1 **Bimodal occupancy-frequency distributions uncover the importance of**
2 **regional dynamics in shaping marine microbial biogeography**
3

4 Markus V. Lindh[1][¶], Johanna Sjöstedt[1][§], Börje Ekstam[2], Michele Casini[3], Daniel Lundin[1], Luisa W.
5 Hugerth[4], Yue O. O. Hu[4], Anders F. Andersson[4], Agneta Andersson[5], Catherine Legrand[1], Jarone
6 Pinhassi[1][*]
7

8 [1]Centre for Ecology and Evolution in Microbial model Systems - EEMiS, Linnaeus
9 University, SE-39231 Kalmar, Sweden.
10 [2]Department of Biology and Environmental Science, Linnaeus University, SE-39231 Kalmar,
11 Sweden.
12 [3]Swedish University of Agricultural Sciences, Department of Aquatic Resources, Institute of
13 Marine Research, SE-45330, Lysekil, Sweden.
14 [4]KTH Royal Institute of Technology, Science for Life Laboratory, School of Biotechnology,
15 SE-10691 Stockholm, Sweden
16 [5]Department of Ecology and Environmental Science, Umeå University, Umeå SE-90187,
17 Sweden.
18 [¶]Present address: Department of Oceanography, Center for Microbial Oceanography
19 Research and Education (C-MORE), University of Hawaii at Manoa, US-96822, Honolulu,
20 USA.
21 [§]Present address: Department of Biology/Aquatic Ecology, Lund University, SE-22362 Lund,
22 Sweden/Department of Ecology and Genetics, Uppsala University, SE-75236 Uppsala,
23 Sweden
24 *To whom correspondence should be addressed. Centre for Ecology and Evolution in Microbial
25 model Systems - EEMiS, Linnaeus University, Barlastgatan 11, SE-39182 Kalmar, Sweden.
26 Phone: +46-480-446212; E-mail: jarone.pinhassi@lnu.se
27

28 **Classification:** Biological Sciences, Environmental Sciences
29 **Keywords**: Bimodal, species patterns, theoretical ecology, regional effects, metapopulation,
30 mechanistic models, occupancy-frequency, marine bacteria, spatio-temporal, bacterial diversity
31 **Short title:** Regional dynamics of marine bacterioplankton
32

37
38

Abstract

Metapopulation theory developed in terrestrial ecology provides applicable frameworks for interpreting the role of local and regional processes in shaping species distribution patterns. Yet, empirical testing of metapopulation models on microbial communities is essentially lacking. Here we determined regional bacterioplankton dynamics from monthly transect sampling in the Baltic Sea Proper (16 sites, 11 occasions, 2010-2011) using 16S rRNA gene pyrosequencing. A strong positive correlation was found between local relative abundance and occupancy of populations. Notably, the occupancy-frequency distributions (the number of populations occupying different number of sites) were significantly bimodal with a satellite mode of mostly rare endemic populations and a core mode of abundant cosmopolitan populations (e.g. *Synechococcus*, SAR11 and SAR86 clade members). Observed temporal changes in population distributions supported theoretical predictions that stochastic variation in local extinction and colonization rates accounted for observed bimodality. Moreover, bimodality was found for bacterioplankton across the entire Baltic Sea, and was also frequent in globally distributed datasets where average Bray-Curtis distances were significantly different between bimodal and non-bimodal datasets. Still, datasets spanning waters with distinct physicochemical characteristics or environmental gradients, e.g. brackish and marine or surface to deep waters, typically lacked significant bimodal patterns. When such datasets were divided into subsets with coherent environmental conditions, bimodal patterns emerged, highlighting the importance of positive feedbacks between local abundance and occupancy within specific biomes. Thus, metapopulation theory applied to microbial biogeography can provide novel insights into the mechanisms governing shifts in biodiversity resulting from natural or anthropogenically induced changes in the environment.

Significance statement

Marine bacteria regulate global cycles of elements essential to life and respond rapidly to environmental change. Yet, the ecological factors that determine distribution and activity patterns of microbial populations across different spatial scales and environmental gradients remain basically unconstrained. Our metapopulation model-based analyses show that dispersal-driven processes contribute to structuring the biogeography of marine microorganisms from small to large geographical areas. Discovery of bimodal distribution patterns pinpointed satellite microbial populations with highly restricted ranges and defined abundant core populations widely distributed in coherence with environmental conditions. Thus, application of metapopulation models on microbial community structure may allow the definition of biogeographic regions critical for interpreting the outcome of future ocean changes.

Introduction

Marine microorganisms regulate ecosystem services essential to life through their influence on fluxes of matter and energy in the ocean (1). Conspicuously, these fluxes are ever dynamic due to the pronounced potential of microbial communities to rapidly respond to environmental change, both through adjustments in species composition and metabolic activity (2-4). Although microbial species or populations have been inferred to be active in a variety of distinct biogeochemical processes (see e.g. 5-8), very little is known about the ecological factors that contribute to determining biogeographical distribution patterns of these key organisms. Recent developments in molecular genetics methodologies, i.e. high-throughput DNA sequencing technologies, now allow for detailed spatial and temporal investigations of complex assemblages of microorganisms (9, 10). Thus, it is now not only possible to address questions regarding existing taxonomic diversity but also what ecological

89   processes that potentially influence the structure of natural marine microbial communities.

90   Historically, local communities have been regarded as independent ecological units
91   with individual integrity, determined by local interactions among coexisting species and
92   environmental factors (11). However, over the past decades, the profound influence of
93   regional factors, e.g. dispersal, in structuring local species assemblages has been established
94   (12-15). A positive relationship between local abundance and regional distribution of species
95   is a pattern characteristic for a wide range of organisms and ecosystems, including both
96   macroorganisms and highly diverse phytoplankton and bacterioplankton assemblages (16-
97   18). Although this relationship is considered to be one of the most robust patterns in
98   community ecology, significant variation in the level of correlation between local abundance
99   and regional distribution occurs among observed assemblages, suggesting that different
100   ecological processes contribute to the shape of the relationship (19-21). There is now an
101   increased awareness of the importance of regional dynamics also for the structuring of
102   aquatic bacterioplankton communities (14, 15, 21). Regional dynamics are linked to a second
103   pattern - the species rank-abundance distribution that describes the abundance of a species in
104   an area or sample as a function of its abundance rank. Especially at larger spatial scales,
105   species abundances exhibit a log-normal distribution, where a few species make up a high
106   proportion of all extant individuals while a majority of species are rare (11, 18, 22, 23, 24).
107   High throughput analyses of linkages between spatial and temporal patterns of
108   bacterioplankton populations across biomes, through assessment of relative sequence
109   abundances, have provided thorough knowledge of the vast genetic diversity dwelling
110   beyond the ocean surface (25, 26). These studies establish that prokaryotic assemblages share
111   much of the same characteristics of biogeographic distribution patterns as macroorganisms.
112   Such high-throughput data from biogeographical studies may disentangle whether
113   mechanistic concepts of distribution patterns in terrestrial systems are applicable also to

114   assemblages of marine pelagic prokaryotes.

115   Two major hypotheses for explaining ecological mechanisms shaping species

116   distribution patterns have attracted important attention; Levin's model and the core-satellite

117   (CSH) hypothesis (11, 16, 27). When applied in terrestrial ecology, these models and their

118   modified versions have provided explanatory power to identify the drivers of distribution

119   patterns among insects, plants and animals; this has had particular significance for developing

120   strategies in conservation biology (28). In essence, metapopulation theory describes the

121   dynamics of populations in a patchy environment, i.e. how local colonization and extinction

122   rates interact with species distribution at a regional scale. The two original models differ in

123   how colonization and extinction rates are calculated to predict occupancy-frequency

124   distributions (the number of sites occupied by different number of species), and have been

125   elaborated over time (29, 30).

126   Levin's model predicts a unimodal occupancy-frequency distribution characterized by

127   a skewed pattern where most species in a region occupy a single site and the number of

128   species decreases with increasing number of sites (27). Hanski's metapopulation model or the

129   CSH hypothesis (16-18) predicts a quadratic function of both colonization and extinction

130   rates and takes into account positive feedback mechanisms between local abundance and

131   occupancy. Extinction rates are therefore predicted to be low for populations with high and

132   low occupancy but high for populations with intermediate occupancy. In contrast to Levin's

133   model, the CSH predicts bimodal distributions where the decrease in number of species with

134   increasing number of study sites is followed by an increase in species occupying all or most

135   sites (16). Thus, the predicted occupancy-frequency pattern in CSH is characterized by a

136   mode of many endemic satellite populations present at only a few sites, and a second mode of

137   cosmopolitan core populations present at all sites. It has been suggested that such bimodal

138   occupancy-frequency patterns should be the norm in aquatic habitats, not least for free-living

139 easily dispersible prokaryotes (31). However, there has been little empirical testing of these

140 models in aquatic habitats in general and for aquatic microorganisms in particular (but see

141 (32)).

142   In the present study, we used metapopulation models to analyze local and regional

143 dynamics of bacterioplankton populations (operational taxonomic units [OTUs], defined at

144 97% 16S rRNA gene sequence identity). We investigated regional effects on monthly

145 sampled local communities in the Baltic Sea Proper by determining: (i) the relationship

146 between local OTU abundance and OTU distribution, i.e. occupancy, (ii) the type of

147 occupancy-frequency distribution pattern, i.e. the number of OTUs occupying different

148 number of sites, and (iii) the relationship between OTU occupancy and colonization and

149 extinction rates. We further identified OTUs showing distinct distributional and

150 colonization/extinction patterns. To validate observed distribution patterns we extended the

151 analyses of occupancy-frequency distributions to complementary datasets spanning the entire

152 Baltic Sea as well as global International Census of Marine Microbes (ICoMM) datasets. We

153 also evaluated observed OTU distributions with beta-diversity patterns and variability in

154 environmental conditions available as metadata in the ICoMM database.

155

156 **Results and Discussion**

157 **Bacterioplankton community composition in the Baltic Sea Proper.** We investigated

158 bacterioplankton population dynamics in a grid of 16 stations in the Baltic Sea Proper over

159 two years (2010-2011) to elucidate the role of regional dynamics in shaping biogeographical

160 patterns (Fig. S1). Bacterioplankton community composition showed pronounced clustering

161 of samples from the two years according to season (Fig. S2). This recurrent seasonal

162 succession in bacterioplankton agrees with the recognized importance of seasonal shifts in

163 environmental conditions in shaping community composition of marine bacteria (23, 24, 25).

164 There was also a tendency towards differentiation of community composition between

165 coastal and offshore sites (Fig. S2).

166   Between 122 and 279 OTUs were detected per sampling point, out of which between

167 10 and 22 OTUs were abundant while a grand majority (typically 100-257 OTUs) were rare,

168 when defining abundant and rare OTUs as having relative abundances >1% and <0.1%,

169 respectively (33). All sampled communities displayed rank-abundance curves following a

170 log-normal distribution (Fig. S3), typical of aquatic microbial communities, and widespread

171 also among macroorganism communities in terrestrial environments (24, 33-36).

172   The pattern of many rare compared to a few abundant populations for organisms in a

173 variety of habitats may be inherently linked with strong regional dynamics, i.e. a positive

174 relationship between local abundance and regional occupancy. Such regional dynamics have

175 been established for a wide range of organisms in different ecosystems, including

176 bacterioplankton in aquatic environments (19-21). We therefore plotted the number of sites

177 occupied by specific OTUs versus their average local relative abundance (Fig. 1). We used

178 locally weighted regression spline smoothing (LOESS) to indicate the positive relationship

179 between average local relative abundance of OTUs with average occupancy. This plot also

180 showed that a majority of OTUs had low abundance and low occupancy compared to

181 relatively few OTUs with high abundance and high occupancy. This suggests that the

182 mechanisms and regional dynamics typically observed among macroorganisms outside the

183 marine realm also shape bacterioplankton assemblages in surface waters.

184

185 **Evaluation of metapopulation models in the Baltic Sea Proper.** The relationship between

186 local relative abundance and regional occupancy may be derived from and be directly linked

187 to OTU abundance patterns. Analysis of distribution patterns among OTUs showed a

188 recurring occupancy-frequency distribution that was significantly bimodal in nine out of

189 eleven sampled months (Mitchell-Olds & Shaw's test, p<0.01, n=6-12; Fig. 2). In these

190 months, the bimodal pattern had one mode with most species found at a single site. This

191 mode was followed by an initial monotonic decrease in the number of populations detected

192 with increasing number of sites surveyed. We then noted a second mode with an increase in

193 the number of populations that occurred at all sites. Thus, at a majority of sampling occasions,

194 bimodal patterns were observed matching predictions of bimodality in the CSH model

195 proposed by Hanski (16) rather than predictions of unimodality in Levin's model (27). To our

196 knowledge, our observed bimodal pattern represents the first finding of bimodality for

197 prokaryotic communities as well as for organisms in marine environments in general.

198 Metapopulation models that include positive feedback mechanisms provide testable

199 predictions of colonization and extinction rates (28). To elucidate how our observed data fit

200 with current models, we calculated rates of colonization and extinction by comparing

201 occupancies of OTUs between succeeding months, i.e. how many new sites an OTU had

202 colonized compared to from how many previously occupied sites it had disappeared (Fig. 3;

203 Table S1; Table S2). As in previous work (37), we considered the terms colonization and

204 extinction in a broad sense as analogous to gain and loss of OTUs, i.e. local extinction is

205 going from presence to absence. Accordingly, absence is indicated by local relative

206 abundance <0.006%, given the detection limit of our sequencing methodology. In an

207 analogous manner, colonization represents the shift from absence (i.e. below detection limit)

208 to presence, irrespectively of whether this presence resulted from on-site growth of microbes

209 from initially very low abundance or from immigration of populations from neighboring

210 waters. Observed patterns in regional colonization rates followed a quadratic curve, where

211 colonization (C) at first increased with fraction of sites occupied (P) until about 50-65%

212 occupancy, followed by a decrease in C until maximum P (Fig. 3). Extinction rates followed

213 a similar quadratic pattern, where extinction (E) first increased with P, but until a higher

8

214   occupancy compared to C (between 65-75%), followed by a decrease of E until maximum P

215   (Fig. 3). We examined how metapopulation models could explain variations in colonization

216   and extinction rates by using non-linear least squares analysis. Predictions based on the CSH

217   model proposed by Hanski (16) agreed well with the field results for extinction rates (See

218   blue dashed lines in Fig. 3), with on average low residual standard error, whereas

219   colonization rates had higher residual standard error (Table S1). The CSH model significantly

220   explained both colonization and extinction rates. Predictions based on Levin's model (27)

221   showed overall higher residual standard error for extinction rates compared to the CSH model

222   (Table S1; S2). Notably, both models significantly explained observed extinction rates,

223   although extinction rates predicted by Levin's model had lower parameter estimates

224   compared to the CSH predictions (Table S1 and S2). Thus, both models produced significant

225   fits for quadratic and linear colonization and extinction rates. However, since the CSH

226   prediction of bimodal occupancy-frequency patterns matched our data on OTU distributions,

227   whereas Levin's predictions of unimodality did not, the CSH yielded an overall better

228   agreement with the observed data. This agreement of theoretical predictions and measured

229   bacterioplankton population dynamics favored the CSH model for interpreting regional

230   effects of OTU distributions in the Baltic Sea Proper.

231

232   **Linking metapopulation dynamics and specific populations in the Baltic Sea Proper.**

233   Beyond the recognition of metapopulation models for estimating colonization and extinction

234   rates of populations at a regional scale, the models function to identify ecological processes

235   that are likely to be involved in determining the distribution of individual populations. In the

236   Baltic Sea Proper dataset, the number of endemic satellite bacterioplankton populations (i.e.

237   only found in one site each month) varied between 940 and 2820 OTUs. In contrast, the

238   much fewer cosmopolitan core populations (i.e. found in all sites each month) ranged

239   between 26 and 69 OTUs. A striking consequence of the steep rank abundance distribution

240   patterns was that the OTUs found at all sites during one sampling month contributed to

241   between 62 and 78% of the sequence abundance that month (see yellow filled circles in Fig.

242   4). Thus, although a grand majority of OTUs occupied a single sampling site each, these

243   OTUs contributed only a very limited proportion of the total sequence abundance compared

244   to the fairly few OTUs occupying all sites. It is therefore reasonable to assume that a

245   community consists of a set of cosmopolitan core populations with high local and regional

246   abundance compared to endemic satellite populations with low local and regional abundance.

247   These findings reveal the importance of assessing core and satellite metapopulation dynamics

248   to understand the genomic potential among bacterioplankton in biogeochemical processes

249   such as carbon cycling. For example, in environmental metagenomic, transcriptomic or

250   single-cell genomic approaches it is not possible to determine satellite or core characteristics

251   per se within the data if samples were collected at a single local site. We infer that

252   interpretations of such –omics approaches could be much aided by complementary analyses

253   of 16S rRNA gene amplicon data at relevant spatio-temporal scales to couple regional

254   dynamics of OTUs for assessment of satellite and core properties. Alternatively, conducting –

255   omics approaches on many samples distributed over relevant spatiotemporal scales may be

256   necessary to distinguish regional dynamics in microbial communities and interpretations

257   thereof.

258   In the present study we aimed to couple the identity of particular OTUs found in the

259   analysis of local abundance and regional occupancy with observed metapopulation dynamics.

260   Thus, populations were defined according to their local relative abundance (i.e.

261   populations/OTUs >1% as abundant, <0.1% as rare (33), and 0.1-1% as common (38-40)).

262   First, abundant populations with little variance in occupancy and relative abundance (SE

263   <0.025 and <0.2%, respectively) were classified as "core"; corresponding to Hanski's core

264   populations, (n=6) (Fig. 1; Table S3). "Core" populations such as those belonging to the

265   *Synechococcus* (PF_000001)*,* CL500-29 (PF_000000), SAR11 (PF_000002), SAR86

266   (PF_000007), hgcI (PF_000008) and NS3a clades (PF_000009) were almost never found in

267   the rare fraction in our study, and were at nearly all times part of the rightmost column in

268   each of the panels in Figure 2. The presence of such cosmopolitan core populations is in

269   agreement with the dominance of abundant lineages such as SAR11 and SAR86 clades

270   typically found in marine environments (40-43). We have no immediate knowledge as to the

271   factors that make such particular "core" populations successful. However, it is possible that

272   "core" OTUs have advantages in utilization of widespread resources compared to other

273   OTUs (17, 30). Hence, "core" OTUs may be linked to the concept of generalist dynamics (4,

274   7). Interestingly, for soil bacteria there is a significant coupling between habitat breadth and

275   larger genome size, metabolic adaptability and occupancy, indicating advantages for versatile

276   generalist populations (44). In the marine environment bacterial populations can also be

277   widespread despite significant genome streamlining, e.g. the SAR11 and SAR86 lineages,

278   suggesting that other mechanisms than genome size can shape microbial biogeography (45).

279   In fact, many core populations found in the present study are predicted by phylogenetic

280   affiliation to have streamlined genomes. We propose that positive feedback mechanisms

281   between local abundance and regional occupancy can contribute to explain the widespread

282   distribution of core OTUs, such as populations in the SAR11 and SAR86 clades and

283   *Synechococcus*. This relative stability may be maintained despite potential local selection

284   pressures by habitat filtering and predation/grazing. In contrast, stochastic variation in

285   colonization and extinction rates has a large effect on populations with intermediate

286   occupancy. Taken together, our data suggest that core OTUs depend largely on maintaining

287   high levels of occupancy, implying an increased probability of colonizing new sites and

288   sustaining large population sizes through positive feedback mechanisms.

289   Second, populations fluctuating between being abundant, rare and common with large

290   variation in occupancy (SE >0.100) were classified as "transient" (n=38) (Fig. 1; Table S3).

291   Interestingly, members of the *Cyclobacteriaceae* (PF_000047) and *Sporichthyaceae*

292   (PF_000021) and the NS11-12 clade (PF_000030) mostly entered and exited between the

293   abundant and common fractions of the bacterioplankton, and did not become rare. In contrast,

294   the SAR92 (PF_000017) and *Spartobacteria* (PF_000011) OTUs fluctuated between the

295   abundant and rare fractions.

296   Third, rare populations (<0.1% of bacteria) with little variance in relative abundance

297   (SE < 0.2%), and high occupancy, were classified as "continuously rare" (n=9) (Fig. 1; Table

298   S3). It is noteworthy that these continuously rare populations, e.g. CL500-29 (PF_000176),

299   GOBB3-C201 (PF_000254), *Reichenbachiella* (PF_000068)*, Thiobacillus* (PF_000187) and

300   *Owenweeksia* (PF_000073), had a high occupancy. This potentially suggests that some

301   populations had a permanent and widespread ecological strategy of being rare. In agreement,

302   Galand and colleagues (46) observed how rare phylotypes remained rare when surface

303   seawater samples were compared with deep waters from the Arctic, highlighting that most

304   rare phylotypes were continuously rare within that ecosystem.

305   Fourth, rare populations with low occupancy were classified as "satellite"; equivalent

306   to Hanski's satellite populations, (n=4369) (Fig. 1; Table S3). Most "satellite" populations

307   were only found in a single site, e.g. *Sulfurimonas, Thermoplasmatales* (PF_012272), SAR11

308   (PF_002803), SAR86 (PF_003406), *Brevundimonas* (PF_001083) and *Pedobacter*

309   (PF_008706). The satellite characteristics of rare populations indicate a clear endemic

310   biogeography of the rare biosphere, likely caused by the high extinction rates observed. In

311   accordance, the rare biosphere has a distinct biogeography despite the potential for high

312   dispersal and low loss rates (46). Thus, in contrast to core populations, rare and

313   geographically more restricted satellite populations will be less likely to colonize new sites

314   and more likely to become locally extinct.

315   We note that the observed extinction rates were always higher than colonization rates

316   (>12.5 and <12.5, respectively) and that the parameter estimates of extinction and

317   colonization rates predicated by the core-satellite model were significantly different (Fig. 3;

318   Table S1; two-sample t-test, p=4.02x10^{-11}, n=21). Higher rates of extinction compared to

319   colonization imply that a few populations with higher colonization compared to extinction

320   rates cause bimodal occupancy-frequency patterns. However, most populations are headed

321   towards local extinction and regional rarity. If a few core populations did not exhibit higher

322   colonization rates than extinction rates through positive feedback mechanism, most

323   populations would eventually disappear. The quadratic relationships for colonization and

324   extinction rates at different occupancies highlight the importance of regional dynamics for

325   local populations with a decreasing probability for extinction with increasing occupancy.

326   Overall, these results substantiate the hypothesis that the effect of stochastic variation is

327   highest for populations with intermediate occupancy. A corollary of this scenario is that it is

328   far more likely for common populations than for rare ones to become more common, and

329   more likely for rare populations to become even more rare. Hence, we validated the core-

330   satellite hypothesis, showing that positive feedback mechanisms regulate bacterioplankton

331   population dynamics in the Baltic Sea Proper.

332   In our Baltic Sea Proper dataset we also observed special cases of population

333   dynamics characterized by high colonization or extinction events. Firstly, several OTUs

334   showed a pattern of being absent one month, and present at nearly all sites the following

335   month, thus having a large colonization rate (Fig. 3). We denote this special case of

336   colonization dynamics as "microbial rain" pattern, as a reformulation of the concept

337   "propagule rain" coined by Gotelli et al. (29) for dispersal patterns of seed propagules from

338   trees. The detected "microbial rain" populations, i.e. with very large variation in occupancy,

339   included *Roseobacter* (PF_000153)*, Fluviicola* (PF_000171) and *Brevundimonas*

340   (PF_000126). A reciprocal pattern was observed for extinction rates, where some OTUs were

341   present at all sites one month, next to be absent from all sites the following month - we term

342   this "microbial evanescence", indicating large extinction rates. "Microbial evanescence"

343   populations were exemplified by *Owenweeksia* (PF_000313) and two *Synechococcus* OTUs

344   (PF_000143, PF_000060). We suggest that "microbial rain" and "microbial evanescence"

345   types of OTU distributions could result from linkages between strong seasonal shifts in

346   environmental conditions and niche differentiation (7). Alternatively, wind-driven upwellings

347   or other hydrological events may also cause rapid shifts in bacterial community composition

348   (47, 48). Still, such hydrological events could not have been common in our study system,

349   considering the relatively stable seasonal succession and smaller impact of spatial differences

350   over the two studied years (Fig. S2). Similarly, analysis bacterioplankton population

351   dynamics in an adjacent Baltic Sea Proper sampling site showed distinct seasonal succession

352   related to environmental conditions such as temperature, but only occasionally influenced

353   directly by hydrological disturbances (38).

354   Altogether, we show that the core-satellite hypothesis can be applied to prokaryotic

355   assemblages and be used to characterize biogeographical patterns providing insight into the

356   importance of regional dynamics. It is noteworthy that the core-satellite hypothesis also

357   provides a mechanistic understanding of the division between rare and abundant

358   bacterioplankton populations in the local environment. Thus, the core-satellite hypothesis can

359   be used to interpret the common observation of log-normal OTU distribution patterns

360   potentially caused by stochastic variation in colonization and extinction rates.

361

362

363 **OTU abundance distribution in a Baltic Sea - Kattegatt transect.** To extend the

364 exploration of the applicability of the CSH and its consequences for interpreting

365 bacterioplankton distribution patterns, we investigated complementary datasets that covered a

366 range of marine environments and geographic distances. By applying our analysis to a

367 separate dataset of 16S rRNA genes from the Baltic Sea we could assess bacterioplankton

368 distribution patterns over the entire 2000 km salinity gradient of this semi-enclosed brackish

369 water system (Fig. S4A; 19 sites). We did not find a significant bimodal occupancy-

370 frequency pattern among surface samples when including marine sites from the Kattegat Sea

371 (Fig. S4B; Mitchell-Olds & Shaw's test, $p>0.01$, n=19). However, when excluding the

372 marine sites from the brackish water sites we found a significant bimodal pattern (Fig. S4C;

373 Mitchell-Olds & Shaw's test, $p<0.01$, n=17). The differentiation between marine and

374 brackish water sites indicated that changes in environmental conditions, in this case likely a

375 salinity shift (from 5.62 ± 2.06 to 22.00 ± 3.10, Baltic Sea and Kattegat Sea respectively),

376 potentially influenced the metapopulation dynamics and separated bacterioplankton

377 assemblages into distinct biomes in the Baltic Sea compared to the Kattegat Sea. In fact,

378 salinity regulates the distribution of bacterial populations and their functional potential in the

379 Baltic Sea (49, 50), and was recently suggested to constitute part of a global brackish

380 microbiome (10). We also note that strong shifts in other marine biota occur at the border

381 between brackish–marine conditions, also known as the Darss Sill, as shown in e.g.

382 palaeoecological analyses of diatoms (51). We conclude that Baltic Sea bacterioplankton

383 communities exhibit core and satellite population dynamics as observed in the Baltic Sea

384 Proper data. Thus, species-abundance patterns among Baltic Sea bacterioplankton may

385 provide clues to how bimodality is linked with specific biomes.

386

387   **OTU abundance distribution in global ICoMM datasets.** To further extend the analyses of

388   bimodal occupancy-frequency distributions, we carried out analyses of distribution patterns

389   on broad global biogeographic datasets provided in the frame of ICoMM (52). This showed

390   that bacterial communities in 17 of 31 oceanic regions displayed a significant bimodal pattern

391   (Table 1). In Fig. 4A we exemplify some of the statistically significant bimodal patterns by

392   the samples from the Arctic Chukchi and Beaufort Seas (ACB), Deep Arctic Ocean (DAO),

393   New Zealand sediments (NZS) and open ocean off Cape Cod (SSD) (Mitchell-Olds &

394   Shaw's test, $p<0.01$, n=8-16; Fig. 4A; Table 1). It is noteworthy that bacterial community

395   composition in several vast oceanic regions (e.g. ACB and NZS) displayed bimodality. Still,

396   in datasets from several regions, OTU occupancy-frequency distribution patterns were not

397   significantly bimodal (Table 1), which can be exemplified by the data sets from the northwest

398   Mediterranean (BMO) and Sponge communities from the Great Barrier Reef (SPO)

399   (Mitchell-Olds & Shaw's test, $p>0.01$, n=11-16; Fig. 4B; Table 1). However, some of these

400   datasets stand out by the heterogeneity of the samples they encompass. For example, it is well

401   established that bacterioplankton community composition differs significantly between

402   depths in a range of ecosystems, including the northwest Mediterranean Sea (see e.g. (53).

403   Still, the BMO transect comprised samples from different depths. We therefore separated

404   samples according to depth and analyzed surface waters separately; we thus found a

405   significant bimodal occupancy-frequency pattern for these surface samples (too few samples

406   were available to test bimodality at other depths). In a similar manner, it is established that

407   bacterial community composition differs considerably between different sponge species (54).

408   Thus, when accounting for sponge species, we found a significant bimodal occupancy-

409   frequency pattern for OTUs associated with the sponge *Rhopaloeides odorabile* in the SPO

410   dataset. Taken together, analyses of datasets from several oceanic regions provided support

411   for the widespread prevalence of bimodality among marine prokaryotic assemblages.

16

412   From the above we deduce that colonization of bacterioplankton populations could be

413   regulated by physicochemical factors, thus structuring community composition into specific

414   biomes. In such a scenario, large differences in environmental conditions resulting from e.g.

415   differences in depth, salinity or nutrient concentrations and/or physical oceanography within

416   or between oceanic regions should lead to distinct patterns of bacterioplankton community

417   structure, i.e. no bimodality. In contrast, regions within which dispersal is high compared to

418   local environmental conditions bimodality should be detected. We therefore evaluated the

419   relationship between bimodal occupancy-frequency patterns and beta-diversity and

420   environmental conditions. To analyze such potential dependencies, we calculated average

421   Bray-Curtis distances for all ICoMM datasets and separated these distances into two groups

422   for datasets with significant or non-significant bimodal patterns, respectively. This revealed a

423   significant difference between groups with and without significant bimodal patterns (two-

424   sample t-test, $p=8.09 \times 10^{-7}$, n=31); datasets with significant bimodal patterns on average had

425   lower Bray-Curtis dissimilarity (Fig. 5A). In contrast, datasets with higher average Bray-

426   Curtis dissimilarity were linked to non-significant bimodal patterns. Moreover, when

427   comparing the ratio between core and satellite populations (percentage of OTUs occupying

428   all sites compared to a single site) and Bray-Curtis dissimilarities, the ratio was significantly

429   negatively correlated with average Bray-Curtis distance (linear regression, $p=6.42 \times 10^{-4}$,

430    $R^2=0.42$, n=31; Fig. 5B). A second-degree polynomial curve fit indicated a saturation where

431   the ratio leveled off at around 0.04 ($p=1.4 \times 10^{-5}$, n=31; Fig. 5B). Similarly, the ratio between

432   core and satellite populations was on average 0.05±0.03 in the Baltic Sea Proper dataset and

433   monthly data coupled with average Bray-Curtis distances followed the negative slope (see

434   open circles Fig. 5B). In terms of absolute changes in environmental conditions - i.e. analyses

435   made on average Euclidean distances - temperature, salinity and depth had an overall

436   tendency toward lower absolute differences for datasets with significant bimodal patterns

17

437 (Fig. 5C-E). However, only depth (two-sample t-test, p=0.045, n=24) and temperature (two-

438 sample t-test, p=0.036, n=20) were significantly different between datasets with and without

439 significant bimodal patterns. Notably, absolute shifts in geographic distance were not

440 significantly different between datasets with or without significantly bimodal patterns (Fig.

441 5F; two-sample t-test, p=0.292, n=26). This suggests that shifts in environmental conditions,

442 and not spatial distance *per se*, are critical for regulating positive feedback mechanisms

443 between local abundance and regional occupancy, thus affecting the occupancy-frequency

444 pattern. Collectively, these data and changes in occupancy-frequency patterns may be highly

445 valuable for distinguishing where shifts in biomes occur across environmental gradients.

446   We are aware of only one study that has found bimodality in aquatic systems, and that

447 study concerned fish in Amazonian lakes (55). The CSH has also been tested in a set of

448 boreal stream systems, where it was found that diatoms did not exhibit bimodality (32). The

449 discovery in the present study of bimodal distribution patterns, linking local abundance and

450 regional occupancy of marine planktonic bacteria, may seem surprising, especially given the

451 common perception that pelagic marine habitats are relatively homogeneous and allow

452 efficient dispersal of prokaryotic assemblages. Nevertheless, differential distribution of

453 marine bacterioplankton populations is frequently found to be linked with shifts in

454 environmental conditions, such as salinity, temperature, phytoplankton biomass and DOC (25,

455 49, 56, 57). Interestingly, it has been suggested that due to the nature of free-living and easily

456 dispersible prokaryotes, bimodal occupancy-frequency patterns should be relatively common

457 (31). It is also noteworthy that explorations of occupancy-frequency patterns in highly

458 diverse communities is lacking; typically <100 species are considered, whereas only a few

459 have investigated communities with >400 species (28, 32). Thus, in the present paper we

460 show for the first time linkages between highly diverse prokaryotic communities in marine

461 environments and metapopulation dynamics.

462   Although the bimodality patterns observed in the current study conforms with

463   Hanski's metapopulation model, alternative hypotheses have been suggested to explain the

464   shape of different occupancy-frequency distributions, including for example sampling effects

465   (28, 29, 58). Still, in the datasets studied here, bimodality was found over a variety of

466   temporal scales and over a diversity of geographical distances, indicating a pronounced

467   robustness in this occupancy-frequency pattern. Alternatively, neutral mechanisms in

468   combination with dispersal could shape a distribution pattern with locally abundant taxa also

469   being frequently occurring (35, 36, 59). However, colonization and extinction rates predicted

470   by the CSH agreed well with our field data and point toward that stochastic variation in rates

471   of local extinction and/or colonization can account for observed bimodality. Taken together,

472   our findings on the biogeography of microbial populations at distances ranging from <10 km

473   to >10,000 km indicate that bimodal OTU distributions are an important recurring pattern in

474   several marine ecosystems across the world's oceans, stressing the importance of regional

475   dynamics.

476

477   **Conclusions.** It is becoming increasingly clear that community-structuring processes acting

478   on macroorganisms also determine distribution patterns of marine bacterioplankton

479   populations (60-62). Our extensive analyses suggest that, as generally recognized in

480   terrestrial ecology, positive feedback mechanisms between local abundance and regional

481   occupancy are important drivers of bacterioplankton community composition. As such, our

482   findings demonstrate that metapopulation theory, developed for terrestrial ecology, also

483   provides a tool for interpreting patterns of microbial biogeography. The application of the

484   core-satellite hypothesis on prokaryotic assemblages in the sea forms a framework for

485   interpreting biogeographical patterns and defining the division between rare and abundant

486   bacterioplankton populations in the local environment. Characterization of metapopulation

487   dynamics in marine bacterioplankton communities would not only allow for comprehensive

488   biogeographical analyses of marine microbes, but also contribute to improve the definition of

489   cosmopolitan and endemic populations and the mechanisms by which they are influenced.

490   Our findings suggest that shifts in microbial biogeography, in particular microbial biomes as

491   defined by bimodal occupancy-frequency patterns compared to unimodal patterns, may be

492   used to describe and pinpoint changes in the environment likely caused by dispersal

493   limitation for use in molecular biomonitoring.

494

495   **Material and Methods**

496   **Field sampling.** In 2010-2011 (April-October), spatio-temporal dynamics of

497   bacterioplankton communities were investigated at 16 stations off the east coast of Sweden in

498   the Baltic Sea Proper. The distance between stations was typically less than 10 km (Fig. S1).

499   The study area was chosen as a part of collaboration between the projects PLANFISH and

500   EcoChange to study food-web dynamics in a coastal to open ocean gradient. For details on

501   hydrological conditions in the study area please see (63, 64). Seawater from each station was

502   collected in acid washed Milli-Q rinsed polycarbonate bottles, at discrete depths (2, 4, 6, 8

503   and 10 m) that were pooled and filtered shipboard. All 16 stations were represented in 2010

504   and classified according to their spatial distribution (coastal-offshore) as coastal, middle and

505   open sea stations. During 2011, only coastal and open stations were represented. Not all

506   stations could be included every month, and we used instead a subset of 5-12 stations (in total

507   117 stations).

508

509   **DNA extraction, 454 PCR, sequence processing and analysis.** Collection of DNA from the

510   117 samples was carried out by filtering 1-2 liter of seawater onto 0.2 µm pore size, 47-mm

511   diameter Supor filters (PALL Life Sciences). The filters were immediately frozen at - 80 °C

512   in 1.8ml TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0) until further processing. DNA

513   extraction was performed according to the phenol-chloroform protocol in Riemann et al. (65).

514   Bacterial 16S rRNA gene fragments was amplified with bacterial primers 341F and 805R

515   (V3-V4 hypervariable region) containing adapters and barcodes following the protocol of

516   Herlemann et al. (49). The resulting purified barcoded amplicons were normalized in

517   equimolar amounts and sequenced on a Roche GS-FLX 454 automated pyrosequencer

518   (Roche Applied Science, Branford, CT, USA) at SciLifeLab, Stockholm, Sweden. Raw

519   sequence data were processed following the bioinformatical pipeline in Lindh et al. (66).

520   Briefly, sequences from all samples were clustered together into operational taxonomic units

521   (OTU) at the 97% identity level (sequence similarity) using USEARCH (67) and

522   taxonomically identified using the SINA/SILVA database. The 454 runs resulted in 300,000

523   reads (2010) and 396,000 reads (2011). After denoising and chimera removal, samples

524   contained on average 4176 (±1716 SD) sequence reads (2010) and 4432 (±1445 SD)

525   sequence reads (2011) for each sample. The final OTU table consisted of 4437 different

526   OTUs (excluding singletons). For alpha-diversity measures we subsampled the sequences to

527   2500 reads per sample. Rarefaction curves are provided in Figure S5. DNA sequences have

528   been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read

529   Archive under accession number SRP023607.

530   As for the Baltic Sea Proper dataset surface water during the full Baltic Sea transect

531   was collected on 0.2 μm pore size, 47-mm diameter Supor filters (PALL Life Sciences) and

532   16S rRNA genes were amplified with 341F-805R. Sequencing of the Baltic Sea transect were

533   performed on Illumina MiSeq. Sequences were clustered into OTUs at 99% identity level

534   using USEARCH (67). For the Baltic Sea dataset we subsampled the sequences to 10,000

535   reads per sample.

536   OTU tables (97% 16S rRNA gene fragment sequence identity) of bacterial sequences

537   (V6 hypervariable region) from ICoMM projects were downloaded from the Visualization

538   and   Analysis   of   Microbial   Population   Structures   (VAMPS)   database

539   (http://vamps.mbl.edu/index.php). ICoMM datasets were subsampled at 10,000 reads per

540   sample.   A   map   of   the   ICoMM   stations   is   available   at

541   http://vamps.mbl.edu/google_earth/ge_icomm.php showing the location and spatial scale of

542   the oceanic regions included in the analyses performed in the present paper.

543

544   **Statistical analyses and graphical outputs.** All statistical tests were performed in R 3.2.2

545   (68), using the package Vegan (69). An equivalent for Tokeshi's test of bimodality was

546   performed using Mitchell-Olds & Shaw's test for the location of quadratic extremes. Non-

547   linear least squares analysis was performed using the equations of Levin's model and the

548   CSH hypothesis. Levin's original model (27) is calculated as follows:

549         $dP/dt = CP(1 - P) - EP$ (1)

550   where P is the fraction of occupied sites, C is colonization rate and E is extinction rate. When

551    $P = 1$, occupancy is 100% and all sites are occupied and when $P = 0$, occupancy is 0%,

552   meaning regional extinction. Colonization rate is the number of colonized empty sites over

553   time. If C is greater than E plus the variance in E ($C > E + s^2_E$), then Levin's model predicts a

554   unimodal distribution. Hanski's model (16) is calculated as follows:

555         $dP/dt = CP(1 - P) - EP(1 - P)$ (2)

556   Thus, if the variance in S ($C - E$) is greater than S (i.e. $\sigma^2_S > S$) the model predicts a bimodal

557   species distribution.

558   Graphical outputs was made in R 3.1.2 using the package ggplot2 (70). Figures S1

559   and S4A were made using ODV (Version 4.5.0).

560

561

22

574

575

576   **References**

577   1.   Falkowski PG, Fenchel T, & Delong EF (2008) The Microbial Engines That Drive

578   Earth's Biogeochemical Cycles. *Science* 320(5879):1034-1039.

579   2.   Allison SD & Martiny JB (2008) Colloquium paper: resistance, resilience, and

580   redundancy in microbial communities. *Proc Natl Acad Sci U S A* 105 Suppl 1:11512-

581   11519.

582   3.   Logue JB*, et al.* (2015) Experimental insights into the importance of aquatic bacterial

583   community composition to the degradation of dissolved organic matter. *ISME J.*

584   4.   Gomez-Consarnau L, Lindh MV, Gasol JM, & Pinhassi J (2012) Structuring of

585   bacterioplankton communities by specific dissolved organic carbon compounds.

586   *Environ Microbiol* 14(9):2361-2378.

587   5.   Moran MA, *et al.* (2004) Genome sequence of *Silicibacter pomeroyi* reveals

588   adaptations to the marine environment. *Nature* 432:910-913.

589   6.   Mou X, Sun S, Edwards RA, Hodson RE, & Moran MA (2008) Bacterial carbon

590   processing by generalist species in the coastal ocean. *Nature* 451(7179):708-711.

591   7.   Teeling H, *et al.* (2012) Substrate-Controlled Succession of Marine Bacterioplankton

592   Populations Induced by a Phytoplankton Bloom. *Science* 336(6081):608-611.

593   8.   Lauro FM, *et al.* (2009) The genomic basis of trophic strategy in marine bacteria.

594   *Proc Natl Acad Sci USA* 106(37):15527-15533.

595   9.   Poisot T, Pequin B, & Gravel D (2013) High-throughput sequencing: a roadmap

596   toward community ecology. *Ecol evol* 3(4):1125-1139.

597   10. Hugerth LW, *et al.* (2015) Metagenome-assembled genomes uncover a global

598   brackish microbiome. *Gen Biol* 830.

599   11. Lawton JH (1999) Are there general laws in ecology? *Oikos*:177-192.

600   12. Ricklefs RE (2008) Disintegration of the ecological community. *Am Nat* 172(6):741-

601   750.

602   13. Harrison S & Cornell H (2008) Toward a better understanding of the regional causes

603   of local community richness. *Ecol Lett* 11(9):969-979.

604   14. Logue JB, Mouquet N, Peter H, & Hillebrand H (2011) Empirical approaches to

605   metacommunities: a review and comparison with theory. *Tr ecol evol* 26(9):482-491.

606   15. Lindström ES & Langenheder S (2012) Local and regional factors influencing

607   bacterial community assembly. *Environ Microbiol Rep* 4(1):1-9.

608   16. Hanski I (1982) Dynamics of regional distribution - the core and satellite species

609   hypothesis. *Oikos* 38(2):210-221.

610   17. Brown JH (1984) On the relationship between abundance and distribution of species

611   *Am Nat* 124(2):255-279.

612   18.   Gaston KJ, *et al.* (2000) Abundance–occupancy relationships. *J Appl Ecol* 37:39-59.

613   19.   Blackburn TM, Cassey P, & Gaston KJ (2006) Variations on a theme: sources of

614   heterogeneity in the form of the interspecific relationship between abundance and

615   distribution. *J Anim Ecol* 75(6):1426-1439.

616   20.   Heino J & Virtanen R (2006) Relationships between distribution and abundance vary

617   with spatial scale and ecological group in stream bryophytes. *Freshwater Biol*

618   51(10):1879-1889.

619   21.   Östman Ö, *et al.* (2010) Regional invariance among microbial communities. *Ecol Lett*

620   13(1):118-127.

621   22.   Preston FW (1948) The commonness, and rarity, of species. *Ecology* 29(3):254-283.

622   23.   Magurran AE & Henderson PA (2003) Explaining the excess of rare species in

623   natural species abundance distributions. *Nature* 422(6933):714-716.

624   24.   Rappe MS, Vergin K, & Giovannoni SJ (2000) Phylogenetic comparisons of a coastal

625   bacterioplankton community with its counterparts in open ocean and freshwater

626   systems. *FEMS Microbiol Ecol* 33(3):219-232.

627   25.   Sunagawa S, *et al.* (2015) Structure and function of the global ocean microbiome.

628   *Science* 348(6237).

629   26.   Zinger L, Boetius A, & Ramette A (2014) Bacterial taxa–area and distance–decay

630   relationships in marine environments. *Mol Ecol* 23(4):954-964.

631   27.   Levin SA (1974) Dispersion and population interactions *Am Nat* 108(960):207-228.

632   28.   McGeoch MA & Gaston KJ (2002) Occupancy frequency distributions: patterns,

633   artefacts and mechanisms. *Biol Rev* 77(3):311-331.

634   29.   Gotelli NJ (1991) Metapopulation models: the rescue effect, the propagule rain, and

635   the core-satellite hypothesis. *Am Nat*:768-776.

636   30.   Hanski I & Gyllenberg M (1993) Two General Metapopulation Models and the Core-
637   Satellite Species Hypothesis. *Am Nat* 142(1):17-41.

638   31.   Mehranvar L & Jackson DA (2001) History and taxonomy: their roles in the core-
639   satellite hypothesis. *Oecologia* 127(1):131-142.

640   32.   Soininen J & Heino J (2005) Relationships between local population persistence,
641   local abundance and regional occupancy of species: distribution patterns of diatoms in
642   boreal streams. *J Biogeogr* 32(11):1971-1978.

643   33.   Pedros-Alio C (2006) Marine microbial diversity: can it be determined? *Trends*
644   *Microbiol* 14(6):257-263.

645   34.   Sogin ML*, et al.* (2006) Microbial diversity in the deep sea and the underexplored
646   "rare biosphere". *Proc Natl Acad Sci U S A* 103(32):12115-12120.

647   35.   Sloan W*, et al.* (2006) Quantifying the roles of immigration and chance in shaping
648   prokaryote community structure. *Environ Microbiol* 8(4):732-740.

649   36.   Woodcock S, Curtis TP, Head IM, Lunn M, & Sloan W (2006) Taxa-area
650   relationships for microbes: the unsampled and the unseen. *Ecol Lett* 9(7):805-812.

651   37.   Wardle DA, Bardgett RD, Callaway RM, & Van der Putten WH (2011) Terrestrial
652   Ecosystem Responses to Species Gains and Losses. *Science* 332(6035):1273-1277.

653   38.   Lindh MV*, et al.* (2015) Disentangling seasonal bacterioplankton population
654   dynamics by high-frequency sampling. *Environ Microbiol* 17(7):2459-2476.

655   39.   Andersson AF, Riemann L, & Bertilsson S (2010) Pyrosequencing reveals contrasting
656   seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. *ISME J*
657   4(2):171-181.

658   40.   Nelson JD, Boehme SE, Reimers CE, Sherrell RM, & Kerkhof LJ (2008) Temporal
659   patterns of microbial community structure in the Mid-Atlantic Bight. *FEMS*
660   *Microbiol Ecol* 65(3):484-493.

661   41.   Ghiglione JF, *et al.* (2012) Pole-to-pole biogeography of surface and deep marine

662   bacterial communities. *Proc Natl Acad Sci U S A* 109(43):17633-17638.

663   42.   Pommier T, *et al.* (2007) Global patterns of diversity and community structure in

664   marine bacterioplankton. *Mol Ecol* 16(4):867-880.

665   43.   Gibbons SM, *et al.* (2013) Evidence for a persistent microbial seed bank throughout

666   the global ocean. *Proc Natl Acad Sci U S A* 110(12):4651-4655.

667   44.   Barberán A, *et al.* (2014) Why are some microbes more ubiquitous than others?

668   Predicting the habitat breadth of soil bacteria. *Ecol Lett* 17(7):794-802.

669   45.   Swan BK, *et al.* (2013) Prevalent genome streamlining and latitudinal divergence of

670   planktonic bacteria in the surface ocean. *Proc Natl Acad Sci U S A* 110(28):11463-

671   11468.

672   46.   Galand PE, Casamayor EO, Kirchman DL, & Lovejoy C (2009) Ecology of the rare

673   microbial biosphere of the Arctic Ocean. *Proc Natl Acad Sci U S A* 106(52):22427-

674   22432.

675   47.   Paterson JS, Nayar S, Mitchell JG, & Seuront L (2012) A local upwelling controls

676   viral and microbial community structure in South Australian continental shelf waters.

677   *Estuar Coast Shelf S* 96:197-208.

678   48.   Alonso-Saez L, *et al.* (2014) Winter bloom of a rare betaproteobacterium in the Arctic

679   Ocean. *Front Microbiol* 5.

680   49.   Herlemann DP, *et al.* (2011) Transitions in bacterial communities along the 2000 km

681   salinity gradient of the Baltic Sea. *ISME J*.

682   50.   Dupont CL, *et al.* (2014) Functional Tradeoffs Underpin Salinity-Driven Divergence

683   in Microbial Community Composition. *PLoS ONE* 9(2):e89549.

684   51.   Witkowski A, *et al.* (2005) Darss Sill as a biological border in the fossil record of the

685   Baltic Sea: evidence from diatoms. *Quat Int* 130(1):97-109.

686   52.   Amaral-Zettler L, *et al.* (2010) A Global Census of Marine Microbes. *Life in the*

687   *World's Oceans*, (Wiley-Blackwell), pp 221-245.

688   53.   Pommier T, *et al.* (2010) Spatial patterns of bacterial richness and evenness in the

689   NW Mediterranean Sea explored by pyrosequencing of the 16S rRNA. *Aquat Microb*

690   *Ecol* 61(3):221-233.

691   54.   Schmitt S, *et al.* (2012) Assessing the complex sponge microbiota: core, variable and

692   species-specific bacterial communities in marine sponges. *ISME J* 6(3):564-576.

693   55.   Hercos AP, Sobansky M, Queiroz HL, & Magurran AE (2013) *Local and regional*

694   *rarity in a diverse tropical fish assemblage*. Proc R Soc B 280: 2012-2076.

695   56.   Barberan A & Casamayor EO (2010) Global phylogenetic community structure and

696   beta-diversity patterns in surface bacterioplankton metacommunities. *Aquatic Microb*

697   *Ecol* 59(1):1-10.

698   57.   Pinhassi J, *et al.* (2003) Spatial variability in bacterioplankton community

699   composition at the Skagerrak-Kattegat Front. *Mar Ecol-Prog Ser* 255:1-13.

700   58.   Tokeshi M (1992) Dynamics of distribution in animal communities - theory and

701   analysis. *Res Popul Ecol* 34(2):249-273.

702   59.   Economo EP & Keitt TH (2008) Species diversity in neutral metacommunities: a

703   network approach. *Ecol Lett* 11(1):52-62.

704   60.   Sul WJ, Oliver TA, Ducklow HW, Amaral-Zettler LA, & Sogin ML (2013) Marine

705   bacteria exhibit a bipolar distribution. *Proc Natl Acad Sci USA* 110(6):2342-2347.

706   61.   Logares R, *et al.* (2013) Biogeography of bacterial communities exposed to

707   progressive long-term environmental change. *ISME J* 7(5):937-948.

708   62.   Comte J & Del Giorgio PA (2011) Composition influences the pathway but not the

709   outcome of the metabolic response of bacterioplankton to resource shifts. *PLoS One*

710   6(9):e25266.

711   63.   Diaz-Gil C, *et al.* (2014) Spatio-temporal composition and dynamics of zooplankton

712   in the Kalmar Sound (Western Baltic Sea) in 2009-2010. *Boreal Environ Res* 19(4).

713   64.   Legrand C, *et al.* (2015) Interannual variability of phyto-bacterioplankton biomass

714   and production in coastal and offshore waters of the Baltic Sea. *AMBIO* 44(3):427-

715   438.

716   65.   Riemann L, Steward GF, & Azam F (2000) Dynamics of bacterial community

717   composition and activity during a mesocosm diatom bloom. *Applied and Environ*

718   *Microbiol* 66(2):578-587.

719   66.   Quince C, Lanzen A, Davenport RJ, & Turnbaugh PJ (2011) Removing Noise From

720   Pyrosequenced Amplicons. *Bmc Bioinformatics* 12.

721   67.   Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST.

722   *Bioinformatics* 26(19):2460-2461.

723   68.   Team RDC (2014) R: A Language and Environment for Statistical Computing.

724   https://cran.r-project.org/.

725   69.   Jari Oksanen, *et al.* (2010) vegan: Community Ecology Package. R package version

726   1.17-5. http://CRAN.R-project.org/package=vegan.

727   70.   Wickham H (2009) *ggplot2: elegant graphics for data analysis* (Springer, New York).

728

729   **Figures and Tables**

730   **Figure 1.** Average relative abundance (normalized sequence abundance) against average

731   occupancy. Color denotes standard error in occupancy where red is high and blue is low

732   variation. Size denotes standard error in relative abundance where large size is high and small

733   size is low variation. By moving towards left on the x-axis OTUs display satellite

734   characteristics and towards right OTUs display core characteristics. Examples of OTUs

735   characterized as core, satellite, transient, continuously rare and microbial rain populations are

736   provided in the graph. Red line is smoothing curve (LOESS). Grey line indicates relative

737   abundance above 1 (log -4.6). OTU numbers are given in parenthesis and results from this

738   analysis is further detailed in Table S2.

739

740   **Figure 2.** Occupancy-frequency distribution of OTUs in our spatiotemporal study of the

741   Baltic Sea Proper (Western Gotland basin and Kalmar sound). Filled yellow circles indicate

742   relative sequence abundance (percentage of total sequences in each sample). Asterisks denote

743   significance levels ($*P<0.05$, $**P<0.01$, $***P<0.001$) for Mitchell-Olds & Shaw's test. NS

744   denote non-significant bimodal patterns. The maximum number of sites sampled each month

745   and at each station (i.e. maximum occupancy) is given by the x-axes.

746

747   **Figure 3.** Colonization (C) and extinction (E) rates against occupancy (P). Colonization rates

748   (A), and extinction rates (B). Each point represents the change in fraction of sites occupied

749   from one month to the next for individual OTUs, in total 4437 OTUs for all months. The red

750   line indicates mean and the red shaded area around the mean is standard error. Blue dashed

751   and brown lines indicate the curve of observed data fitted by non-linear least squares to the

752   metapopulation model by Hanski (1982) and Levin (1974), respectively, (Table S1 and

753   Supporting information). Examples are from April, June and September from both years.

754   OTU positions are jittered to reduce overplotting and colored in grey scale according to OTU

755   density.

756

757   **Figure 4.** Examples of occupancy-frequency distribution of OTUs from global ICoMM

758   datasets displaying significant bimodal patterns (A), examples not showing significant

759   bimodality unless applying the analysis on a subset of samples within datasets (B). Oceanic

760   regions are abbreviated; ACB – Arctic Chukchi Beaufort, DAO – Deep Arctic Ocean and

761   NZS – New Zeeland Sediment, SSD – Spatial Scaling Diversity (Cape Cod Atlantic Ocean),

762   BMO – Blanes Bay Microbial Observatory (North West Mediterranean Sea), SPO – Sponge

763   Bacteria (Great Barrier Reef). Please note break in axes for the leftmost bar in each graph.

764   Asterisks denote significance levels (*P<0.05, **P<0.01, ***P<0.001) for Mitchell-Olds &

765   Shaw's test. NS denote non-significant bimodal patterns.

766

767   **Figure 5.** Boxplot of average Bray-Curtis distances from each ICoMM dataset and non-

768   significant or significant bimodality (A), the relationship between average Bray-Curtis

769   distance and the ratio between core and satellite populations (percentage of OTUs occupying

770   all compared to one site) (B), and the relationship between average Euclidean distances of

771   environmental variables, geographic distance and non-significant or significant bimodality

772   (C-F). Open circles in (B) denote monthly Baltic Sea Proper samples collected in this study

773   that were added independently of the regression and polynomial analyses. Asterisks denote

774   significance levels (*P<0.05, **P<0.01, ***P<0.001) for two sample t-tests.

775

776

31

777 **Table 1.** Bimodality in global ICoMM data of marine microbes. Regions with >5

778 spatially different sites (horizontal or vertical) were selected. For each ICoMM dataset,

779 we have included the number of sites, total number of OTUs and the time frame

780 within which the samples were collected, with p-values for Mitchell-Olds & Shaw's

781 test of quadratic extremes. We further provide percentages of OTUs occupying 1 site,

782 intermediary sites (i.e. average percentage of OTUs occupying 2 or more sites but not

783 all) and all sites. The ratio between core (i.e. OTUs occupying all sites) and satellite

784 (i.e. OTUs occupying 1 site) populations is also displayed. For comparison, average

785 Bray-Curtis distances for each dataset is also provided.

ICoMM ID	No. of Sites	Total No. of OTUs	Time frame (mo)	Mitchell-Olds & Shaw's test	1 site %	Intermediary %	All sites %	Ratio Core vs. Satellite	Avg. Bray-Curtis distance
ABR	8	1817	< 3	0,56	68.52	5.10	0.88	0.01	0.66
ACB	11	2098	> 3	**0.001**	63.68	3.69	3.15	0.05	0.38
ALR	10	2885	< 3	0.11	66.55	4.17	0.10	0.00	0.68
AOT	16	4223	< 3	**0.001**	58.11	2.91	1.16	0.02	0.38
ASA	14	1621	< 3	**0.00005**	54.60	3.68	1.23	0.02	0.52
ASV	16	10424	> 3	0.85	80.36	1.40	0.03	0.00	0.79
AWP	13	4471	< 3	0.92	57.95	3.77	0.60	0.01	0.50
BMO	16	3329	< 3	0.06	57.28	3.01	0.63	0.01	0.58
BSP	8	1387	> 3	**0.02**	56.02	6.54	4.76	0.08	0.42
CAM	16	2910	> 3	**0.0002**	64.57	2.41	1.68	0.03	0.40
CAR	14	3449	> 3	**0.04**	52.91	3.76	1.91	0.04	0.33
CCB	8	6956	< 3	0.15	65.45	5.63	0.78	0.01	0.58
CFU	7	10211	> 3	0.18	86.06	2.79	0.01	0.00	0.62
CMM	9	9837	> 3	**0.03**	62.75	5.24	0.58	0.01	0.49
CRS	8	7896	< 3	**0.01**	72.91	4.42	0.56	0.01	0.46
DAO	13	3364	< 3	**0.001**	42.93	4.74	4.96	0.12	0.19
FIS	16	19505	< 3	0.74	58.64	2.94	0.23	0.00	0.47
GMS	10	8974	> 3	0.10	77.72	2.78	0.01	0.00	0.60
GOA	16	3918	> 3	**0.009**	62.66	2.59	1.07	0.02	0.38
HCW	16	3807	> 3	**0.04**	62.10	2.63	1.10	0.02	0.45
ICR	13	16315	> 3	0.20	73.61	2.40	0.01	0.00	0.72
KNX	8	2688	< 3	**0.01**	54.06	6.76	5.39	0.10	0.21
LCR	27	16710	> 3	0.93	74.64	1.01	0.00	0.00	0.74
NADP	6	3883	< 3	0.03	74.56	5.49	3.50	0.05	0.23
NADW	48	12317	> 3	0.31	65.49	0.73	0.00	0.00	0.45
NZS	16	31129	< 3	**0.0003**	59.25	2.80	1.61	0.03	0.18
ODP	8	3935	< 3	0.43	76.87	3.74	0.69	0.01	0.38
SMS	16	12286	< 3	**0.005**	50.09	3.46	1.51	0.03	0.30
SMT	40	13961	> 3	0.22	60.53	1.04	0.04	0.00	0.70
SPO	16	5295	< 3	0.06	67.74	2.30	0.06	0.00	0.69
SSD	8	4523	< 3	**0.006**	63.63	5.25	4.84	0.08	0.47
VAG	16	19750	> 3	**0.01**	58.10	2.93	0.83	0.01	0.39

786   **Supplementary Figures and Tables**

787   **Figure S1.** Map showing the 80-km transect in the Baltic Sea Proper, adapted from

788   Legrand et al. (63, 64). This transect was performed in the PLANFISH and

789   EcoChange frameworks, (see e.g. (63, 64)).

790

791   **Figure S2.** Cluster analysis for comparing beta diversity calculated from Bray-Curtis

792   distance estimation based on 454 pyrosequencing data using 97% 16S rRNA

793   sequence similarity. Color of clusters indicates seasonal communities. Symbols

794   indicate coastal, middle or open ocean sites.

795

796   **Figure S3.** Baltic Sea Proper species rank–abundance curves for separate samples (A),

797   and for all samples collectively (B). All were log-normal, with a few dominant OTUs

798   and a tail of rare OTUs.

799

800   **Figure S4.** A Baltic Sea transect covering the entire 2000 km salinity gradient (A),

801   with the occupancy-frequency distribution of OTUs for all sites of the transect (B),

802   and the occupancy-frequency distribution of OTUs for brackish water sites only (i.e.

803   salinity <15). Asterisks denote significance levels (*$P<0.05$, **$P<0.01$, ***$P<0.001$)

804   for Mitchell-Olds & Shaw's test. NS denote non-significant bimodal patterns.

805

806   **Figure S5.** Rarefaction curves for each sample collected in the Baltic Sea Proper.

807

808

809   **Table S1.** Non-linear least squares analysis and statistics of observed colonization (C)

810   and extinction (E) rates fitted with Hanski's model (16) of $C=cP(1-P)$ and $E=eP(1-P)$.

811   Asterisks denote significance levels (*P<0.05, **P<0.01, ***P<0.001) for NLS tests.

Year	Month	Colonization (C)			Extinction (E)		
		Parameter Estimate	p	Residual std. Error	Parameter Estimate	p	Residual std. Error
2010	April	0.23	***	0.06	-0.77	***	0.05
	Early May	0.39	***	0.08	-0.75	***	0.04
	Late May	0.31	***	0.07	-0.76	***	0.04
	June	0.44	***	0.12	-0.85	***	0.06
	September	0.28	***	0.11	-0.93	***	0.08
2011	April	0.21	***	0.06	-0.86	***	0.06
	May	0.28	***	0.11	-1.06	***	0.07
	June	0.11	***	0.07	-1.05	***	0.06
	July	0.25	***	0.09	-1.00	***	0.06
	August	0.35	***	0.09	-0.74	***	0.05
	September	0.32	***	0.09	-0.82	***	0.05

812

813

814   **Table S2.** Non-linear least squares analysis and statistics of observed extinction (E)

815   rates fitted with Levin's model (27) of $E=eP$. Asterisks denote significance levels

816   (*P<0.05, **P<0.01, ***P<0.001) for NLS tests.

Year	Month	Parameter Estimate	p	Residual std. Error
2010	April	-0.27	***	0.09
	Early May	-0.75	***	0.09
	Late May	-0.24	***	0.09
	June	-0.29	**	0.10
	September	-0.33	**	0.12
2011	April	-0.34	**	0.12
	May	-0.46	**	0.12
	June	-0.44	**	0.12
	July	-0.47	**	0.12
	August	-0.31	**	0.09
	September	-0.31	**	0.10

817

818 **Table S3.** Classification of taxa into "core", "transient", "continuously rare", "satellite", "microbial rain" and "microbial evanescence"

819 populations. Times detected of each individual OTU with times abundant (i.e. >1%), common (i.e. 0.1%-1%), rare (<0.1%), and average

820 occupancy (P) in percentage. Populations with high average relative abundance > 1% and high occupancy with little variance in occupancy and

821 relative abundance were classified as "core". Populations with high average relative abundance and high variance in occupancy and relative

822 abundance were classified as "transient". Populations with low relative abundance but high occupancy and little variance in relative abundance

823 and occupancy were classified as "continuously rare". Populations with low average relative abundance and low occupancy were classified as

824 "satellite". "Microbial rain" and "Microbial evanescence" populations have large variance in occupancy, but have at one or more occasions gone

825 from being undetected at all sites to being present at all sites, and vice versa, respectively. Asterisk (*) indicates if an OTU has been part of the

826 rightmost bar in the bimodal analysis (Fig. 2) at least once.

827

Classsification	OTU	Taxa	Phyla/Class	Detected	Abundant	Common	Rare	Average P [%]
Core	PF_000000	CL500-29*	Actino.	117	98	18	1	100
	PF_000008	hgcI clade*	Actino.	115	94	18	3	87
	PF_000002	SAR11 clade*	Alpha.	117	92	18	7	100
	PF_000009	NS3a*	Bact.	117	106	10	1	100
	PF_000001	Synechococcus*	Cyano.	117	93	17	7	100
	PF_000007	SAR86*	Gamma.	117	96	19	2	100
Transient	PF_000021	Sporichthyaceae*	Actino.	115	48	59	8	97
	PF_000027	hgcI clade*	Actino.	111	66	37	8	81
	PF_000046	CL500-29*	Actino.	109	57	42	10	63
	PF_000062	hgcI clade*	Actino.	95	10	47	38	79
	PF_000047	Cyclobacteriaceae	Bact.	116	28	85	3	64
	PF_000030	NS11-12*	Bact.	112	60	43	9	84
	PF_000032	NS5*	Bact.	100	38	34	28	74
	PF_000035	Synechococcus*	Cyano.	98	30	41	27	82
	PF_000017	SAR92 clade*	Gamma.	100	30	42	28	97
	PF_000011	Candidatus Spartobacteria Baltica*	Verruco.	96	32	34	30	77
Cont. Rare	PF_000051	CL500-29	Actino.	91	0	0	91	95
	PF_000176	CL500-29 marine group	Actino.	78	3	0	75	51
	PF_000254	GOBB3-C201	Alpha.	89	0	3	86	67
	PF_000145	SAR116 clade	Alpha.	76	0	4	72	76
	PF_000068	Reichenbachiella	Bact.	88	0	8	80	75
	PF_000073	Owenweeksia	Bact.	84	0	6	78	53
	PF_000084	Flavobacterium	Bact.	82	1	5	76	55
	PF_000187	Thiobacillus	Beta.	84	0	5	79	54
	PF_000053	Candidatus Methylacidiphilum	Verruco.	82	0	5	77	52
Satellite	PF_012272	Thermoplasmatales	Archaea	1	0	0	1	2
	PF_001083	Brevundimonas	Alpha.	1	0	0	1	1
	PF_001540	Rheinheimera	Gamma.	1	0	0	1	1
	PF_001590	Comamonadaceae	Beta.	1	0	0	1	4
	PF_000553	Sulfurimonas	Epsilon.	1	0	0	1	3
	PF_002803	SAR11 clade	Alpha.	1	0	0	1	1
	PF_003406	SAR86 clade	Gamma.	1	0	0	1	1
	PF_005381	Anabaena	Cyano.	1	0	0	1	2
	PF_006004	hgcI clade	Actino.	1	0	0	1	2

|	PF_008706	*Pedobacter*	Bact.	1	0	0	1	1
Microbial rain	PF_000153	Uncl. *Roseobacter**	Alpha.	51	1	40	10	48
	PF_000126	*Brevundimonas**	Alpha.	6	0	1	5	9
	PF_000171	*Fluviicola**	Bact.	42	3	17	22	41
	PF_000722	NS5 marine group*	Bact.	17	0	2	15	17
	PF_000812	*Owenweeksia**	Bact.	11	0	0	11	15
	PF_000215	NS11-12 marine group*	Bact.	18	0	12	6	32
	PF_000294	VC2.1 Bac22*	Bact.	8	0	3	5	2
	PF_001042	Uncl. Deltaprotebacterium*	Delta.	4	0	0	4	7
	PF_000116	Uncl. Gammaproteobacterium*	Gamma.	14	0	5	9	19
	PF_000263	FukuN18 freshwater group*	Verruco.	34	0	16	18	33
Microbial evanescence	PF_000199	AEGEAN-169 marine group	Alpha.	35	4	27	4	16
	PF_000103	Uncl. *Roseobacter*	Alpha.	11	1	4	6	40
	PF_000383	SAR116 clade	Alpha.	21	0	8	13	20
	PF_000118	*Flavobacteriales*	Bact.	27	1	15	11	50
	PF_000183	*Leadbetterella*	Bact.	40	0	23	17	28
	PF_000313	*Owenweeksia*	Bact.	38	0	15	23	15
	PF_000087	Uncl. Bacteroidetes	Bact.	103	19	69	15	47
	PF_000143	*Synechococcus*	Cyano.	23	2	10	11	43
	PF_000060	*Synechococcus*	Cyano.	48	5	32	11	37
	PF_000091	*Halomonas*	Gamma.	21	3	6	12	77

828
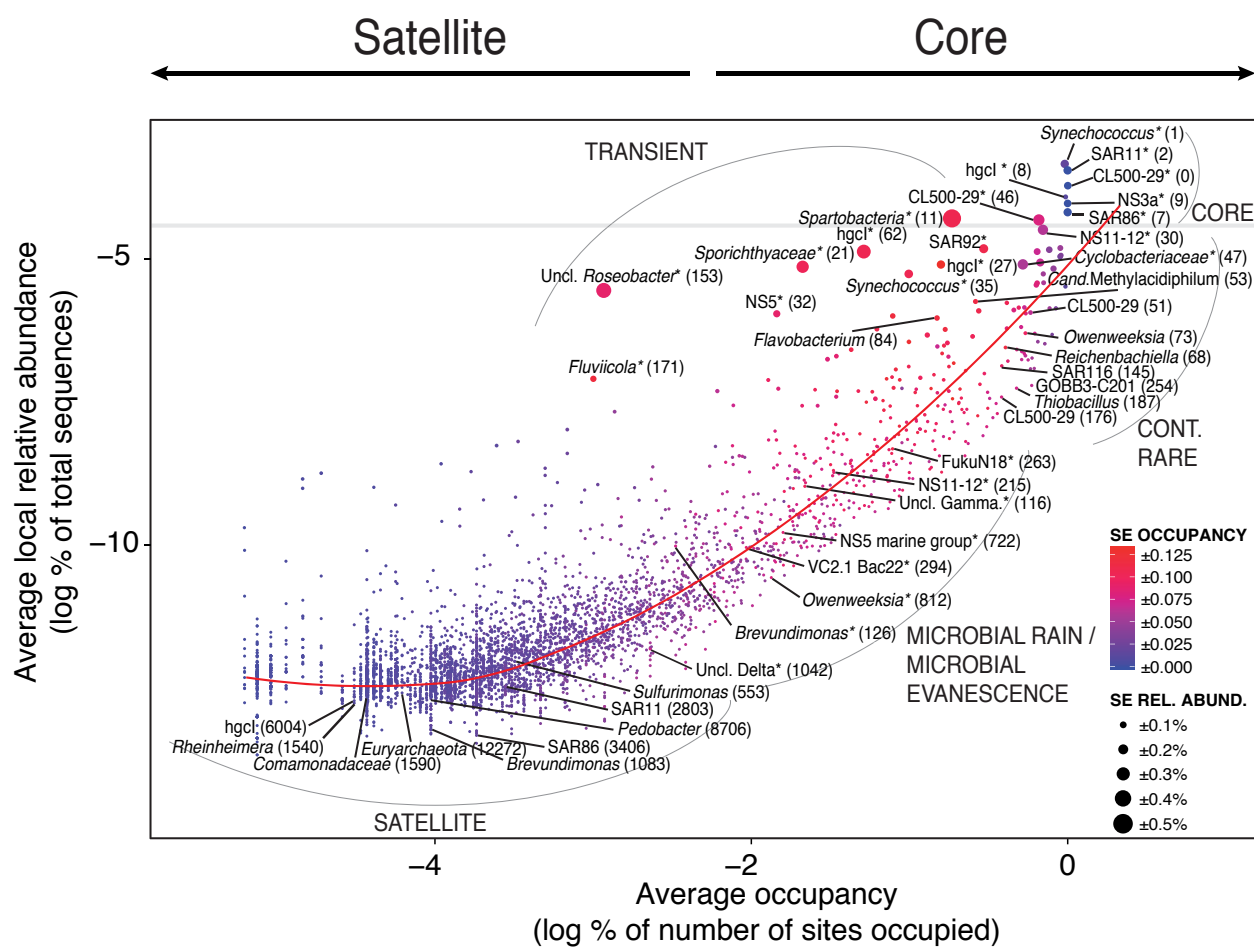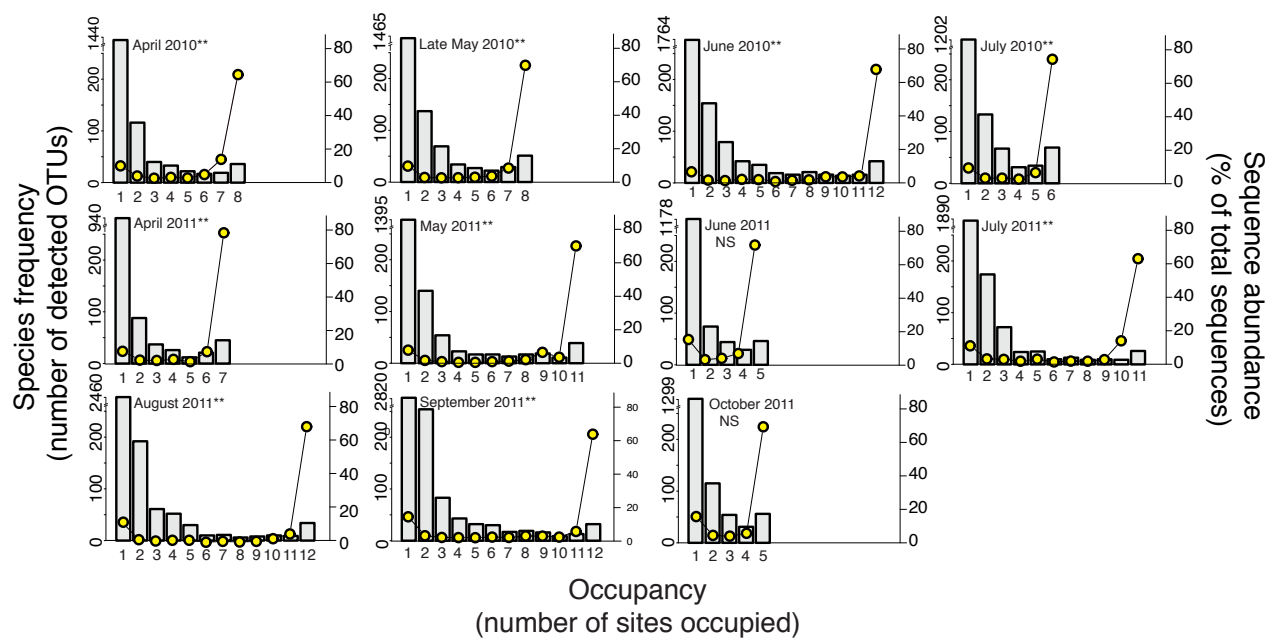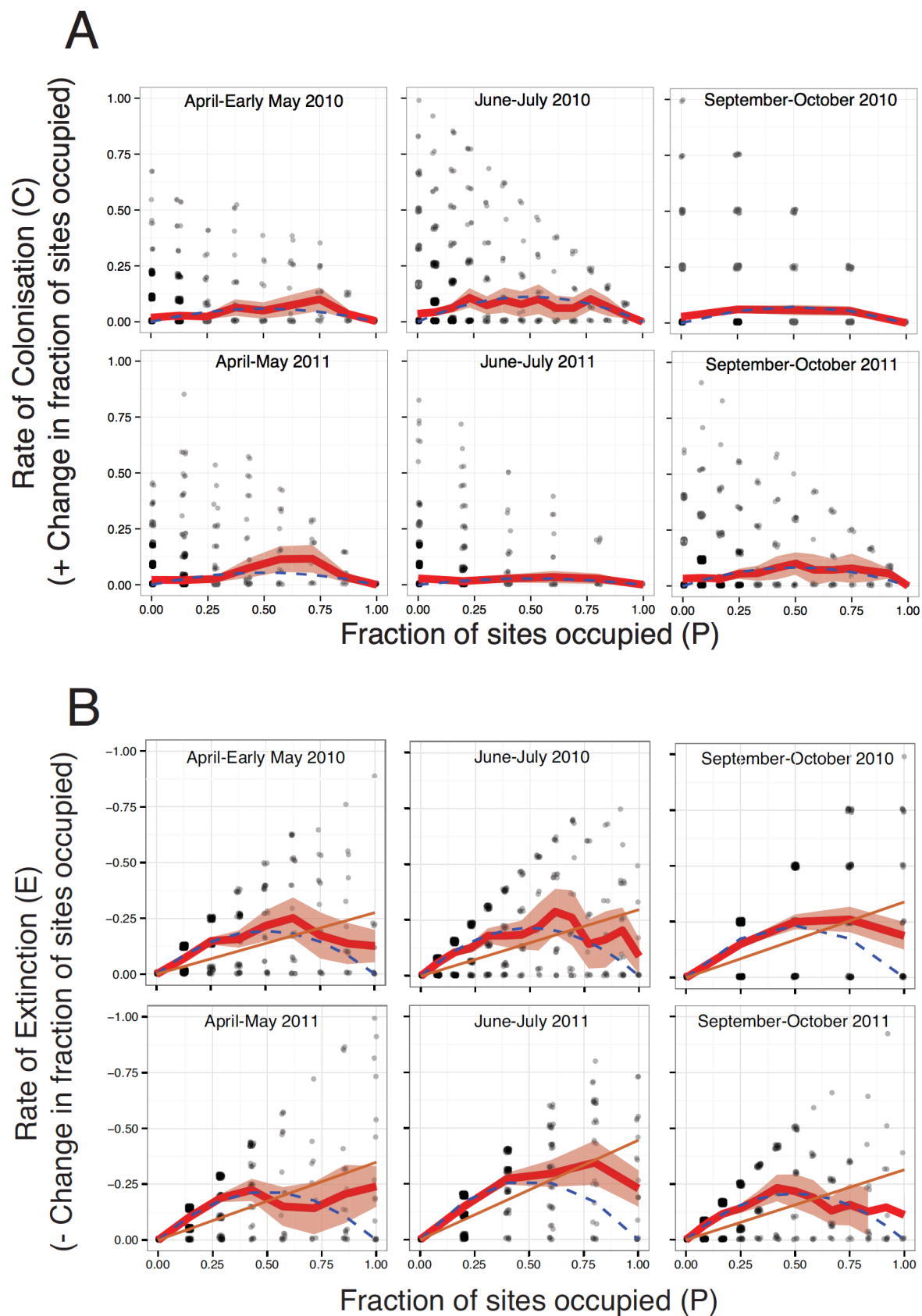
Figure 1

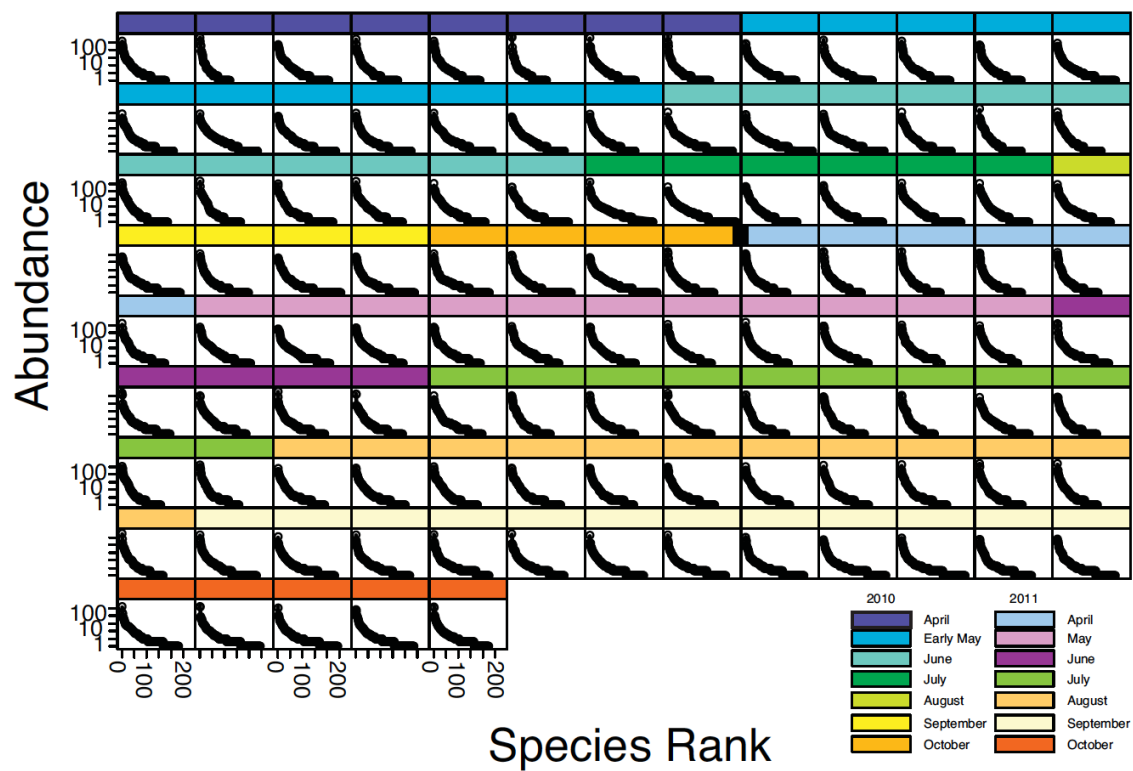*Figure 2*

Figure 3
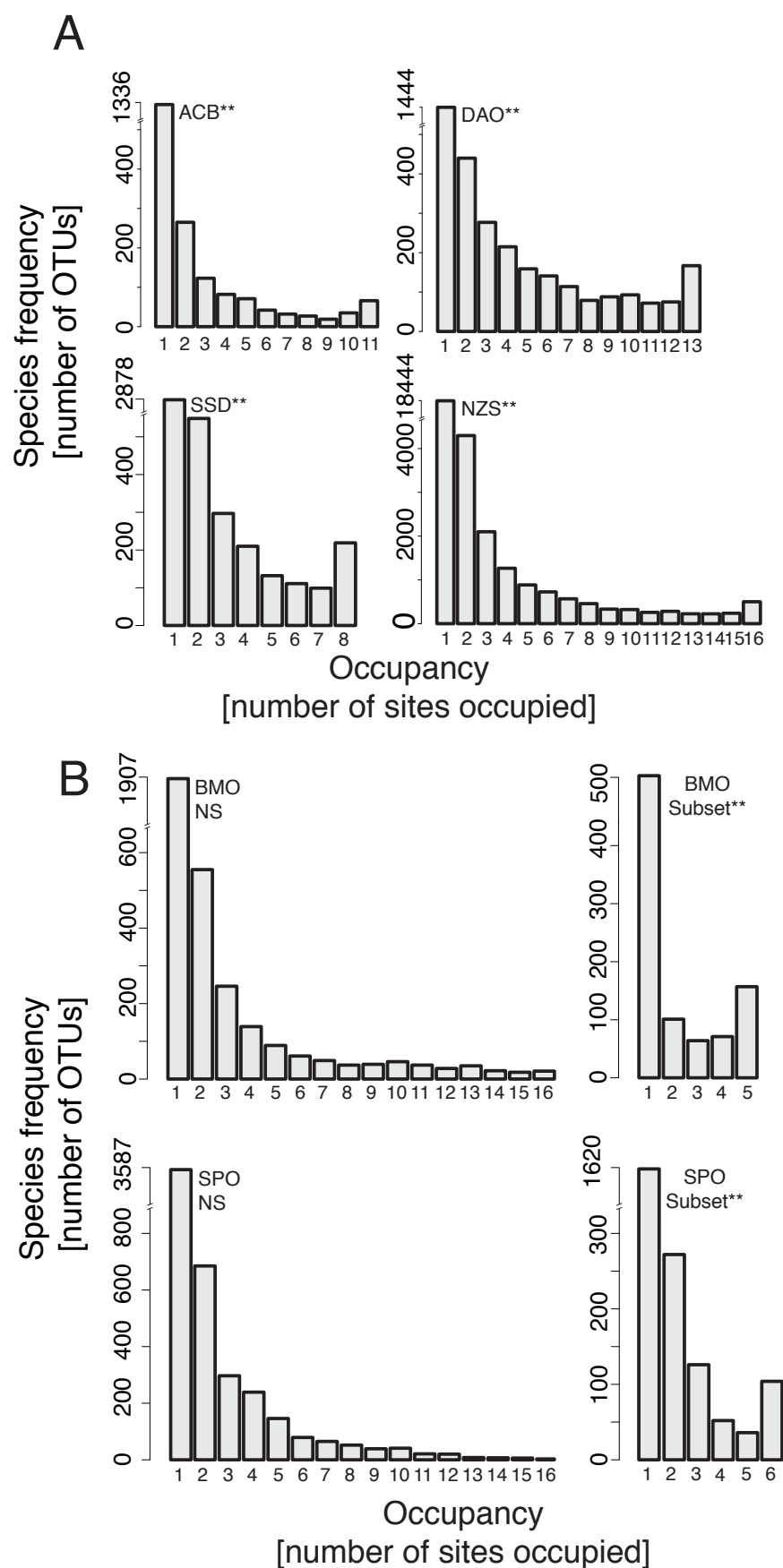
Figure 4
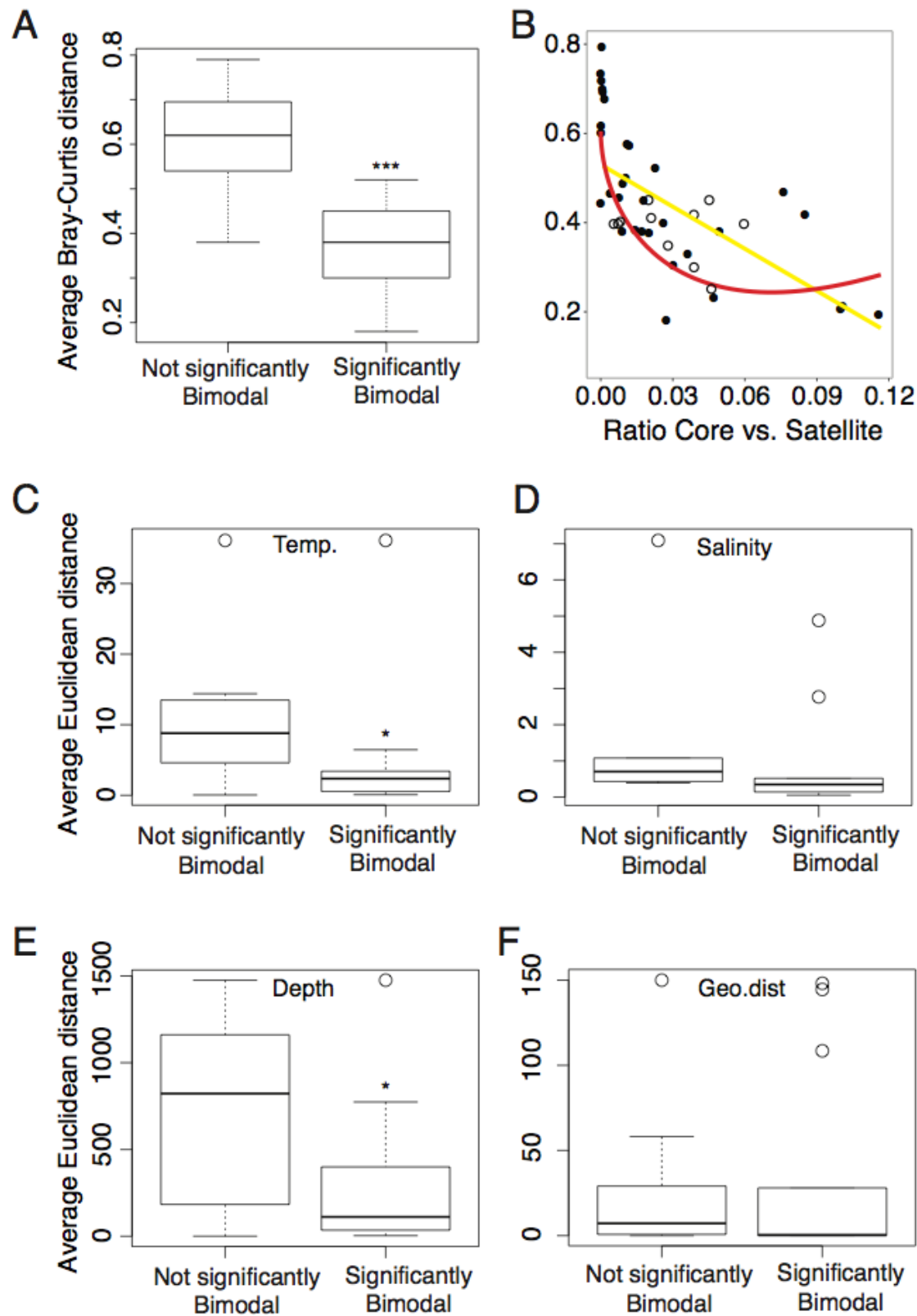
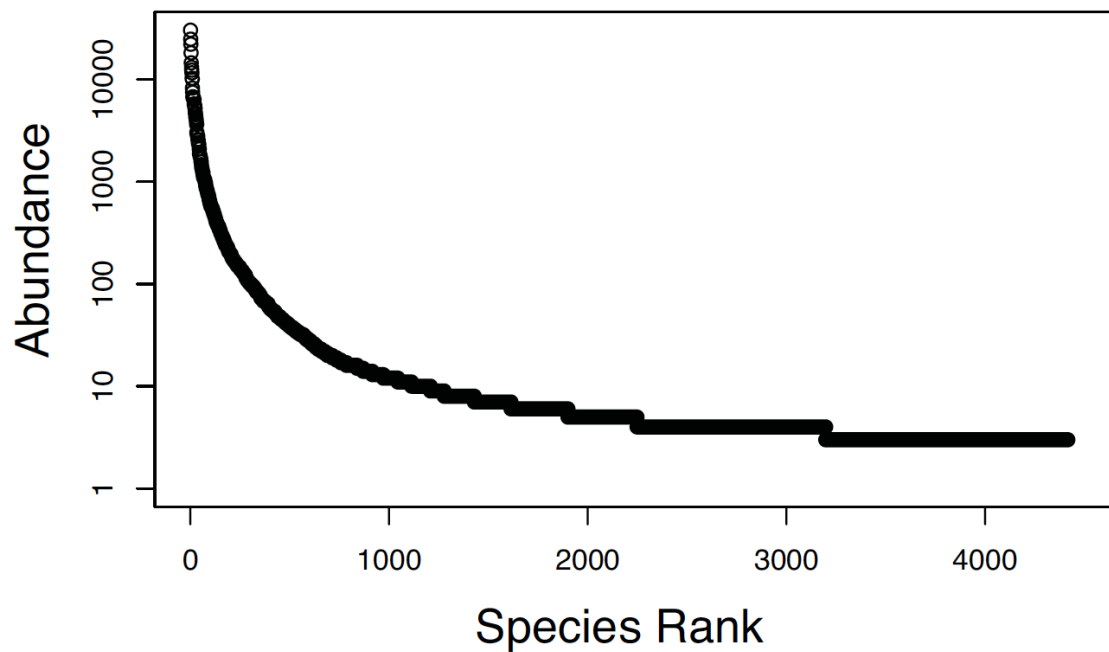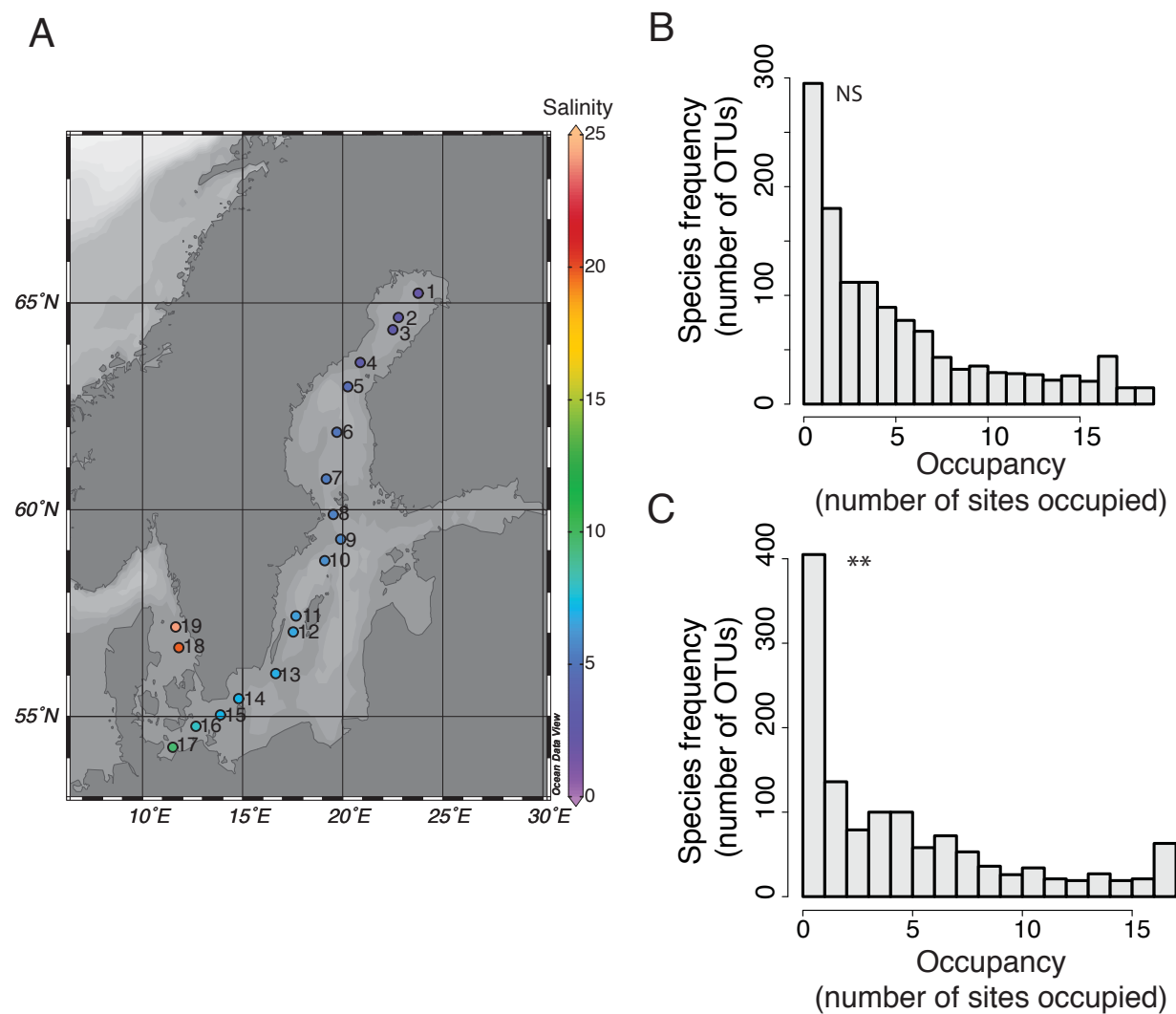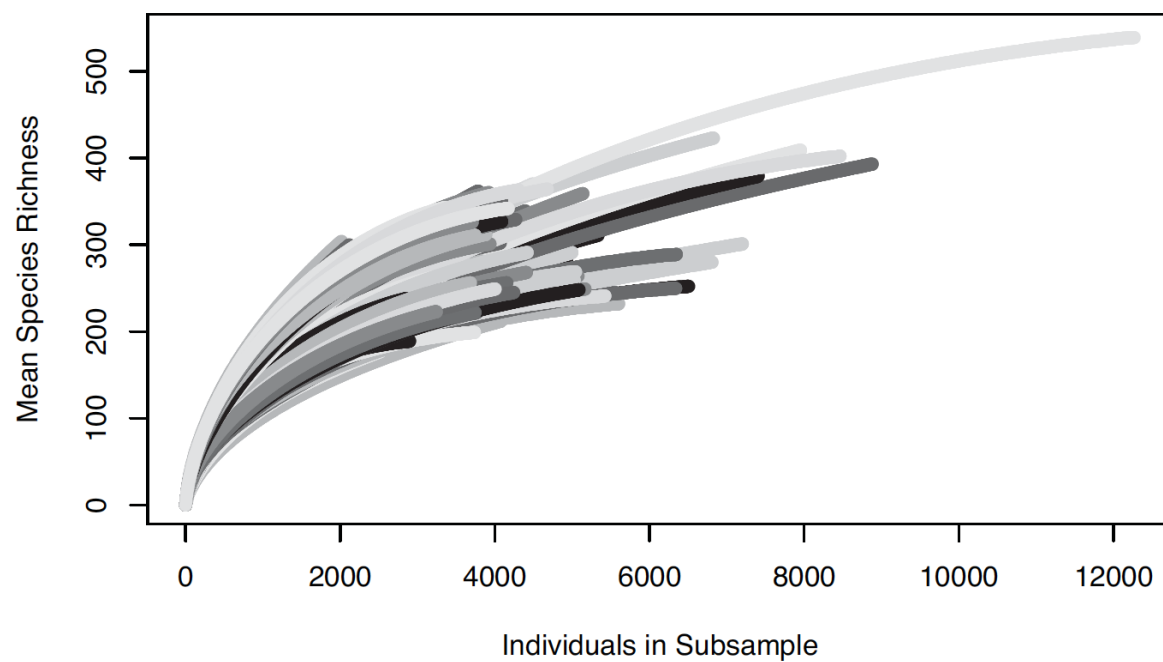*Figure 5*
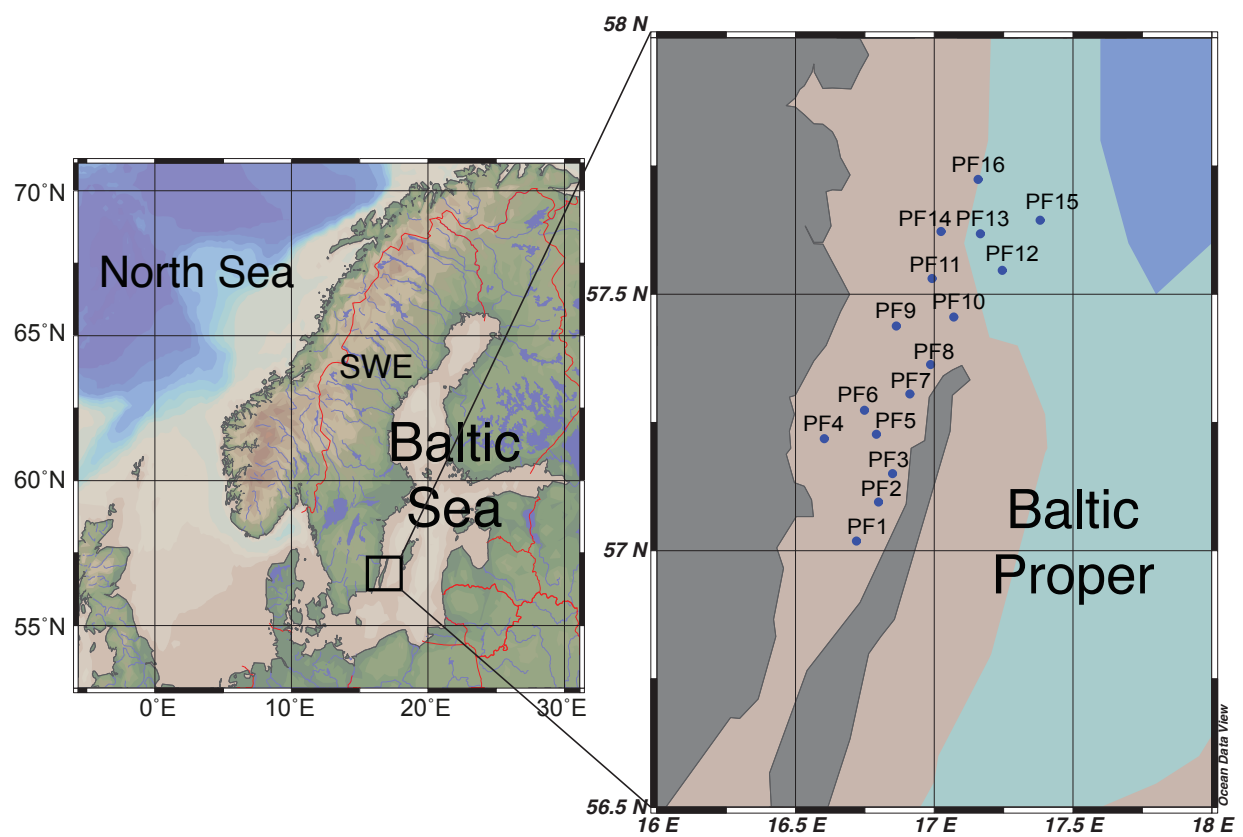
Supplementary Figure 1
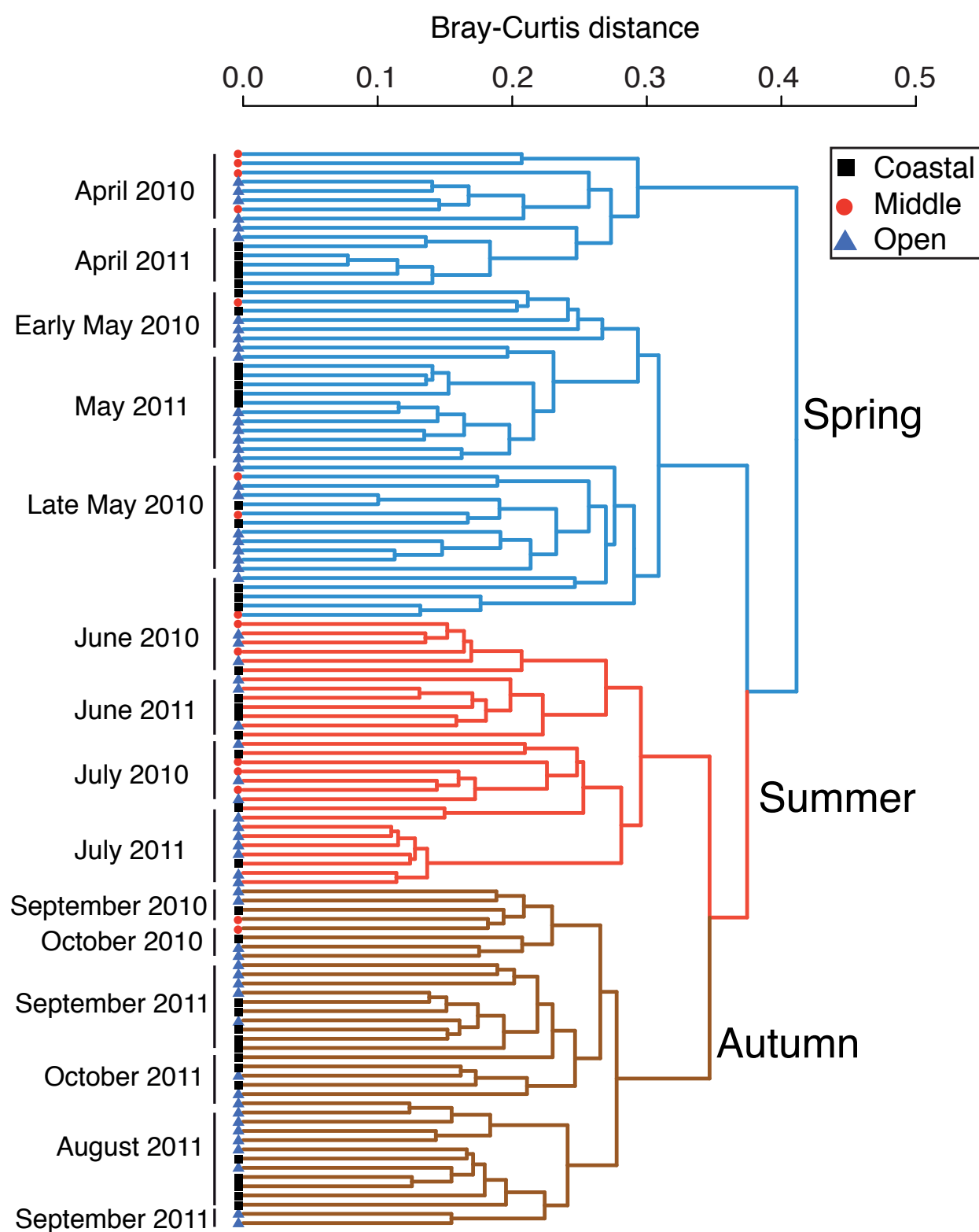
Supplementary Figure 2

Supplementary Figure 3

Supplementary Figure 4

Supplementary Figure 5