovered from KByT.** ASVs with significant DESeq2

1035   normalized differences across environments and seasons within each environment (C=coastal,

1036   T=transition, and O=offshore) are noted.

1037

1038   **Table S9. Summary of KByT ASVs that match isolated SAR11 strains.**

1039
1040

Figure 1

Location and characteristics of the Kāneʻohe Bay Time-series (KByT) sampling stations.

(A) Map of sampling stations within and immediately adjacent to Kāneʻohe Bay on the island of Oʻahu, Hawaiʻi. **(B-G)** Distribution of environmental parameters over two years of sampling across the coastal, transition, and offshore regions of KByT, including **(B)** temperature, **(C)** chlorophyll *a*, **(D)** cellular abundance of heterotrophic bacteria, **(E)** cellular abundance of *Synechococcus*, **(F)** cellular abundance of eukaryotic picophytoplankton, and **(G)** cellular abundance of *Prochlorococcus*. Box plots show the mean and first and third quartile. Map plotted in ArcGIS Pro 2.7.

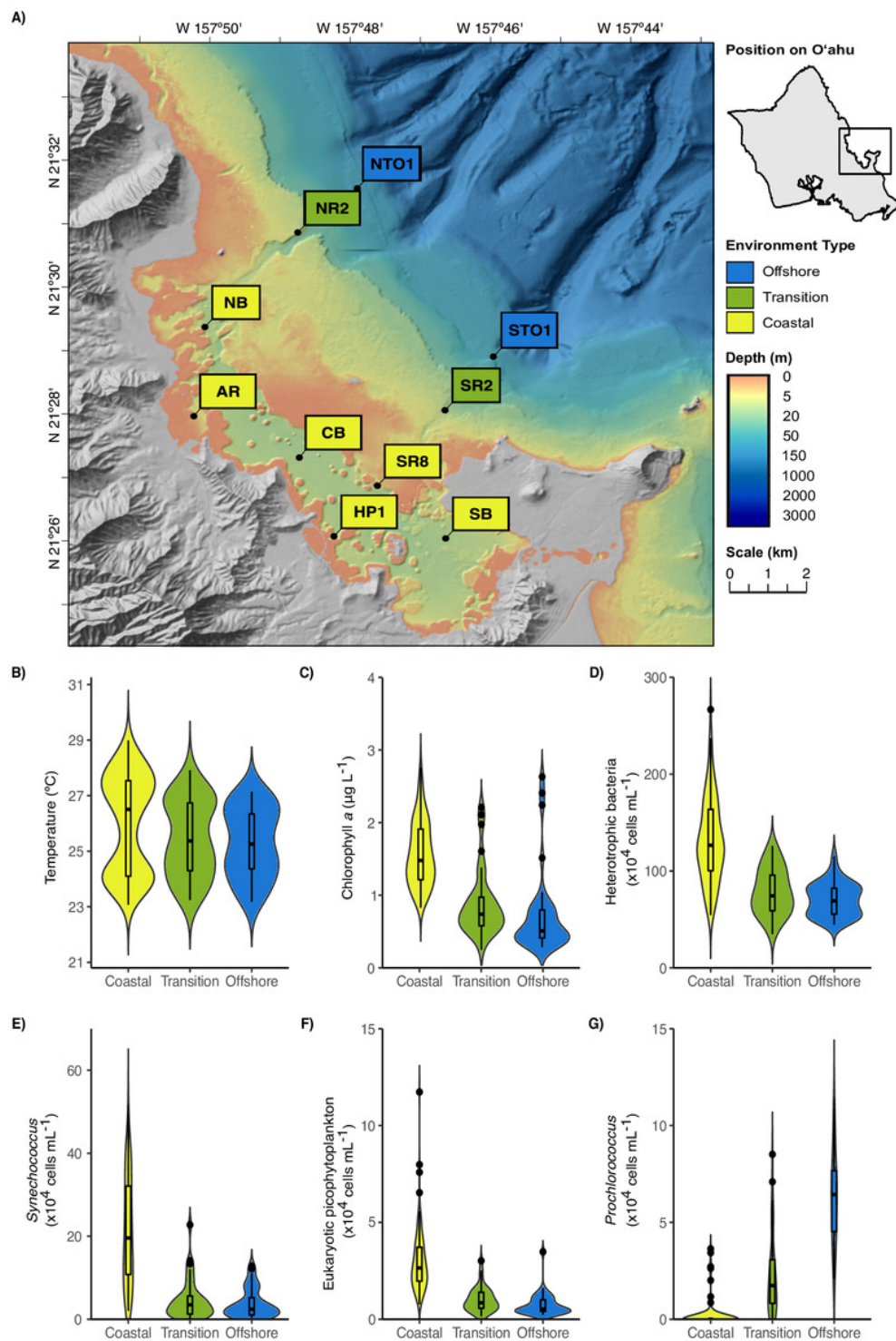



Figure 2

Community structure of dominant groups across environment types in KByT

(**A**) Average relative abundance of the fourteen most abundant order-level 16S rRNA gene-based groups across coastal (n=120), transition (n=40), and offshore (n=40) environments sampled over two years of KByT. Taxonomy based on Silva v132. (**B**) Distribution of *Synechococcus* and *Prochlorococcus* ASVs across KByT. N=20 samples per station. (**C**) Distribution of *Synechococcus* and *Prochlorococus* cellular abundances relative to the total prokaryotic community. N=20 samples per station.

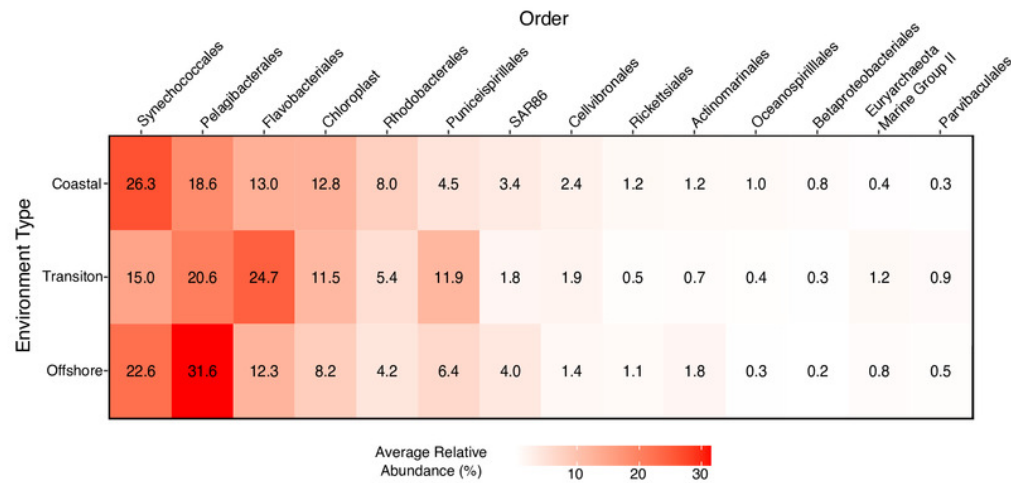
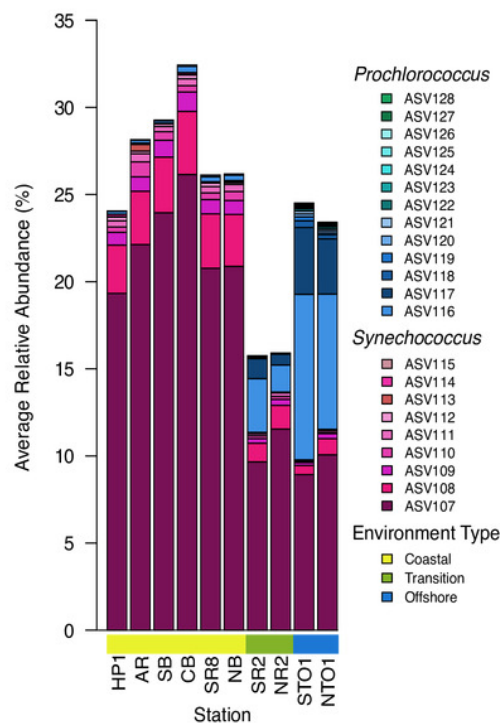
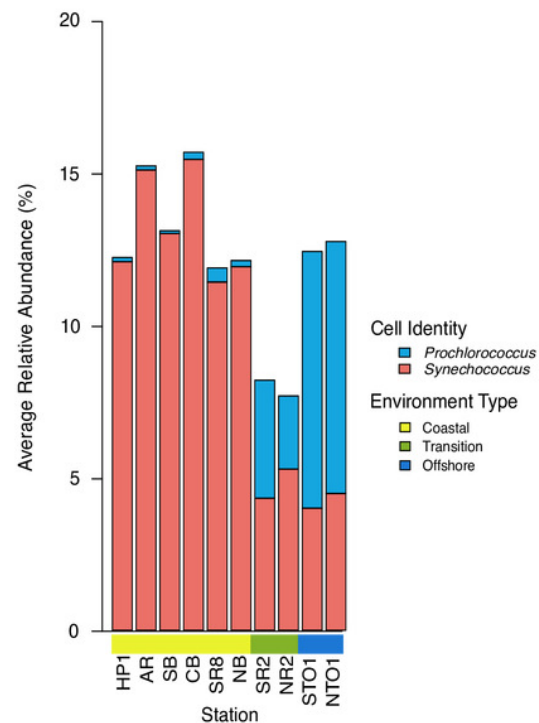



Figure 3

Spatial distribution of SAR11 subclades.

(**A**) Average number of SAR11 ASVs per subclade, and (**B**) subclade relative abundance across KByT stations. n=20 samples per station. (**C**) Average proportion of the total SAR11 relative abundance for the ubiquitous SAR11 ASVs. Seven ASVs are either ubiquitous within coastal (ASV16, ASV96), offshore (ASV34, ASV18, ASV4), or both (ASV47, ASV78) environments. Circle size indicates the relative proportion of SAR11 in the total microbial community. Map plotted in ArcGIS Pro 2.7.

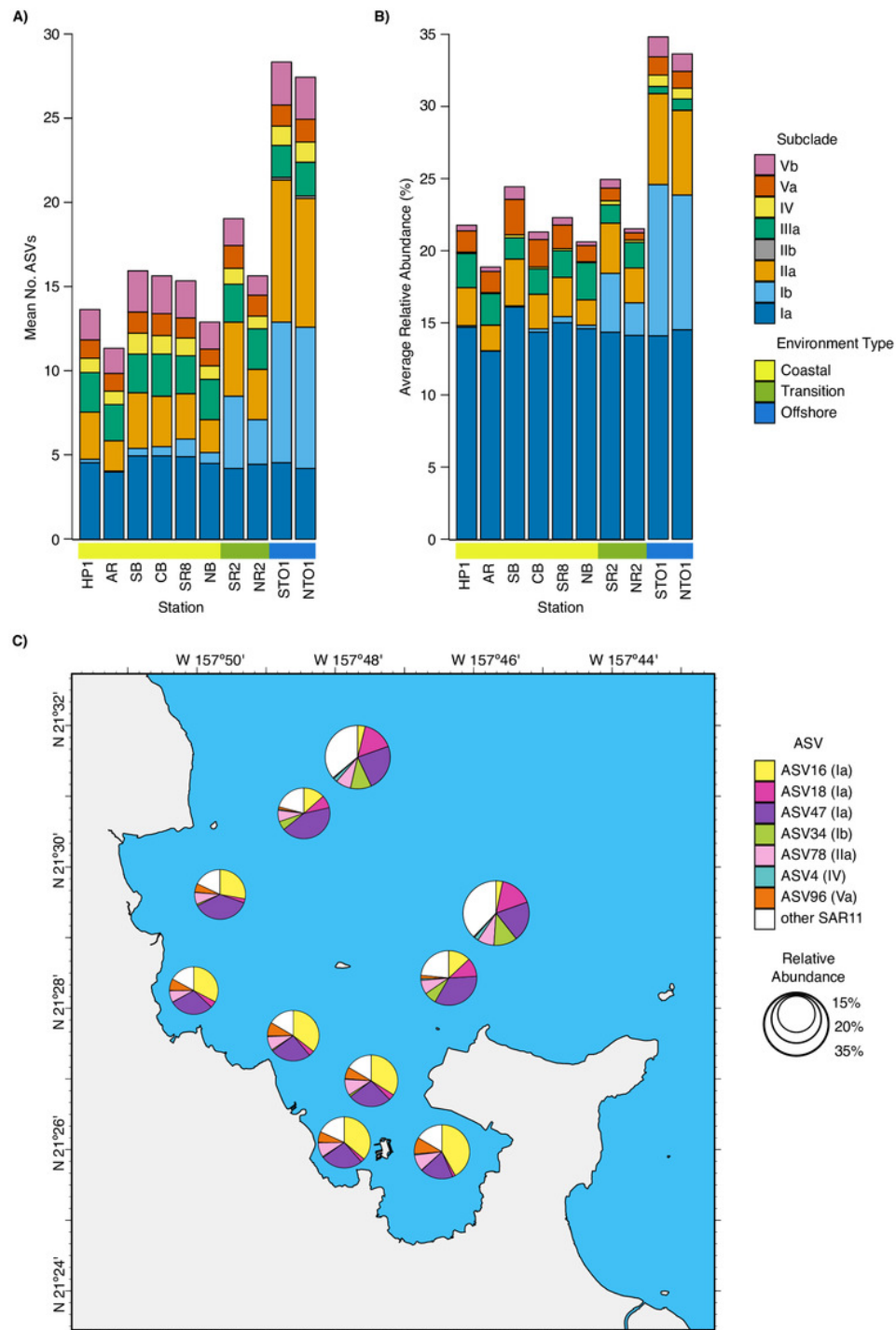



Figure 4

Environmental and seasonal distributions of SAR11 ASVs.

Venn diagrams comparing the distribution of total number of ASVs detected across **(A)** environments and **(B)** seasons.

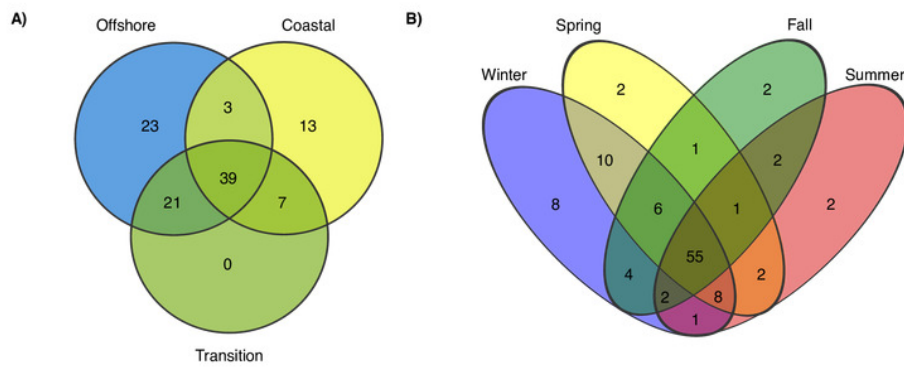
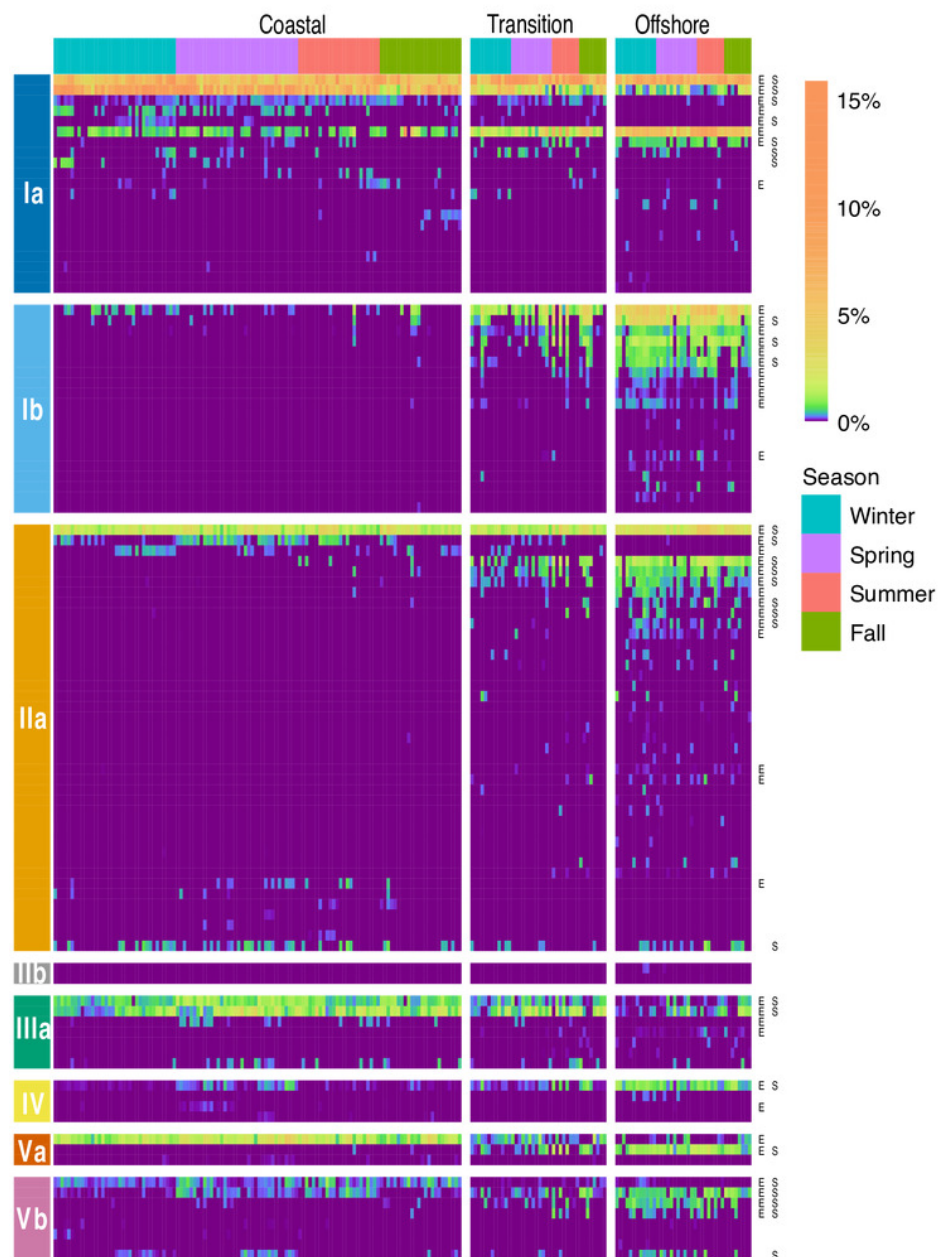



Figure 5

SAR11 ASV distributions.

Heatmap indicating the relative abundance of SAR11 ASVs (rows) per sample (columns). ASVs are ordered vertically by subclade, and samples are ordered horizontally by environment, season, date sampled, and site. "E" next to ASVs denotes a significant difference between environments, while "S" denotes a significant differences among seasons within a given environment. Wald tests to assess the significance of differences among environment and season for each SAR11 ASV were conducted using DESeq2 normalized counts.

Environmental parameters and cellular abundances from KByT.

Samples are averaged (mean ± s.d.) over environmental category and season.

1 **Table 1. Environmental parameters and cellular abundances from KByT.** Samples are averaged (mean ± s.d.) over environmental
2 category and season.

	Coastal				Transition				Offshore			
	Summer (n=24)	Fall (n=24)	Winter (n=36)	Spring (n=36)	Summer (n=8)	Fall (n=8)	Winter (n=12)	Spring (n=12)	Summer (n=8)	Fall (n=8)	Winter (n=12)	Spring (n=12)
Seawater temp. (ºC)	27.6 ± 0.6	26.5 ± 1.3	24.0 ± 0.4	26.5 ± 1.5	27.0 ± 0.6	26.3 ± 0.9	24.0 ± 0.5	25.4 ± 1.0	26.7 ± 0.5	26.2 ± 0.8	23.9 ± 0.6	25.2 ± 0.8
pH	7.9 ± 0.1	7.9 ± 0.2	8.0 ± 0.1	7.9 ± 0.2	8.0 ± 0.1	7.9 ± 0.3	8.0 ± 0.1	7.8 ± 0.2	8.1 ± 0.2	8.0 ± 0.2	8.1 ± 0.1	7.9 ± 0.1
Salinity	34.0 ± 1.2	34.4 ± 0.6	34.2 ± 0.6	35.1 ± 1.3	34.4 ± 0.8	35.2 ± 0.5	34.9 ± 0.5	35.5 ± 1.2	34.4 ± 0.8	35.3 ± 0.7	35.0 ± 0.4	35.6 ± 1.2
Chlorophyll a (μg L^{-1})	1.6 ± 0.4	1.5 ± 0.4	1.7 ± 0.5	1.5 ± 0.4	0.9 ± 0.5	0.7 ± 0.3	0.8 ± 0.3	0.9 ± 0.6	0.5 ± 0.3	0.8 ± 0.7	0.6 ± 0.3	0.8 ± 0.8
Heterotrophic bacteria (x10^6 cells mL^{-1})	1.4 ± 0.5	1.1 ± 0.5	1.3 ± 0.4	1.4 ± 0.4	0.8 ± 0.2	0.7 ± 0.4	0.7 ± 0.2	0.9 ± 0.2	0.7 ± 0.2	0.6 ± 0.2	0.7 ± 0.2	0.8 ± 0.2
Synechococcus (x10^4 cells mL^{-1})	26.6 ± 12.7	16.8 ± 13.5	16.5 ± 12.8	25.6 ± 9.7	4.0 ± 7.8	6.2 ± 7.8	1.8 ± 1.2	6.8 ± 4.5	4.3 ± 3.7	4.3 ± 3.7	1.6 ± 2.0	5.3 ± 4.0
Eukaryotic picophytoplankton (x10^4 cells mL^{-1})	2.9 ± 1.2	2.8 ± 1.2	2.9 ± 2.4	3.2 ± 0.9	0.9 ± 0.4	1.3 ± 1.0	0.9 ± 0.6	1.1 ± 0.6	0.6 ± 0.5	0.7 ± 0.1	0.9 ± 0.9	0.8 ± 0.4
Prochlorococcus (x10^4 cells mL^{-1})	0.1 ± 0.7	0.7 ± 0.1	0.0 ± 0.0	0.0 ± 0.0	2.8 ± 3.3	2.8 ± 2.4	1.8 ± 1.5	2.0 ± 1.4	6.9 ± 2.7	5.7 ± 3.1	6.9 ± 0.2	5.5 ± 1.9