



HAL
open science

Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria

Gregory K. Farrant, Hugo Doré, Francisco M. Cornejo-Castillo, Frédéric Partensky, Morgane Ratin, Martin Ostrowski, Frances D. Pitt, Patrick Wincker, David J. Scanlan, Daniele Iudicone, et al.

► **To cite this version:**

Gregory K. Farrant, Hugo Doré, Francisco M. Cornejo-Castillo, Frédéric Partensky, Morgane Ratin, et al.. Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 2016, 113 (24), pp.E3365-E3374. 10.1073/pnas.1524865113 . hal-01331214

HAL Id: hal-01331214

<https://hal.sorbonne-universite.fr/hal-01331214>

Submitted on 13 Jun 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 Classification: BIOLOGICAL SCIENCES

2

3 **Delineating ecologically significant taxonomic units from global patterns of marine**
4 **picocyanobacteria**

5 Gregory K. Farrant^{1,2†}, Hugo Doré^{1†}, Francisco M. Cornejo-Castillo³, Frédéric Partensky¹, Morgane
6 Ratin¹, Martin Ostrowski⁴, Frances D. Pitt⁵, Patrick Wincker⁶, David J. Scanlan⁵, Daniele Iudicone⁷, Silvia
7 G. Acinas³ and Laurence Garczarek¹

8 ¹Sorbonne Universités, UPMC Université Paris 06, CNRS, UMR 7144, Station Biologique, CS 90074, Roscoff,
9 France. ²Present address: Matis Ltd., Food Safety, Environment, and Genetics, Reykjavík, Iceland. ³Department
10 of Marine Biology and Oceanography, Institute of Marine Sciences (ICM), CSIC, Barcelona ES-08003, Spain.
11 ⁴Macquarie University, Department of Chemistry and Biomolecular Sciences, Sydney, Australia; ⁵University of
12 Warwick, School of Life Sciences, Coventry CV4 7AL, UK; ⁶Commissariat à l'Energie Atomique et aux Energies
13 Alternatives (CEA), Institut de Génomique, Genoscope, 91057 Evry, France. ⁷Stazione Zoologica Anton Dohrn,
14 80121 Naples, Italy.

15

16 †These authors contributed equally to this work

17

18 Correspondence to: laurence.garczarek@sb-roscoff.fr

19

20 Keywords: biodiversity, next-generation sequencing, *Tara* Oceans, cyanobacteria, *Prochlorococcus*,
21 *Synechococcus*, metagenomics, miTags, molecular ecology.

22

23 Submitted to: Proceedings of the National Academy of Sciences of the USA

24

25 \abstract

26 *Prochlorococcus* and *Synechococcus* are the two most abundant and widespread phytoplankton in the
27 global ocean. In order to better understand the factors controlling their biogeography, a reference
28 database of the high resolution taxonomic marker *petB*, encoding cytochrome *b₆*, was used to recruit
29 reads out of 109 metagenomes from the *Tara Oceans* expedition. An unsuspected novel genetic
30 diversity was unveiled within both genera, even for the most abundant and well-characterized clades,
31 and 136 divergent *petB* sequences were successfully assembled from metagenomic reads, significantly
32 enriching the reference database. We then defined Ecologically Significant Taxonomic Units (ESTUs),
33 i.e. organisms belonging to the same clade and occupying a given oceanic niche. Three major ESTU
34 assemblages were identified along the cruise transect for *Prochlorococcus* and eight for
35 *Synechococcus*. While *Prochlorococcus* HLIIIA and HLIVA ESTUs co-dominated in iron-depleted areas of
36 the Pacific Ocean, CRD1 and the yet-to-be cultured EnvB were the prevalent *Synechococcus* clades in
37 this area, with three different CRD1 and EnvB ESTUs occupying distinct ecological niches with regard
38 to iron availability and temperature. Sharp community shifts were also observed over short geographic
39 distances, e.g. around the Marquesas Islands or between southern Indian and Atlantic Oceans,
40 pointing to a tight correlation between ESTU assemblages and specific physico-chemical parameters.
41 Together, this study demonstrates that there is a previously overlooked ecologically meaningful fine-
42 scale diversity within some currently defined picocyanobacterial ecotypes, bringing novel insights into
43 the ecology, diversity and biology of the two most abundant phototrophs on Earth.

44

45 **Significance**

46 Metagenomics has become an accessible approach to study complex microbial communities thanks to
47 the advent of high-throughput sequencing technologies. However, molecular ecology studies often
48 face interpretation issues, notably due to the lack of reliable reference databases for assigning reads
49 to the correct taxa and use of fixed cut-offs to delineate taxonomic groups. Here, we considerably
50 refined the phylogeography of marine picocyanobacteria, responsible for about 25% of global marine
51 productivity, by recruiting reads targeting a high resolution marker from *Tara Oceans* metagenomes.
52 By clustering lineages based on their distribution patterns, we showed that there is significant diversity
53 at a finer resolution than the currently defined 'ecotypes', which is tightly controlled by environmental
54 cues.

55 \body

56 **Introduction**

57 The ubiquitous marine picocyanobacteria *Prochlorococcus* and *Synechococcus* are major contributors
58 to global chlorophyll biomass, together accounting for a quarter of global carbon fixation in marine
59 ecosystems, a contribution predicted to further increase in the context of global change (1-3). Thus,
60 determining how environmental conditions control their global distribution patterns, particularly at a
61 fine taxonomic resolution (i.e., sufficient to identify lineages with distinct traits), is critical for
62 understanding how these organisms populate the oceans, and in turn contribute to global carbon
63 cycling. The availability of numerous strains in culture and sequenced genomes make
64 picocyanobacteria particularly well suited for cross-scale studies from genes to the global ocean (4).
65 Physiological studies of a range of *Prochlorococcus* strains isolated from various depths and
66 geographical regions, notably revealed the occurrence of genetically distinct populations exhibiting
67 different light or temperature growth optima and tolerance ranges (5, 6). These observations are
68 congruent on the one hand, with the well-known depth partitioning of genetically distinct
69 *Prochlorococcus* populations in the ocean, with high light-adapted (hereafter HL) populations in the
70 upper lit layer and low light-adapted (hereafter LL) populations located further down the water
71 column, and on the other hand, with the latitudinal partitioning between *Prochlorococcus* HLI and HLII
72 clades that are adapted to temperate and tropical waters, respectively (5, 7, 8). For *Synechococcus*,
73 although no clear depth partitioning (i.e., phototypes) has been observed so far, the occurrence of
74 different ‘thermotypes’ has been clearly demonstrated among strains isolated from different latitudes
75 (9, 10). This latter finding agrees well with biogeographical patterns of the most abundant
76 *Synechococcus* lineages, with members of clades I and IV restricted to cold and temperate waters,
77 while clade II populations are mostly found in warm, (sub)tropical areas (11-13). Recently, several
78 studies have shown that iron could also be an important parameter controlling the composition of
79 picocyanobacterial community structure since *Prochlorococcus* HLIII/IV ecotypes (14, 15) and
80 *Synechococcus* clade CRD1 (16, 17) were shown to be dominant within high nutrient-low chlorophyll
81 (HLNC) areas, where iron is limiting. Most of these studies considered members of the same clade —
82 i.e. *Prochlorococcus* clades HLI-VI and LLI-VI or *Synechococcus* clades I-IX, which are congruent between
83 different genetic markers (13, 18-21)— as one ecotype, i.e. a group of phylogenetically related
84 organisms sharing the same ecological niche (4, 22). Yet the use of a high taxonomic resolution marker,
85 the core, single copy *petB* gene encoding cytochrome *b₆*, has revealed different spatially structured
86 populations (subclades) within the major *Synechococcus* clades that were adapted to distinct niches
87 (12), suggesting that the ‘clade’ level might not be the most ecologically relevant taxonomic unit.
88 Moreover, the systematic use of probes and/or PCR amplification might have led to overlook some
89 important genetic diversity, a drawback potentially resulting in a poor assessment of the relative

90 proportion of co-occurring populations at any given station. In this context, the occurrence of a huge
91 microdiversity within wild *Prochlorococcus* populations was recently demonstrated by estimating the
92 genomic diversity within coexisting members of the HLII clade using a large-scale single-cell genomics
93 approach (23). Still, the congruency of phylogenies based on whole genome and internally transcribed
94 spacer (ITS) suggests that ITS ribotype clusters coincide, in most cases, with distinct genomic
95 backbones that would have diverged at least a few million years ago and the relative abundance of
96 which vary through temporal and local adjustments (23). Thus, approaches using a single marker gene
97 remain valid but fine spatial, temporal and taxonomic resolution is required to better understand how
98 divergent picocyanobacterial lineages have adapted to different niches in the global ocean.

99 Here, we analyzed 109 metagenomic samples collected during the 2.5-year *Tara* Oceans
100 circumnavigation (24, 25), a project surveying the diversity of marine plankton that produced nearly
101 eleven times more non-redundant sequences than the previous Global Ocean Sampling (GOS)
102 expedition (14). In order to retrieve taxonomically relevant information for picocyanobacteria and to
103 avoid PCR-amplification biases, reads targeting the high resolution *petB* gene (12) were recruited using
104 a *mi*Tag approach (26). Even though this approach did not give us access to the rare biodiversity, these
105 analyses unveiled a previously unsuspected genetic diversity within both *Prochlorococcus* and
106 *Synechococcus* genera. Clustering based on the distribution patterns of picocyanobacterial
107 communities allowed us to define Ecologically Significant Taxonomic Units (ESTUs), i.e., genetically
108 related subgroups within clades that co-occur in the field. Analyses of the biogeography of ESTU
109 assemblages showed that they were strongly correlated with specific environmental cues, allowing us
110 to define distinct realized environmental niches for the major ESTUs.

111

112 **Results**

113 **Revealing novel picocyanobacterial diversity using *petB*-*mi*Tags and newly assembled sequences.** To
114 evaluate the taxonomic resolution potential of *petB* *mi*Tags, for assessing picocyanobacterial genetic
115 diversity, simulated 100 bp reads (i.e., the minimum size of the *Tara* Oceans merged metagenomic
116 reads) were generated by fragmenting sequences from our reference database (**Datasets 1-2**). This
117 analysis showed that *petB* reads can be assigned reliably at the finest taxonomic level, i.e. subclade
118 (12), over most of the gene length (**Fig. S1**). The *petB*-*mi*Tags approach was therefore applied to the
119 whole *Tara* Oceans transect (66 stations, 109 metagenomes, 20.2 ± 9.9 Gb of metagenomic data per
120 sample). With the exception of the Southern Ocean and its vicinity (TARA_082 to TARA_085) for which
121 no *petB* reads were recruited, picocyanobacteria were present at all sampled *Tara* Oceans stations.
122 From 119 to 14,139 picocyanobacterial *petB* reads (average: 3,309; median: 2,545; **Dataset 3**) were
123 recruited per sample using a non-redundant reference database of 585 high quality *petB* sequences,
124 representing most of the genetic diversity identified so far among *Prochlorococcus* and *Synechococcus*

125 isolates and environmental clone libraries (**Fig. 1**). Interestingly, most *petB* sequences in our database
126 recruited at least one read from the *Tara Oceans* metagenome as best hit, with the notable exception
127 of some sequences of the cold-water adapted *Synechococcus* clade I, likely due to the limited sampling
128 performed at high latitudes during the *Tara Oceans* expedition (27). This suggests that most genotypes
129 known so far are sufficiently well represented in the marine environment to be detected by this
130 approach. Still, we cannot exclude that this preliminary analysis provides a somewhat biased picture
131 of the diversity toward the ‘already known’, since most current reference sequence databases are
132 potentially skewed by culture isolation and/or amplification biases.

133 To search for potential hidden genetic diversity within the *Tara Oceans* picocyanobacterial
134 communities, we then examined the percent identity of recruited reads with regard to their best hit in
135 the *petB* database (**Figs. 2A-B and S2**). *Prochlorococcus* and *Synechococcus petB* sequences can be
136 easily differentiated from non-specific signal by selecting reads above 80 % identity to the closest
137 reference *petB* sequence. The diversity within the most abundant *Synechococcus* clades (I-IV) was
138 generally well covered by reference sequences since most reads displayed >94 % identity to their best-
139 hit in the database, a cut-off value previously shown to allow an optimal separation of *Synechococcus*
140 lineages displaying distinct distribution patterns (12). In contrast, for other clades, some of the
141 recruited reads were quite distantly related to reference sequences (i.e., between 80-94% identity),
142 indicating that the *in situ* diversity of these clades was not fully covered by the reference database (**Fig.**
143 **2B**, top panels).

144 To have a more realistic and exhaustive view of this diversity, we assembled 136 distinct nearly
145 complete *petB* sequences from environmental reads (121 *Prochlorococcus* and 15 *Synechococcus*),
146 corresponding to the most divergent genotypes present in the whole *Tara Oceans* dataset. By adding
147 these novel sequences to the reference database (see **Dataset 1** and sequences in white in **Fig. 1**), we
148 significantly improved taxonomic assignments of *petB*-miTags, since 80.3 % of the *Prochlorococcus* and
149 90.2 % of the *Synechococcus* environmental *petB* reads were found to display >94 % identity with their
150 best hits in the enriched reference database, an increase of about 11 and 7 % compared to our initial
151 assessment, respectively (**Figs. 2B and S2**). Interestingly, quite a few highly divergent sequences from
152 *Prochlorococcus* HLIII, HLIV and LLI as well as *Synechococcus* CRD1 were assembled from TARA_052,
153 located East of Madagascar, a station exhibiting a picocyanobacterial community atypical for this
154 oceanic area (see below). Although most of these additional sequences fell into known phylogenetic
155 clades, they allowed us to better assess the extent of genetic diversity within both *Prochlorococcus*
156 and *Synechococcus* (**Fig. 1**). While only a few *petB* sequences, all coming from cultured strains, were
157 available for the *Prochlorococcus* HLI and LLI clades prior to this study, we added 43 novel HLI
158 sequences (within-clade nucleotide identity range: 87-99.6%), 29 LLI sequences (within-clade identity
159 range: 85.5-99.6%) as well as 11 sequences of the uncultured HLIII and IV clades, some of which form

160 distinct monophyletic branches comprised entirely of novel sequences (**Fig. 1 and Dataset 1**). Although
161 many HLII sequences were recently obtained by high throughput single cell genomics focused on this
162 clade (23), assembly of *Tara* Oceans reads allowed us to retrieve several divergent HLII sequences
163 (within-clade identity range: 86.2-99.8%) including a new, well-supported group (corresponding to
164 ESTU HLIIIC, see below), located at the base of the HLII radiation. Similarly for *Synechococcus*, newly
165 assembled sequences allowed us to refine the taxonomy of several taxa, notably for CRD1 and EnvB
166 clades as well as subcluster 5.3, three ecologically important but previously overlooked phylogenetic
167 lineages.

168

169 **Using global picocyanobacterial distribution patterns to define ESTUs.** As expected from previous
170 literature (1, 2, 5, 28), *Prochlorococcus* was the most abundant picocyanobacterium at the global scale,
171 representing ~91% of all *petB* reads from the bacterial size fraction, compared to 9% for *Synechococcus*
172 (**Fig. S3A**). These percentages compare fairly well with the global contribution of *Prochlorococcus* and
173 *Synechococcus* estimated from flow cytometry data as 80.6% ($2.9 \pm 0.1 \times 10^{27}$ cells) and 19.4 % ($7.0 \pm$
174 0.3×10^{26} cells), respectively (1). The apparent lower contribution of *Synechococcus* in our dataset
175 might be due to the fact that the *Tara* Oceans sampling was not made at random in the ocean, since
176 most stations were located in the inter-tropical zone and/or selected for displaying specific traits of
177 interest (e.g., upwelling, fronts, island proximity, etc.), while Flombaum and coworkers' dataset
178 included many data from temperate stations, where *Synechococcus* is often abundant.

179 To study the global distribution of these organisms at a finer taxonomic resolution, we then examined
180 whether *Prochlorococcus* and *Synechococcus* clades and/or subclades were ecologically meaningful.
181 To do this, we analyzed the distribution patterns along the *Tara* Oceans transect of within-clade
182 Operational Taxonomic Units (OTUs), as defined using a cut-off at 94% nucleotide identity (**Figs. 2C**
183 **and S4 and Dataset 4**). Although for some clades, OTUs displayed a homogeneous pattern over their
184 geographical distribution area (e.g., *Prochlorococcus* HLIII and IV, **Fig. S4**) or were too scarce to reliably
185 distinguish ESTUs (*Synechococcus* subcluster 5.2 and clades I, V-VIII, WPC1, EnvA, IX, XVI, XX, UC-A,
186 *Prochlorococcus* clades LLII-IV), most of the prevalent clades encompassed several coherent OTU
187 clusters displaying distinct distribution patterns (and thus likely occupying distinct ecological niches)
188 that were gathered into independent ESTUs (**Fig. 2C, Fig. S4**). For instance, OTUs within *Synechococcus*
189 clade CRD1 could be split into 3 ESTUs (CRD1A-C) based on clustering of their abundance per station.
190 Some of these ESTUs corresponded to previously described clades (e.g., *Prochlorococcus* HLIIIA and
191 HLIVA) or subclades (e.g., *Synechococcus* IVC), while others gathered subclades having similar
192 distribution patterns. For instance, *Synechococcus* ESTU IIA encompasses subclades IIA-d and IIf and
193 ESTU IIB gathers subclades IIe and IIh, as previously defined by Mazard et al. (12). Thus, although most
194 previous field diversity studies on picocyanobacteria focused on clades (5, 13, 17, 20, 21), which were

195 generally considered as distinct 'ecotypes' (*sensu* (19)), our data indicate that ESTUs provide a finer
196 estimate of *Prochlorococcus* and *Synechococcus* ecotypes than do clades. These ESTUs were then used
197 to study the biogeography of marine picocyanobacteria by clustering together stations exhibiting
198 similar ESTU assemblages (**Figs. 3A and 4A**).

199

200 **Biogeography of *Prochlorococcus* reveals the occurrence of minor ESTUs with unexpected**
201 **distribution patterns.** Most major *Prochlorococcus* clades (HLI, HLII and LLI) could be split into several
202 ESTUs, though for the former two, one ESTU was clearly predominant (**Figs. 3A and S5**). Only three
203 major ESTU assemblages were identified in surface samples: i) dominance of HLIA ESTU in temperate
204 waters (above 35°N and 32°S), ii) dominance of HLIIA in warm and iron-replete waters between 30°S
205 and 30°N, with mixed HLIA-HLIIA profiles at intermediate latitudes and iii) co-occurrence of HLIIIA and
206 IVA at a ratio of ca. 1:2.6 (± 0.7) in warm, high nutrient-low chlorophyll (HNLC) areas. The low
207 abundance of LLII-IV clades in the whole *Tara* Oceans dataset (**Fig. S6A-C**) is likely due to the fact that
208 they usually thrive below the DCM (5, 29), i.e. at depths not sampled during the expedition. In contrast,
209 most LLI ESTUs were very abundant in subsurface waters (**Figs. S3 and S5b**) and sometimes even
210 reached the surface (e.g., at TARA_066-070, **Figs. 3A**), as expected from the ability of members of the
211 LLI clade to tolerate a strong mixing rate and short-term exposure to high light (5, 8, 29, 30).

212 HLIIIA and HLIIA ESTUs altogether contributed to 15.5% of the *Prochlorococcus* community in *Tara*
213 Oceans samples, i.e. about as much as HLI (17%) or LLI (15.2%; **Fig. S3A**). This value is slightly higher
214 than the 9% previously estimated for HLIII-IV clades from the analysis of GOS samples (11). Consistent
215 with previous studies (11, 15, 31, 32), we show here that their distribution covers most of the warm
216 (>25°C), low-Fe equatorial Pacific zone from 13°S (TARA_100) to 14°N (TARA_137), where they
217 constitute the vast majority of the *Prochlorococcus* community in surface waters. In the Indian Ocean,
218 we only observed them at two stations near the northern coast of Madagascar (TARA_052 and
219 TARA_056), in agreement with a previous report that found them at two sites located further east (31),
220 all these sites likely being influenced by the Indonesian throughflow originating from the tropical
221 Pacific Ocean (33). Thus, HLIII/IV seemingly occurs over a much thinner latitudinal band (centered
222 around 15°S) in the Indian compared to the Pacific Ocean, and they are apparently very scarce in the
223 part of the Atlantic Ocean explored by the *Tara* schooner, even though the area around stations
224 TARA_072 and TARA_070 is known to be iron-depleted (see **Fig. S1** in (17)). Altogether, the distribution
225 patterns of the dominant *Prochlorococcus* HL ESTUs seem to be mainly driven by temperature and iron
226 availability, as confirmed by non-metric multidimensional scaling (NMDS) analyses (**Fig. 3C**). These
227 results are globally consistent with previous reports that analyzed *Prochlorococcus* clades (5, 8, 15, 29,
228 31), indicating that the latter studies actually targeted the dominant ESTUs.

229 In contrast, a number of minor ESTUs were found to display distribution patterns very different from
230 the major ESTU of the same clade. For instance, the relative contribution of the above mentioned novel
231 HLIIC ESTU was highest at the DCM in the equatorial Indian Ocean (TARA_041-042; **Fig. S5b**),
232 suggesting that members of this ESTU are adapted to mid-depth waters, much like members of the LLI
233 clade (5, 29). Similarly, ESTUs HLIB and D can sometimes take over the prevalent HLIA populations and
234 become abundant in surface waters at specific locations (e.g., at TARA_093 and TARA_094,
235 respectively). In contrast, HLIC, which comprises a complex microdiversity (10 OTUs; **Fig. S4**), was
236 found to exhibit a particularly large niche, co-occurring with HLIA at high latitude but also being present
237 as the major HLI population in warm oligotrophic waters, where HLIIA dominated the *Prochlorococcus*
238 community (e.g., in the Indian Ocean, **Fig. S6A**). This suggests that members of the HLIC ESTU might
239 have a larger tolerance to temperature than the globally dominant HLIA. It is also worth noting that
240 among the four ESTUs defined within the LLI clade, LLIB, which is entirely comprised of newly
241 assembled *petB* sequences, dominates the LLI population in surface iron-limited HNLC areas in both
242 the equatorial/tropical Pacific (TARA_110 to 128) and Indian Ocean (TARA_052, **Fig. S6B**). Thus,
243 adaptation to low iron conditions in *Prochlorococcus* might not be an exclusive trait of HLIIIA and
244 HLIVA.

245

246 **CRD1 and EnvB ESTUs are the dominant *Synechococcus* lineages in the Pacific Ocean.** *Synechococcus*
247 assemblages were much more diverse than *Prochlorococcus* with 8 distinct ESTU clusters observed
248 along the *Tara* Oceans transect (**Fig. 4A-B**). None of these assemblages were specific of a given oceanic
249 region, though cluster 2 was mainly found in the Mediterranean Sea. ESTUs IA and IVA, IVB and/or IVC
250 dominated at most stations within clusters 4, 5 and 8 that were typical of cold, coastal or mixed open
251 ocean waters at high latitude, in agreement with previous reports on the distribution of clades I and
252 IV (11-13, 17). In contrast, ESTU IIA, dominated by a single OTU (OTU003; **Fig. 2C**), was by far the major
253 component of cluster 1, an assemblage characteristic of most warm, mesotrophic and oligotrophic iron
254 replete waters that encompass the vast majority of the Atlantic and Indian Oceans (**Fig. 4B**).
255 Consistently, NMDS analysis showed that the occurrence of clusters 4, 5, 8 on the one hand, and cluster
256 1 on the other hand, were associated both with temperature and Chl *a*, but in opposite ways (**Figs. 4C**
257 **and S7**). Interestingly, while ESTU IIA was typical of warm waters, the minor ESTU IIB was found to be
258 restricted to fairly cold (14.1 to 17.5°C), mixed waters and to co-occur with IVA-B (**Fig. 4**).

259 Several other salient features arose from analyses of the *Tara* Oceans metagenomes. First, ESTU IIIA,
260 the major contributor of cluster 2, was found only in the Mediterranean Sea (TARA_007 to 030) and
261 the Gulf of Mexico (TARA_142; **Fig. 4A-B**). Both areas are known to be P-depleted (34, 35), suggesting
262 that the dominance of this ESTU could be linked to a specific adaptation to P limitation, as confirmed
263 by the inverse correlation of cluster 2 with P concentrations (**Fig. 4C**) and correlation analyses between

264 IIIA and individual physico-chemical parameters (**Fig. S7**). The differential availability of this nutrient
265 on both sides of the Suez Canal is therefore probably responsible for the strong community shift from
266 a IIIA- to a IIA-dominated assemblage between the Mediterranean and Red Sea (**Fig. S5a**), although
267 one cannot exclude that other specific characteristics of the Mediterranean Sea, such as the presence
268 in the eastern basin of copper, a trace metal toxic to a number of phytoplankton species (36), might
269 also be involved. While the dominance of clade III in the Mediterranean Sea is consistent with previous
270 studies (13, 37), it was also reported in fair abundance along a N-S transect in the northern Atlantic
271 Ocean in fall 2004 (AMT15) as well as in sub-tropical waters of the Pacific and Atlantic oceans (12, 13),
272 whereas we found it only as a minor component of the *Synechococcus* community in these areas. It is
273 possible that the relative contribution of clade III might have been overestimated using PCR-based or
274 dot-blot hybridization approaches. A more likely explanation is that this clade is subject to seasonality,
275 as suggested by a year-round survey in the Red Sea, showing that clade III abundance peaks occur
276 during summer, stratified conditions, and remains at low concentrations over the rest of the year (19,
277 38). In this context, it is important to note that during *Tara* Oceans, the north and south Atlantic as
278 well as the southern Indian Ocean were all sampled during winter or early spring, while the
279 Mediterranean Sea was sampled in fall (**Dataset 3**). Hence, this warrants future global metagenomic
280 studies at various seasons as well as finer-scale studies looking at seasonal variations in community
281 structure.

282 Also unexpected was the large global abundance (6% of total *Synechococcus* reads, Fig. S3) of
283 subcluster 5.3 (formerly clade X; (39)). Members of ESTU 5.3A (mostly co-occurring with ESTU IIIA)
284 were found mostly along the transect from Panama to Bermuda (TARA_140-149), in the Mozambique
285 Channel (TARA_057 and TARA_062) as well as at all stations of the Red Sea and Mediterranean Sea,
286 where they contributed up to ca. 30 % of the local *Synechococcus* community, e.g., at the Gibraltar
287 strait (TARA_007, **Fig. 4A-B**). In contrast, ESTU 5.3B (co-occurring with ESTU IIA) was always present in
288 low relative abundance. Members of subcluster 5.3 have only been sporadically detected in previous
289 studies mostly in open-ocean habitats in the northwestern Atlantic and Pacific Ocean and in the
290 Mediterranean Sea (11-13, 16, 20, 37), reaching significant abundances only in transitional waters,
291 such as the Amazon plume or the Benguela upwelling (17). These specific localizations might explain
292 why only a few sequences of this subcluster were previously detected in the GOS database (11).

293 Another striking result of this study was the strong global contribution of the co-occurring clades
294 CRD1 and EnvB (8.4% and 5.4% of total *Synechococcus* reads, respectively; **Fig. S3D-E**). Recently, low
295 Fe regions of the western equatorial Pacific (5°S-10°N) and southeastern Atlantic Oceans (15-20°S)
296 were shown to be dominated by CRD1 (16, 17), a clade that was previously thought to be specific to
297 the Costa Rica dome, where *Synechococcus* cell densities are known to be the highest worldwide (40,
298 41). Here, we show that CRD1 and EnvB ESTUs actually co-dominate the *Synechococcus* community

299 over most of the Pacific Ocean from 33°S to 35°N and can also be prevalent in both the South
300 (TARA_068-072) and North Atlantic (TARA_150-152) as well as in the Indian Ocean (TARA_052) but are
301 seemingly absent from the Mediterranean Sea (**Fig. 4A-B**). So, it seems that, in contrast to
302 *Prochlorococcus* HLIII/IV, the distribution of CRD1 in the Pacific Ocean extends way beyond HNLC areas.
303 Furthermore, we show here that both the CRD1 and EnvB clades actually encompassed 3 distinct
304 ESTUs, displaying partially overlapping niches and falling into five clusters (3, 5-8; **Fig. 4A**) that were
305 also split far apart by NMDS analyses (**Fig. 4C**). CRD1B and EnvBB were restricted to high latitude, cold,
306 mixed waters (cluster 8), where they systematically co-dominated with ESTU IA, IVA and IVC. This
307 includes TARA_093 located in the Chilean upwelling, TARA_152 in North Atlantic as well as TARA_068
308 in South Atlantic corresponding to a young Agulhas ring (42). In contrast, CRD1C and EnvBC
309 preferentially thrived in warm HNLC regions (cluster 3 and the warmest stations of cluster 6), with
310 CRD1C largely dominating the *Synechococcus* population in the Pacific inter-tropical area as well as at
311 the Indian Ocean station TARA_052. Comparatively, CRD1A and EnvBA that were found in both kinds
312 of environments, appear to be much more ubiquitous and to tolerate a much wider temperature
313 range, not only than other CRD1 and EnvB ESTUs, but also more generally than all other *Synechococcus*
314 strains characterized so far in culture (9, 10). Several previous studies also reported the presence of
315 CRD2, co-occurring with CRD1 mainly in the Costa Rica dome area and in equatorial waters and
316 generally constituting around 10-15 % of the total *Synechococcus* surface population (16, 17). It is
317 tempting to speculate that the *petB*-defined EnvB clade, which had so far only been reported at one
318 station in the middle of the North Atlantic basin (12), corresponds to the ITS-defined CRD2 clade.
319 However, the different proportions of EnvB and CRD2 relative to CRD1 strongly suggests that the qPCR
320 primers used in these studies targeted only a fraction of the CRD2/EnvB population, possibly
321 corresponding to EnvBC, which like CRD2, is positively correlated with temperature ((17) and **Fig. S7**).
322 Alternatively, seasonal variations might also explain the differences observed between these two
323 datasets.

324

325 **Discussion**

326 The comprehensive nature of the *Tara* Oceans dataset, analyzed here at high taxonomic resolution,
327 has markedly improved our current knowledge of the global phylogeography of marine
328 picocyanobacteria, and highlighted the key role of environmental parameters in shaping their
329 distribution patterns. Indeed, by assigning *petB*-miTags recruited for each clade to narrow OTUs, then
330 clustering those sharing a similar ecological distribution into the same ESTU, we showed that despite
331 a wide genetic diversity, *Prochlorococcus* and *Synechococcus* communities can be split into a fairly
332 limited number of characteristic ESTU assemblages, often dominated by one or two major ESTU(s). This

333 includes the co-dominating *Prochlorococcus* HLIIIA-HLIVA, which occurred at a fairly constant ratio
334 (1:2.6) throughout low Fe regions (**Fig. 3A**), *Synechococcus* IIIA that was abundant all over the
335 Mediterranean Sea or CRD1 and EnvB ESTUs, co-dominating the *Synechococcus* community in vast
336 expanses of the Pacific Ocean (**Fig. 4A**). Interestingly, we also showed that most picocyanobacterial
337 clades encompass minor ESTUs that occupy niches distinct from dominant ones. This indicates that
338 there is ecologically meaningful fine-scale diversity within currently defined *Synechococcus* or
339 *Prochlorococcus* clades, even though the latter have often be referred to as 'ecotypes' (5, 29). In this
340 context, it is important to note that the *Prochlorococcus* genus is thought to have occurred
341 concomitantly to the major diversification event that also led to the splitting of *Synechococcus*
342 subcluster 5.1 into about fifteen distinct clades (20, 43, 44), suggesting that, from a phylogenetic point
343 of view, the whole *Prochlorococcus* genus is actually equivalent to a single *Synechococcus* clade,
344 explaining why linking clades to a given ecological niche is trickier for the latter genus. In
345 *Prochlorococcus*, several physico-chemical parameters have seemingly played a decisive role in the
346 genetic diversification of this genus, at distinct periods of its evolutionary history, starting with light
347 (split between LL and HL lineages), then iron availability (HLIII/IV vs. other HL) and temperature (HLI
348 vs. HLII; (18, 21, 45)). In contrast, nitrogen and phosphorus availability influenced genetic
349 diversification only in the 'leaves' of the *Prochlorococcus* radiation, through lateral transfers of gene
350 cassettes conferring on populations the ability to adapt to local N or P-depleted niches (46, 47). Despite
351 this apparent solid relationship between *Prochlorococcus* phylogeny and community structure, a
352 recent study looking at the genomic diversity of individual *Prochlorococcus* cells in a single water
353 sample highlighted a huge microdiversity within the HLII clade (23). This microdiversity seemingly
354 allows cells to adapt to slightly different selective pressures, such as biotic factors (phages, grazing,
355 etc). Here, we also observed a large microdiversity within the HLII lineage, with 25 OTUs comprising 4
356 ESTUs, but in agreement with a recent study (48), there were only subtle differences between the
357 distribution patterns of these intra-clade groups (except for ESTU HLIIC, represented by a single OTU;
358 **Fig. 2C**), confirming that abiotic factors have only marginally affected the genetic diversification within
359 this clade. In contrast, the microdiversity that we identified within HLI and LLI has seemingly allowed
360 members of these clades to colonize ecological niches clearly different from that of the dominant
361 ESTUs, extending the global niche occupied by these lineages. This includes LLIB, which seems to be
362 adapted to Fe-limited surface waters, much like HLIIIA-IVA, as well as HLIC, which thrives not only in
363 cold temperate waters, as do the more typical HLIA, but also in warm sub-tropical waters, where it co-
364 occurs with the dominant HLIIA (**Fig. S6**). This is consistent with the recent finding that HLI sub-clades
365 are driven by distinct environmental traits (48) and that even in HLII-dominated waters, HLI is never
366 competed to extinction (7).

367 Similarly, splitting *Synechococcus* clades into ESTUs revealed that this genus comprises a number of
368 specialists, mostly characterized by their respective temperature and Fe requirements (**Fig. 5**). While
369 CRD1B/EnvBB, CRD1A/EnvBA/EnvAA and CRD1C/EnvBC were respectively found in cold, intermediate
370 and warm waters with various degrees of Fe limitation, other ESTUs preferentially thrive in regions
371 where this nutrient is not limiting in either cold (IA, IVA, IIB), intermediate (IIIA, 5.3A) or warm (IIA)
372 waters. The third most discriminating parameter appears to be P-limitation that only ESTUs IIIA and
373 5.3A can stand, but only in Fe-replete conditions. It is also worth noting that several ESTUs, such as
374 those classified as ‘temperature intermediate’, display a larger tolerance range with regard to
375 temperature than their ‘cold’ and ‘warm’ counterparts (**Fig. 5**). Altogether, these results temper the
376 paradigm of *Synechococcus* being a generalist and physiologically more plastic than *Prochlorococcus*,
377 which mainly relied on the ability of the former to colonize much wider ecological niches than the
378 latter and on the apparent absence of genome streamlining in *Synechococcus* compared to
379 *Prochlorococcus* (18, 49-51). Thus, our results demonstrate that the observed ubiquity of the
380 *Synechococcus* genus as a whole (1, 2) in fact rests on a complex suite of specialists adapted to fairly
381 narrow niches, as is the case for *Prochlorococcus*.

382 Focusing on shifts in community composition associated to changes in local environmental conditions
383 or to physical barriers (**Fig. S5a-b**) provided additional insights into this global picture and revealed
384 that some ESTUs behave as opportunists. For instance, this is the case off the Marquesas Islands, where
385 the proximity of the coast induced an iron enrichment at TARA_123 and 124 as compared to a typical
386 HNLC situation at TARA_122 and TARA-128. While CRD1C dominated at the latter stations, ESTU IIA
387 took over this local population in these iron-replete patches (with an intermediate situation at
388 TARA_125; **Fig S5a**). By comparison, the *Prochlorococcus* abundance drastically dropped at TARA_123
389 but without any significant change in the community structure, suggesting that the minor HLIIA
390 component of this assemblage was not responsive enough to local Fe enrichment to outcompete the
391 dominant HLIIIA/IVA population. Another abrupt shift in community composition occurred at the
392 Agulhas choke point off the southern tip of Africa, where huge anticyclonic rings (i.e., Agulhas rings)
393 are formed in the Indian Ocean and then drift across the South Atlantic (42, 52). The strong drop in
394 temperature, occurring within the youngest ring (TARA_068), was likely responsible for a large part in
395 the shift from a typical subtropical ESTU assemblage in the Indian Ocean, dominated by
396 *Prochlorococcus* HLIIA-B and *Synechococcus* IIA (TARA_064-065), to a cold water ESTU assemblage
397 (HLIA, LLIA, CRD1A, EnvBA and IVA-B) at TARA_068 (**Fig. S5a**), suggesting that the latter ESTUs might
398 also have an opportunistic behavior with regard to their warm waters counterparts. Although these
399 two examples correspond to biogeochemical processes likely occurring at different time scales, the
400 observed ESTU assemblage changes likely result from differences in the intrinsic dynamics of ESTUs

401 within both genera, the most adapted one outcompeting others in favorable ecological conditions,
402 with *Synechococcus* displaying a more opportunistic behavior than *Prochlorococcus*.

403 Our results also raise several questions that can only be addressed in the laboratory or *in silico*. From
404 a physiological point of view, the fact that some ESTUs seemingly get counter-selected in response to
405 nutrient enrichment (e.g., iron in the case of CRD1C) suggests that, as proposed for *Prochlorococcus*
406 HLIII/IV (31), their growth capacity in nutrient replete conditions is lower than that of opportunistic
407 ESTUs (e.g. IIA) and this could be checked by comparing representative strains of these two lifestyles
408 in single or co-cultures. It is also unclear yet whether differences between these two behaviors is due
409 to the loss of genes costly to maintain for the cells, to a better affinity of core enzymes (e.g., for nutrient
410 scavenging) and/or to the acquisition of specific gene sets by lateral gene transfer, as reported for
411 *Prochlorococcus* regarding phosphate and nitrogen uptake and assimilation (46, 47). Adaptation to low
412 Fe is particularly striking in this context since our study showed that this ability seems to have appeared
413 several times during evolution in quite distantly related ESTUs, namely *Prochlorococcus* HLIIIA/HLIVA
414 —that likely occurred via a single diversification event— and LLIB as well as *Synechococcus* CRD1A,
415 CRD1C, EnvBA, EnvBC and EnvAA (**Fig. 5**). Although no *Prochlorococcus* isolates of HLIIIA/IVA are
416 available in culture yet, sequencing of single amplified genomes suggested that these organisms have
417 adapted to Fe-limited environments by lowering their cellular Fe requirement through loss of genes
418 encoding Fe-rich proteins and by acquiring siderophore transporters for efficient scavenging of
419 organic-bound forms of this element (31, 32). Genomic comparison of *Synechococcus* strains, including
420 representatives of the different CRD1 ESTUs, as well as whole genome recruitment of metagenomic
421 data should allow to check whether a similar adaptation process has occurred in this genus.

422 In conclusion, although very few studies have so far combined information from high resolution
423 phylogenetic markers and geographical distribution to detect ecologically coherent taxonomic groups
424 (e.g., (48, 53)), we show here that this approach can bring invaluable insights for deciphering the links
425 between genetic diversity and niche occupancy. Indeed, the definition of within-clade ESTUs using a
426 reference *petB* database enriched with ecologically relevant and distantly related sequences
427 assembled from *Tara* Oceans reads, has allowed us to obtain clear-cut spatial distribution patterns for
428 taxa within both *Prochlorococcus* and *Synechococcus* genera, indicating that we explored the diversity
429 of the picocyanobacterial community at the right taxonomic resolution. Additionally, in contrast to
430 other phytoplankton groups, such as diatoms (54), these biogeographical patterns were found to be
431 tightly controlled by environmental factors. Besides helping to refine models of picocyanobacterial
432 distributions and predicting their behavior in response to ongoing climate change, knowledge of the
433 oceanic areas where poorly characterized ESTUs predominate, will also guide future strain isolation
434 (e.g., for the yet uncultured EnvA and EnvB) and sequencing efforts. Characterizing and comparing
435 such ecologically representative strains will help further unveil the basis of niche partitioning.

436 **Materials and methods**

437 **Genomic material.** This study focused on 109 *Tara* Oceans metagenomes corresponding to 66 stations
438 along the *Tara* Oceans transect for which a ‘bacterial size fraction’ was available (i.e. 0.2-1.6 μm for
439 TARA_004 to TARA_052 and 0.2-3 μm for TARA_056 to TARA_152). Water samples were collected at
440 two depths, surface (SUR) and deep chlorophyll maximum (DCM), the latter sample sometimes being
441 merely collected in the upper mixed layer, when the DCM was not clearly delineated (**Dataset 3**).
442 Metagenomes were sequenced using the Illumina® technology as overlapping paired reads of
443 $\sim 100/108$ bp with various sequencing depths, ranging from 16×10^6 to 258×10^6 reads after quality
444 control, corresponding to an average 20.2 ± 9.9 Gb of sequence data per sample. Reads were merged
445 using FLASH v1.2.7 with default parameters (55) and cleaned based on quality using CLC QualityTrim
446 v4.10.86742 (CLC Bio, Aarhus, Denmark), resulting in 100 to 215 bp fragments. **Dataset 3** describes all
447 metagenomic samples with location and sequencing effort. All metagenomes and corresponding
448 environmental parameters measured during the *Tara* Oceans expedition are available at
449 www.pangea.de, except for the iron and ammonium data that were simulated with the ECCO2-Darwin
450 model and the iron limitation index Φ_{sat} (56) and are available in **Dataset 3**.

451 **Building of the PetB-DB database.** To recruit and taxonomically assign metagenomics reads targeting
452 the high resolution *petB* gene marker, we analyzed 1,091 sequences of the *petB* gene from cultured
453 isolates and environmental samples and built a reference database including all non-redundant high
454 quality sequences of this marker available for the marine picocyanobacteria *Prochlorococcus* (69
455 sequences covering 7 clades) and *Synechococcus* (399 sequences covering 3 subclusters, 22 clades and
456 30 subclades). The dataset also includes outgroup sequences from publicly available cyanobacteria,
457 including marine (13 sequences) and freshwater isolates (40 sequences), as well as representatives of
458 the main marine eukaryotic phytoplankton taxa and eukaryotic cyanobionts (64 plastid *petB*
459 sequences), raising the number of *petB* sequences to 585 (**Datasets 1 and 2**). To avoid differential
460 alignment effects at the edge of the reference sequences, all sequences were aligned and trimmed to
461 557 bp. This database was secondarily complemented by 136 *petB* sequences assembled from selected
462 *Tara* Oceans reads displaying less than 94 % identity with previously known *petB* sequences (yet some
463 of these new sequences could exhibit more than 94 % identity with one another).

464 **Read recruitments.** Targeted *petB* fragment recruitments were performed using a two-step protocol.
465 In order to maximize the diversity while reducing the weight of the resulting tabulated files, translated
466 sequences of the non-redundant *petB* database were used to recruit candidate *petB* gene fragments
467 by BLASTX (v2.2.28+) using default parameters but by limiting the results to 1 target sequence. These
468 *petB* candidates were then compared to the full reference *petB* database using BLASTN (v2.2.28+) with

469 sensitive configuration (`-task blastn -gapopen 8 -gapextend 6 -reward 5 -penalty -4 -word_size 8`)
470 and cut-offs to reduce the weight of resulting tabulated files (`-perc_identity 50 -evaluate 0.0001`).

471 Reads with more than 90 % of their sequence aligned and with more than 80 % sequence identity to
472 their BLASTN best-hit (see result section for the determination of this cut-off) were selected as genuine
473 picocyanobacterial *petB*, taxonomically assigned to their BLASTN best-hit and subsequently used to
474 build per-strain read counts tables. Counts were then aggregated by clade or ESTU and subsequently
475 used to build pie charts or community structure profiles.

476 **Phylogenetic and statistical analyses.** Phylogenetic reconstructions were based on multiple
477 alignments of *petB* nucleotide sequences generated using MAFFT v7.164b with default parameters
478 (57). A maximum likelihood tree was inferred using PHYML v3.0 – 20120412, (58) with the HKY + G
479 substitution model, as determined using jModeltest v2.1.4 (59), and the estimation of the gamma
480 distribution parameter of the substitution rates among sites and of the proportion of invariables sites.
481 Confidence of branch points was determined by performing bootstrap analyses including 1000
482 replicate data sets. Phylogenetic trees were edited using the Archaeopteryx v0.9901 beta program (60)
483 and drawn using iTOL (<http://itol.embl.de>; (61)). Operational taxonomical units (OTUs) for the *petB*
484 reference data set at 94% were defined by nucleotide identity using Mothur v1.34.4 (62).

485 In each clade, ESTUs were defined using a type 3 SIMPROF approach (53) by considering: i) for
486 *Prochlorococcus*, stations with more than 100 reads and OTUs recruiting more than 150 reads and ii)
487 for *Synechococcus*, stations with more than 20 reads and OTUs recruiting more than 25 reads.
488 Hierarchical clustering was performed on the remaining stations and OTUs using the Bray-Curtis
489 distance between relative abundance profiles using *heatmap.3* function in GMD v0.3.1.1 R package
490 (ward algorithm; (63)). Statistical significance of the difference between clusters was first assessed by
491 a permutation analysis using the *clustsig* v1.1 R package (alpha=0.05, Bray-Curtis distance, otherwise
492 default parameters). ESTU delineation was then manually refined, e.g. ESTUs were sometimes defined
493 from single OTUs if the Bray-Curtis distance was >0.65 or if pairs of OTUs were not defined as coherent
494 groups because all OTUs within a clade were equally distant from each other. In contrast, some
495 potential ESTUs were not considered as reliable, e.g. if high Bray-Curtis distances were due to
496 differences in abundance and not in distribution.

497 Hierarchical clustering and NMDS analyses of stations were performed using R packages *cluster* v1.14.4
498 (64) and *MASS* v7.3-29 (65), respectively. *petB*-_{mi}Tag contingency tables aggregated at the ESTU level
499 were filtered as above and normalized using Hellinger transformation that gives lower rates to rare
500 ESTUs. Bray-Curtis distance was then used for both clustering (*agnes* function, default parameters) and
501 ordination (*isoMDS* function, `maxit=100`, `k=2`). All displayed clusters were significant ($p < 0.01$,
502 permutation tests). Fitting of environmental parameters on NMDS ordination was performed with

503 function *envfit* in *vegan* v2.2-1 package and p-value based on 999 permutations was used to assess the
504 significance of the fit and only environmental parameters showing an adjusted p-value below 0.05
505 were used.

506 **Visualization of realized environmental niches.** In order to visualize the tolerance range of each ESTU
507 with regard to physico-chemical parameters, values were scaled and reduced before analysis. For each
508 ESTU, *Tara* Oceans stations were sorted by order of abundance, and stations gathering 80% of all reads
509 of the given ESTU were kept. A boxplot was then computed for each parameter taking into account
510 the values of this parameter in the kept stations.

511

512 **Acknowledgements**

513 We warmly thank M. Follows and O. Jahn for providing us with ECCO2-Darwin simulation values for
514 iron and S. Speich for fruitful discussions on oceanographic context. This work was supported by the
515 French “Agence Nationale de la Recherche” Programs SAMOSA (ANR-13-ADAP-0010) and France
516 Génomique (ANR-10-INBS-09), the French Government 'Investissements d'Avenir' programs
517 OCEANOMICS (ANR-11-BTBR-0008), UK Natural Environment Research Council grants NE/I00985X/1
518 and NE/J02273X/1, and the European Union's Seventh Framework Programs FP7 MicroB3 (grant
519 agreement 287589) and MaCuMBA (grant agreement 311975). We also thank the support and
520 commitment of the *Tara* Oceans coordinators and consortium, Agnès b. and E. Bourgois, the Veolia
521 Environment Foundation, Region Bretagne, Lorient Agglomeration, World Courier, Illumina, the EDF
522 Foundation, FRB, the Prince Albert II de Monaco Foundation, the *Tara* schooner and its captains and
523 crew. *Tara* Oceans would not exist without continuous support from 23 institutes
524 (<http://oceans.taraexpeditions.org>).

525

526 **References**

- 527 1. Flombaum P, *et al.* (2013) Present and future global distributions of the marine cyanobacteria
528 *Prochlorococcus* and *Synechococcus*. *Proc Natl Acad Sci U S A* 110(24):9824-9829.
- 529 2. Partensky F, Hess WR, & Vaulot D (1999) *Prochlorococcus*, a marine photosynthetic prokaryote
530 of global significance. *Microbiol Mol Biol Rev* 63(1):106-127.
- 531 3. Dutkiewicz S, *et al.* (2015) Impact of ocean acidification on the structure of future
532 phytoplankton communities. *Nat Clim Change* 5:10002-11009.
- 533 4. Coleman ML & Chisholm SW (2007) Code and context: *Prochlorococcus* as a model for cross-
534 scale biology. *Trends Microbiol* 15(9):398-407.
- 535 5. Johnson ZI, *et al.* (2006) Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale
536 environmental gradients. *Science* 311(5768):1737-1740.

- 537 6. Moore LR, Rocap G, & Chisholm SW (1998) Physiology and molecular phylogeny of coexisting
538 *Prochlorococcus* ecotypes. *Nature* 393(6684):464-467.
- 539 7. Chandler JW, *et al.* (2016) Variable but persistent coexistence of *Prochlorococcus* ecotypes
540 along temperature gradients in the ocean's surface mixed layer. *Environ Microbiol Rep*
541 8(2):272-284.
- 542 8. Zinser ER, *et al.* (2007) Influence of light and temperature on *Prochlorococcus* ecotype
543 distributions in the Atlantic Ocean. *Limnol Oceanogr* 52(5):2205-2220.
- 544 9. Mackey KR, *et al.* (2013) Effect of temperature on photosynthesis and growth in marine
545 *Synechococcus* spp. *Plant Physiol* 163(2):815-829.
- 546 10. Pittera J, *et al.* (2014) Connecting thermal physiology and latitudinal niche partitioning in
547 marine *Synechococcus*. *ISME J* 8(6):1221-1236.
- 548 11. Huang S, *et al.* (2012) Novel lineages of *Prochlorococcus* and *Synechococcus* in the global
549 oceans. *ISME J* 6(2):285-297.
- 550 12. Mazard S, Ostrowski M, Partensky F, & Scanlan DJ (2012) Multi-locus sequence analysis,
551 taxonomic resolution and biogeography of marine *Synechococcus*. *Environ Microbiol*
552 14(2):372-386.
- 553 13. Zwirgmaier K, *et al.* (2008) Global phylogeography of marine *Synechococcus* and
554 *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ*
555 *Microbiol* 10(1):147-161.
- 556 14. Rusch DB, *et al.* (2007) The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic
557 through Eastern Tropical Pacific. *PLoS Biol* 5(3):398-431.
- 558 15. West NJ, Lebaron P, Strutton PG, & Suzuki MT (2010) A novel clade of *Prochlorococcus* found
559 in high nutrient low chlorophyll waters in the South and Equatorial Pacific Ocean. *ISME J*
560 5(6):933-944.
- 561 16. Ahlgren NA, *et al.* (2014) The unique trace metal and mixed layer conditions of the Costa Rica
562 upwelling dome support a distinct and dense community of *Synechococcus*. *Limnol Oceanogr*
563 59:2166–2218.
- 564 17. Sohm JA, *et al.* (2015) Co-occurring *Synechococcus* ecotypes occupy four major oceanic
565 regimes defined by temperature, macronutrients and iron. *ISME J* 10:333-345.
- 566 18. Kettler G, *et al.* (2007) Patterns and implications of gene gain and loss in the evolution of
567 *Prochlorococcus*. *PLoS Genet* 3:e231.
- 568 19. Post AF, *et al.* (2011) Long term seasonal dynamics of *Synechococcus* population structure in
569 the gulf of aqaba, northern red sea. *Front Microbiol* 2(2):131.

- 570 20. Ahlgren NA & Rocap G (2012) Diversity and distribution of marine *Synechococcus*: Multiple
571 gene phylogenies for consensus classification and development of qPCR Assays for sensitive
572 measurement of clades in the ocean. *Front Microbiol* 3:213.
- 573 21. Biller SJ, Berube PM, Lindell D, & Chisholm SW (2015) *Prochlorococcus*: the structure and
574 function of collective diversity. *Nat Rev Microbiol* 13(1):13-27.
- 575 22. Koeppel AF, *et al.* (2013) Speedy speciation in a bacterial microcosm: new species can arise as
576 frequently as adaptations within a species. *ISME J* 7(6):1080-1091.
- 577 23. Kashtan N, *et al.* (2014) Single-cell genomics reveals hundreds of coexisting subpopulations in
578 wild *Prochlorococcus*. *Science* 344(6182):416-420.
- 579 24. Armbrust EV & Palumbi SR (2015) Marine biology. Uncovering hidden worlds of ocean
580 biodiversity. *Science* 348(6237):865-867.
- 581 25. Karsenti E, *et al.* (2011) A holistic approach to marine eco-systems biology. *Plos Biol* 9(10).
- 582 26. Logares R, *et al.* (2014) Metagenomic 16S rDNA Illumina tags are a powerful alternative to
583 amplicon sequencing to explore diversity and structure of microbial communities. *Environ*
584 *Microbiol* 16(9):2659-2671.
- 585 27. Sunagawa S, *et al.* (2015) Ocean plankton. Structure and function of the global ocean
586 microbiome. *Science* 348(6237):1261359.
- 587 28. Bouman HA, *et al.* (2006) Oceanographic basis of the global surface distribution of
588 *Prochlorococcus* ecotypes. *Science* 312(5775):918-921.
- 589 29. Malmstrom RR, *et al.* (2010) Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic
590 and Pacific oceans. *ISME J* 4(10):1252-1264.
- 591 30. Partensky F & Garczarek L (2010) *Prochlorococcus*: advantages and limits of minimalism. *Ann*
592 *Rev Mar Sci* 2:305-331.
- 593 31. Rusch DB, *et al.* (2010) Characterization of *Prochlorococcus* clades from iron-depleted oceanic
594 regions. *Proc Natl Acad Sci USA* 107(37):16184-16189.
- 595 32. Malmstrom RR, *et al.* (2013) Ecology of uncultured *Prochlorococcus* clades revealed through
596 single-cell genomics and biogeographic analysis. *ISME J* 7(1):184-198.
- 597 33. Song Q, Gordon AL, & Visbeck M (2004) Spreading of the Indonesian Throughflow in the Indian
598 Ocean. *J Phys Oceanogr* 34(4):772-792.
- 599 34. Moutin T, *et al.* (2002) Does competition for nanomolar phosphate supply explain the
600 predominance of the cyanobacterium *Synechococcus*? *Limnol Oceanogr* 47(5):1562-1567.
- 601 35. Pendorf KJ & Duhamel S (2015) Variable phosphorus uptake rates and allocation across
602 microbial groups in the oligotrophic Gulf of Mexico. *Environ Microbiol* 17(10):3992-4006.
- 603 36. Paytan A, *et al.* (2009) Toxicity of atmospheric aerosols on marine phytoplankton. *Proc Natl*
604 *Acad Sci USA* 106:4601-4605.

- 605 37. Mella-Flores D, *et al.* (2011) Is the distribution of *Prochlorococcus* and *Synechococcus* ecotypes
606 in the Mediterranean Sea affected by global warming? *Biogeosciences* 8:2785–2804.
- 607 38. Fuller NJ, *et al.* (2005) Dynamics of community structure and phosphate status of
608 picocyanobacterial populations in the Gulf of Aqaba, Red Sea. *Limnol Oceanogr* 50(1):363-375.
- 609 39. Dufresne A, *et al.* (2008) Unraveling the genomic mosaic of a ubiquitous genus of marine
610 cyanobacteria. *Genome Biol* 9(5):R90.
- 611 40. Saito MA, Rocap G, & Moffett JW (2005) Production of cobalt binding ligands in a
612 *Synechococcus* feature at the Costa Rica upwelling dome. *Limnol Oceanogr* 50(1):279-290.
- 613 41. Gutierrez-Rodríguez A, *et al.* (2014) Fine spatial structure of genetically distinct
614 picocyanobacterial populations across environmental gradients in the Costa Rica Dome. *Limnol*
615 *Oceanogr* 59(3):705–723.
- 616 42. Villar E, *et al.* (2015) Ocean plankton. Environmental characteristics of Agulhas rings affect
617 interocean plankton transport. *Science* 348(6237):1261447.
- 618 43. Urbach E & Chisholm SW (1998) Genetic diversity in *Prochlorococcus* populations flow
619 cytometrically sorted from the Sargasso Sea and Gulf Stream. *Limnol Oceanogr* 43(7):1615-
620 1630.
- 621 44. Fuller NJ, *et al.* (2003) Clade-specific 16S ribosomal DNA oligonucleotides reveal the
622 predominance of a single marine *Synechococcus* clade throughout a stratified water column in
623 the Red Sea. *Appl Environ Microbiol* 69(5):2430-2443.
- 624 45. Martiny JB, Jones SE, Lennon JT, & Martiny AC (2015) Microbiomes in light of traits: A
625 phylogenetic perspective. *Science* 350(6261):aac9323.
- 626 46. Martiny AC, Huang Y, & Li W (2009) Occurrence of phosphate acquisition genes in
627 *Prochlorococcus* cells from different ocean regions. *Environ Microbiol* 11(6):1340-1347.
- 628 47. Martiny AC, Kathuria S, & Berube PM (2009) Widespread metabolic potential for nitrite and
629 nitrate assimilation among *Prochlorococcus* ecotypes. *Proc Natl Acad Sci USA* 106(26):10787-
630 10792.
- 631 48. Larkin AA, *et al.* (2016) Niche partitioning and biogeography of high light adapted
632 *Prochlorococcus* across taxonomic ranks in the North Pacific. *ISME J*
633 doi:10.1038/ismej.2015.244.
- 634 49. Palenik B, *et al.* (2003) The genome of a motile marine *Synechococcus*. *Nature*
635 424(6952):1037-1042.
- 636 50. Scanlan DJ, *et al.* (2009) Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol*
637 *Rev* 73(2):249-299.
- 638 51. Dufresne A, Garczarek L, & Partensky F (2005) Accelerated evolution associated with genome
639 reduction in a free-living prokaryote. *Genome Biol* 6(2):R14.

- 640 52. Biastoch A, Boning CW, & Lutjeharms JR (2008) Agulhas leakage dynamics affects decadal
641 variability in Atlantic overturning circulation. *Nature* 456(7221):489-492.
- 642 53. Somerfield PJ & Clarke KR (2013) Inverse analysis in non-parametric multivariate analyses:
643 distinguishing of groups of associated species which covary coherently across samples. *J Exp*
644 *Mar Biol Ecol* 449:261 – 273
- 645 54. Malviya S, *et al.* (2015) Insights into global diatom distribution and diversity in the world's
646 ocean. *Proc Natl Acad Sci USA* 113(11):E1516-E1525.
- 647 55. Magoc T & Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome
648 assemblies. *Bioinformatics* 27(21):2957-2963.
- 649 56. Behrenfeld MJ, *et al.* (2009) Satellite-detected fluorescence reveals global physiology of ocean
650 phytoplankton. *Biogeosciences* 6(5):779-794.
- 651 57. Katoh K & Standley DM (2014) MAFFT: iterative refinement and additional methods. *Methods*
652 *Mol Biol* 1079:131-146.
- 653 58. Guindon S & Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large
654 phylogenies by maximum likelihood. *Syst Biol* 52(5):696-704.
- 655 59. Darriba D, Taboada GL, Doallo R, & Posada D (2012) jModelTest 2: more models, new heuristics
656 and parallel computing. *Nat Methods* 9(8):772.
- 657 60. Han MV & Zmasek CM (2009) phyloXML: XML for evolutionary biology and comparative
658 genomics. *BMC Bioinfo* 10:356.
- 659 61. Letunic I & Bork P (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree
660 display and annotation. *Bioinformatics* 23(1):127-128.
- 661 62. Schloss PD, *et al.* (2009) Introducing mothur: open-source, platform-independent, community-
662 supported software for describing and comparing microbial communities. *Appl Environ*
663 *Microbiol* 75(23):7537-7541.
- 664 63. Zhao X, Valen E, Parker BJ, & Sandelin A (2011) Systematic clustering of transcription start site
665 landscapes. *PLoS One* 6(8):e23409.
- 666 64. Maechler M, Rousseeuw P, Struyf A, Hubert M, & Hornik K (2015) cluster: cluster analysis
667 basics and extensions. R package version 2.0.3).
- 668 65. Venables WN & Ripley BD (2002) *Modern applied statistics with S* (Springer, New York) 4th Ed.
669 495 pp.
- 670 66. Biller SJ, *et al.* (2014) Genomes of diverse isolates of the marine cyanobacterium
671 *Prochlorococcus*. *Nature Scient Data* 1:140034.
- 672 67. Choi DH & Noh JH (2009) Phylogenetic diversity of *Synechococcus* strains isolated from the
673 East China Sea and the East Sea. *FEMS Microbiol Ecol* 69(3):439-448.

674 **Figure Legends**

675

676 **Figure 1. Maximum likelihood tree of *Synechococcus* and *Prochlorococcus* lineages based on *petB***
677 **gene sequences from both isolates and environmental sequences.** Diamonds at nodes indicate
678 bootstrap support over 70%. Taxonomic assignments are given by the color codes at clade level for
679 *Prochlorococcus* (top left) and clade (e.g. V, CRD1) or subclade (e.g. Ia-c) for *Synechococcus* (bottom
680 right). Sequences were named after ID_subcluster_clade_subclade_ESTU for *Synechococcus* ID_LL or
681 HL_clade_ESTU for *Prochlorococcus*. The outer pink ring indicates that the corresponding sequence in
682 the tree was the best-hit of at least one *Tara* Oceans picocyanobacterial read and the inner blue bar
683 plot shows the \log_2 of the number of metagenomic reads recruited for this sequence (range: 0-10.84).
684 Sequences in black letters correspond to the initial reference database and those in white or light grey
685 letters to newly assembled *petB* sequences from *Tara* Oceans metagenome reads. The scale bar
686 represents the number of substitutions per nucleotide position. For improved readability, the length
687 of three *Prochlorococcus* branches was reduced, as indicated by double slashes. *Prochlorococcus* clade
688 assignment is as in (66), while for *Synechococcus* subcluster 5.1, subclade assignments are as in (67)
689 for WPC1 and WPC2 and as in (12) for all other clades.

690

691 **Figure 2. Percent identity of *Tara* Oceans *petB* _{mi} tags vs. sequences of the reference database and**
692 **abundance at different stations along the transect of operational taxonomic units (OTUs) clustered**
693 **into ESTUs. (A)** Distribution of the percent identity of best-hits of all *petB* candidate reads recruited
694 from the *Tara* Oceans bacterial-size fraction metagenomes against the *petB* reference database.
695 Populations 1 and 2 correspond respectively to genuine *petB* reads and to non-specific signal, due
696 either to *petB* reads from organisms not included in the reference database or to *petB*-related genes.
697 The grey part in population 1 corresponds to *petB* reads attributable to photosynthetic organisms of
698 the reference database other than *Prochlorococcus* and *Synechococcus*. The red arrow shows the 80%
699 cut-off used to separate the *petB* signal from noise. The top and bottom panels correspond to
700 recruitments made before and after addition of the 136 newly assembled environmental *petB*
701 sequences, respectively. **(B)** Same as above but for some selected *Synechococcus* taxa (see **Fig. S2** for
702 all other picocyanobacterial taxa). **(C)** Determination of ESTUs based on the distribution patterns of
703 within-clade 94% OTUs. At each station, the number of reads assigned to a given OTU is normalized by
704 the total number of reads assigned to the clade in this station. Stations and OTUs are filtered based on
705 the number of reads recruited and hierarchically clustered (Bray-Curtis distance) according to
706 distribution pattern. Only *Synechococcus* clades split into different ESTUs are shown (see Fig. S4 for

707 *Prochlorococcus*). Stars indicate nodes supported by p-value < 0.05 (SIMPROF test not applicable to
708 pair comparisons).

709

710 **Figure 3. Biogeography of *Prochlorococcus* ESTUs in surface *Tara* Oceans metagenomes and relation**
711 **to physico-chemical parameters. (A)** Histograms of the relative abundance of *Prochlorococcus* ESTUs
712 at each station sorted by similarity, as determined by hierarchical clustering (Bray-Curtis distance). Left
713 panel indicates seawater temperature (°C) at each station. **(B)** Distribution of the ESTU assemblages,
714 color-coded as in A, along the *Tara* Oceans transect. **(C)** NMDS analysis of stations according to Bray-
715 Curtis dissimilarity between *Prochlorococcus* assemblages, with fitted statistically significant (adjusted
716 p-value < 0.05) physico-chemical parameters. Samples that belong to the same ESTU assemblage have
717 been colored according to the color-code defined in A and contours of the same color gather all
718 samples comprised within each cluster. NMDS stress value: 0.0985.

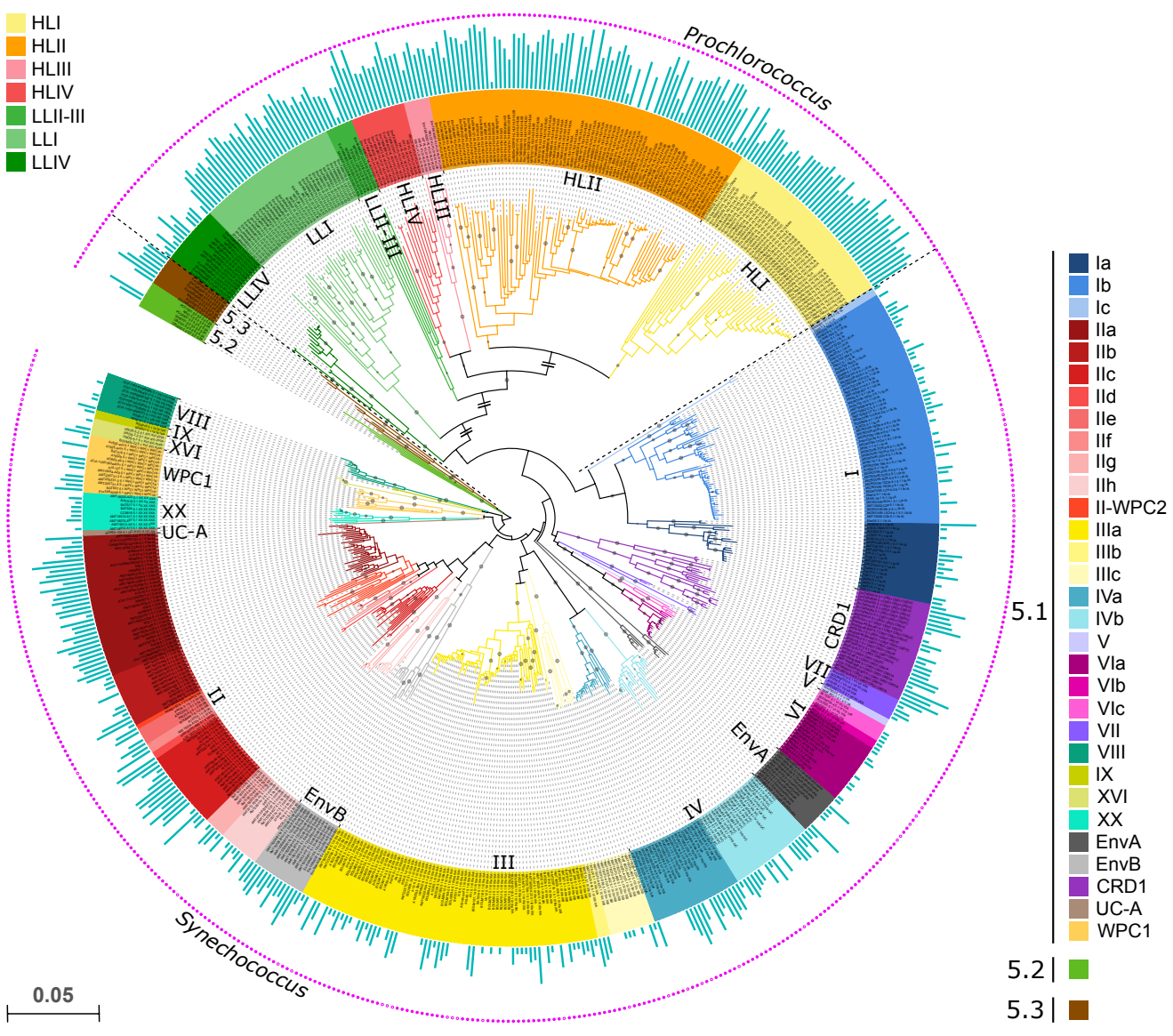
719

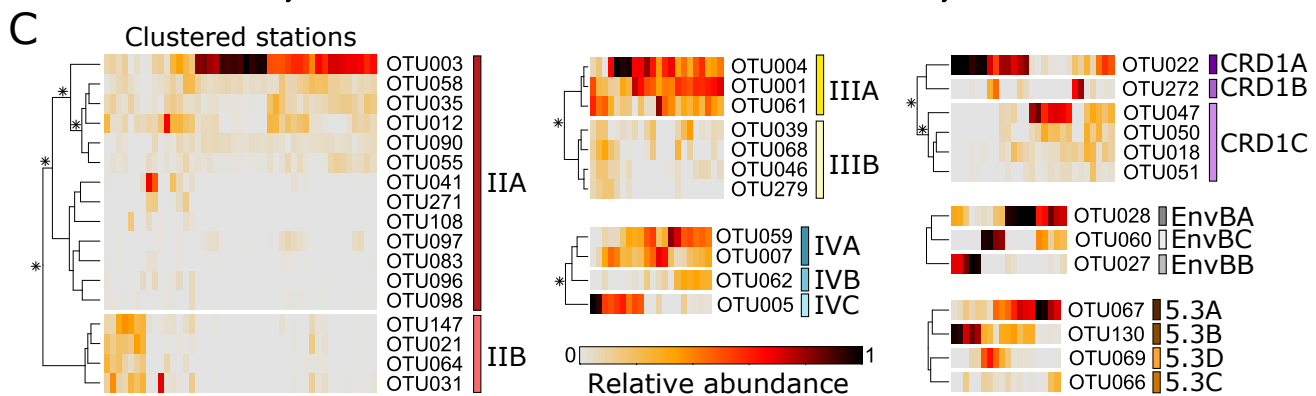
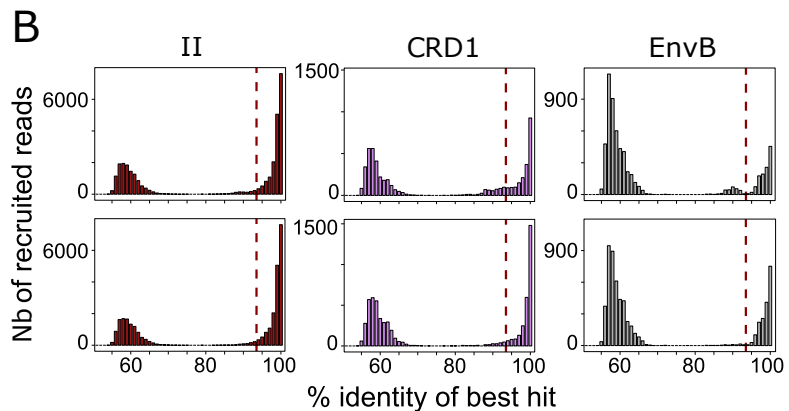
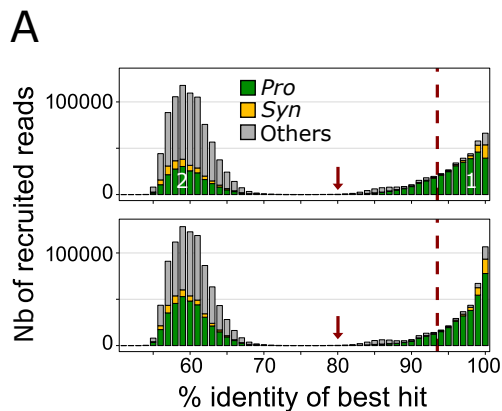
720 **Figure 4. Same as Fig. 3 but for *Synechococcus*.** NMDS stress value: 0.1369.

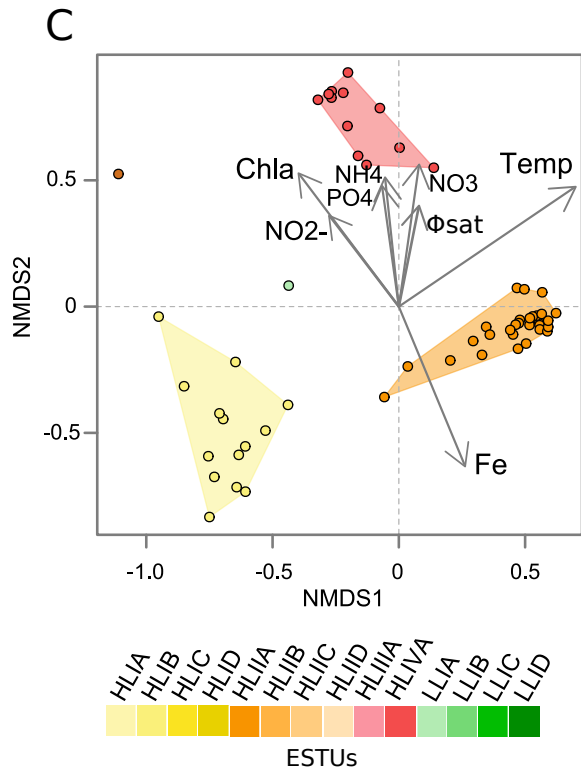
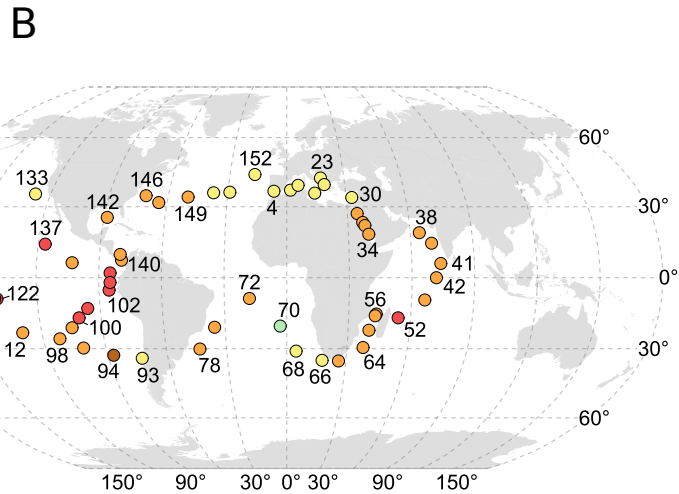
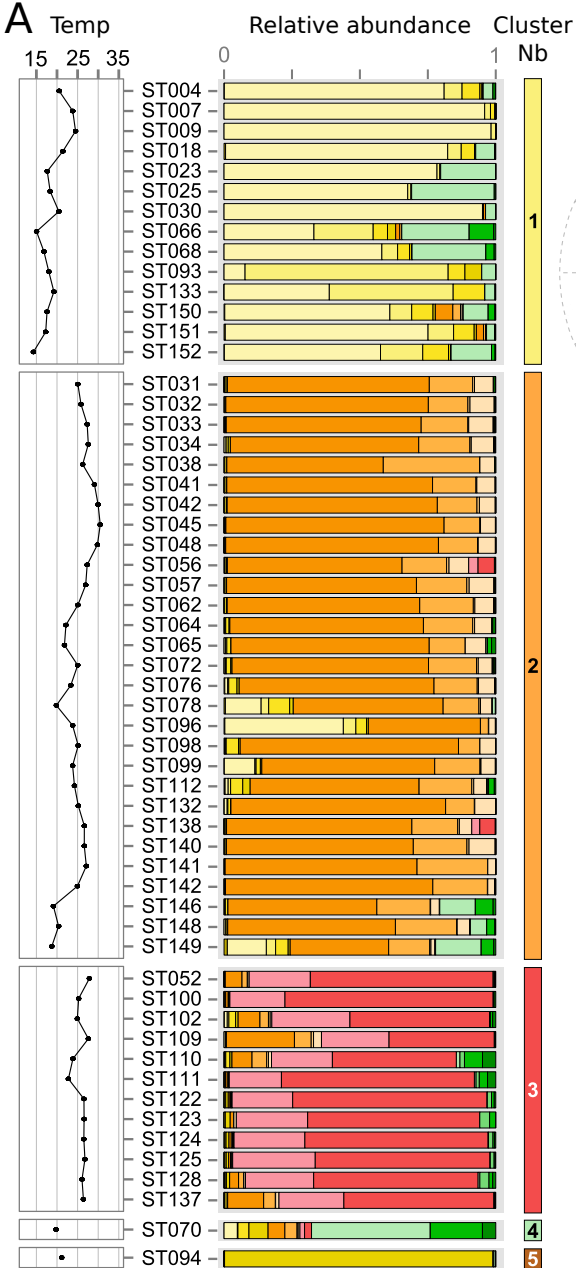
721

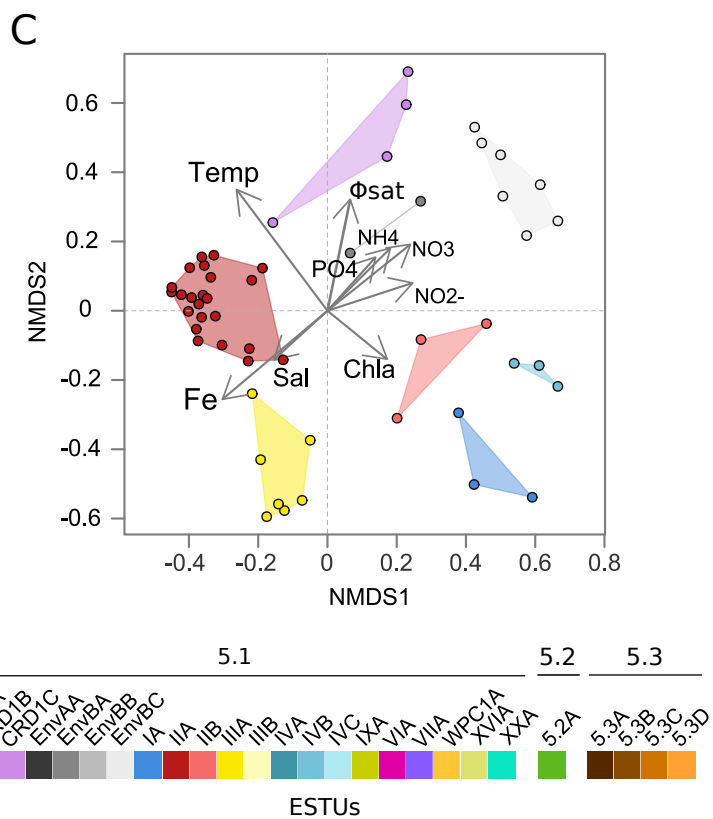
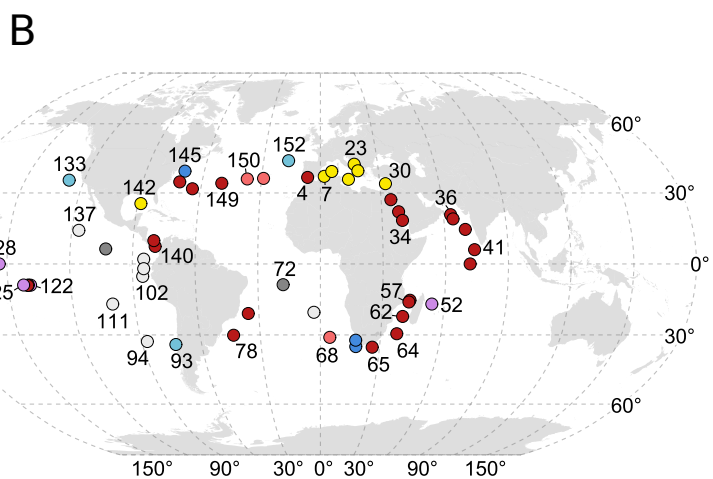
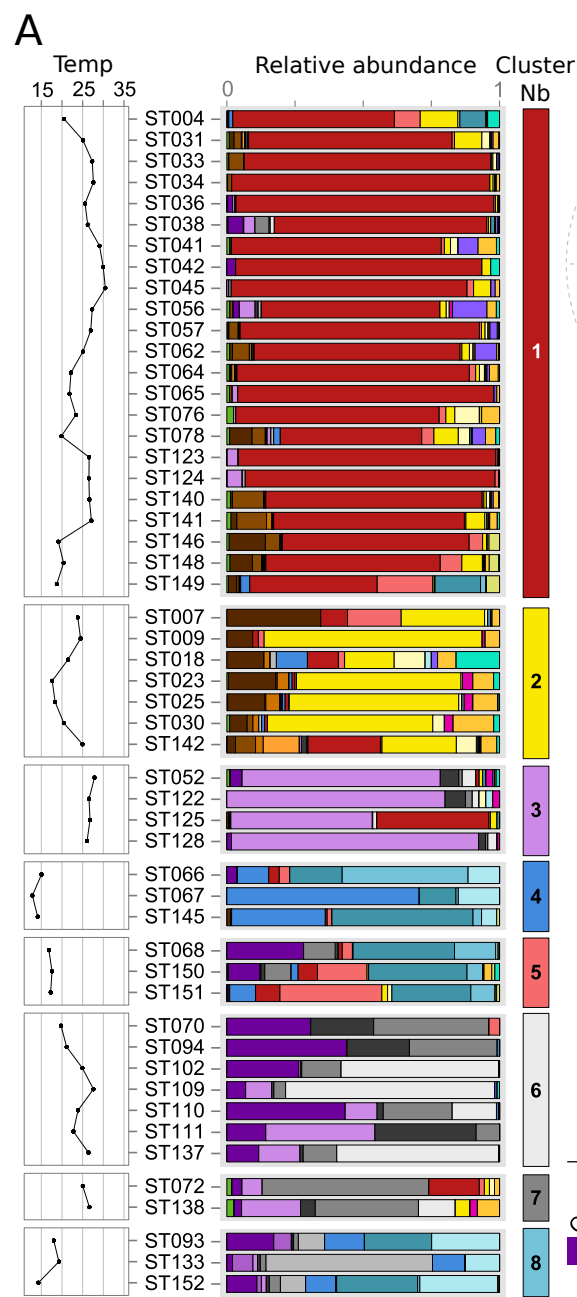
722 **Figure 5. Realized environmental niche of the major *Synechococcus* ESTUs in surface waters.**

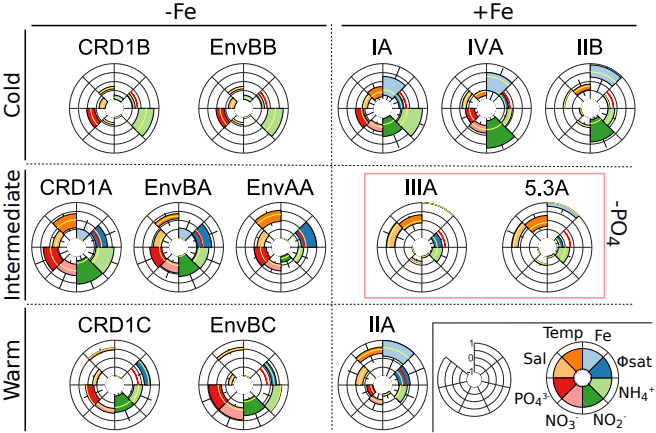
723 For each ESTU, stations were sorted by order of normalized abundance and only stations cumulating
724 80% of the total abundance were used to draw the graph. Boxplots represent the range of each
725 parameter (in relative units) tolerated by any given ESTU and the median is indicated by a yellow line.
726 ESTUs are organized according to their relative temperature range (cold, intermediate or warm),
727 tolerance to iron limitation (-Fe, +Fe) and tolerance to phosphate limitation (-PO₄). Please note that
728 the two proxies used to estimate Fe-limitation ([Fe] derived from the ECCO2-Darwin model and the
729 Φ_{sat} index; the red line indicates the 1.4 % value above which iron is considered limiting; (56)) are
730 sometimes contradictory e.g., for CRD1B and EnvBB.











Supporting information

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34

Figure S1: Variation of the assignment ability of each individual 100 bp gene fragment along the sequence of *petB* gene using reference databases for *Prochlorococcus* (A) or *Synechococcus* (B). Simulated reads were generated by 100 bp sliding windows along the marker sequences and the lowest taxonomic level at which they could be assigned is shown by a different blue tone (as indicated in the insert; for *Prochlorococcus*, the subcluster level actually corresponds to a LL or HL assignment, while the clade level corresponds to HLI-IV and LLI-IV, the lowest taxonomic level available for this genus).

Figure S2a: Distribution of the percent identity of *petB*_{mi} tags recruited from the bacterial-size fraction of the *Tara* Oceans metagenomes with regard to their best-hits in the reference database for each *Prochlorococcus* clade (top 7 graphs) and *Synechococcus* subclade (bottom 18 graphs) before addition of the 136 newly assembled environmental *petB* sequences. Note that clade XX was formerly called EnvC (12) but the name was changed here because there is at least one representative isolate (i.e., strain CC9616).

Figure S2b: Same as Fig. S2a but after addition of the 136 newly assembled environmental *petB* sequences.

Figure S3: Global recruitments of marine picocyanobacteria *petB*_{mi} tags in the bacterial size fraction of the *Tara* Oceans metagenomes. (A) All picocyanobacterial clades at both sampled depths; (B-C) percentage of each *Prochlorococcus* clade in surface (B) and at the deep chlorophyll maximum (DCM; C). (D-E) percentage of each *Synechococcus* clade in surface (D) and at the DCM (E). Note that clade XX was formerly called EnvC (12) but the name was changed here because there is now at least one representative isolate (i.e., strain CC9616).

Figure S4: *Prochlorococcus* ESTUs based on the distribution patterns of within-clade 94% OTUs. At each station, the number of reads assigned to a given OTU is normalized by the total number of reads assigned to the clade in this station. Stations and OTUs are filtered based on the number of reads recruited. OTUs are hierarchically clustered (Bray-Curtis distance) according to their distribution pattern. Stars indicate nodes supported by p-value < 0.05 (SIMPROF test not applicable to pair comparisons).

35

36 **Figure S5a:** Marine picocyanobacteria community structure in *Tara* Oceans surface metagenomes
37 based on *petB*-miTags recruitments. (A) Surface water temperature along the *Tara* Oceans transect. (B)
38 Relative abundances of *Prochlorococcus* and *Synechococcus* normalized to the total number of reads
39 at each station. (C-D) Relative abundances of *Prochlorococcus* and *Synechococcus* ESTUs, respectively.
40 White, grey and black dots indicate the number of reads used to build the profile, as detailed in the
41 insert. For readability, temperature for stations TARA_082 (7.3°C), TARA_084 (1.8°C) and TARA_085
42 (0.7°C) are not shown on graph A. Abbreviations: IO, Indian Ocean; MS; Mediterranean Sea; NAO:
43 North Atlantic Ocean; NPO, North Pacific Ocean; RS, Red Sea; SAO, South Atlantic Ocean; SO, Southern
44 Ocean.

45

46 **Figure S5b:** Same as Fig. S5a but at the DCM. A depth profile along the *Tara* Oceans transect was added.
47 For readability, temperature for stations TARA_082 (7.0°C) and TARA_085 (-0.8°C) are not shown on
48 graph A, while temperature is missing for station TARA_007.

49

50 **Figure S6: Distribution of minor *Prochlorococcus* ESTUs with regard to major ESTUs in the *Tara***
51 **Oceans metagenomes.** Relative abundance normalized to the total number of reads per ESTU of (A)
52 ESTUs HLI A and HLI C with regard to HLI A in surface waters and (B-C) ESTUs LLIA-C with regard to HLI A
53 in surface waters and the DCM, respectively. For graph A, stations were sorted from the lowest to
54 highest temperatures and for graph B by sampling date.

55

56 **Figure S7:** Correlation analysis between marine picocyanobacterial ESTUs and environmental
57 parameters measured along the *Tara* Oceans transect for all sampled depths. (A) *Prochlorococcus*
58 ESTUs, (B) *Synechococcus* ESTUs. The scale shows the degree of correlation (blue) or anti-correlation
59 (red) between the two sets of data. Correlations with adjusted p-value > 0.05 are indicated by grey
60 crosses. Abbreviations: Sal, salinity; Temp, temperature; fCDOM, fluorescence, colored dissolved
61 organic matter; MLD, mixed layer depth; DCM, deep chlorophyll maximum; Φ_{sat} , satellite-based NPQ-
62 corrected quantum yield of fluorescence.

63

64 **Dataset 1:** Summary data for picocyanobacterial *petB* reference sequences used in this study, including
65 newly assembled sequences. The table includes subclade designation based on (12).

66

67 **Dataset 2:** Summary data for *petB* reference sequences for photosynthetic organisms other than
68 marine picocyanobacteria used in this study.

69

70 **Dataset 3:** *Tara* Oceans sample description including the number of recruited *petB* reads per station.
71 Iron and ammonium concentrations were simulated using the ECCO2-Darwin model and an
72 independent parameter to assess iron limitation (Φ_{sat}) was obtained using Behrenfeld et al.'s formula
73 (56) applied to monthly averaged satellite data (AMODIS chl_{ocx}, nflh and ipar) retrieved from the
74 NASA website (<http://oceandata.sci.gsfc.nasa.gov/>) for each station and corresponding sampling date.
75 Other environmental parameters measured during the *Tara* Oceans expedition and the methods used
76 to acquire them are available at <http://www.pangea.de>.

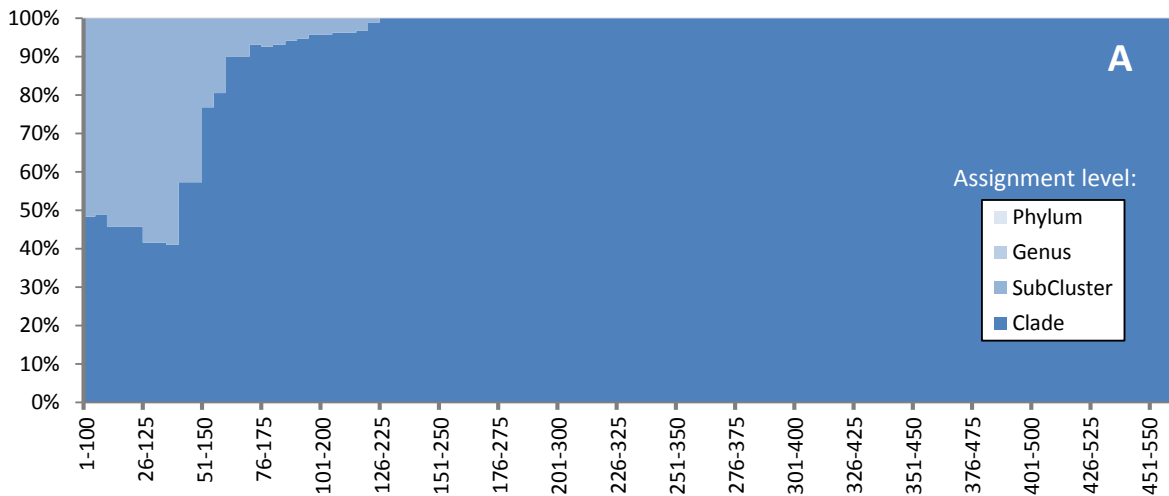
77

78 **Dataset 4:** Sequence names of the members of each Operational Taxonomical Unit (OTU) defined for
79 *petB* at 94% nucleotide sequence identity.

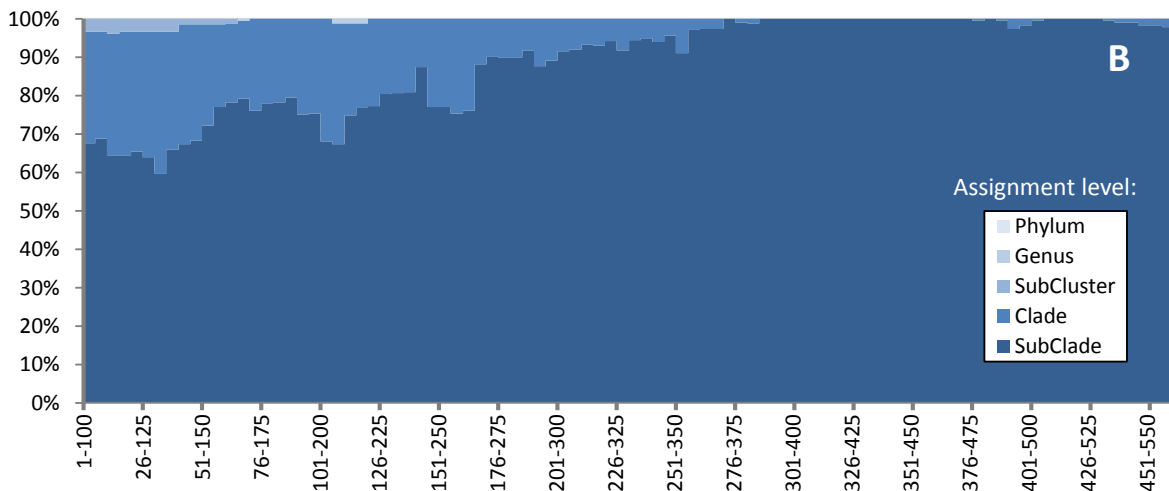
80

81

Prochlorococcus

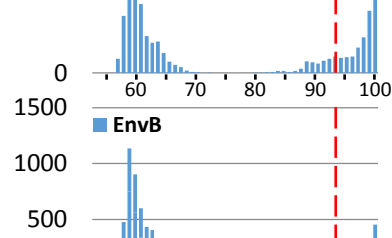
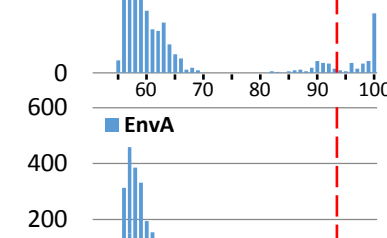
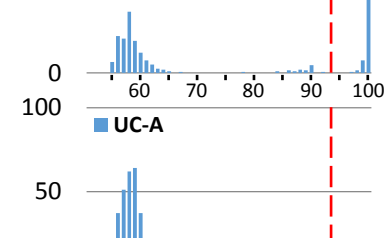
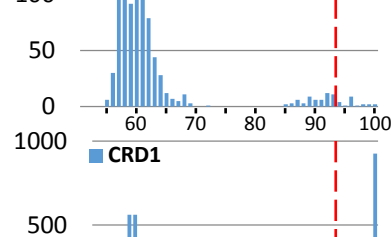
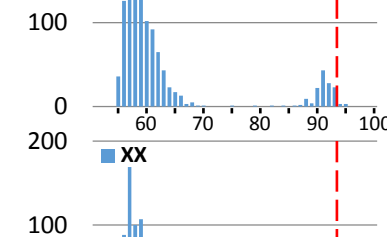
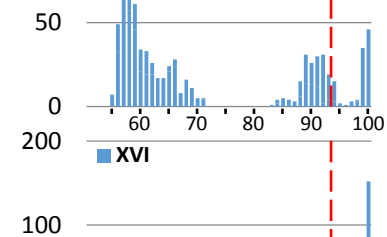
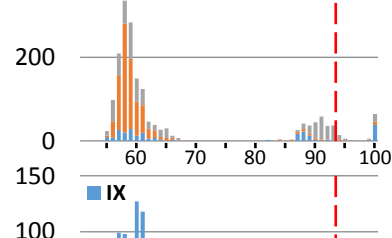
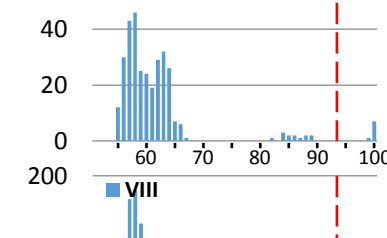
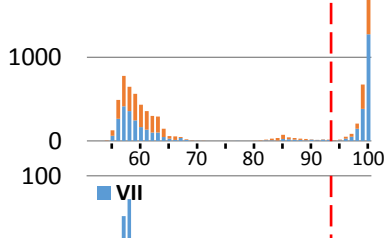
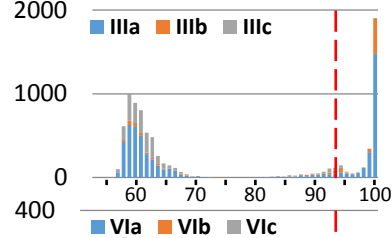
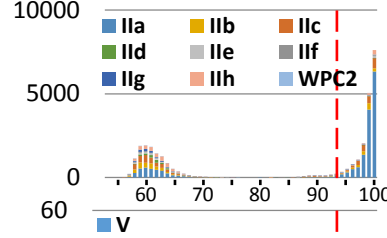
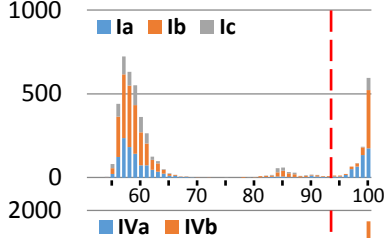
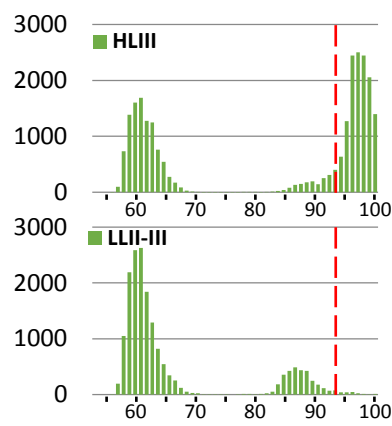
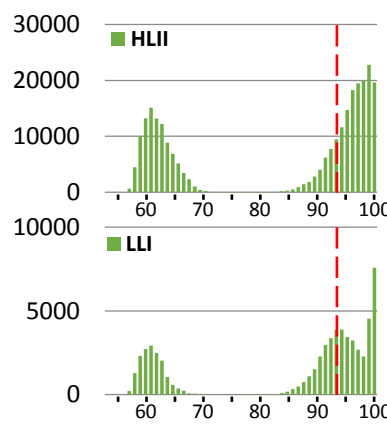
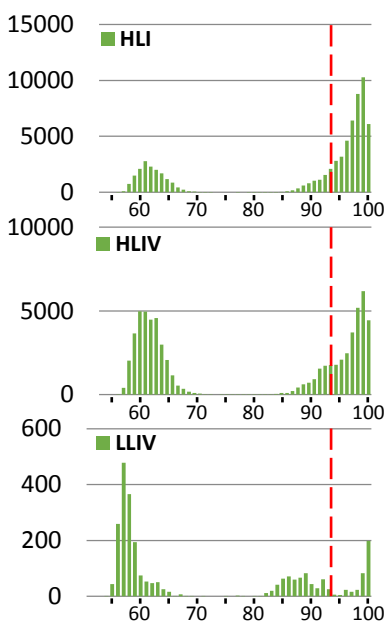


Synechococcus

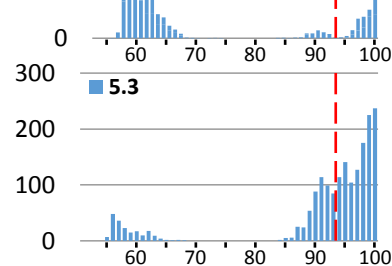
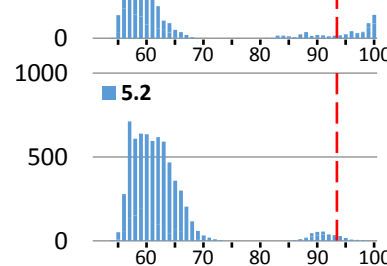
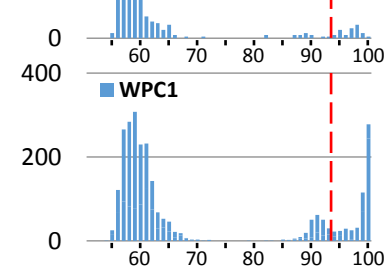


Position of 100 bp simulated reads along the *petB* sequence

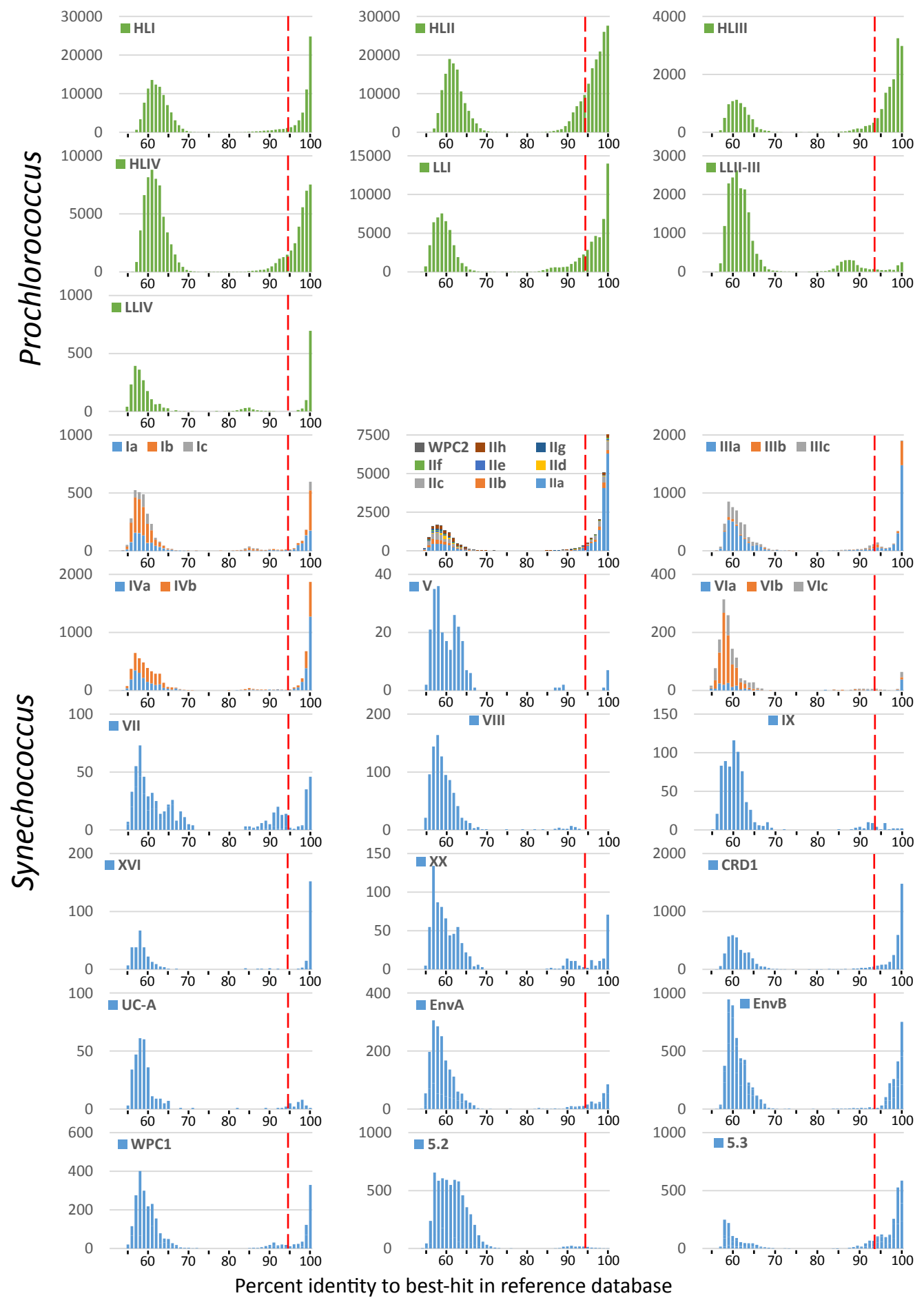
Prochlorococcus



Synechococcus

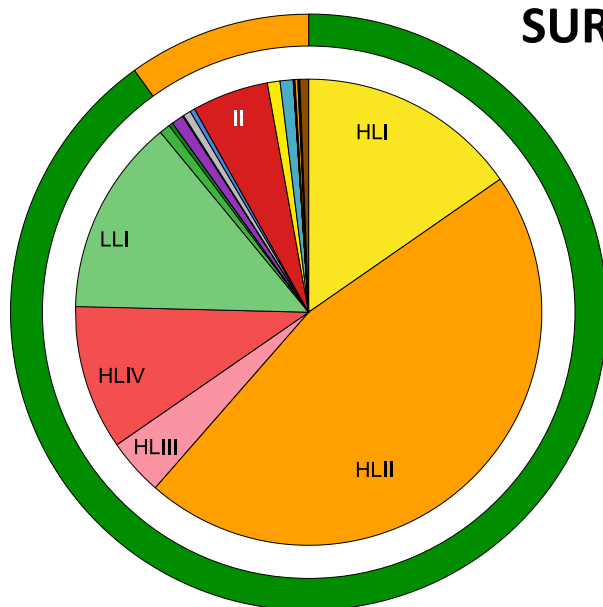


Percent identity to best-hit in reference database



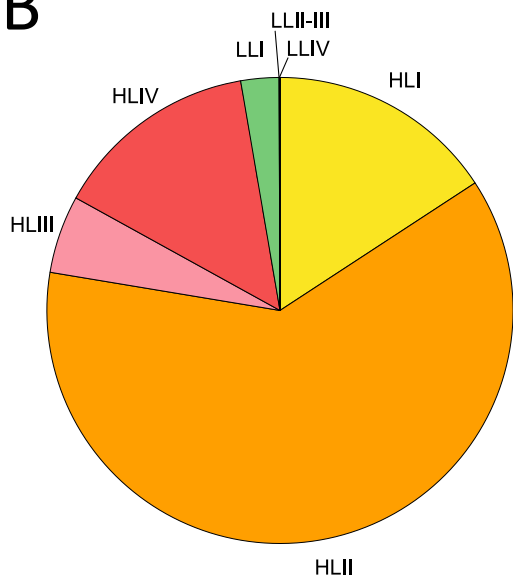
SURFACE+DCM

A



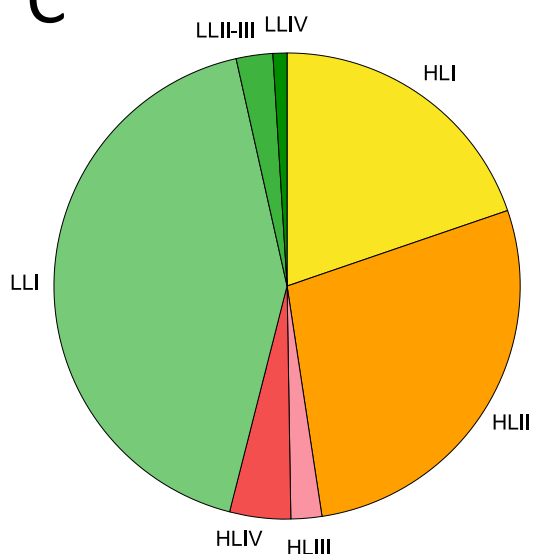
SURFACE

B

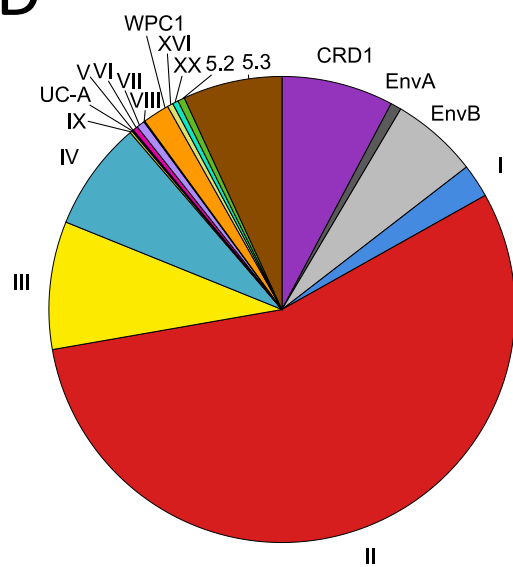


DCM

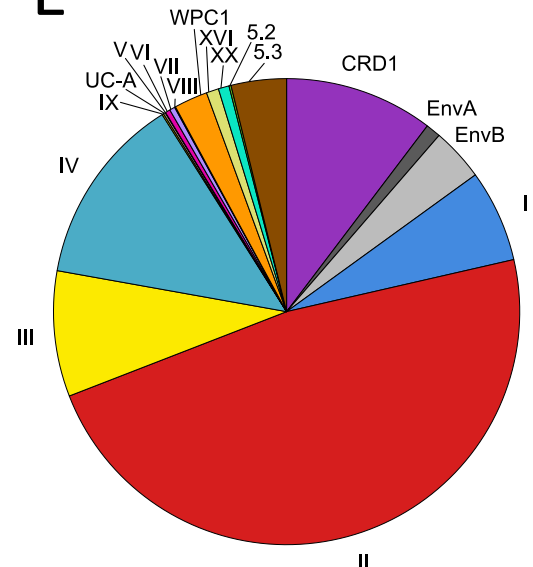
C



D



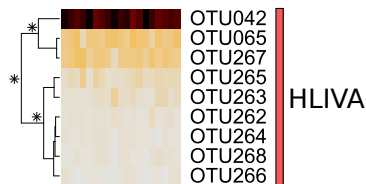
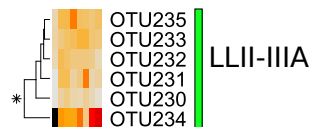
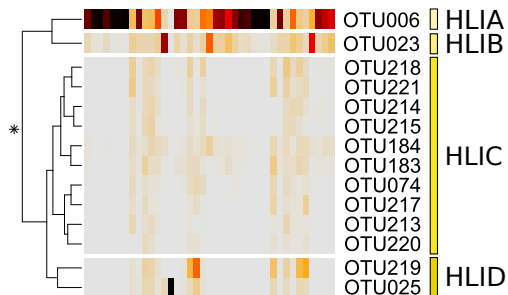
E



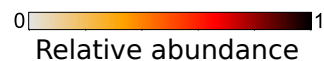
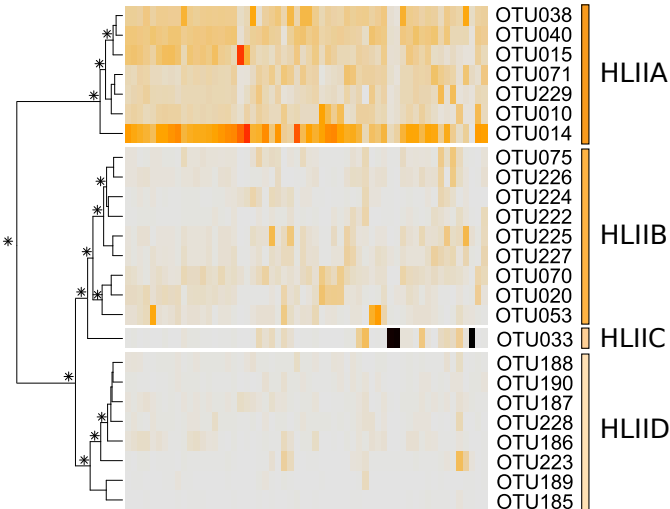
Prochlorococcus

Synechococcus

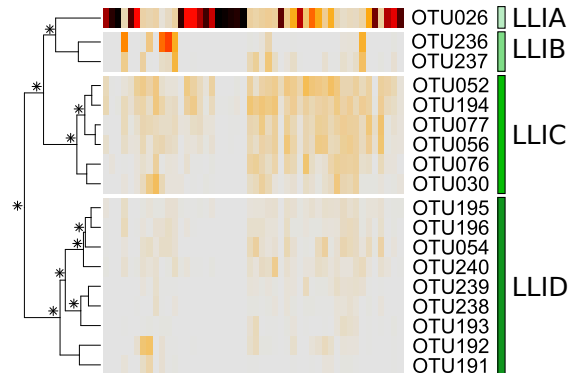
Stations

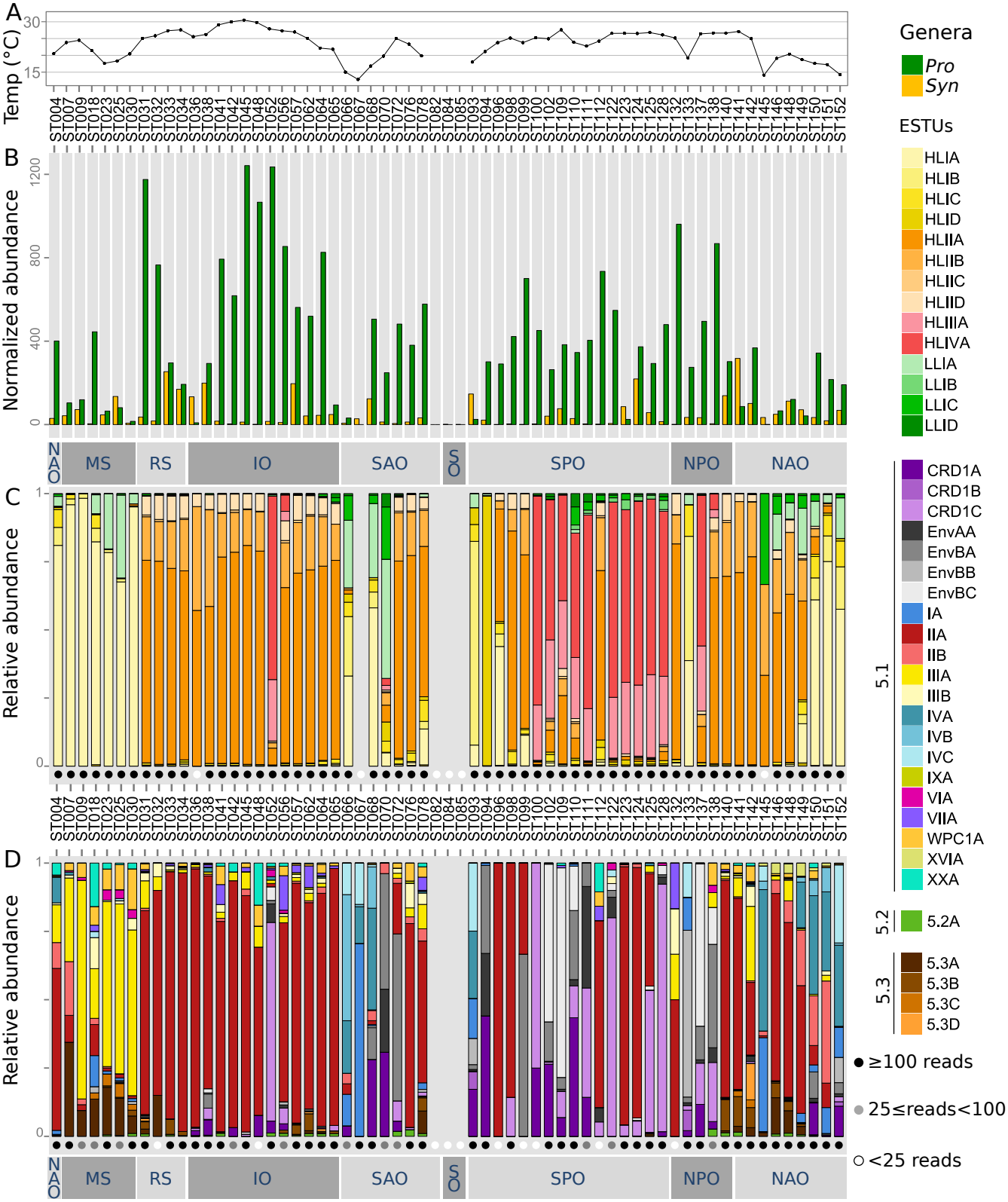


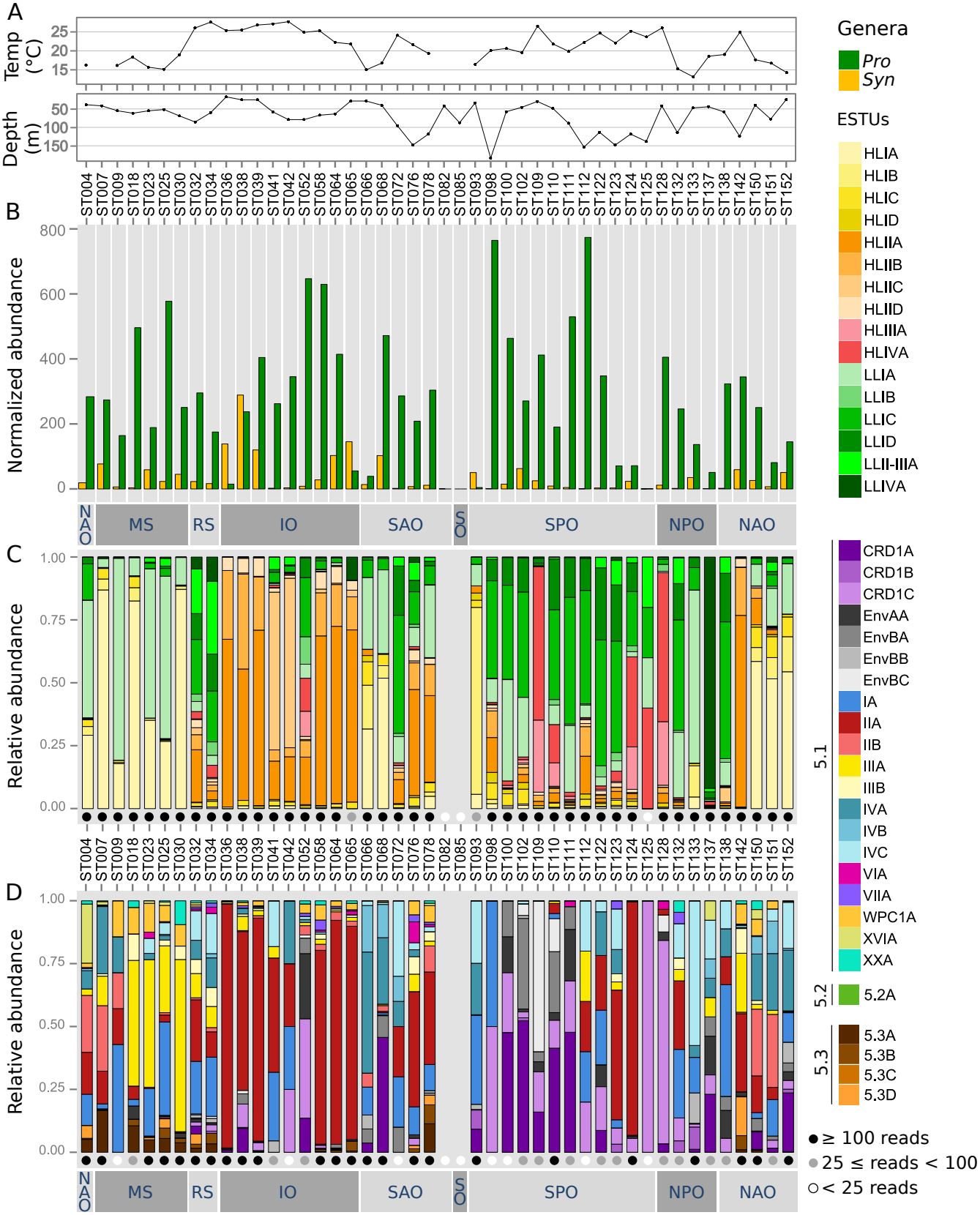
Stations

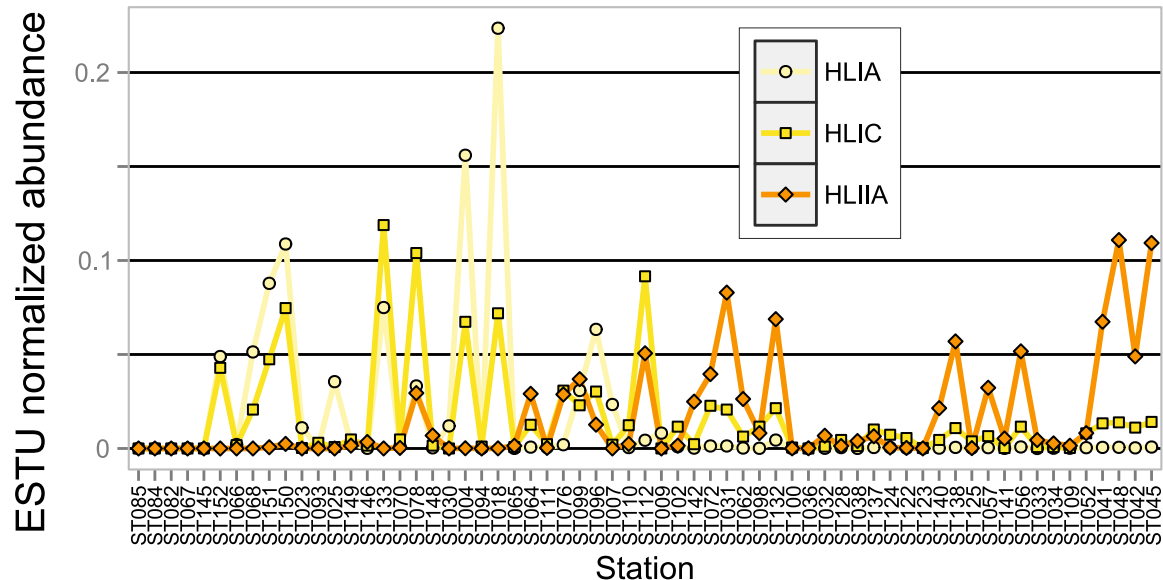
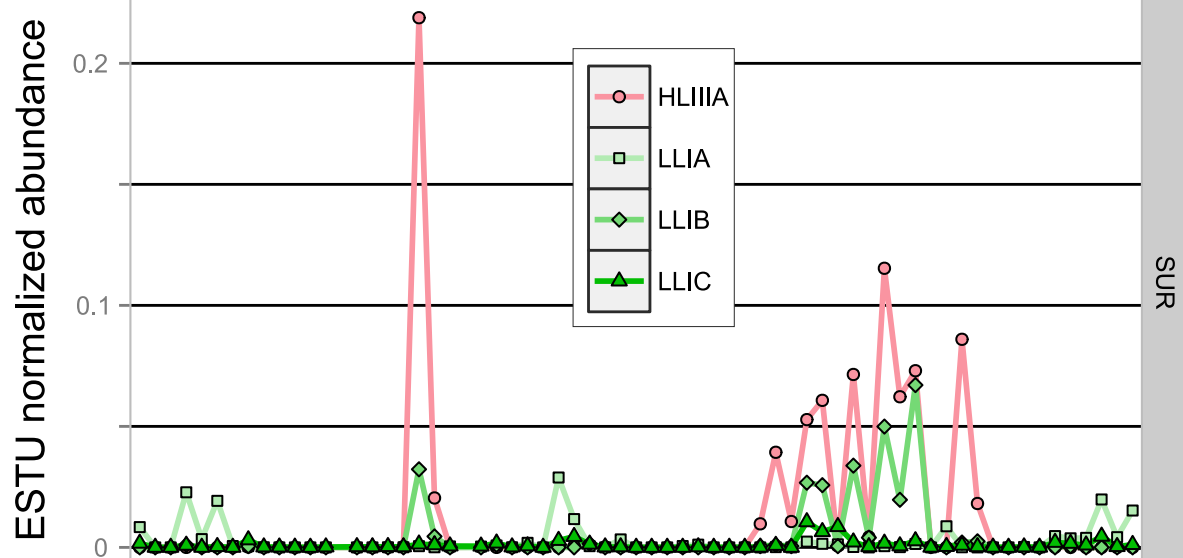


Stations







A**B****C**