Lawrence Berkeley National Laboratory

Title

In-depth Spatiotemporal Characterization of Planktonic Archaeal and Bacterial Communities in North and South San Francisco Bay

Permalink

https://escholarship.org/uc/item/85d9r9b0

Authors

Rasmussen, Anna N
Damashek, Julian
Eloe-Fadrosh, Emiley A
et al.

1 Diversity metrics of pelagic archaeal and bacterial communities in San Francisco Bay depend on

2 scale

3

4 Anna Rasmussen[1], Julian Damashek[1†], Emiley Eloe-Fadrosh[2], and Christopher A. Francis[1]*

5

6 [1]Department of Earth System Science, Stanford University, Stanford, CA 94305, USA

7 [2]Department of Energy Joint Genome Institute, Walnut Creek, California 94598, USA

8

9 *Correspondence: Christopher Francis, Department of Earth System Science, 473 Via Ortega,

10 Y2E2 Bldg Rm 140, Stanford University, Stanford, CA 94305, USA. Email: caf@stanford.edu

11

12 Running Title: Pelagic bacteria and archaea of San Francisco Bay

13

14

[†]Present Address: Department of Biology, Utica College, Utica, NY 13502, USA

15   **Originality-Significance Statement**

16   This is the first assessment of the biogeography of pelagic bacteria and archaea in San Francisco

17   Bay using high throughput sequencing. We amplified the V4 and V4-V5 regions of the 16S rRNA

18   gene in 174 samples collected during a two-year monthly time series along a 150-km transect of

19   San Francisco Bay using two different 'universal' primer pairs. We analyze diversity metrics at

20   several taxonomic levels and different physicochemical extents to reveal patterns in richness,

21   nestedness, turnover, and site-specificity and gain further insight into estuarine microbial

22   community structure. At the 97% OTU level, site-specificity and turnover are high while at the

23   phylum level organisms are more broadly distributed and the community is more nested.

24

25   **SUMMARY**

26   This study uses high throughput sequencing (HTS) to examine bacterioplankton and

27   archaeoplankton communities in 174 samples collected along a 150-km transect in San

28   Francisco Bay over a two-year monthly time-series. To better understand microbial

29   biogeography in San Francisco Bay, we analyzed communities using two different sets of 16S

30   rRNA primer sets at several taxonomic levels to reveal patterns in richness, nestedness, and

31   site-specificity. Our analysis reveals that both updated V4 and V4-V5 primers similarly describe

32   diverse estuarine microbial communities. We find that OTUs (97% identity) show high site-

33   specificity, occurring in a small subset of samples either defined by narrow salinity or temporal

34   ranges. At the OTU level, turnover is high along the salinity gradient and distinct brackish

35   communities are observed. At coarser taxonomic levels (e.g. phylum, class) taxa are broadly

36   distributed across salinity zones and communities appear to be a mix of fresh and marine end-

37   member communities. However, differential abundance testing shows that, despite high

38   prevalence across salinity zones, most phyla have preferences for a narrower salinity range.

39   While salinity is the dominant force shaping community structure, seasonal variations in

40   communities are observed within salinity zones. In addition, suspended particulate matter

41   concentrations are linked to patterns in alpha and beta diversity.

INTRODUCTION

San Francisco Bay (SFB) is the largest estuary on the west coast of the continental United States and is surrounded by approximately 7.6 million people (US Census Bureau 2017). SFB consists of two arms, generally referred to as North and South Bay, which have differing freshwater sources and water residence times (Walters *et al.*, 1985; Kimmerer, 2004). North Bay is river-dominated and includes San Pablo bay, Suisun bay, and the Delta region while South Bay is a weakly mixed marine lagoon. Intense urban development along the shores and other human activities such as damming, diking, dredging, historic mining, and pollution have led SFB to be considered one of the most anthropogenically altered estuaries in the United States (Nichols *et al.*, 1986).

Long term monitoring projects from both federal and state agencies have also made SFB one of the most studied estuaries in the world (Kimmerer, 2004) and, consequently, SFB has served as a model for understanding physical, chemical, and biological estuarine dynamics (Cloern, 1996, 2001; Lucas *et al.*, 1998, 2009; Cloern and Jassby, 2012; Raimonet and Cloern, 2017). For over four decades, water quality in SFB has been monitored regularly by the United States Geological Survey (USGS) (Schraga and Cloern, 2017), showing both gradual and abrupt changes in water quality due to human activity (Cloern *et al.*, 2017; Beck *et al.*, 2018; Cloern, 2019) and leading to a thorough characterization of phytoplankton dynamics (Cloern, 1987; Cloern and Dufford, 2005; Cloern and Jassby, 2010; Sutula *et al.*, 2017). In contrast to the in-depth monitoring of phytoplankton in SFB, bacterioplankton and archaeoplankton populations are remarkably understudied in this system. In fact, only three studies to date have examined pelagic microbial

64 community structure in SFB (Murray *et al.*, 1996; Hollibaugh *et al.*, 2000; Stepanauskas *et al.*,

65 2003), all of which used molecular approaches with limited phylogenetic resolution (i.e. DGGE,

66 T-RFLP) or sequencing depth. Here, we build considerably on this literature, using deep 16S

67 rRNA amplicon sequencing at a large spatial and temporal scale to understand bacterial and

68 archaeal ecology in the turbid estuarine waters of SFB.

69

70 SFB is an excellent model system for understanding microbial community dynamics because it

71 encompasses both spatial (e.g. salinity) and temporal (e.g. temperature) physicochemical

72 gradients (Cloern *et al.*, 2017). Salinity has been identified as a universal driver of prokaryotic

73 community structure (Lozupone and Knight, 2007; Thompson *et al.*, 2017), as well as a key

74 driver of estuarine bacterial community composition in numerous studies (Crump *et al.*, 1999,

75 1999, 2004, 1999; Fortunato and Crump, 2011; Herlemann *et al.*, 2011; Fortunato *et al.*, 2012;

76 Mason *et al.*, 2016; Doherty *et al.*, 2017). However, studies in estuarine environments have also

77 identified drivers of microbial community structure besides salinity, including temperature, pH,

78 dissolved oxygen, water residence time, organic carbon or nutrient availability (Murrell *et al.*,

79 1999; Hollibaugh *et al.*, 2000; Crump *et al.*, 2004; Herlemann *et al.*, 2011; Liu *et al.*, 2014;

80 Satinsky *et al.*, 2014). To further assess broader scale patterns in microbial biogeography, we

81 investigate diversity metrics at varying taxonomic grain sizes (Thompson *et al.*, 2017; Ladau and

82 Eloe-Fadrosh, 2019).

83

84 While we expect salinity to be the main driver of community structure in SFB, we address the

85 following questions: What other environmental variables influence alpha and beta diversity

5

86   metrics throughout the bay and within different salinity zones? Are samples nested along the

87   salinity gradient? How does taxonomic grain size impact diversity metrics? Do most organisms

88   have broad or narrow distributions? How does primer choice impact the characterization of

89   microbial community composition and structure? We sampled 12 stations ranging from fresh

90   riverine inputs to brackish mixing zones to highly marine-influenced regions along a ~150-km

91   transect of the SFB channel. We sequenced bottom water samples collected monthly over two-

92   years, capturing microbial communities across several seasonal gradients (e.g. temperature,

93   freshwater flow rate) as well. Amplicon libraries were generated using two updated 16S rRNA

94   primer sets, which amplify variable region 4 (Apprill *et al.*, 2015; Parada *et al.*, 2016) and

95   regions 4 and 5 (Parada *et al.*, 2016), respectively. Both primer sets were designed to remedy

96   known biases against SAR11 and Thaumarchaeota—two of the most abundant microbial taxa

97   on Earth and important organisms in marine and estuarine environments, including San

98   Francisco Bay.

99   **RESULTS & DISCUSSION**

100   **Environmental Data**. Water column samples within this dataset correspond to salinities ranging

101   from fresh (minimum 0.07) to euhaline (maximum 32.42) and temperatures from 6.8 to 22.7 °C

102   (Fig.1). Examples of stations typically falling into the Venice salinity zones (Battaglia, 1959) are

103   as follows: station 657 is fresh (<0.5), 649 is oligohaline (.05 to <5), 6 is mesohaline (5 to <18),

104   13 is polyhaline (18 to <30), and 18 is euhaline (30 to <40) (Fig. 1). Chlorophyll a concentrations

105   were typically low (~3µg/L) with peaks concentrations occurring in late winter/spring and in

106   South Bay (Fig. S1). Ammonium concentrations were highest in riverine samples due to inputs

107   from the Sacramento Regional WWTP, while nitrate concentrations were highest in Lower

108   South Bay (Fig. S1). Nitrite concentrations occasionally reached high concentrations (>9 μM) in

109   South Bay. High nitrite concentrations in this region of the bay have been reported previously

110   (Wankel *et al.*, 2006; Damashek *et al.*, 2016). The wet season generally has lower temperatures,

111   higher suspended particulate matter (SPM) concentrations, higher delta outflow, and lower

112   salinities in each region of SFB when compared to the dry season (Fig S2).

113

114   **Updated 16S rRNA V4 and V4-V5 primers similarly describe estuarine microbial communities**.

115   For the remainder of this text, 16S rRNA libraries amplified by the 515F-Y (Parada *et al.*, 2016)

116   and 806RB (Apprill *et al.*, 2015) primers will be referred to as the 'V4 dataset' and libraries

117   amplified by 515F-Y and 926R (Parada *et al.*, 2016) will be referred to as the 'V4-V5' dataset,

118   based on which variable regions they amplify. While library sizes vary between sequencing runs

119   and primer pairs, the general description of microbial communities is strikingly similar. Despite

120   some differences in taxonomy between datasets (Table S1), large-scale interpretation of alpha

121   and beta diversity is essentially the same for both primer pairs (Fig. S3). The relative abundance

122   of phyla that occur in both primer sets is strongly correlated (Fig. S3A; r^2 = 0.99, $p <$ 0.001), as is

123   the richness within phyla (Fig. S3B; r^2 = 0.99, $p <$ 0.001). There is also a correlation between the

124   relative abundance of top genera (Fig. S3C; r^2 = 0.74, $p <$ 0.0001). Sample richness is correlated

125   between primer pairs (r^2 = 0.83, $p <$ 0.001) and mostly strongly correlated to SPM for both

126   primer pairs (Table S2). Procrustes analyses of PCoA ordinations reveal very similar patterns

127   (Fig. S3D; m^2 = 0.007, r^2 = 0.996) and relationships between PCoA axes and environmental

128   variables are also consistent (Table S2).

129

130   Since updated primers were paired and tested with mock communities lacking both SAR11 and

131   *Thaumarchaeota* (Walters *et al.*, 2016) and the updated V4 primer pair has subsequently been

132   adopted by the Earth Microbiome Project (EMP), we are particularly interested in comparing

133   diversity metrics of these environmentally-relevant groups. In this study, SAR11 and

134   *Thaumarchaeota* are abundant in amplicon libraries constructed with both primer pairs, and

135   display similar distribution patterns (Fig. S4-S5).

136

137   While communities are generally similar in both datasets, there were some minor differences.

138   We note that the V4-V5 dataset contains more phyla but fewer OTUs (Table S1) and there are

139   differences in the relative abundances of certain groups of organisms (Fig. S3A,C). For example,

140   some archaea are more relatively abundant in the V4-V5 dataset (*Thaumarchaeota* and

141   *Bathyarchaeota*) while others are more relatively abundant in the V4 (*Woesarchaeota and*

142   *Euryarchaeota*) (Fig. S6). Despite these differences in relative abundance, beta diversity metrics

143   are very similar for each dataset including the whole community (Fig. S3D) or just archaeal

144   OTUs (Fig. S6). Thus, depending on the target organisms or analyses of interest, one primer pair

145   may be better than the other, but for overall community diversity analysis of estuarine samples

146   both primer pairs work similarly. Most diversity metrics are similar regardless of primer pair

147   (Fig. S3, Table S2) and differences occurring in classification of OTUs could in part be due to

148   differing amplicon lengths.

149

150 **Salinity, SPM, seasonality, and region influence community structure in SFB.** We used a

151 variety of analyses to assess beta diversity. We expected salinity to be a key factor in shaping

152 communities, as observed in a wide variety of estuaries (Crump *et al.*, 2004; Hewson and

153 Fuhrman, 2004; Herlemann *et al.*, 2011; Fortunato *et al.*, 2012; Aguirre *et al.*, 2017), and the

154 findings in our study agree with this body of literature. First we used ordination analyses to

155 assess beta diversity based on Bray-Curtis dissimilarity between communities. The first two

156 axes explain 54.3% of total variation in PCoA plots (Fig. 2). Salinity is strongly correlated to Axis

157 1 in PCoA plots (r^2 = 0.94, $p <$ 0.001; Table S2). Based on a "goodness of clusters" gap statistic,

158 PCoA ordination forms 4 clusters that generally correspond with salinity zone [fresh (<0.5),

159 oligohaline (.05 to <5), mesohaline (5 to <18), polyhaline (18 to <30), and euhaline (30 to <40)],

160 with polyhaline and euhaline samples forming less distinct groupings (Fig. S7). Using the

161 *betapart* function, we confirm that turnover is a much larger component than nestedness in

162 community dissimilarity metrics (Table S3). A heatmap based on Bray-Curtis dissimilarity and

163 hierarchical clustering based on Jaccard dissimilarity both show that samples generally group by

164 salinity zone (Fig. S8). Graph-based testing using Friedman-Rafsky tests on a minimum spanning

165 tree, a distance-threshold graph, and a k-nearest neighbor graph all show that samples within

166 salinity zones contain more pure edges than the null distribution ($p <$ 0.001, Fig. S9), indicating

167 samples within a salinity zone are more similar to one another than expected by chance. Taken

168 together, ordination, hierarchical clustering, and graph-based tests all support that salinity zone

169 definitions are ecologically relevant.

170

171 Spatial/salinity gradients have overwhelmed seasonal variation in other estuarine communities

172 (Fortunato *et al.*, 2012) so we examined diversity metrics within salinity zones. PCoA

173 ordinations within a given salinity zone reveal stronger seasonal groupings. In fresh, oligohaline,

174 and mesohaline ordination plots, there is a separation between samples from the wet versus

175 the dry season (Fig. S10). Axes correlate strongly with variables that vary seasonally, including

176 temperature and SPM (Table S4).

177

178 Inspection of ordinations, networks, and heatmaps shows that polyhaline and euhaline samples

179 may cluster together more than other zones (Figs. 2, S7-S9), but also indicates that South and

180 North Bay samples are more distinct from one another even when samples have similar

181 salinities. This may be unsurprising given the differences in residence times and hydrology in

182 the two arms of the bay (Walters *et al.*, 1985; Kimmerer, 2004), leading them to often be

183 classified as two separate ecosystems. We investigated South Bay and North Bay samples

184 separately. While beta diversity in the North Bay is predominantly influenced by salinity,

185 ordinations of South Bay samples show a strong influence of SPM and temperature on axes 1

186 and 2, respectively (Fig. 2, Table S4).

187

188 SPM concentration influences both alpha diversity measures of richness (Fig. 3) and beta

189 diversity (Fig. 2) of pelagic bacteria and archaea in SFB. Community richness most strongly

190 correlates with SPM concentrations at all taxonomic levels (i.e. phylum through OTU) and does

191 not appear to strongly correspond to salinity (Table S2). Richness of samples within each salinity

192 zone and in either North or South Bay is most strongly correlated to SPM (Fig. S10, Table S4).

193 SPM varies over space and time, which can make other spatial and temporal patterns in

194 richness difficult to observe (Fig. 3). In PCoA ordinations of all SFB samples, axis 2 most strongly

195 correlates to SPM (Fig. 2, Table S2). NODF-based nestedness analyses support that less-

196 rich/low-SPM communities are a nested subset of richer, high-SPM associated communities at

197 the phylum through genus level but not the OTU level (Fig. 4). Thus, at the finest taxonomic

198 scale (OTU 97%), less-rich/low-SPM communities do not appear to be a nested subset of more-

199 rich/high-SPM communities (Fig. 4). Our findings that communities are nested at coarser

200 taxonomic levels is in agreement with a previous study that found free-living and particle-

201 associated communities were very similar using DGGE (Hollibaugh *et al.*, 2000), a technique

202 that may have identified organisms at a coarser taxonomic scale than the 97% identity OTU

203 level.

204

205 SPM could be important for microbial community composition because it provides nutrients for

206 organisms and/or has a long residence time that allows unique, particle-associated

207 communities to form on and around particles (Hollibaugh *et al.*, 2000). Indeed, phytoplankton

208 dynamics and bacterioplankton activity have been linked to SPM dynamics in SFB (Cloern, 1987;

209 Hollibaugh and Wong, 1999; Murrell *et al.*, 1999; Hollibaugh *et al.*, 2000, 200). Microbes

210 associated with resuspended sediments could also explain the potential distinction between

211 low- and high-SPM associated communities. This idea is supported by the negative correlation

212 between richness and the ratio of active chlorophyll *a* to degraded phaeopigments (Table S2). A

213 low ratio of active chlorophyll *a* to phaeopigments indicates strong resuspension of bottom

214 sediments, which are rich in degraded algal material. Our study does not differentiate between

215 'free-living' and 'particle-associated' organisms using the same size fractions as in Hollibaugh *et*

216 *al.* (2000) and includes all organisms small enough to pass through a 10 μm filter but captured

217 by a 0.22 μm filter. While our findings indicate that particle-associated communities could be

218 more diverse and distinct from free-living communities, further studies are necessary to directly

219 address differences between various size fractions of particle-associated versus free-living

220 communities at the OTU level as well as tease apart the relationships between microbial

221 community composition and SPM quantity, quality, and source.

222

223 **Salinity tolerance varies at different taxonomic grain sizes.** To further understand how

224 communities change along the salinity gradient, we looked more deeply at community diversity

225 nestedness and taxa entropy. We used the NODF statistic to calculate if less rich communities

226 were a subset of richer communities. We find that communities are highly nested at coarser

227 taxonomic levels (e.g. phylum, class) and become gradually less nested at finer taxonomic

228 levels, before becoming substantially less nested at the OTU level (Fig. 4). All NODF statistics are

229 significantly greater than the null model values except at the OTU level, indicating that at most

230 taxonomic scales less diverse communities are to some extent a nested subset of more diverse

231 communities (Fig. 4). The lack of significance and low NODF values indicate that OTUs are not

232 nested.

233

234 We also calculated entropy using the Shannon index as a measure of site-specificity (i.e. how

235 many different samples a given taxa occurs in) of a given taxa. Entropy of taxa decreases from

236 coarse to finer taxonomic scales (Fig. 4), indicating that the site-specificity of taxa increases at

237   finer scales such as at the OTU level. Higher entropy values at the phylum level indicate less

238   site-specificity, while lower entropy values at the OTU level indicate greater site-specificity. This

239   can also be observed visually in relative abundance bar plots, with most abundant phyla

240   occurring across a broad range of samples while abundant genera or OTUs occur in a smaller

241   subset of samples (Fig. 5, Fig. S11-S16).

242

243   Nestedness and taxa entropy are both greater at coarse taxonomic levels and gradually

244   decrease from the phylum to genus level, then decrease substantially at the OTU level (Fig. 4).

245   These findings support that at the phylum level, organisms have broader salinity tolerance but

246   organisms at the 97% identity level are adapted to a specific salinity range and turnover is high

247   along the salinity gradient, potentially with closely related organisms replacing their more or

248   less salt-tolerant relatives along the gradient. Bar plots reveal clear patterns in each salinity

249   zone, with a gradual distinction between freshwater and saline sites at the class and order level,

250   followed by more specific mesohaline communities emerging at the genus and OTU level (Figs.

251   S11-16). Because richness is not correlated strongly with salinity at any taxonomic level,

252   communities are not necessarily nested along the salinity gradient. Rather, the low nestedness

253   of OTUs and partitioning of beta diversity metrics indicate that turnover dominates along the

254   salinity gradient (Fig. 2, Fig. 4, Table S3). The high nestedness of samples at the phylum level

255   and low nestedness at the OTU level agrees with recent findings from the EMP dataset

256   (Thompson *et al.*, 2017) and support similar findings in the brackish Baltic Sea (Herlemann *et*

257   *al.*, 2011). Interestingly, differential abundance testing shows most phyla (33/46) vary

258   significantly in abundance across salinity zones (Fig. 4, S17). At all taxonomic levels, many taxa

259 (between 40 and 75%) show significant variation in abundance between the five salinity zones

260 (Fig. 4). OTUs show the greatest magnitude of abundance change (Fig. 4). Thus, while phyla

261 have broader distributions along the salinity gradient in terms of presence/absence there are

262 significant variations in abundance between salinity zones, supporting that even at the phylum

263 level there are ecologically meaningful adaptations based on salinity.

264

265 **Abundant microbial taxa in SFB**

266 The top 10 phyla in the V4 and V4-V5 datasets are *Proteobacteria*, *Bacteroidetes*,

267 *Actinobacteria*, *Verrucomicrobia*, *Cyanobacteria*, *Thaumarchaeota*, *Planctomycetes*,

268 *Euryarchaeota*, *Marinimicrobia*, and *Acidobacteria* (Fig. S11). OTUs in these 10 phyla account

269 for over 90% of reads in each sample (Fig. S11). These abundant phyla have broad distributions

270 across samples, though differential abundance testing shows that only *Bacteroidetes*,

271 *Thaumarchaeota*, and *Cyanobacteria* do not have significant changes in abundance between

272 salinity zones (Figs. S17). Of these abundant phyla, *Actinobacteria*, *Verrucomicrobia*, and

273 *Acidobacteria* are more abundant near fresh end-member sites while *Marinimicrobia*,

274 *Euryarchaeota*, *Planctomycetes*, and *Proteobacteria* are more abundant closer to marine end-

275 member sites.

276

277 *Proteobacteria* is the most abundant and richest phylum in the SFB dataset (Fig. S2). Common

278 patterns observed in estuaries such as decreasing *Betaproteobacteria* and increasing *Alpha-*

279 and *Gammaproteobacteria* with increasing salinity (Murray *et al.*, 1996; Crump *et al.*, 1999,

280 1999; Bouvier and Giorgio, 2002; Zhang *et al.*, 2006; Kan *et al.*, 2008; Herlemann *et al.*, 2011)

281 were also observed in our study (Fig. S11,). At the order level, *Oceanospiralles* and

282 *Rhodobacterales* are more abundant in marine-influenced samples and *Burkholderales*

283 dominates in freshwater end-member stations (S13). In general, OTUs have lower entropy

284 (higher site-specificity) and abundant OTUs have narrower salinity ranges than observed at

285 coarser taxonomic levels within the *Proteobacteria* (e.g. class, order, family) (Fig. S11). Richness

286 of *Proteobaceteria* OTUs is most strongly correlated with SPM concentrations (r^2 = 0.42) and

287 ordinations highlight the prominent role of salinity in shaping the *Proteobacteria* community

288 (Fig. S11).

289

290 Unlike other orders of *Proteobacteria*, SAR11 clade bacteria (*Alphaproteobacteria*) are

291 abundant along the entire salinity gradient. SAR11 are the most abundant organisms at the

292 order level (generally ~10-30% of reads in a sample), constitute one of the most abundant

293 genera (*Pelagibacter*), and correspond to many of the top OTUs in our dataset (Figs. 5, S5, S11).

294 SAR11 are highly structured along the salinity gradient in SFB, a pattern observed in other

295 estuaries (Kan *et al.*, 2008; Campbell and Kirchman, 2013; Herlemann *et al.*, 2014). The

296 abundance of SAR11 in our dataset bolsters previous findings in SFB indicating that SAR11 clade

297 bacteria are ubiquitous (Murray *et al.*, 1996) and LD12 organisms are abundant in fresh sites

298 (Stepanauskas *et al.*, 2003). The distribution of 'LD12 clade' organisms in our study aligns with

299 findings from recent cultured isolates (Henson *et al.*, 2018), with peak LD12 abundances

300 occurring in primarily fresh and oligohaline sites (salinity < 5; Fig. S11). LD12 are proposed to be

301 specialized to freshwater environments through the loss of key compatible solute genes

302 (Henson *et al.*, 2018), highlighting one way this group may have differentiated itself from its

15

303   relatives and adapted to a specific niche. A family identified as the 'Chesapeake Delaware Bay

304   clade' [SAR11 IIIa (Kan *et al.*, 2008)] is most abundant at mesohaline and oligohaline sites, while

305   'Surface 1' [SAR11 Ia (Vergin *et al.*, 2013)] OTUs dominate at poly- and euhaline sites. This

306   partitioning of groups along the salinity gradient can also be observed from the family to the

307   OTU level (Figs. 5 & S5). Seasonal variations in SAR11 communities are also observed in

308   ordination plots (Figs. S5).

309

310   *Bacteroidetes*, and particularly *Flavobacteriia*, which are important for the breakdown of

311   organic matter in estuaries and coasts (Williams *et al.*, 2013; Smith *et al.*, 2017, 201), are also

312   abundant in SFB waters (Fig. S11). *Bacteroidetes* is the second most abundant and second

313   richest phylum in the dataset (Fig. S3), in agreement with previous findings that marine

314   *Flavobacteria* have high global and local diversity (Alonso *et al.*, 2007). *Flavobacteriales* is one

315   of the most abundant orders in the dataset. The NS5 marine clade genus of *Flavobacteriaceae*

316   is among the top genera (Figs. S3 & S11). In fresh end-member stations the families

317   *Chitinophagaceae*, *Cytophagaceae*, and 'NS11-12 marine group' (Alonso *et al.*, 2007) dominate

318   (Fig. S14). Brackish stations are dominated by the family *Cryomorphaceae* while marine end-

319   member stations are dominated by *Flavobacteraceae*. Bacteroidetes OTUs are more site-

320   specific and have narrower salinity tolerance than at coarser taxonomic levels and richness is

321   most strongly correlated to SPM concentrations (Fig. S14, r^2 = 0.49).

322

323   *Actinobacteria* is the third most abundant phylum in SFB. *Candidatus Actinomarina* and 'hgcl

324   clade' (Glöckner *et al.*, 2000) are among the top genera (Figs. S3 & S11), with diverse 'hgcl

16

325   clade' organisms dominating freshwater end member stations and *Ca. Actinomarina*-like OTUs

326   dominating in more saline samples (Fig. S12 & S15). Other estuarine studies have found high

327   diversity and specialization of *Actinobacteria* with differing environmental gradients (Kirchman

328   *et al.*, 2005; Holmfeldt *et al.*, 2009; Campbell and Kirchman, 2013), which we observe as well.

329   *Actinobacteria* richness is most strongly correlated with salinity and ordination plots show

330   strong separation of communities based on wet versus dry season (Fig. S15). While there

331   appears to be more OTU-level diversity in the freshwater stations, only one *Ca. Actinomarina-*

332   *like* OTU (OTU 0) dominates the most marine-influenced station (18). Poly/Euhaline stations 13

333   and 27 are dominated by OTUs 0 and 8, indicating microdiversity may exist below the 97% OTU

334   definition (Fig S12 & S16). Metagenomes of *Ca. Actinomarina* organisms suggest they are very

335   small, free-living photoheterotrophic organisms co-occuring with *Synechococcus* in marine

336   photic zones and containing streamlined and low-GC genomes with novel rhodopsins (Ghai *et*

337   *al.*, 2013; Mizuno *et al.*, 2015). *Ca. Actinomarina*-like OTU 8 can reach over 10% relative

338   abundance in late summer in South Bay station 27 and shows smaller but noticeable peaks in

339   station 13 as well (Fig. S12 & S16). Both fresh and marine planktonic *Actinobacteria* have been

340   described as photoheterotrophs, and some are capable of degradation of recalcitrant organic

341   matter (Ghai *et al.*, 2014), highlighting their potential functional role in SFB.

342

343   *Cyanobacteria* are generally low abundance in non-summer months in SFB and are dominated

344   by *Synecchococcus*-like organisms (Fig. 5). Ordination plot shows strong seasonal variation in

345   *Cyanobacteria* communities (Fig. S18). While salinity most strongly impacts community

346   structure for most taxa, we do observe seasonal variation within groups such as *Cyanobacteria*,

347   *Actinobacteria*, and SAR11. Stronger temporal variations within specific taxonomic groups as

348   compared to the whole community have been observed in other estuaries (Fortunato *et al.*,

349   2013, 2; Li *et al.*, 2017).

350

351   Bacteria are relatively more abundant than archaea, which make up only ~2% of the overall

352   reads in our dataset. In freshwater sites, archaea generally have a low relative abundance (0.02

353   – 0.2%) and are predominantly *Woesearchaeota* and *Thaumarchaeota* (Fig. S6). However, in

354   poly- and euhaline sites *Thaumarchaeota* and *Euryarchaeota* can become abundant, comprising

355   over 20% and roughly 6% of reads, respectively (Fig. S6)*.*

356

357   *Thaumarchaeota* genera and OTUs are distinct at the fresh and marine end-member sites, with

358   organisms generally considered fresh (*Nitrosoarchaeum*-like) (Blainey *et al.*, 2011; Mosier *et al.*,

359   2012) populating station 657 and marine (*Nitrosopelagicus*) (Santoro *et al.*, 2015) populating

360   station 18 (Fig. 5). The apparent 'bloom' of *Thaumarchaeota* in South Bay, with putative

361   *Nitrospumilus*-like sequences comprising as much as 25% of the total reads, has not been

362   previously described for this system. Blooms of *Thaumarchaeota* have been reported in other

363   estuaries in summer (Hollibaugh *et al.*, 2014; Schaefer and Hollibaugh, 2017), but also in coastal

364   areas in fall (Hu *et al.*, 2013; Kim *et al.*, 2019) or winter (Wuchter *et al.*, 2006; Pitcher *et al.*,

365   2011). Interestingly, this bloom is associated with nitrite accumulation (Fig. 5, $r^2 = 0.79$), a

366   phenomenon observed in other coastal and estuarine sites (Schaefer and Hollibaugh, 2017; Kim

367   *et al.*, 2019). Warmer temperatures between 20°C and 30°C have been proposed to explain the

368   decoupling of ammonia and nitrite oxidation in other systems; however, this SFB bloom occurs

369 in fall when temperatures are generally at or below 20°C and decreasing (Fig. 1). This apparent

370 decoupling of ammonia and nitrite oxidation in South San Francisco Bay is intriguing and in

371 need of further investigation. The high abundance of these *Thaumarchaeota* could also have

372 implications for nitrogen cycling; indeed, nitrification rate measurements for this area of the

373 bay are limited but could potentially be high (Damashek *et al.*, 2016).

374

375 **Brackish communities are distinct at the OTU (97% similarity) level.** While at the class and

376 family level microbial communities appear to be a mix of freshwater and marine end-members

377 (Figs. S11), at the OTU level communities display distinct fresh, brackish, and marine

378 communities along the salinity gradient (Fig. S11-S16). Classes such as *Sphingobacteriia*,

379 *Cytophagia*, *Actinobacteria*, and *Betaproteobacteria* are abundant at fresh end-member sites

380 but decrease along the salinity gradient, as classes such as *Alphaproteobacteria*,

381 *Gammaproteobacteria*, *Flavobacteriia*, and *Acidimicrobiia* become more abundant towards

382 marine end-member sites (Fig. S11). At fresh and oligohaline sites, families such as

383 *Comamonadaceae*, *Chitinophagaceae*, *Sporichthyaceae*, and 'LD freshwater clade' (SAR11) are

384 abundant, while at polyhaline and euhaline sites *Rhodobacteraceae* and OM1 (*Actinobacteria*)

385 are abundant (Fig. S11). At mesohaline sites, families such as 'Chesapeake Delaware Bay clade'

386 [SAR11 IIIa (Kan *et al.*, 2008)] dominate. More distinct brackish communities start to emerge at

387 the genus and OTU levels (Fig. S11). Some of the OTUs with highest abundance in brackish sites

388 belong to taxa such as the SAR11 'Chesapeake Delaware Bay clade' (OTU 2), *Oceanospirillaceae*

389 (OTU 13), *Hydrogenophilaceae* (OTU 55), *Ca. Actinomarina* (OTU 8), *Owenweeksia* (OTU 25),

390 AEGEAN-169 clade of the *Rhodospirillaceae* family (OTU 14), and the NS5 marine group of the

391   *Flavobacteriaceae* family (OTU 13) (Figs. S12, S17). The uncultivated groups NS5 and AEGEAN-

392   169 have been associated with phytoplankton blooms or the breakdown of dissolved organic

393   matter (Gómez-Pereira *et al.*, 2010; Yang *et al.*, 2015; Bennke *et al.*, 2016; Seo *et al.*, 2017;

394   Ngugi and Stingl, 2018).

395

396   Regardless of variations in season, estuary type (North Bay is river-dominated versus South Bay,

397   which is more of a marine lagoon), or taxonomic group (e.g. *Proteobacteria*, *Actinobaceria*,

398   SAR11), a distinct brackish community emerges at the OTU level. Previous studies have posited

399   that distinct brackish microbial communities arise in systems where bacterial doubling rates are

400   faster than water residence times (Crump *et al.*, 2004; Herlemann *et al.*, 2011). Given the

401   variety of seasons and systems sampled, the findings of our study suggest that defined brackish

402   communities may also arise over shorter time scales and survive in waters with a variety of

403   water residence times. A recent, high-resolution daily time series found that coastal microbial

404   communities are constantly and quickly changing, with rapid transitions between distinct but

405   transient communities as coastal conditions change (Martin-Platero *et al.*, 2018). The

406   apparently robust and defined estuarine communities along the salinity gradient observed in

407   our dataset suggest that communities have adapted to a specific salinity regime and may be

408   capable of quickly transitioning with changing salinity gradient structures. Our study strongly

409   supports previous findings that distinct brackish bacterial communities exist and that closely

410   related organisms specialize/adapt to distinct salinity regimes (Herlemann *et al.*, 2011; Liu *et*

411   *al.*, 2015; Mehrshad *et al.*, 2016). Despite the unprecedented spatiotemporal sampling (and

412   sequencing) of SFB pelagic communities in this study, more fine-scale temporal sampling is

413 necessary to confirm how rapidly communities shift with changing salinity gradient structure

414 and to tease apart if communities are ephemeral (quickly turning over) or if residence times are

415 generally long enough to allow for distinct estuarine communities to develop regularly in SFB.

416 **EXPERIMENTAL PROCEDURES**

417 **Sampling and DNA extraction.** Bacterioplankton and archaeoplankton biomass was collected

418 for DNA extraction approximately monthly between April 2012 and March 2014 during USGS

419 Water quality monitoring cruises (https://sfbay.wr.usgs.gov/access/wqdata/index.html) in the

420 channel of the San Francisco Bay estuary. Microbial cells were collected from bottom waters

421 (1m above estuary floor) by pressure-filtering 150-1000mL of water from CTD casts through a

422 10-μm pore size polycarbonate Isopore membrane filter (47-mm diameter; EMD Millipore,

423 Darmstadt, Germany) in line with a 0.22 μm polyethersulfone Supor-200 membrane filter (47-

424 mm diameter; Pall, Port Washington, NY), followed by flash freezing on liquid nitrogen prior to

425 storage at -80°C. Only 0.22 μm filters were frozen at -80°C and saved for extraction. DNA was

426 extracted with the FastDNA SPIN Kit for Soil (MP Biomedicals, Santa Ana, CA), following the

427 manufacturer's instructions, with the following modifications: bead tubes were homogenized

428 for 40 seconds at speed 6.0 in a FastPrep bead beater, and final DNA was eluted into 75 μL 55°C

429 sterile DNase-free water. DNA was quantified using the Qubit dsDNA Broad Range assay (Life

430 Technologies, Grand Island, NY). DNA was stored at -80°C.

431

432 **Environmental Data**. Corresponding water quality data from sampling cruises was downloaded

433 from the USGS San Francisco Bay Water Quality and California Day Flow websites, available at

434   the   following   links: (https://sfbay.wr.usgs.gov/access/wqdata/query/index.html   and

435   http://www.water.ca.gov/dayflow/). Additional ammonium, nitrate, and nitrite measurements

436   were made using filtered (0.2 µm pore size) water that was frozen on dry ice prior to storage at

437   -20°C. Ammonium was measured using the salicylate-hypochlorite method (Bower and Holm-

438   Hansen, 1980). Nitrate and nitrite were measured using a SmartChem200 Discrete Analyzer

439   (Unity Scientific, Brookfield, CT) following standard procedures. Nutrients were measured

440   within one week of sample collection.

441

442   **Sequencing**. Through a DOE Joint Genome Institute (JGI) Community Science Program (CSP)

443   project, we have assembled 348 16S rRNA Illumina amplicon libraries from bottom water

444   samples (collected 1m above the estuary floor) on 20 approximately monthly cruises at 12

445   USGS monitoring stations spanning a 150 km transect (Fig. 1). In total, 174 samples were

446   sequenced with both new 16S V4 primers (Apprill *et al.*, 2015; Walters *et al.*, 2016) (515F-Y

447   GTGYCAGCMGCCGCGGTAA and 806RB GGACTACNVGGGTWTCTAAT) and 16S V4-V5 primers

448   (Parada *et al.*, 2016) (515F-Y and 926R CCGYCAATTYMTTTRAGTTT) and were included in the

449   following analyses. Samples were pooled and sequenced on an Illumina MiSeq. Sequences are

450   available on NCBI SRA under the BioProject PRJNA577706.

451

452   **Data Processing.** Raw reads were processed by JGI using the iTagger v2.2 method

453   (https://bitbucket.org/berkeleylab/jgi_itagger), which uses Usearch (Edgar, 2010), MAFFT

454   (Katoh *et al.*, 2002) and QIIME (Caporaso *et al.*, 2010). Reads were clustered at decreasing

455   levels of identity until the 97% cutoff to create OTUs. Low abundance sequences were not used

456   to cluster and were later mapped back to cluster centroids. Cluster centroid sequences were

457   evaluated with the reference database SILVA 128.

458

459   In general, amplicon library processing followed established protocols for filtering low

460   abundance reads, transforming read counts, and normalizing library sizes (McMurdie and

461   Holmes, 2014; Callahan *et al.*, 2016). To prevent high-variance-low-abundance OTUs from

462   strongly influencing downstream analyses, OTUs with less than 16 reads occurring in less than 3

463   samples were removed from libraries. Libraries were normalized in two ways depending on the

464   analysis, either through transforming counts to relative abundances or using the DESeq2

465   method, which uses a variance stabilizing transformation to adjust counts based on variation in

466   library size, and transformed using a geometric mean to account for large variation in OTU

467   counts (McMurdie and Holmes, 2014). Similar results were yielded with rarefied data (data not

468   shown). All analyses were conducted with R using primarily *phyloseq* (McMurdie and Holmes,

469   2013, 2014; Callahan *et al.*, 2016) and *vegan* (Oksanen *et al.*, 2018). Richness was calculated for

470   unfiltered amplicon libraries and libraries with OTUs filtered at a cutoff of greater than 15 reads

471   in greater than 2 samples, yielding similar results [data not shown]. Beta diversity metrics using

472   Bray-Curtis dissimilarity were calculated for VST or relative abundance transformed libraries.

473

474   Environmental data was centered and scaled prior to correlation testing (*scale*, *cor.test*, and

475   *ggpairs* functions). The natural log was used to transform variables where values ranged several

476   orders of magnitude (e.g. suspended particulate matter, chlorophyll). Samples were assigned

477   into salinity zones defined as fresh, oligohaline, mesohaline, polyhaline, and euhaline (Venice

478   system, (Battaglia, 1959)). The wet season was defined as December through May and the dry

479   season as June through November.

480

481   **Community Diversity Analysis.** Alpha and beta diversity measures were performed using

482   *phyloseq* and *vegan*. Heatmaps were made using Bray-Curtis dissimilarity and the *pheatmaps*

483   function. Friedman-Rafsky tests were used to test for graph-based sample segregation between

484   samples factored by salinity regime (fresh, oligohaline, mesohaline, polyhaline, or euhaline) and

485   based on Jaccard distance and Bray-Curtis dissimilarity between samples. Pure edges were

486   defined for samples within the same level (fresh, oligohaline, mesohaline, etc.). Edges were

487   defined using skeleton graphs including a minimum spanning tree (MST), distance threshold

488   (max jaccard distance = 0.4), and k-nearest neighbors (k=1.) Graphs were made using *igraph*,

489   *ggnetwork* and *phyloseqGraph Test* libraries. Differential abundance was analyzed through

490   DESeq2 using VST libraries. Principal coordinates analyses (PCoAs) use Bray-Curtis dissimilarity

491   and the V4 dataset unless otherwise noted. The *clusGap* function was used to calculate the gap

492   statistic for "goodness of clusters". Analysis of variance of ordinations was calculated using the

493   *adonis* function. Procrustes tests were used to compare ordinations quantitatively using the

494   *protest* function [*vegan*], which rotates and stretches ordinations until the distance between

495   corresponding points is minimized. M^2 represents one minus the squared correlation

496   coefficient (R) between the coordinates of corresponding points between the two ordinations

497   being tested. Differential abundance was analyzed through DESeq2 using VST libraries.

498   Nestedness versus turnover components of beta diversity analyses were calculated using the

499   *betapart* function and Bray-Curtis dissimilarity.

500

501 Nestedness of samples was calculated using the NODF statistic (Nestedness metric based on

502 Overlap and Decreasing Fill), which describes combined column and row nestedness based on

503 decreasing fill and paired overlap of presence data in a matrix (Almeida-Neto *et al.*, 2008). To

504 calculate NODF, count tables were converted to a presence absence matrix and ordered by

505 decreasing row sums, then decreasing column sums (decreasing prevalence and decreasing

506 richness, respectively). The *oecosimu* function [*vegan*] was used to calculate the NODF statistic,

507 which approximates the average percent of taxa from less diverse samples that occur in more

508 diverse samples. The null model used was the "c0" model, which holds constant the column

509 sums (sample richness) but allows row sums to vary (phyla prevalence) from the original data.

510 Significance testing was based on default for *oecosimu* package and set to test if data values

511 were greater than the null model. Entropy was calculated using the *diversity* function [*vegan*]

512 and defined as the Shannon index of specific taxa (instead of by sample). Entropy, nestedness,

513 and ordinations analyses were conducted using the full dataset and also using a random

514 subsampling of samples evenly distributed across the five salinity zones (fresh, oligohaline,

515 mesohaline, polyhaline, and euhaline), yielding similar results.

516

517 **Acknowledgements**

527

528   **Conflict of Interest**

529   The authors declare no conflict of interest.

References

Aguirre, M., Abad, D., Albaina, A., Cralle, L., Goñi-Urriza, M.S., Estonba, A., and Zarraonaindia, I. (2017) Unraveling the environmental and anthropogenic drivers of bacterial community changes in the Estuary of Bilbao and its tributaries. *PLOS ONE* **12**: e0178755.

Ahlgren, N.A., Chen, Y., Needham, D.M., Parada, A.E., Sachdeva, R., Trinh, V., et al. (2017) Genome and epigenome of a novel marine Thaumarchaeota strain suggest viral infection, phosphorothioation DNA modification and multiple restriction systems. *Environ Microbiol* **19**: 2434–2452.

Almeida-Neto, M., Guimarães, P., Guimarães, P.R., Loyola, R.D., and Ulrich, W. (2008) A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* **117**: 1227–1239.

Alonso, C., Warnecke, F., Amann, R., and Pernthaler, J. (2007) High local and global diversity of Flavobacteria in marine plankton. *Environ Microbiol* **9**: 1253–1266.

Apprill, A., McNally, S., Parsons, R., and Weber, L. (2015) Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat Microb Ecol* **75**:.

Battaglia, B. (1959) Final resolution of the symposium on the classification of brackish waters. *Archo Oceanogr Limnol* **11(suppl)**: 243–8.

Beck, M.W., Jabusch, T.W., Trowbridge, P.R., and Senn, D.B. (2018) Four decades of water quality change in the upper San Francisco Estuary. *Estuar Coast Shelf Sci* **212**: 11–22.

Bennke, C.M., Krüger, K., Kappelmann, L., Huang, S., Gobet, A., Schüler, M., et al. (2016) Polysaccharide utilisation loci of Bacteroidetes from two contrasting open ocean sites in the North Atlantic. *Environ Microbiol* **18**: 4456–4470.

Blainey, P.C., Mosier, A.C., Potanina, A., Francis, C.A., and Quake, S.R. (2011) Genome of a Low-Salinity Ammonia-Oxidizing Archaeon Determined by Single-Cell and Metagenomic Analysis. *PLOS ONE* **6**: e16626.

Bouvier, T.C. and Giorgio, P.A. del (2002) Compositional changes in free-living bacterial communities along a salinity gradient in two temperate estuaries. *Limnol Oceanogr* **47**: 453–470.

Callahan, B.J., Sankaran, K., Fukuyama, J.A., McMurdie, P.J., and Holmes, S.P. (2016) Bioconductor workflow for microbiome data analysis: from raw reads to community analyses. *F1000Research* **5**: 1492.

Campbell, B.J. and Kirchman, D.L. (2013) Bacterial diversity, community structure and potential growth rates along an estuarine salinity gradient. *ISME J* **7**: 210–220.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.

Cloern, J.E. (2001) Our evolving conceptual model of the coastal eutrophication problem. *Mar Ecol Prog Ser* **210**: 223–253.

Cloern, J.E. (2019) Patterns, pace, and processes of water-quality variability in a long-studied estuary. *Limnol Oceanogr* **64**: S192–S208.

Cloern, J.E. (1996) Phytoplankton bloom dynamics in coastal ecosystems: A review with some general lessons from sustained investigation of San Francisco Bay, California. *Rev Geophys* **34**: 127–168.

Cloern, J.E. (1987) Turbidity as a control on phytoplankton biomass and productivity in estuaries. *Cont Shelf Res* **7**: 1367–1381.

Cloern, J.E. and Dufford, R. (2005) Phytoplankton community ecology: principles applied in San Francisco Bay. *Mar Ecol Prog Ser* **285**: 11–28.

Cloern, J.E. and Jassby, A.D. (2012) Drivers of change in estuarine-coastal ecosystems: Discoveries from four decades of study in San Francisco Bay. *Rev Geophys* **50**: RG4001.

Cloern, J.E. and Jassby, A.D. (2010) Patterns and Scales of Phytoplankton Variability in Estuarine–Coastal Ecosystems. *Estuaries Coasts* **33**: 230–241.

Cloern, J.E., Jassby, A.D., Schraga, T.S., Nejad, E., and Martin, C. (2017) Ecosystem variability along the estuarine salinity gradient: Examples from long-term study of San Francisco Bay. *Limnol Oceanogr* **62**: S272–S291.

Crump, B.C., Armbrust, E.V., and Baross, J.A. (1999) Phylogenetic Analysis of Particle-Attached and Free-Living Bacterial Communities in the Columbia River, Its Estuary, and the Adjacent Coastal Ocean. *Appl Environ Microbiol* **65**: 3192–3204.

Crump, B.C., Hopkinson, C.S., Sogin, M.L., and Hobbie, J.E. (2004) Microbial Biogeography along an Estuarine Salinity Gradient: Combined Influences of Bacterial Growth and Residence Time. *Appl Environ Microbiol* **70**: 1494–1505.

Damashek, J., Casciotti, K.L., and Francis, C.A. (2016) Variable Nitrification Rates Across

Environmental Gradients in Turbid, Nutrient-Rich Estuary Waters of San Francisco Bay.

Estuaries Coasts **39**: 1050–1071.

Doherty, M., Yager, P.L., Moran, M.A., Coles, V.J., Fortunato, C.S., Krusche, A.V., et al. (2017)

Bacterial Biogeography across the Amazon River-Ocean Continuum. *Front Microbiol* **8**:.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*

26: 2460–2461.

Fortunato, C.S. and Crump, B.C. (2011) Bacterioplankton Community Variation Across River to

Ocean Environmental Gradients. *Microb Ecol* **62**: 374–382.

Fortunato, C.S., Eiler, A., Herfort, L., Needoba, J.A., Peterson, T.D., and Crump, B.C. (2013)

Determining indicator taxa across spatial and seasonal gradients in the Columbia River

coastal margin. *ISME J* **7**: 1899–1911.

Fortunato, C.S., Herfort, L., Zuber, P., Baptista, A.M., and Crump, B.C. (2012) Spatial variability

overwhelms seasonal patterns in bacterioplankton communities across a river to ocean

gradient. *ISME J* **6**: 554–563.

Ghai, R., Mizuno, C.M., Picazo, A., Camacho, A., and Rodriguez-Valera, F. (2014) Key roles for

freshwater Actinobacteria revealed by deep metagenomic sequencing. *Mol Ecol* **23**:

6073–6090.

Ghai, R., Mizuno, C.M., Picazo, A., Camacho, A., and Rodriguez-Valera, F. (2013) Metagenomics

uncovers a new group of low GC and ultra-small marine Actinobacteria. *Sci Rep* **3**:

srep02471.

Glöckner, F.O., Zaichikov, E., Belkova, N., Denissova, L., Pernthaler, J., Pernthaler, A., and

Amann, R. (2000) Comparative 16S rRNA Analysis of Lake Bacterioplankton Reveals

Globally Distributed Phylogenetic Clusters Including an Abundant Group of

Actinobacteria. *Appl Env Microbiol* **66**: 5053–5065.

Gómez-Pereira, P.R., Fuchs, B.M., Alonso, C., Oliver, M.J., van Beusekom, J.E.E., and Amann, R.

(2010) Distinct flavobacterial communities in contrasting water masses of the North

Atlantic Ocean. *ISME J* **4**: 472–487.

Happel, E., Bartl, I., Voss, M., and Riemann, L. (2018) Extensive nitrification and active ammonia

oxidizers in two contrasting coastal systems of the Baltic Sea. *Environ Microbiol* **20**:

2913–2926.

Henson, M.W., Lanclos, V.C., Faircloth, B.C., and Thrash, J.C. (2018) Cultivation and genomics of

the first freshwater SAR11 (LD12) isolate. *ISME J* **12**: 1846–1860.

Herlemann, D.P., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J.J., and Andersson, A.F. (2011)

Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic

Sea. *ISME J* **5**: 1571–1579.

Herlemann, D.P.R., Woelk, J., Labrenz, M., and Jürgens, K. (2014) Diversity and abundance of

"Pelagibacterales" (SAR11) in the Baltic Sea salinity gradient. *Syst Appl Microbiol* **37**:

601–604.

Hewson, I. and Fuhrman, J.A. (2004) Richness and Diversity of Bacterioplankton Species along

an Estuarine Gradient in Moreton Bay, Australia. *Appl Env Microbiol* **70**: 3425–3433.

Hollibaugh, J., Gifford, S., Moran, M.A., Ross, M., Sharma, S., and Tolar, B. (2014) Seasonal variation in the metratranscriptomes of a Thaumarchaeota population from SE USA coastal waters. *ISME J* **8**: 685–698.

Hollibaugh, J.T. and Wong, P.S. (1999) Microbial processes in the San Francisco Bay estuarine turbidity maximum. *Estuaries* **22**: 848–862.

Hollibaugh, J.T., Wong, P.S., and Murrell, M.C. (2000) Similarity of particle-associated and free-living bacterial communities in northern San Francisco Bay, California. *Aquat Microb Ecol* **21**: 103–114.

Holmfeldt, K., Dziallas, C., Titelman, J., Pohlmann, K., Grossart, H.-P., and Riemann, L. (2009) Diversity and abundance of freshwater Actinobacteria along environmental gradients in the brackish northern Baltic Sea. *Environ Microbiol* **11**: 2042–2054.

Hu, A., Yang, Z., Yu, C.-P., and Jiao, N. (2013) Dynamics of Autotrophic Marine Planktonic Thaumarchaeota in the East China Sea. *PLoS ONE* **8**:.

Kan, J., Evans, S.E., Chen, F., and Suzuki, M.T. (2008) Novel estuarine bacterioplankton in rRNA operon libraries from the Chesapeake Bay. *Aquat Microb Ecol* **51**: 55–66.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**: 3059–3066.

Kim, J.-G., Gwak, J.-H., Jung, M.-Y., An, S.-U., Hyun, J.-H., Kang, S., and Rhee, S.-K. (2019) Distinct temporal dynamics of planktonic archaeal and bacterial assemblages in the bays of the Yellow Sea. *PLOS ONE* **14**: e0221408.

Kimmerer, W. (2004) Open Water Processes of the San Francisco Estuary: From Physical Forcing to Biological Responses. *San Franc Estuary Watershed Sci* **2**:.

Kirchman, D.L., Dittel, A.I., Malmstrom, R.R., and Cottrell, M.T. (2005) Biogeography of major

bacterial groups in the Delaware Estuary. *Limnol Oceanogr* **50**: 1697–1706.

Ladau, J. and Eloe-Fadrosh, E.A. (2019) Spatial, Temporal, and Phylogenetic Scales of Microbial

Ecology. *Trends Microbiol* **0**:

Li, J., Jiang, X., Jing, Z., Li, G., Chen, Z., Zhou, L., et al. (2017) Spatial and seasonal distributions of

bacterioplankton in the Pearl River Estuary: The combined effects of riverine inputs,

temperature, and phytoplankton. *Mar Pollut Bull* **125**: 199–207.

Liu, J., Fu, B., Yang, H., Zhao, M., He, B., and Zhang, X.-H. (2015) Phylogenetic shifts of

bacterioplankton community composition along the Pearl Estuary: the potential impact

of hypoxia and nutrients. *Front Microbiol* **6**:.

Liu, J., Yu, S., Zhao, M., He, B., and Zhang, X.-H. (2014) Shifts in archaeaplankton community

structure along ecological gradients of Pearl Estuary. *FEMS Microbiol Ecol* **90**: 424–435.

Lozupone, C.A. and Knight, R. (2007) Global patterns in bacterial diversity. *Proc Natl Acad Sci*

104: 11436–11440.

Lucas, L.V., Cloern, J.E., Koseff, J.R., Monismith, S.G., and Thompson, J.K. (1998) Does the

Sverdrup critical depth model explain bloom dynamics in estuaries? *J Mar Res* **56**: 375–

415.

Lucas, L.V., Koseff, J.R., Monismith, S.G., and Thompson, J.K. (2009) Shallow water processes

govern system-wide phytoplankton bloom dynamics: A modeling study. *J Mar Syst* **75**:

70–86.

Martin-Platero, A.M., Cleary, B., Kauffman, K., Preheim, S.P., McGillicuddy, D.J., Alm, E.J., and

Polz, M.F. (2018) High resolution time series reveals cohesive but short-lived

communities in coastal plankton. *Nat Commun* **9**: 266.

Mason, O.U., Canter, E.J., Gillies, L.E., Paisie, T.K., and Roberts, B.J. (2016) Mississippi River

Plume Enriches Microbial Diversity in the Northern Gulf of Mexico. *Front Microbiol* **7**:.

McMurdie, P.J. and Holmes, S. (2013) phyloseq: An R Package for Reproducible Interactive

Analysis and Graphics of Microbiome Census Data. *PLOS ONE* **8**: e61217.

McMurdie, P.J. and Holmes, S. (2014) Waste Not, Want Not: Why Rarefying Microbiome Data Is

Inadmissible. *PLOS Comput Biol* **10**: e1003531.

Mehrshad, M., Amoozegar, M.A., Ghai, R., Fazeli, S.A.S., and Rodriguez-Valera, F. (2016)

Genome Reconstruction from Metagenomic Data Sets Reveals Novel Microbes in the

Brackish Waters of the Caspian Sea. *Appl Environ Microbiol* **82**: 1599–1612.

Mizuno, C.M., Rodriguez-Valera, F., and Ghai, R. (2015) Genomes of Planktonic

Acidimicrobiales: Widening Horizons for Marine Actinobacteria by Metagenomics. *mBio*

6: e02083-14.

Mosier, A.C., Lund, M.B., and Francis, C.A. (2012) Ecophysiology of an ammonia-oxidizing

archaeon adapted to low-salinity habitats. *Microb Ecol* **64**: 955–963.

Murray, A.E., Hollibaugh, J.T., and Orrego, C. (1996) Phylogenetic compositions of

bacterioplankton from two California estuaries compared by denaturing gradient gel

electrophoresis of 16S rDNA fragments. *Appl Environ Microbiol* **62**: 2676–2680.

Murrell, M.C., Hollibaugh, J.T., Silver, M.W., and Wong, P.S. (1999) Bacterioplankton dynamics

in northern San Francisco Bay: Role of particle association and seasonal freshwater flow.

Limnol Oceanogr **44**: 295–308.

Ngugi, D.K. and Stingl, U. (2018) High-Quality Draft Single-Cell Genome Sequence of the NS5

Marine Group from the Coastal Red Sea. *Genome Announc* **6**: e00565-18.

Nichols, F.H., Cloern, J.E., Luoma, S.N., and Peterson, D.H. (1986) The Modification of an

Estuary. *Science* **231**: 567–573.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2018) vegan:

Community Ecology Package.

Parada, A.E., Needham, D.M., and Fuhrman, J.A. (2016) Every base matters: assessing small

subunit rRNA primers for marine microbiomes with mock communities, time series and

global field samples. *Environ Microbiol* **18**: 1403–1414.

Pitcher, A., Wuchter, C., Siedenberg, K., Schouten, S., and Sinninghe Damsté, J.S. (2011)

Crenarchaeol tracks winter blooms of ammonia-oxidizing Thaumarchaeota in the coastal

North Sea. *Limnol Oceanogr* **56**: 2308–2318.

Raimonet, M. and Cloern, J.E. (2017) Estuary–ocean connectivity: fast physics, slow biology.

Glob Change Biol **23**: 2345–2357.

Santoro, A.E., Dupont, C.L., Richter, R.A., Craig, M.T., Carini, P., McIlvin, M.R., et al. (2015)

Genomic and proteomic characterization of "Candidatus Nitrosopelagicus brevis": An

ammonia-oxidizing archaeon from the open ocean. *Proc Natl Acad Sci* **112**: 1173–1178.

Satinsky, B.M., Crump, B.C., Smith, C.B., Sharma, S., Zielinski, B.L., Doherty, M., et al. (2014) Microspatial gene expression patterns in the Amazon River Plume. *Proc Natl Acad Sci* **111**: 11085–11090.

Schaefer, S.C. and Hollibaugh, J.T. (2017) Temperature Decouples Ammonium and Nitrite Oxidation in Coastal Waters. *Environ Sci Technol*.

Schraga, T.S. and Cloern, J.E. (2017) Water quality measurements in San Francisco Bay by the U.S. Geological Survey, 1969–2015. *Sci Data* **4**: 170098.

Seo, J.-H., Kang, I., Yang, S.-J., and Cho, J.-C. (2017) Characterization of spatial distribution of the bacterial community in the South Sea of Korea. *PLoS ONE* **12**:.

Smith, M.W., Herfort, L., Fortunato, C.S., Crump, B.C., and Simon, H.M. (2017) Microbial players and processes involved in phytoplankton bloom utilization in the water column of a fast-flowing, river-dominated estuary. *MicrobiologyOpen* **6**: e00467.

Stepanauskas, R., Moran, M.A., Bergamaschi, B.A., and Hollibaugh, J.T. (2003) Covariance of bacterioplankton composition and environmental variables in a temperate delta system. *Aquat Microb Ecol* **31**: 14.

Sutula, M., Kudela, R., Hagy, J.D., Harding, L.W., Senn, D., Cloern, J.E., et al. (2017) Novel analyses of long-term data provide a scientific basis for chlorophyll-a thresholds in San Francisco Bay. *Estuar Coast Shelf Sci* **197**: 107–118.

Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Locey, K.J., et al. (2017) A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**: 457–463.

Vergin, K.L., Beszteri, B., Monier, A., Cameron Thrash, J., Temperton, B., Treusch, A.H., et al. (2013) High-resolution SAR11 ecotype dynamics at the Bermuda Atlantic Time-series Study site by phylogenetic placement of pyrosequences. *ISME J* **7**: 1322–1332.

Walters, R.A., Cheng, R.T., and Conomos, T.J. (1985) Time scales of circulation and mixing processes of San Francisco Bay waters. *Hydrobiologia* **129**: 24.

Walters, W., Hyde, E.R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., et al. (2016) Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *mSystems* **1**: e00009-15.

Wankel, S.D., Kendall, C., Francis, C.A., and Paytan, A. (2006) Nitrogen sources and cycling in the San Francisco Bay Estuary: A nitrate dual isotopic composition approach. *Limnol Oceanogr* **51**: 1654–1664.

Williams, T.J., Wilkins, D., Long, E., Evans, F., DeMaere, M.Z., Raftery, M.J., and Cavicchioli, R. (2013) The role of planktonic Flavobacteria in processing algal organic matter in coastal East Antarctica revealed using metagenomics and metaproteomics. *Environ Microbiol* **15**: 1302–1317.

Wuchter, C., Abbas, B., Coolen, M.J.L., Herfort, L., van Bleijswijk, J., Timmers, P., et al. (2006) Archaeal nitrification in the ocean. *Proc Natl Acad Sci U S A* **103**: 12317–12322.

Yang, C., Li, Y., Zhou, B., Zhou, Y., Zheng, W., Tian, Y., et al. (2015) Illumina sequencing-based analysis of free-living bacterial community dynamics during an *Akashiwo sanguine* bloom in Xiamen sea, China. *Sci Rep* **5**: 8476.

Zhang, Y., Jiao, N.Z., 焦念志, Cottrell, M.T., and Kirchman, D.L. (2006) Contribution of major

bacterial groups to bacterial biomass production along a salinity gradient in the South

China Sea.

Zou, D., Li, Y., Kao, S.-J., Liu, H., and Li, M. (2019) Genomic adaptation to eutrophication of

ammonia-oxidizing archaea in the Pearl River estuary. *Environ Microbiol* **21**: 2320–2332.

Figures Captions

Fig. 1 Map of San Francisco Bay and USGS water quality stations sampled in this study (A) and tile plots of environmental variables, including salinity (B), temperature in °C (C), and suspended particulate matter (SPM) in mg/L (D). For (B-D) the x-axis corresponds to date of 20 approximately monthly cruises from April 2012 to March 2014 and y-axis to USGS station sampled going from Lower South Bay (36) to the Sacramento River (657).

Fig. 2 Beta diversity of the V4 dataset shown through PCoA plots based on Bray-Curtis dissimilarity for the V4 dataset. Whole bay ordinations include all 174 samples (A, B). North Bay ordinations only include samples from Stations 657 to 18 (C, D). And South Bay ordinations include samples from stations 24 through 36 (E, F). Point color is based on the environmental variable with the highest correlation (*r* value) with axis 1 (A, C, E) or axis 2 (B, D, F), respectively.

Fig. 3 Richness (# of OTUs) of microbial communities shown for all samples (A) with the x-axis corresponds to date of 20 approximately monthly cruises from April 2012 to March 2014 and the y-axis corresponding to station number going from Lower South Bay (36) to the Sacramento River (657). (B) shows the correlation of richness and SPM ($r^2 = 0.55$).

Fig. 4 Diversity metrics at different taxonomic levels. (A) The NODF statistic was used to measure nestedness of each sample and compared to a null model with constant column sums (sample richness) but changing rows (taxa prevalence). * indicate a value is greater than null NODF with a $p < 0.05$. (B) Entropy calculated as the Shannon index of taxa. Violin plots are scaled to have the same width and horizontal lines indicate the 50th quartile based on point distribution. (C) DESeq2 was used to calculate log2-fold change in taxa abundance across salinity zones. Only taxa with significant variation ($p < .001$) across salinity zones are included in plots.

Fig. 5 The % relative abundance of SAR11 families (A) and Cyanobacteria genera (B) at 6 stations representative for the five salinity zones and South Bay. High Thaumarchaeota abundance (C) corresponds with high nitrite concentrations in South Bay (D) ($r^2 = 0.79$). The y-axis shows the relative abundance of taxa in the whole community (A-C). The x-axis of each panel corresponds to date of 20 approximately monthly cruises from April 2012 to March 2014 and station is indicated at the top of the panel.

Figure 1
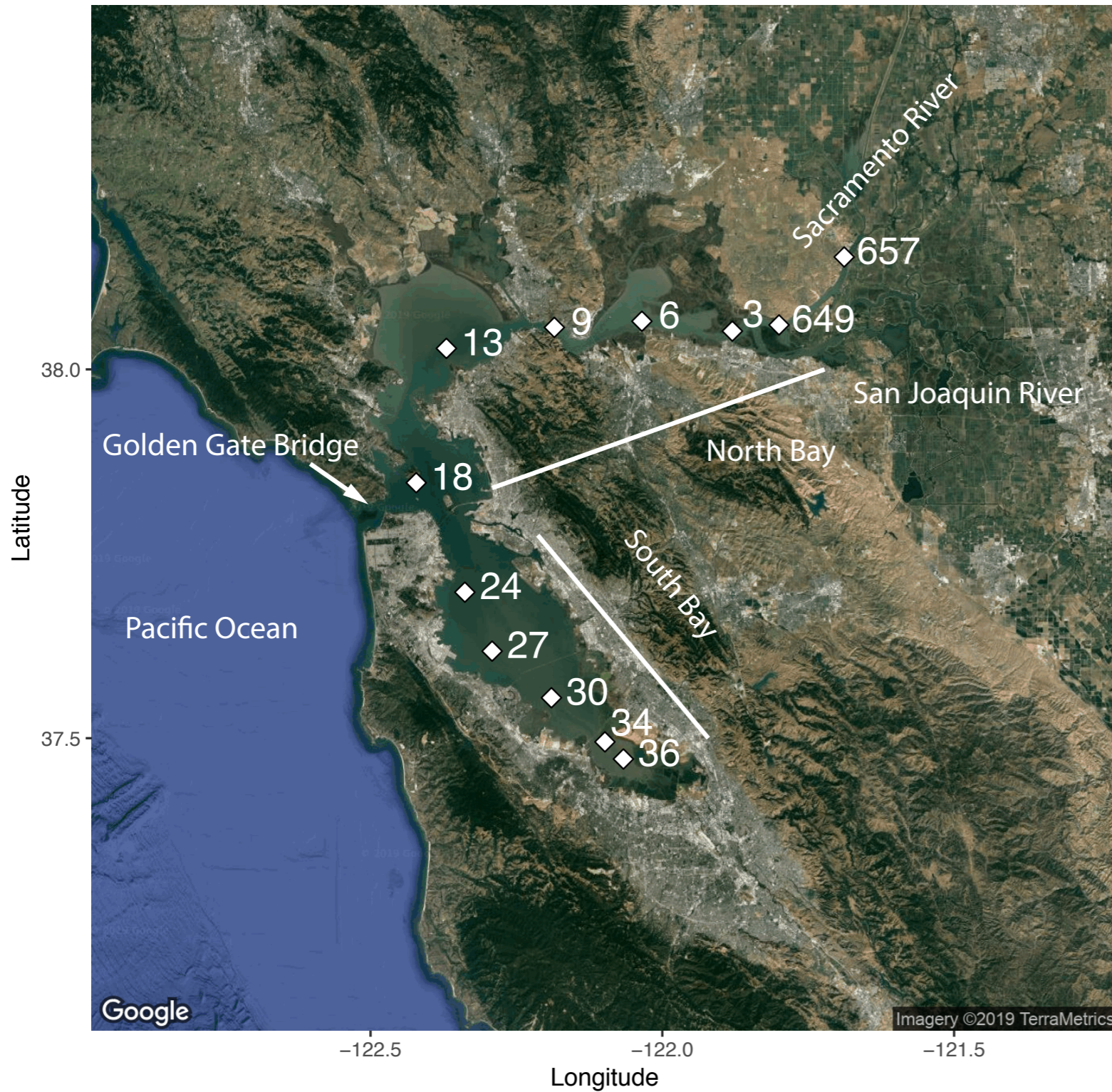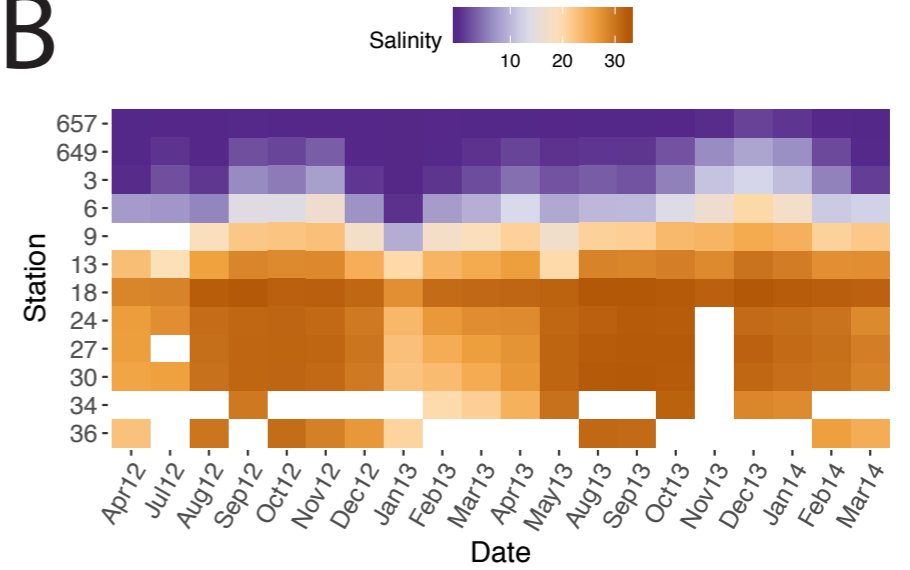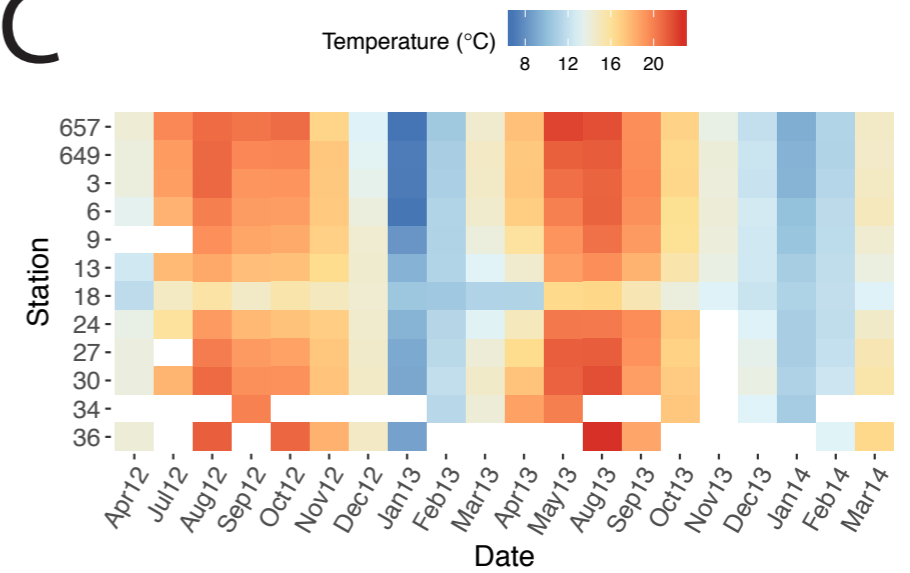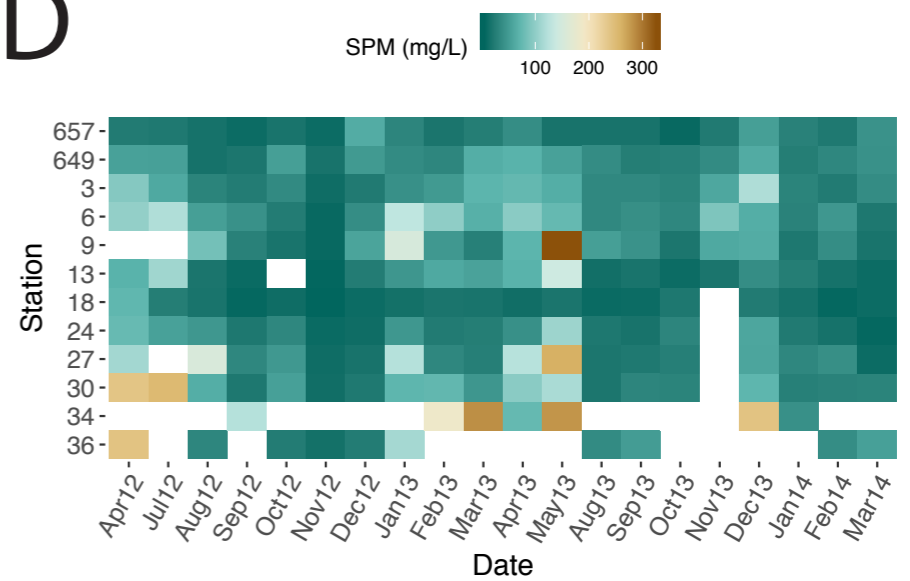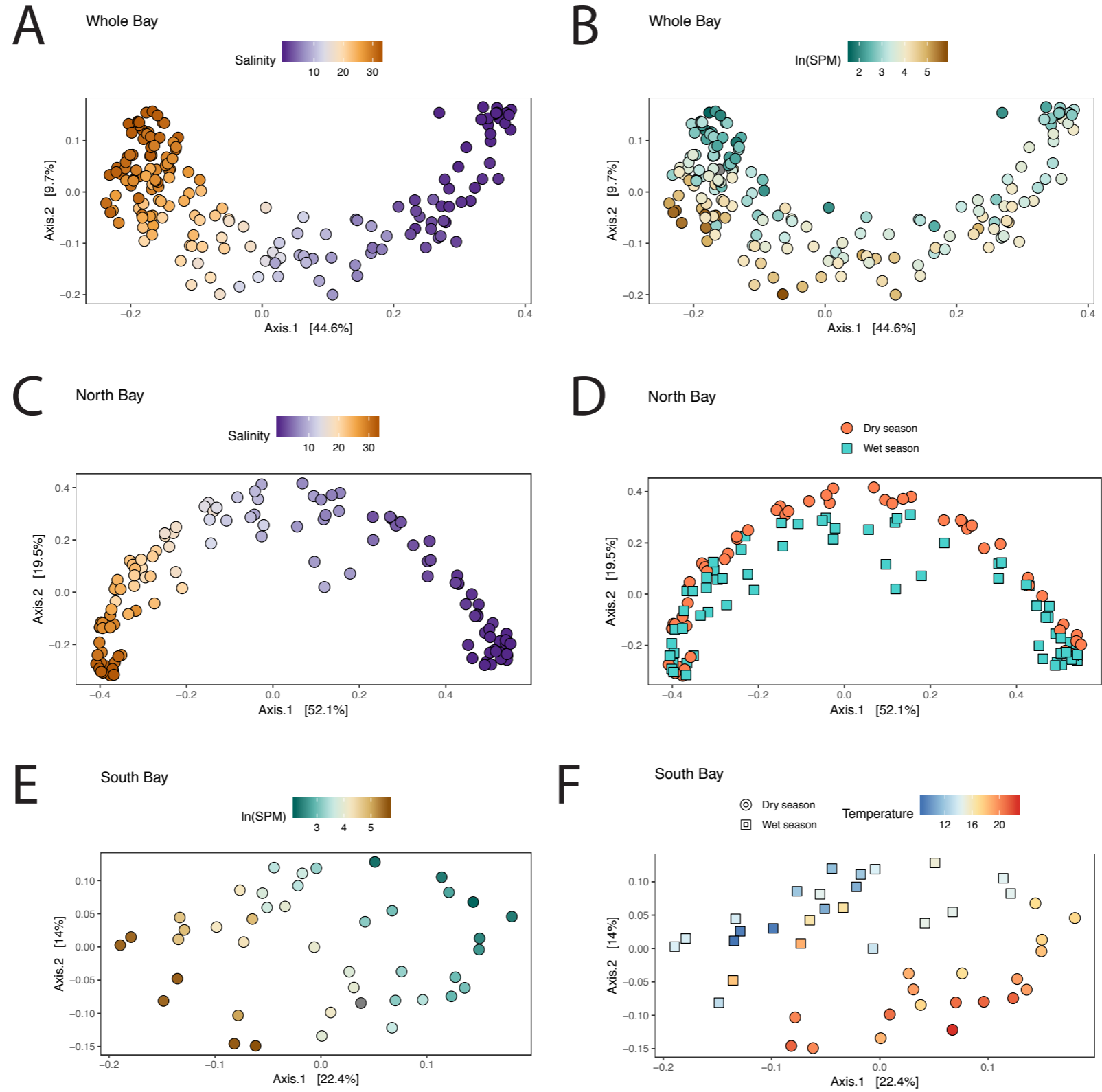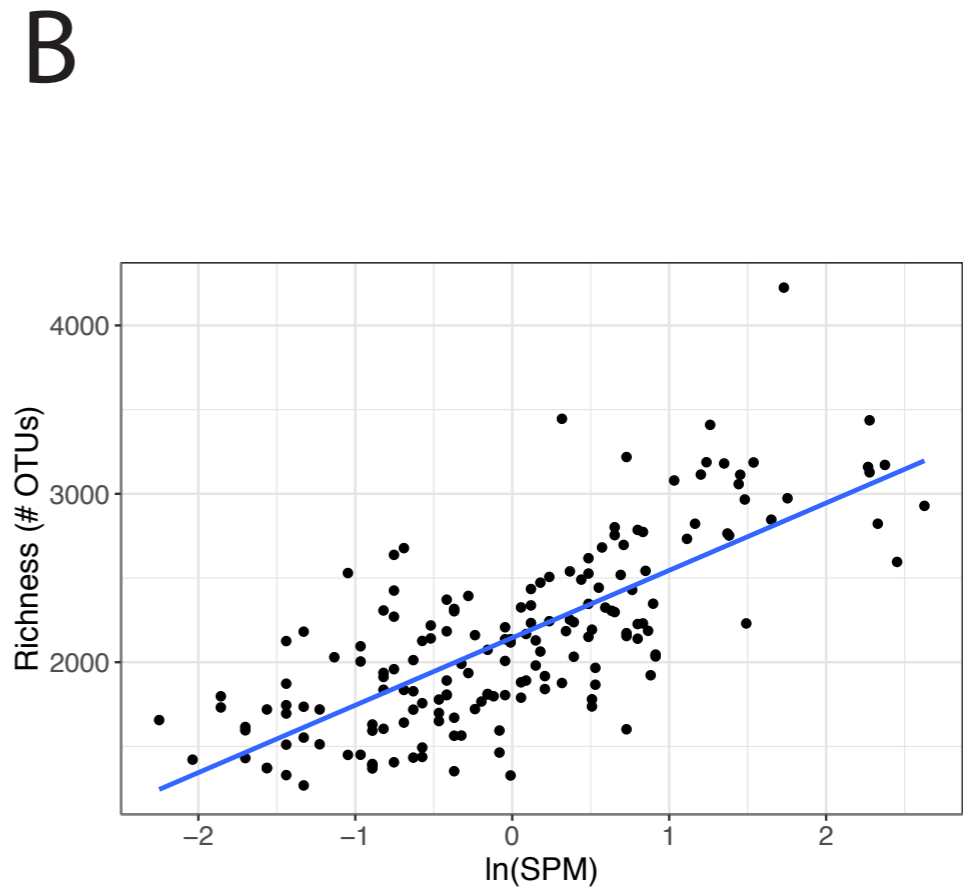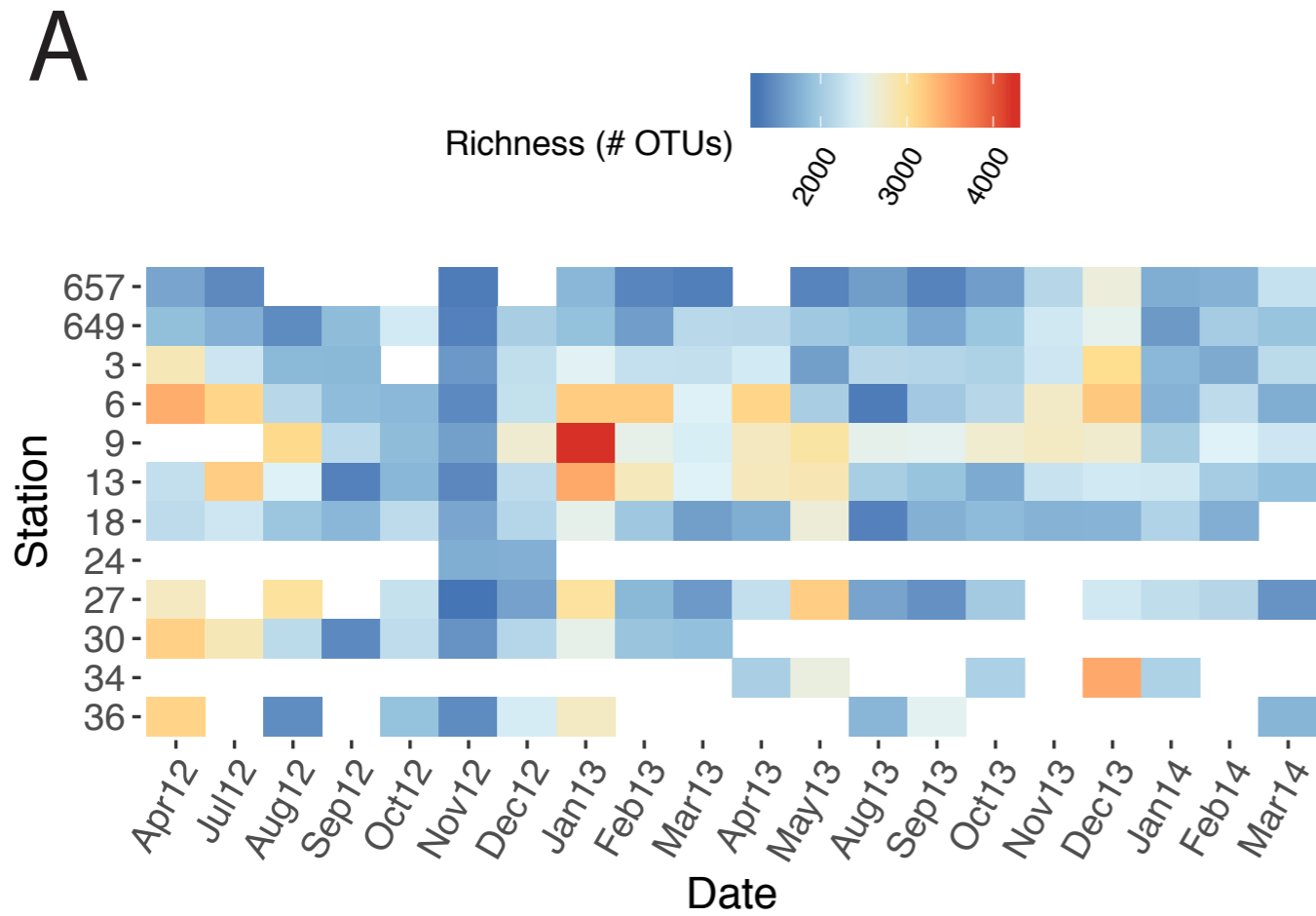
Figure 2

Figure 3

Figure 4
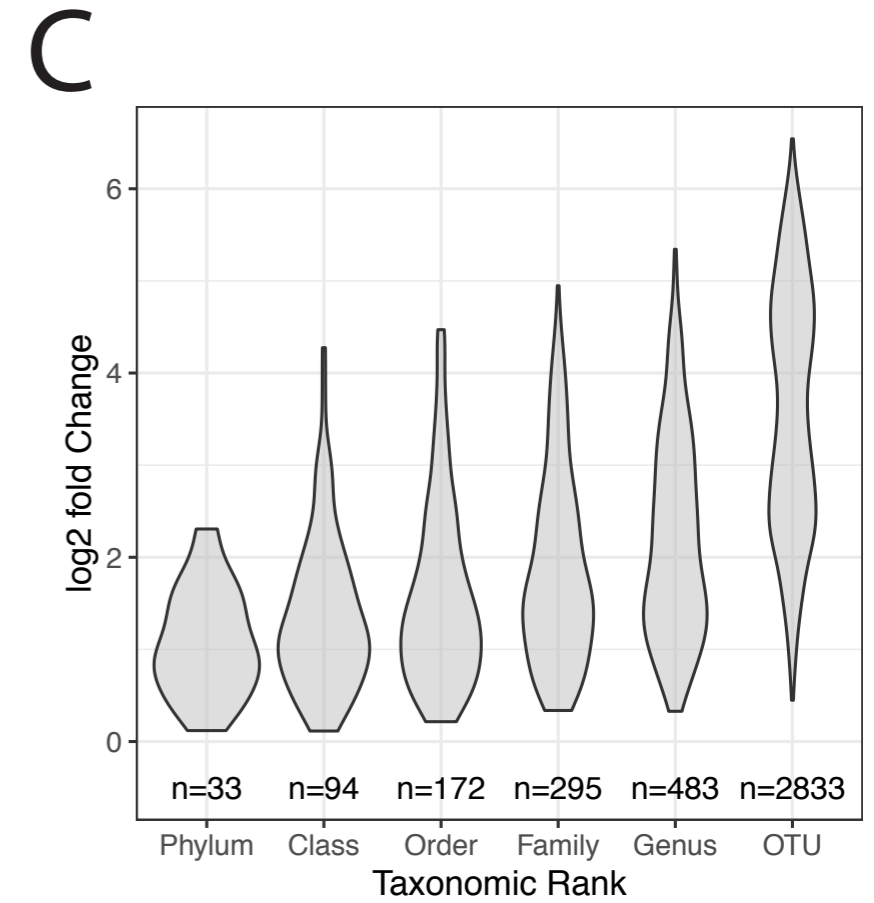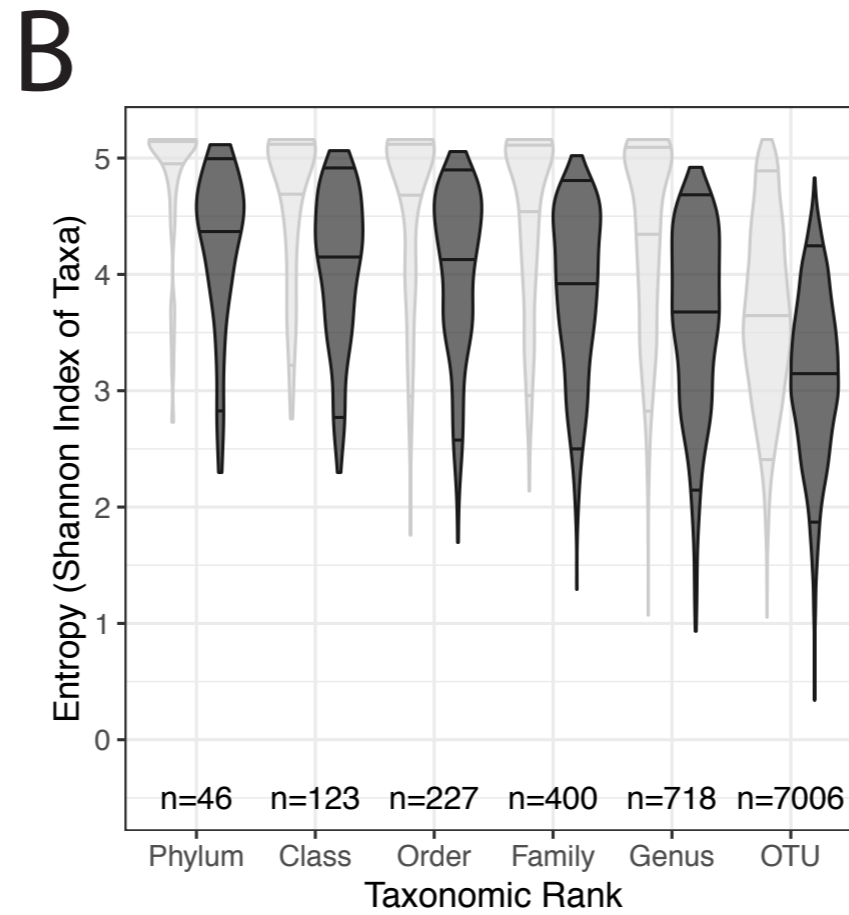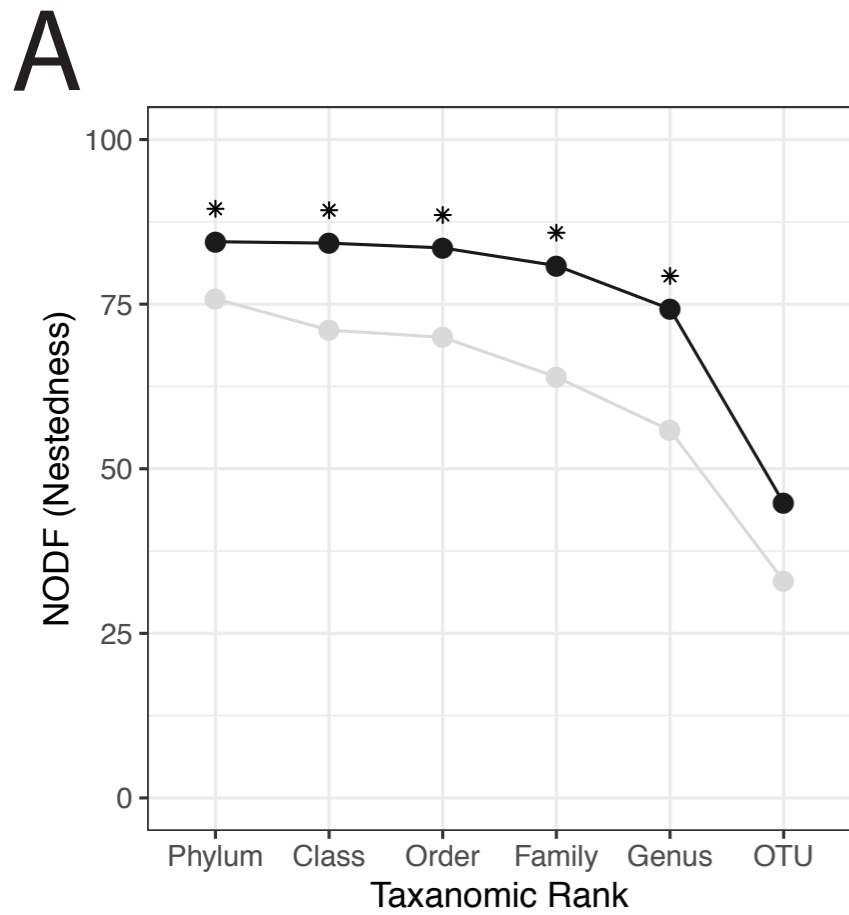
Figure 5
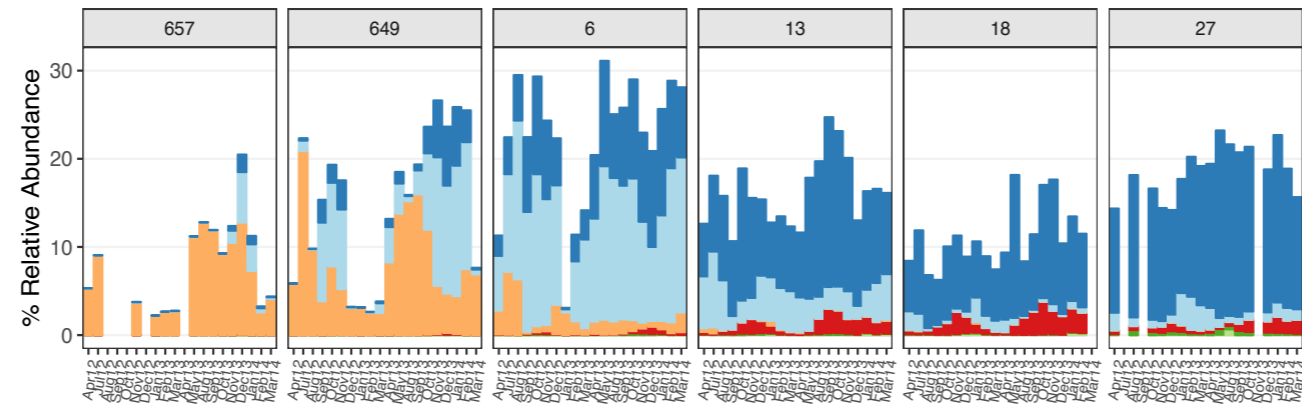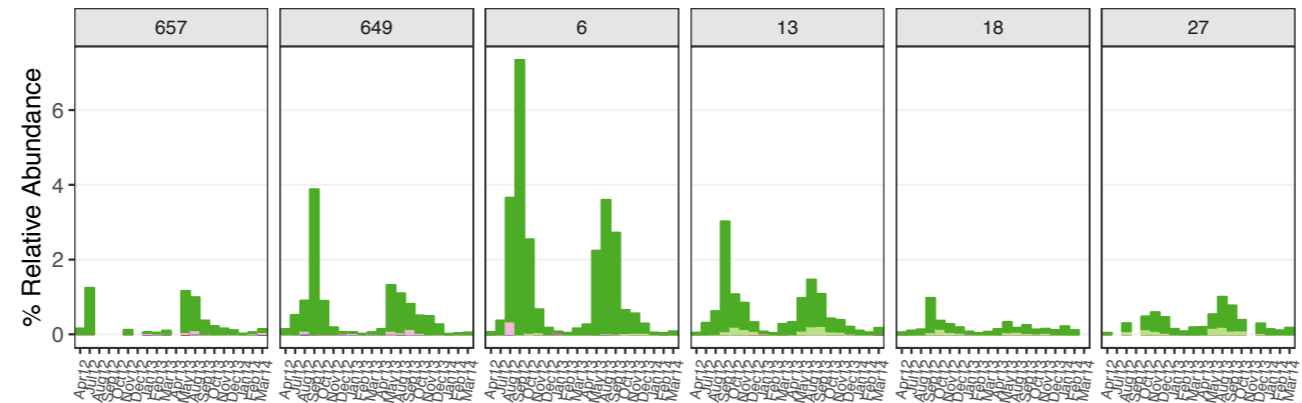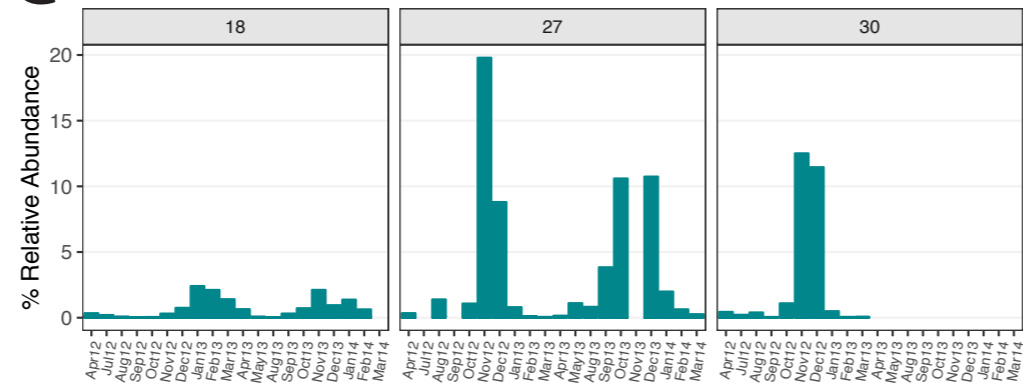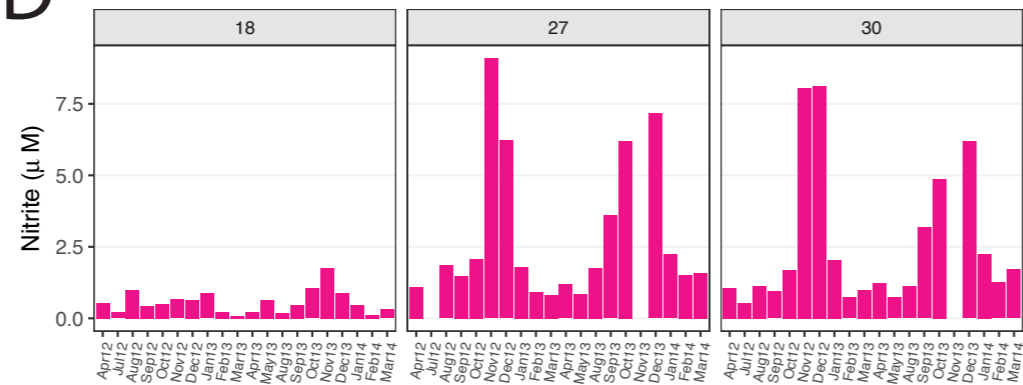
Supplemental Information

Diversity metrics of pelagic archaeal and bacterial communities in San Francisco Bay

depend on scale

Anna Rasmussen[1], Julian Damashek[1†], Emiley Eloe-Fadrosh[2], and Christopher A.

Francis[1]*

[1]Department of Earth System Science, Stanford University, Stanford, CA 94305, USA

[2]Department of Energy Joint Genome Institute, Walnut Creek, California 94598, USA

*Correspondence: Christopher Francis, Department of Earth System Science, 473 Via

Ortega, Y2E2 Bldg Rm 140, Stanford University, Stanford, CA 94305, USA. Email:

caf@stanford.edu

Running Title: Pelagic bacteria and archaea of San Francisco Bay

List of Supporting Information

Figure S1 Additional environmental data including: (A) chlorophyll a, (B) ammonium, (C) nitrate, and (D) nitrite
Figure S2 Differences in environmental data in wet versus dry season
Figure S3 Comparison of diversity metrics between V4 and V4-V5 primer sets
Figure S4 Comparison of Thaumarchaeota (A,C) distribution and (B,D) beta-diversity in V4 and V4-V5 datasets
Figure S5 Comparison of SAR11 (A,C) distribution and (B,D) beta-diversity in V4 and V4-V5 datasets
Figure S6 Comparison of Archaea (A,C) distribution and (B,D) beta-diversity in V4 and V4-V5 datasets
Figure S7 (A) Adonis test between salinity zone cluster centroids and (B) "goodness of cluster" statistics for V4 PCoA
Figure S8 (A) Heatmap and (B) hierarchical clustering of communities
Figure S9 Graph based testing of salinity zone clusters using (A) a minimum spanning tree, (B) distance thresholding graph, and (C) nearest-neighbor based graph
Figure S10 (A-E) PCoA ordinations for the 5 different salinity zones and (F) richness
Figure S11 Barcharts showing the relative abundance of the most abundant phyla through genera in V4 dataset
Figure S12 Relative abundance of the 30 most abundant OTUs
Figure S13 Proteobacteria (A) relative abundance of the most abundant orders, (B) beta diversity, and (C) entropy of taxa at different taxonomic levels
Figure S14 Bacteroidetes (A) relative abundance of the most abundant orders, (B) beta diversity, and (C) entropy of taxa at different taxonomic levels
Figure S15 Actinobacteria (A) relative abundance of the most abundant orders, (B) beta diversity, and (C) entropy of taxa at different taxonomic levels
Figure S16 Relative abundance of 24 most abundant OTUs in poly-/euhaline sites 13,18, and 27.
Figure S17 DESeq2 differential abundance testing of (A) phyla and (B) OTUs.
Figure S18 Cyanobacteria (A) relative abundance of the most abundant orders, (B) beta diversity, and (C) entropy of taxa at different taxonomic levels
Table S1 Taxonomic and read count data
Table S2 Correlations between environmental variables and diversity metrics for V4 and V4-V5 datasets
Table S3 Nestedness and turnover components of PCoA ordinations
Table S4 Correlations between environmental variables and diversity metrics for samples in North or South bay or various salinity zones
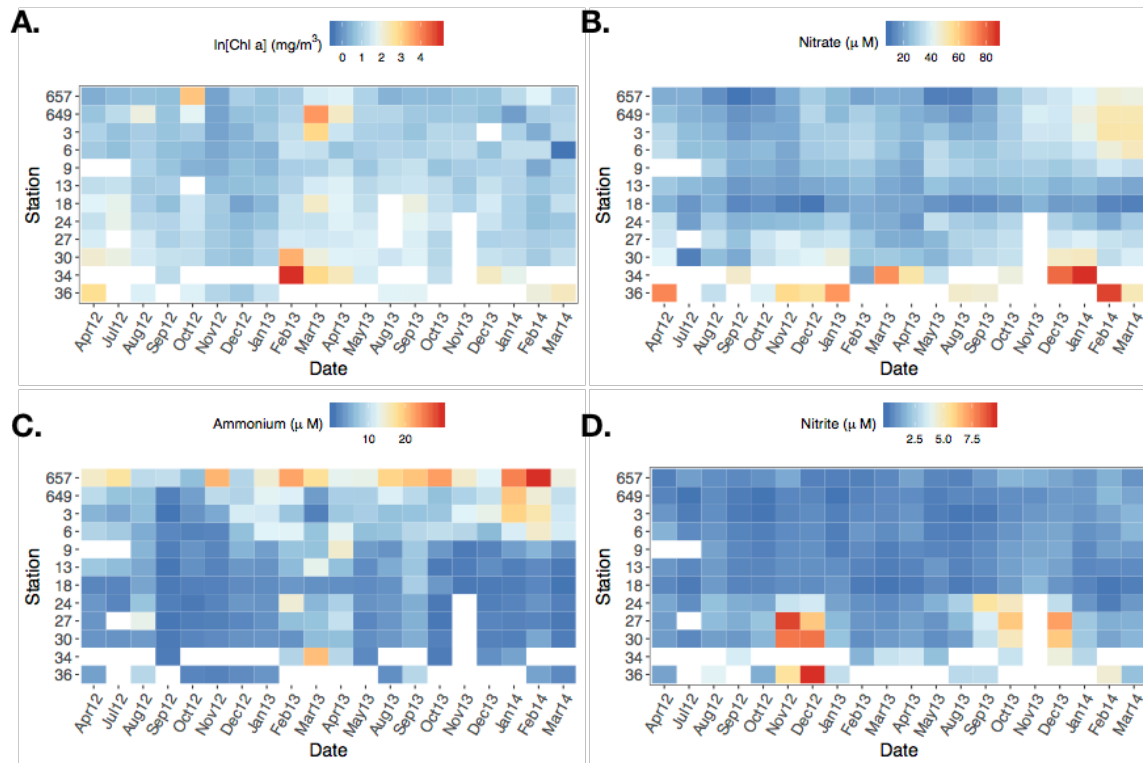
Fig. S1 Environmental data for the natural log of chlorophyll a (A), nitrate (B), ammonium (C), and nitrite (D). The x-axis corresponds to date of 20 approximately monthly cruises from April 2012 to March 2014 and the y-axis to USGS station sampled going from Lower South Bay (36) to the Sacramento River (657).
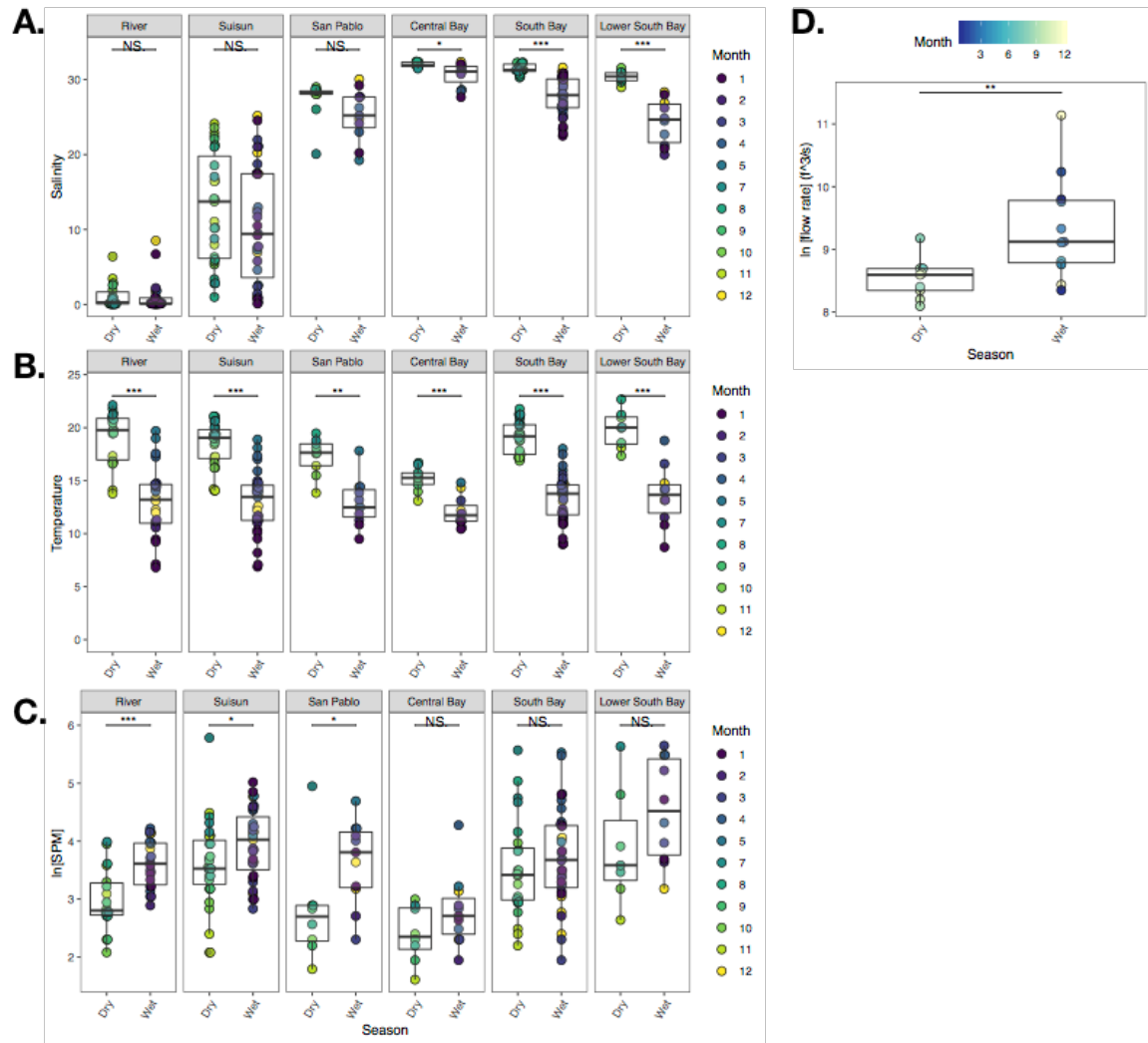
Fig. S2 Environmental variables shown in different regions of San Francisco Bay for wet and dry seasons. * indicate p-value < .05, ** p-value < .01, *** p-value <.001 and NS is not significant.

Fig. S3 Comparison of taxonomic and beta diversity between 16S rRNA V4 and V4-V5 primer sets. (A) Comparison between primer sets of relative abundances of phyla. Empty circles indicate phyla found in only one dataset. (B) Of the top 100 genera, the relative abundance of shared genera is shown (excludes "unknown" genera). Point color indicates class. (C) Comparison of richness within each phylum for both primer sets. (D) A procrustes plot of the first 2 axes of PCoA ordinations for both primer sets based on Bray-Curtis dissimilarity. Points indicate sample coordinates for V4 ordination and blue arrows indicate minimized distance to the corresponding V4-V5 ordination coordinates. In A-C the 1:1 line is indicated by the dashed black line.
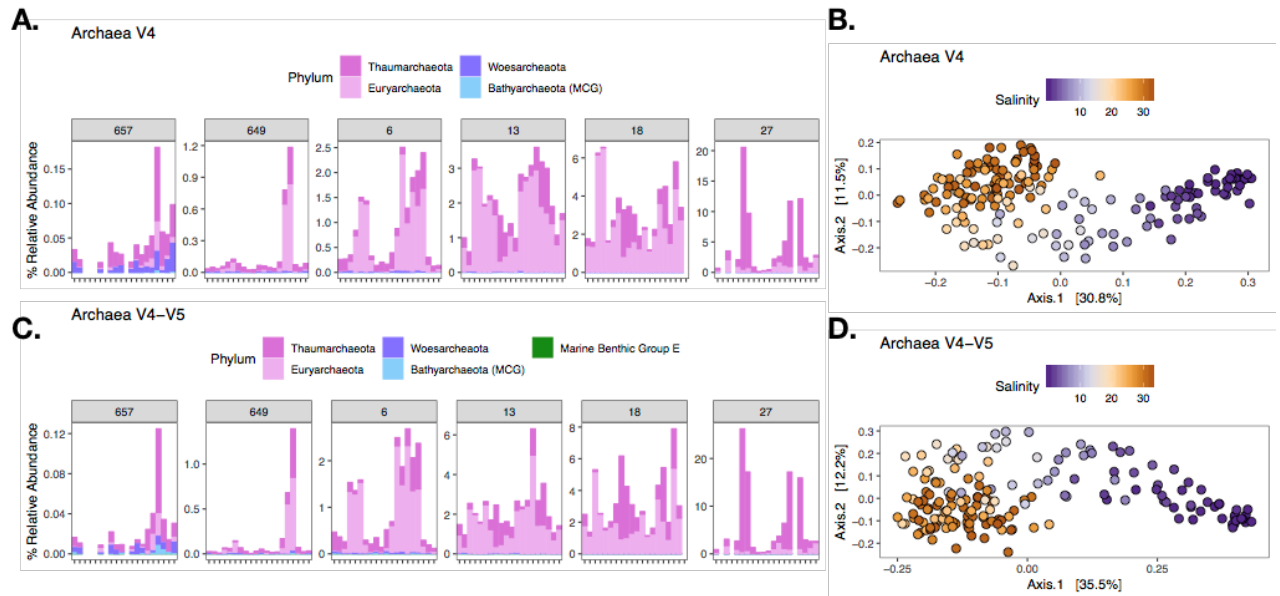
Fig. S4 V4 relative abundance of Thaumarchaeota (A) and beta diversity shown by PCoA based on Bray-Curtis dissimilarity (B) and V4-V5 relative abundance of Thaumarchaeota (C) and beta diversity shown by PCoA based on Bray-Curtis dissimilarity (D). For (A) and (C), panels correspond to 6 representative stations for the five salinity zones and South Bay. The x-axis of each panel corresponds to date of 20 approximately monthly cruises from April 2012 to March 2014. The y-axis is the % relative abundance. Note difference in scale on y-axis between each panel. In (B) and (D), point color corresponds to environmental variable with strongest correlation to axis 1.
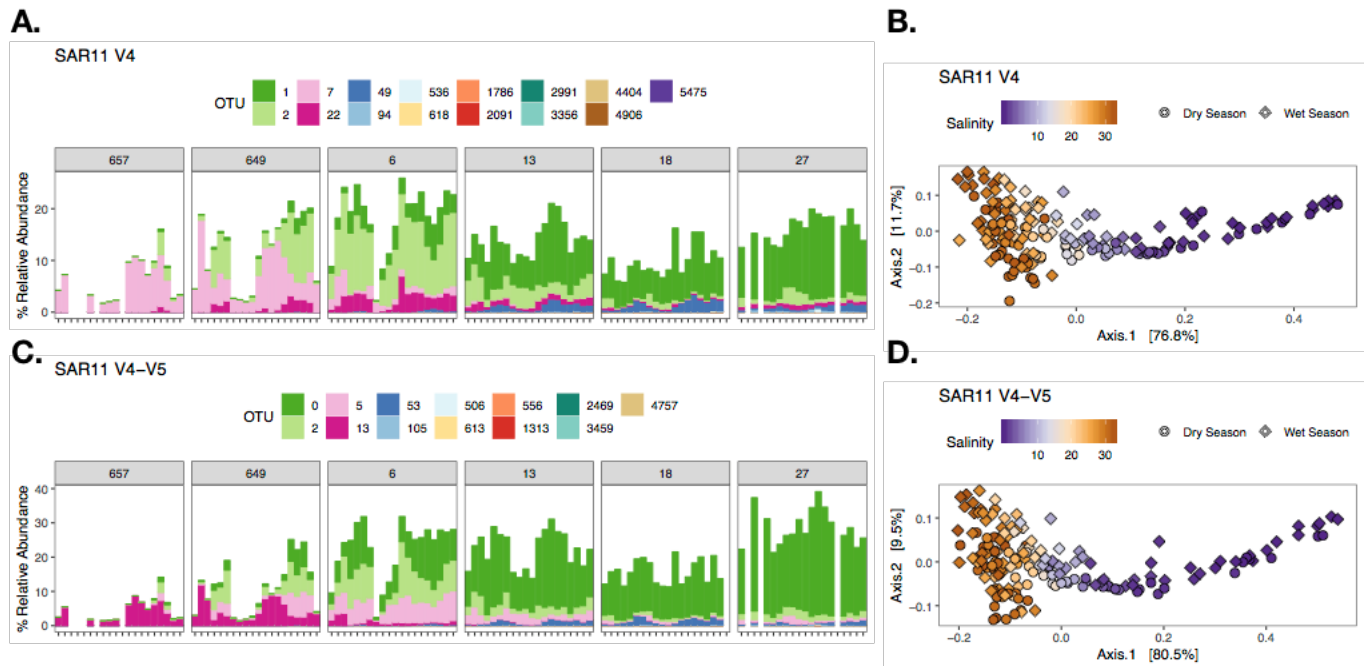
Fig. S5 V4 relative abundance of SAR11 (A) and beta diversity shown by PCoA based on Bray-Curtis dissimilarity (B) and V4-V5 relative abundance (C) and beta diversity shown by PCoA based on Bray-Curtis dissimilarity (D). For (A) and (C), panels correspond to 6 representative stations for the five salinity zones and South Bay. The x-axis of each panel corresponds to date of 20 approximately monthly cruises from April 2012 to March 2014. The y-axis is the % relative abundance. In (B) and (D), point color corresponds to environmental variable with strongest correlation to axis 1 and point shape highlights separation of wet versus dry season samples.

Fig. S6 V4 relative abundance of Archaea (A) and beta diversity shown by PCoA based on Bray-Curtis dissimilarity (B) and V4-V5 relative abundance (C) and beta diversity shown by PCoA based on Bray-Curtis dissimilarity (D). For (A) and (C), panels correspond to 6 representative stations for the five salinity zones and South Bay. The x-axis of each panel corresponds to date of 20 approximately monthly cruises from April 2012 to March 2014. The y-axis is the % relative abundance. In (B) and (D), point color corresponds to environmental variable with strongest correlation to axis 1 and point shape highlights separation of wet versus dry season samples.
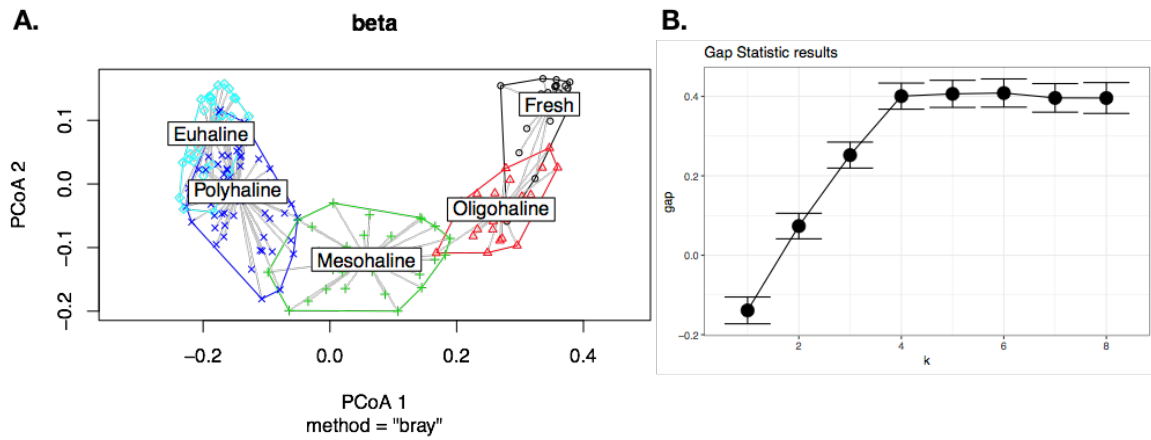
Fig. S7 (A) PCoA ordination with Adonis test of difference between centroids for each salinity zone and (B) gap statistic for goodness of clusters for V4 PCoA ordination based on Bray-Curtis dissimilarity.
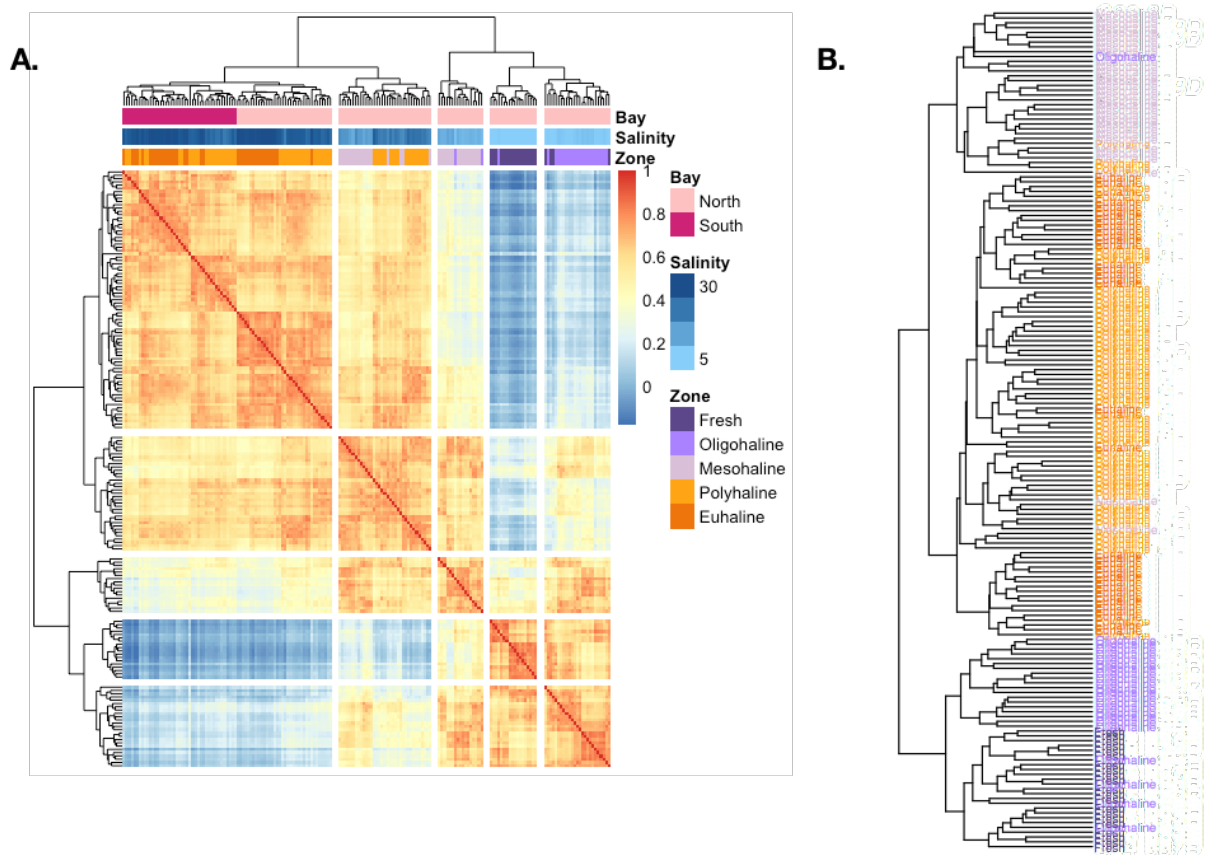
Fig. S8 (A) Heatmap based on Bray Curtis dissimilarity between samples and (B) hierarchical clustering based on Jaccard distance. In (B) tip labels are colored by salinity zone according to the legend in (A).

Fig. S9 Graph based on (A) a minimum spanning tree, which connects all samples and minimizes edge lengths based on Bray-Curtis dissimilarity. (B) Graph based on distance-based threshold, making connections between samples with a Bray-Curtis dissimilarity less than 0.25. Solid lines indicate pure edges (samples are in the same salinity zone) and dashed lines indicate mixed edges (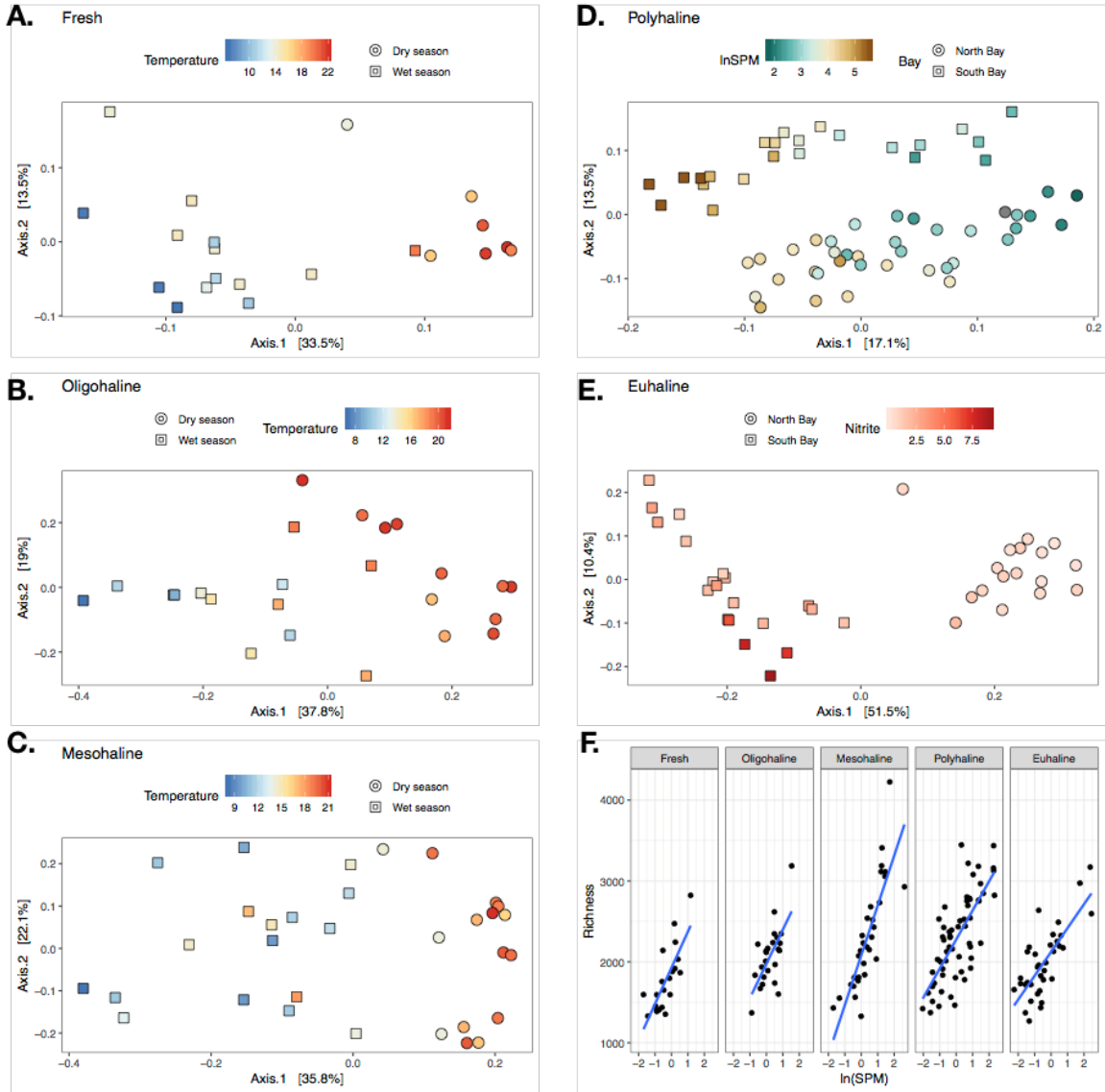connection is between samples of neighboring salinity zone). (C) Graph based on a knn =1, makes connections between nearest neighbor. Solid lines indicate pure edges (samples are in the same salinity zone) and dashed lines indicate mixed edges (connection is between samples of neighboring salinity zone).

Fig. S10 (A-E) PCoA ordinations based on Bray-Curtis dissimilarity for each salinity zone. Points are colored by strongest environmental correlate with axis 1. Point shape in corresponds to season in (A-C) and bay in (D-E). (F) Shows correlation between SPM concentration and richness within each salinity zone.
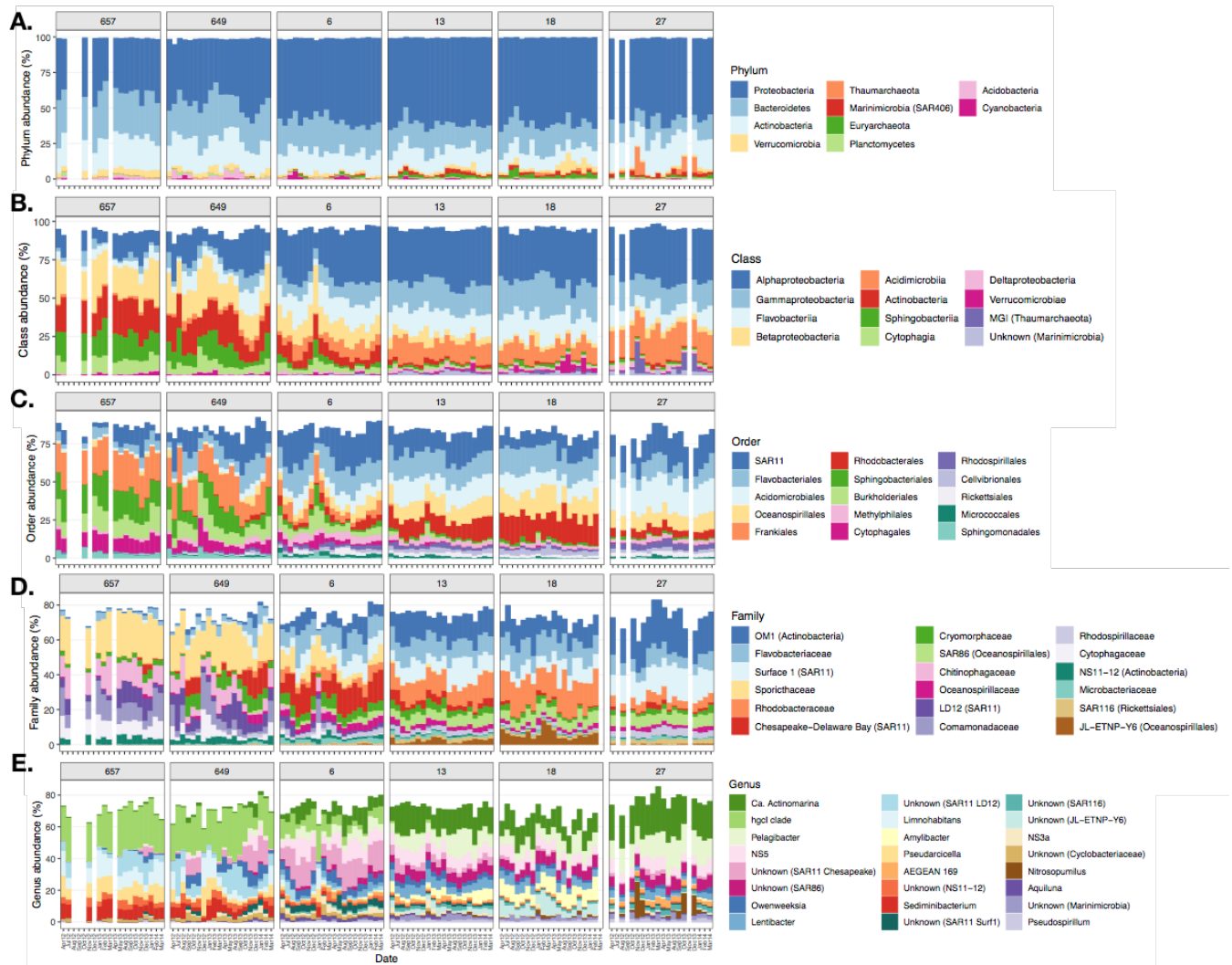
Fig. S11 Relative abundance of the most abundant 10 phyla (A), 12 classes (B), 15 orders (C), 18 families (D), and top 24 genera (E). Panels correspond to 6 representative stations for the five salinity zones and South Bay. The x-axis of each panel corresponds to date of 20 approximately monthly cruises from April 2012 to March 2014. The y-axis is the % relative abundance.
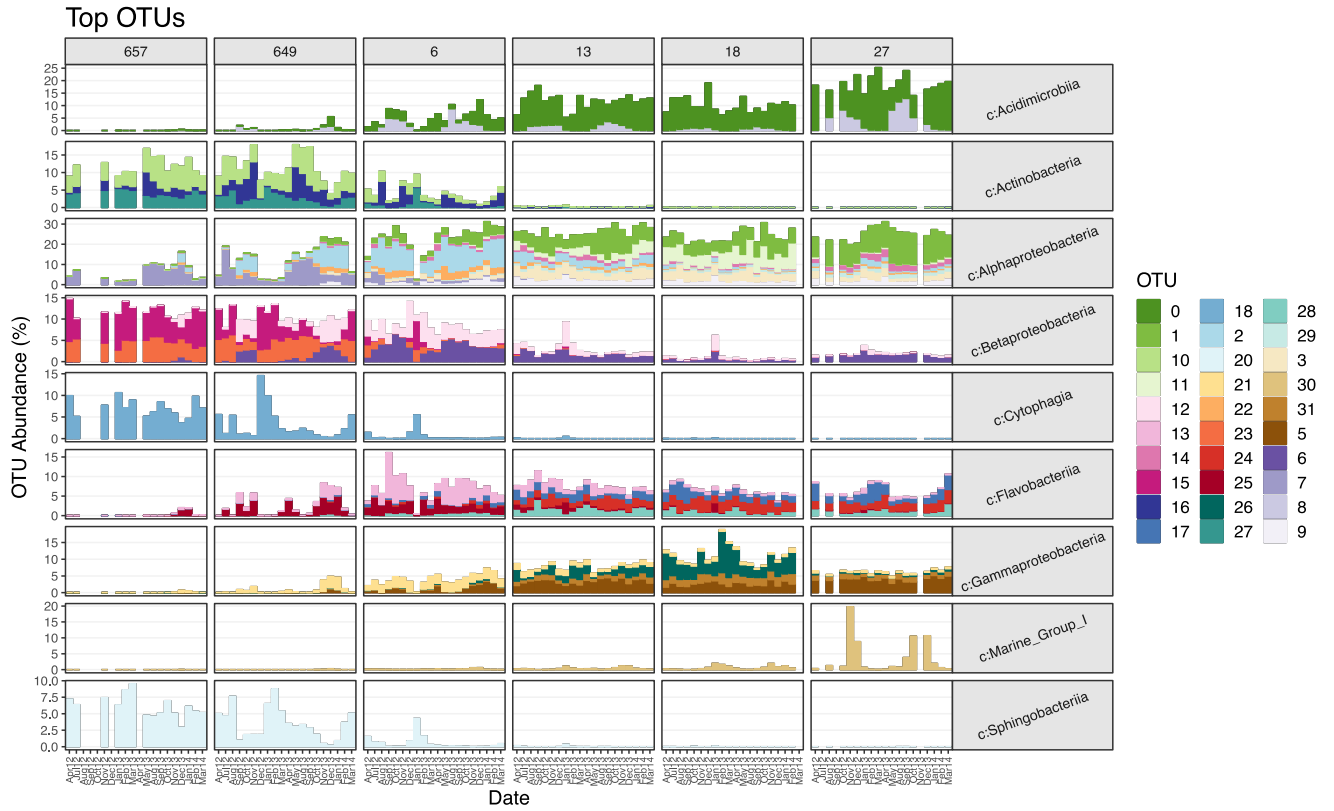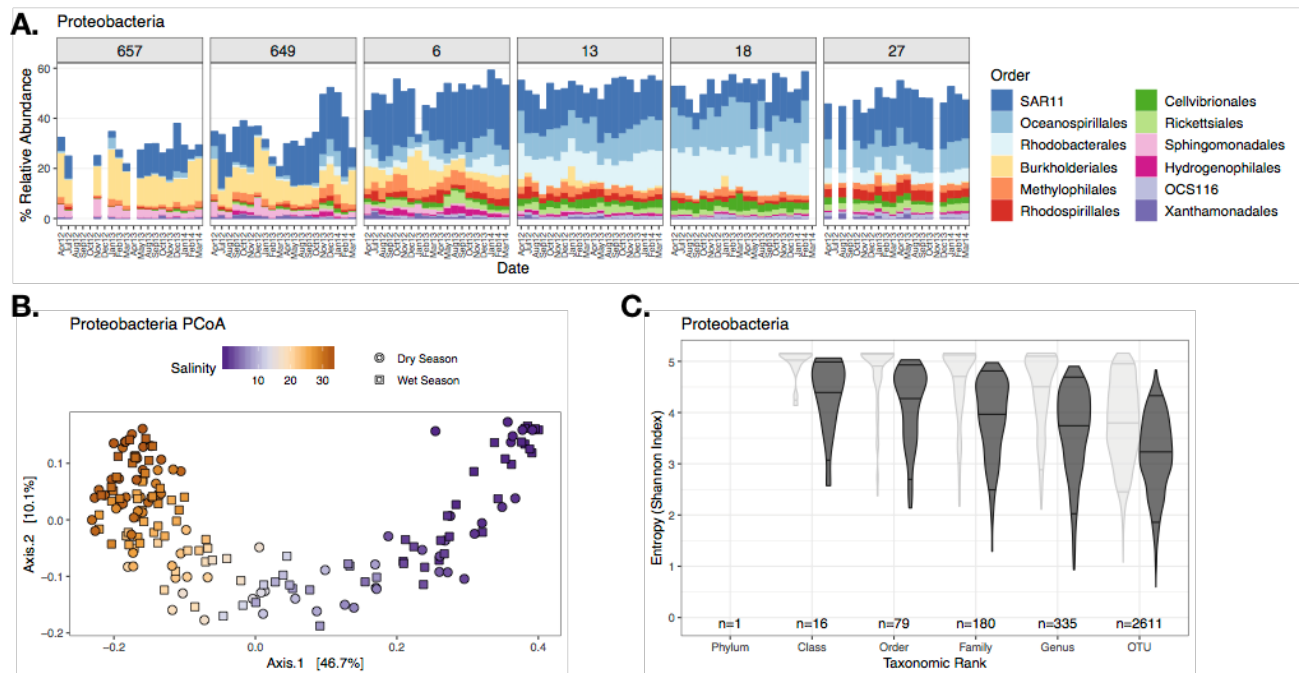
13

Fig. S12 Relative abundance of the 40 most abundant OTUs in SFB. Panels correspond to 6 representative stations for the five salinity zones and South Bay. The x-axis of each panel corresponds to date of 20 approximately monthly cruises from April 2012 to March 2014. The y-axis is the % relative abundance.

Fig. S13 Proteobacteria diversity metrics. Relative abundance of (A) 10 most abundant Proteobacteria orders. Panels correspond to 6 representative stations for the five salinity zones and South Bay. The x-axis of each panel corresponds to date of 20 approximately monthly cruises from April 2012 to March 2014. The y-axis is the % relative abundance. (B) PCoA ordination based on bray-curtis dissimilarity shows beta diversity is heavily influenced by salinity. (C) Entropy of taxa decreases at finer taxonomic levels, indicating high site-specificity of OTUs.
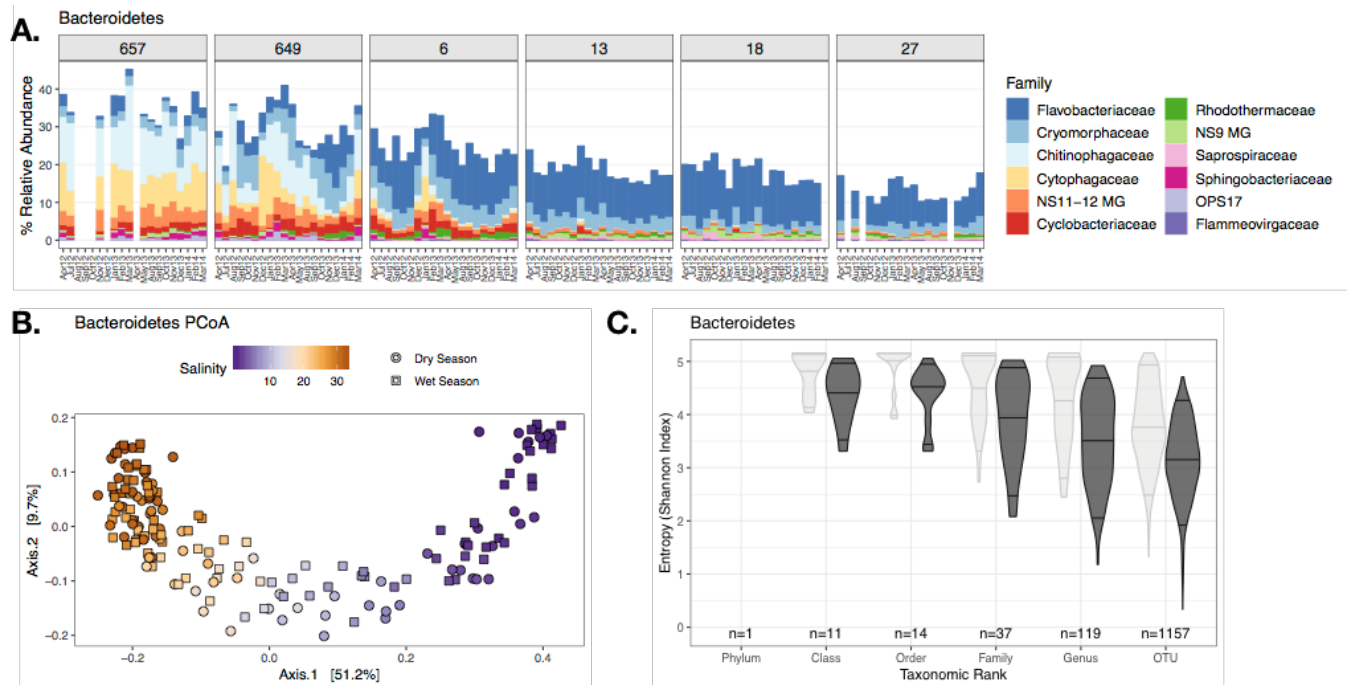
Fig. S14 Bacteroidetes diversity metrics. Relative abundance of (A) 10 most abundant Bacteroidetes families. Panels correspond to 6 representative stations for the five salinity zones and South Bay. The x-axis of each panel corresponds to date of 20 approximately monthly cruises from April 2012 to March 2014. The y-axis is the % relative abundance. (B) PCoA ordination based on bray-curtis dissimilarity shows beta diversity is heavily influenced by salinity. (C) Entropy of taxa decreases at finer taxonomic levels, indicating high site-specificity of OTUs.
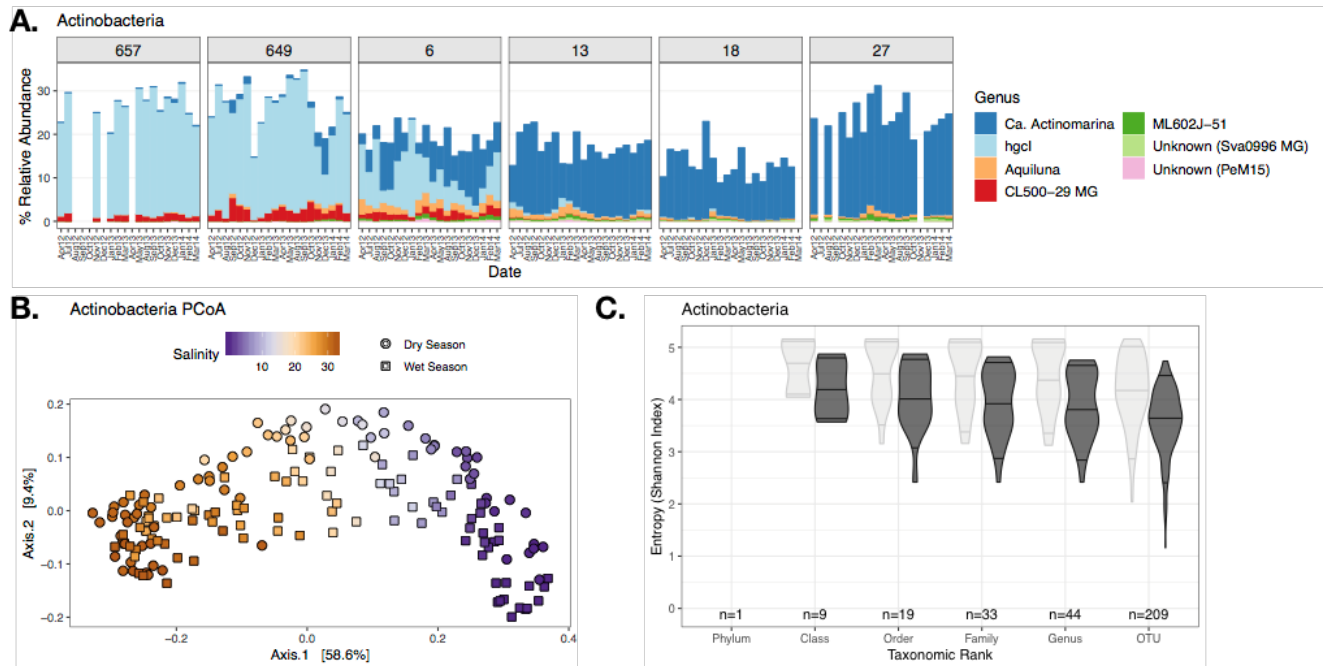
Fig. S15 Actinobacteria diversity metrics Relative abundance of (A) 10 most abundant Bacteroidetes families. Panels correspond to 6 representative stations for the five salinity zones and South Bay. The x-axis of each panel corresponds to date of 20 approximately monthly cruises from April 2012 to March 2014. The y-axis is the % relative abundance. (B) PCoA ordination based on bray-curtis dissimilarity shows beta diversity is heavily influenced by salinity and season (wet versus dry). (C) Entropy of taxa decreases at finer taxonomic levels, indicating high site-specificity of OTUs.
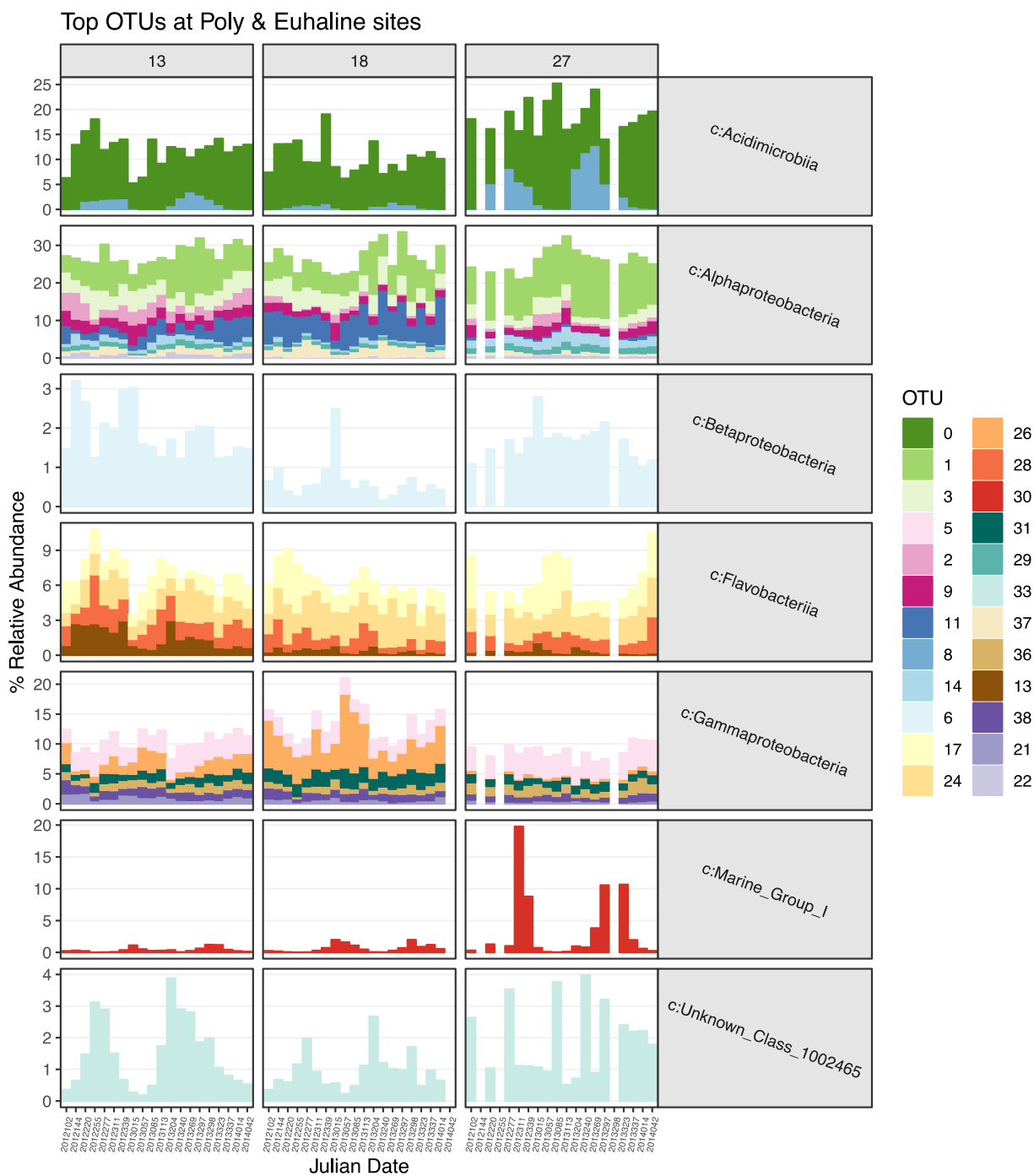
Fig. S16 OTUs in poly and euhaline sites. Panels correspond to 3 generally poly- and euhaline stations, 13 in North Bay, 18 at the mouth of the Golden Gate Bridge, and 27 in South Bay. The x-axis of each panel corresponds to date of 20 approximately monthly cruises from April 2012 to March 2014. The y-axis is the % relative abundance of OTUs.
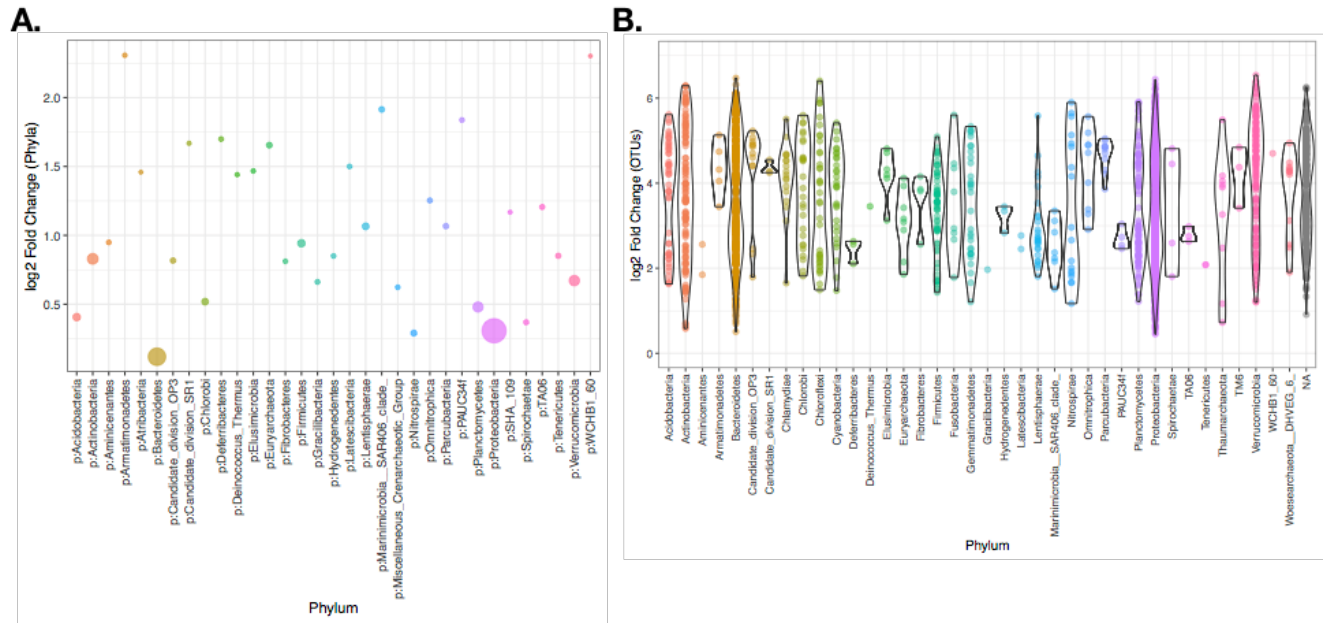
Fig. S17 Differential abundance plots show the log2 fold change of a taxa across salinity zones. Size of points is based on the mean abundance of the taxa.
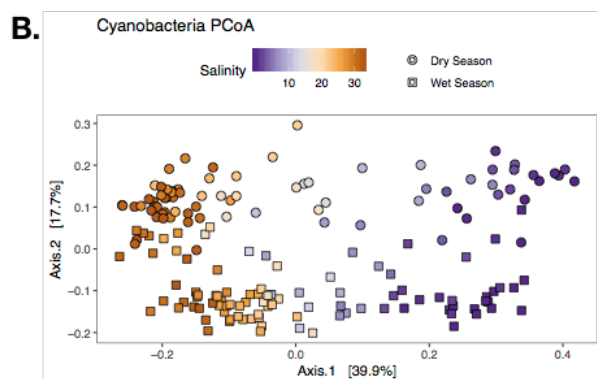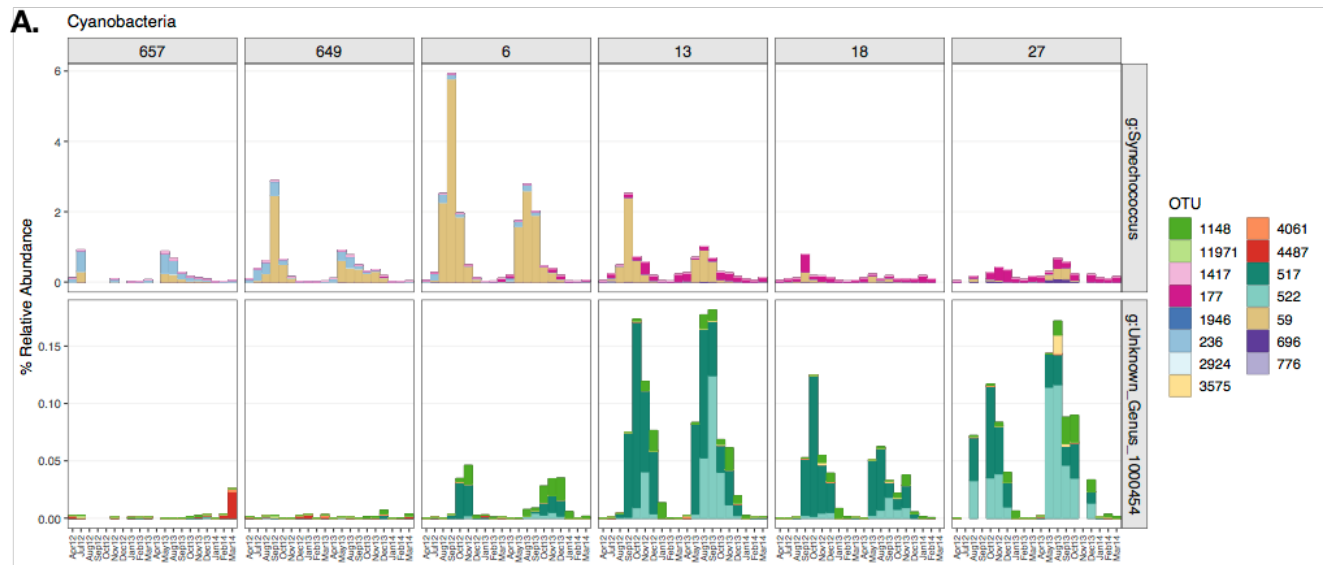
Fig. S18 Cyanobacteria diversity metrics. Relative abundance of (A) the two most abundant genera (*Synechococcus* and an Unknown *Sericytochromatia*) and (B) PCoA ordination showing beta diversity is most strongly influenced by salinity on axis 1 and season on axis 2. Cyanobacteria show strong summer peaks in abundance and are dominated by *Synechococcus*. Panels correspond to 6 representative stations for the five salinity zones and South Bay. The x-axis of each panel corresponds to date of 20 approximately monthly cruises from April 2012 to March 2014. The y-axis is the % relative abundance. Note the difference in y-axis scale in each row of panels.

Table S1 Taxonomic and read count data for libraries processed through the itagger method and after various filtering stages. Data includes two combined 16S rRNA amplicon (iTag) plates for each primer set.

	V4		V4-V5	
	Raw	>15 reads in >2 samples	Raw	>15 reads in >2 samples
Phyla	54	46	63	50
Class	143	123	169	124
Order	292	227	293	207
Family	549	400	510	342
Genera	1064	718	951	610
OTUs	21703	7006	15097	4913
Reads	155,573,985	154,724,570	107,986,733	107,350,210

Table S2 Correlations between environmental variables and diversity metrics for V4 and V4-V5 datasets with the highest and most significant coefficients or of most interest. Referring to PCoAs that include all samples.

	V4		V4-V5	
Linear Regression test	r	r^2	r	r^2
OTU				
PCoA Axis 1 ~ Salinity	-0.97**	0.94	-0.97**	0.94
PCoA Axis 2 ~ ln(SPM)	0.60**	0.36	0.57**	0.32
Richness ~ ln(SPM)	0.74**	0.55	0.70**	0.49
Richness ~ chla/(chla+phaeo)	-0.59**	0.35	-0.52**	0.27
Richness ~ Salinity	0.056	0.003	-0.11	0.012
Genera				
Richness ~ ln(SPM)	0.59**	0.35		
Richness ~ chla/(chla+phaeo)	-0.53**	0.28		
Richness ~ Salinity	-0.24*	0.06		
Phyla				
Richness ~ ln(SPM)	0.44**	0.19		
Richness ~ chla/(chla+phaeo)	-0.31**	0.10		
Richness ~ Salinity	0.40**	0.16		

**p-value < 0.0001
*p-value < 0.001
values without * have p-value > 0.05

Table S3 Nestedness and turnover components of PCoA ordinations based on Jaccard Dissimilarity. Values were calculated using the beta.core and beta.multi functions which compute multiple-site dissimilarities for spatial turnover and nestedness components of beta diversity, and the sum of both values

Ordination	Turnover fraction	Nestedness-resultant fraction	Overall beta diversity
Whole community	0.972	0.007	0.979
North Bay	0.981	0.005	0.986
South Bay	0.915	0.031	0.946
Fresh	0.840	0.050	0.890
Oligohaline	0.881	0.027	0.908
Mesohaline	0.898	0.036	0.934
Polyhaline	0.944	0.018	0.962
Euhaline	0.917	0.023	0.940

Table S4 Strongest correlation coefficients between environmental variables and diversity metrics for samples in North or South bay or various salinity zones.		
	r	*r²*
North Bay		
OTU Richness ~ ln(SPM)	0.71	0.50
PCoA Axis 1 ~ Salinity	-0.96	0.92
PCoA Axis 2 ~ lnSPM	0.37	0.14
South Bay		
OTU Richness ~ ln(SPM)	0.88	0.77
PCoA Axis 1 ~ ln(SPM)	-0.89	0.79
PCoA Axis 2 ~ Temperature	-0.70	0.49
Fresh		
PCoA Axis 1 ~ Temperature	0.85	0.72
PCoA Axis 2 ~ Salinity	0.73	0.53
Oligohaline		
PCoA Axis 1 ~ Temperature	0.80	0.64
PCoA Axis 2 ~ Salinity	-0.65	0.42
Mesohaline		
PCoA Axis 1 ~ Temperature	0.76	0.58
PCoA Axis 2 ~ Salinity	0.93	0.86
Polyhaline		
PCoA Axis 1 ~ ln (SPM)	-0.86	0.74
PCoA Axis 2 ~ Distance from 36	-0.86	0.74
Euhaline		
PCoA Axis 1 ~ Distance from 36	0.93	0.86
PCoA Axis 2 ~ Nitrite	-0.57	0.32

all p < 0.0001