



HAL
open science

Differential global distribution of marine picocyanobacteria gene clusters reveals distinct niche-related adaptive strategies

Hugo Doré, Ulysse Guyet, Jade Leconte, Gregory Farrant, Benjamin Alric, Morgane Ratin, Martin Ostrowski, Mathilde Ferrieux, Loraine Brillet-Guéguen, Mark Hoebeke, et al.

► To cite this version:

Hugo Doré, Ulysse Guyet, Jade Leconte, Gregory Farrant, Benjamin Alric, et al.. Differential global distribution of marine picocyanobacteria gene clusters reveals distinct niche-related adaptive strategies. *The International Society of Microbiological Ecology Journal*, 2023, 17 (5), pp.720-732. 10.1038/s41396-023-01386-0 . hal-04238382

HAL Id: hal-04238382

<https://hal.science/hal-04238382v1>

Submitted on 12 Oct 2023

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 Differential global distribution of marine picocyanobacteria gene
2 clusters reveals distinct niche-related adaptive strategies

3 Hugo Doré^{a,1}, Ulysse Guyet^{a,1}, Jade Leconte^a, Gregory K. Farrant^a, Benjamin Alric^a, Morgane
4 Ratin^a, Martin Ostrowski^{b,2}, Mathilde Ferrieux^a, Loraine Brillet-Guéguen^{c,d}, Mark Hoebeke^c, Jukka
5 Siltanen^c, Gildas Le Corguillé^c, Erwan Corre^c, Patrick Wincker^{f,g}, David J. Scanlan^b, Damien
6 Eveillard^{h,g}, Frédéric Partensky^a, and Laurence Garczarek^{a,g,*}

7 ^aSorbonne Université, CNRS, UMR 7144 Adaptation and Diversity in the Marine Environment
8 (AD2M), Station Biologique de Roscoff (SBR), Roscoff, France; ^bSchool of Life Sciences, University
9 of Warwick, Coventry CV4 7AL, UK; ^cCNRS, FR 2424, ABiMS Platform, Station Biologique de Roscoff
10 (SBR), Roscoff, France; ^dSorbonne Université, CNRS, UMR 8227, Integrative Biology of Marine
11 Models (LBI2M), Station Biologique de Roscoff (SBR), Roscoff, France; ^eGenoscope, Institut de
12 biologie François-Jacob, Commissariat à l’Energie Atomique (CEA), Université Paris-Saclay, Evry,
13 France; ^fGénomique Métabolique, Genoscope, Institut de biologie François Jacob, CEA, CNRS,
14 Université d’Evry, Université Paris-Saclay, Evry, France; ^gResearch Federation (FR2022) Tara
15 Océans GO-SEE, Paris, France; ^hNantes Université, Centrale Nantes, CNRS, LS2N, UMR 6004,
16 Nantes, France.

17 ¹H.D. and U.G. contributed equally to this work

18 ²Current address: Climate Change Cluster, University of Technology, Broadway NSW 2007,
19 Australia

20 *To whom correspondence should be addressed. Email: laurence.garczarek@sb-roscoff.fr. phone
21 number: +33 2 98 29 25 38

22 **Running Title:** Niche adaptation strategies in marine picocyanobacteria

23 **Competing Interest Statement:** The authors declare no competing interests.

24 **Keywords:** *Prochlorococcus*, *Synechococcus*, niche partitioning, Tara Oceans, metagenomics

25 **Abstract**

26 The ever-increasing number of available microbial genomes and metagenomes provides new
27 opportunities to investigate the links between niche partitioning and genome evolution in the
28 ocean, especially for the abundant and ubiquitous marine picocyanobacteria *Prochlorococcus* and
29 *Synechococcus*. Here, by combining metagenome analyses of the *Tara* Oceans dataset with
30 comparative genomics, including phyletic patterns and genomic context of individual genes from
31 256 reference genomes, we show that picocyanobacterial communities thriving in different
32 niches possess distinct gene repertoires. We also identify clusters of adjacent genes that display
33 specific distribution patterns in the field (eCAGs) and are thus potentially involved in the same
34 metabolic pathway and may have a key role in niche adaptation. Several eCAGs are likely involved
35 in the uptake or incorporation of complex organic forms of nutrients, such as guanidine, cyanate,
36 cyanide, pyrimidine, or phosphonates, which might be either directly used by cells, for example
37 the biosynthesis of proteins or DNA, or degraded to inorganic nitrogen and/or phosphorus forms.
38 We also highlight the enrichment of eCAGs involved in polysaccharide capsule biosynthesis in
39 *Synechococcus* populations thriving in both nitrogen- and phosphorus-depleted areas vs. low-iron
40 (Fe) regions, suggesting that the complexes they encode may be too energy-consuming for
41 picocyanobacteria thriving in the latter areas. In contrast, *Prochlorococcus* populations thriving in
42 Fe-depleted areas specifically possess an alternative respiratory terminal oxidase, potentially
43 involved in the reduction of Fe(III) to Fe(II). Altogether, this study provides insights into how
44 phytoplankton communities populate oceanic ecosystems, which is relevant to understanding
45 their capacity to respond to ongoing climate change.

46

47 **Introduction**

48 Although phytoplankton communities play a crucial role in marine biogeochemical cycles [1, 2],
49 the relative contribution of different species or ecotypes to these cycles remains difficult to assess
50 due to a lack of knowledge of specific metabolic traits. Indeed, trait-based functional diversity is
51 thought to be a better predictor of ecosystem functioning than species diversity [3, 4], and
52 understanding which metabolic traits have facilitated the adaptation of an ecotype to a particular
53 environment is key to understanding each species' ecological role. Comparative genomics has
54 sometimes been used to try to identify the genetic basis of niche adaptation. However, this

55 approach has revealed only a few genes specific to particular ecotypes and thus potentially
56 involved in niche adaptation, perhaps due to the fairly low number of genomes available (even
57 for major phytoplankton groups) and the poor ecological representation and physiological
58 characterization of available isolates [5–9]. An alternative to better deciphering the links between
59 functional diversity and niche partitioning involves exploiting the rapidly growing number of
60 metagenomes. These can be used to generate metagenome-assembled genomes (MAGs) that can
61 fill the gaps for yet uncultured microbial taxa as well as to identify niche-specific genes, i.e. genes
62 enriched in specific spatial and/or temporal environmental conditions, by recruiting metagenomic
63 reads onto reference genomes [10–16].

64 Due to their abundance and ubiquity in the field and the large number of available genomes,
65 including single amplified genomes (SAGs) and MAGs [6, 17–20], the marine picocyanobacteria
66 *Prochlorococcus* and *Synechococcus* constitute highly pertinent models to study how
67 phytoplankton cells adapt to their variable physico-chemical environment. These genera are
68 indeed the two most abundant members of the phytoplankton community, *Prochlorococcus* being
69 restricted to the 40°S–50°N latitudinal band, whilst the distribution of *Synechococcus* extends from
70 the equator to subpolar waters [21–23]. By combining laboratory and environmental studies
71 scientists have managed to decipher their genetic diversity and delineate specific ecotypes or
72 “ecologically significant taxonomic units” (ESTUs), i.e. genetic groups within clades occupying a
73 specific ecological niche [24–28]. While three major ESTU assemblages were identified for
74 *Prochlorococcus* in surface waters, whose distribution was found to be mainly driven by
75 temperature and iron (Fe) availability, eight distinct assemblages were identified for
76 *Synechococcus* depending on three main environmental parameters: temperature, Fe, and
77 phosphorus (P) availability. Nevertheless, few studies have so far linked knowledge of the
78 distribution of the different ecotypes to their functional diversity in order to identify potential
79 niche-specific genes, based on gene relative abundance in different ecosystems [10, 14, 15, 29,
80 30]. Furthermore, most of these previous studies have focused on the abundance of individual
81 genes, or more rarely, on just a few genomic regions with known function, for example those
82 involved in nitrogen (N) or P uptake and assimilation [31–33].

83 Here, in order to better understand the relationship between biogeochemistry and metabolic
84 traits of marine picocyanobacteria, we searched for global patterns of picocyanobacterial gene

85 distributions. To do so, we used a network approach to integrate metagenome analyses of the
86 oceanwide *Tara* Oceans dataset and synteny of individual accessory genes in 256 reference
87 genomes, SAGs, or MAGs covering the wide diversity existing within *Prochlorococcus* and
88 *Synechococcus*. Using this approach, we identified many clusters of adjacent genes that display
89 distinctive global distribution patterns *in situ* and thus likely play important roles in the adaptation
90 of these bacteria to the main ecological niches they occupy in the ocean. Given that gene synteny
91 is commonly used as an indicator of shared function [34, 35], delineation of these gene clusters
92 should also help to identify the putative function of numerous unknown genes, based on their
93 occurrence alongside functionally annotated genes in the same cluster. Overall, this study
94 provides novel insights into the genetic basis of niche partitioning in key members of the
95 phytoplankton community.

96

97 **Results and Discussion**

98

99 **Different picocyanobacterial communities exhibit distinct gene repertoires**

100 To analyze the distribution of *Prochlorococcus* and *Synechococcus* reads along the *Tara* Oceans
101 transect, metagenomic reads corresponding to the bacterial size fraction were recruited against
102 256 picocyanobacterial reference genomes, including SAGs and MAGs representative of
103 uncultured lineages (e.g. *Prochlorococcus* HLIII-IV, *Synechococcus* EnvA or EnvB). This yielded a
104 total of 1.07 billion recruited reads, of which 87.7% mapped onto *Prochlorococcus* genomes and
105 12.3% onto *Synechococcus* genomes, which were then functionally assigned by mapping them
106 onto the manually curated Cyanorak v2.1 CLOG database [19]. In order to identify
107 picocyanobacterial genes potentially involved in niche adaptation, we analyzed the distribution
108 across the oceans of flexible (i.e. non-core) genes. Clustering of *Tara* Oceans stations according to
109 the relative abundance of flexible genes resulted in three well-defined clusters for
110 *Prochlorococcus* (Fig. 1A), which matched those obtained when stations were clustered according
111 to the relative abundance of *Prochlorococcus* ESTUs, as assessed using the high-resolution marker
112 gene *petB*, encoding cytochrome *b₆* (Fig. 1A; [24]). Only a few discrepancies can be observed
113 between the two trees, including stations TARA-070 that displayed one of the most disparate
114 ESTU compositions and TARA-094, dominated by the rare HLID ESTU (Fig. 1A). Similarly, for
115 *Synechococcus*, most of the eight assemblages of stations discriminated based on the relative

116 abundance of ESTUs (Fig. 1B) were also retrieved in the clustering based on flexible gene
117 abundance, except for a few intra-assemblage switches between stations, notably those
118 dominated by ESTU IIA (Fig. 1B). Despite these few variations, four major clusters can be clearly
119 delineated in both *Synechococcus* trees, corresponding to four broadly defined ecological niches,
120 namely i) cold, nutrient-rich, pelagic or coastal environments (blue and light red in Fig. 1B), ii) Fe-
121 limited environments (purple and grey), iii) temperate, P-depleted, Fe-replete areas (yellow) and
122 iv) warm, N-depleted, Fe-replete regions (dark red). This correspondence between taxonomic and
123 functional information was also confirmed by the high congruence between distance matrices
124 based on ESTU relative abundance and on CLOG relative abundance (p value $< 10^{-4}$, mantel test
125 $r=0.84$ and $r=0.75$ for *Synechococcus* and *Prochlorococcus*, respectively; dataset 1-4). Altogether,
126 this indicates that distinct picocyanobacterial communities, as assessed based on a single
127 taxonomic marker, also display different gene repertoires. As previously suggested for
128 *Prochlorococcus* [36], this strong correlation between taxonomy and gene content strengthens
129 the idea that, in both genera, the evolution of the accessory genome mainly occurs by vertical
130 transmission, with a relatively low extent of lateral gene transfers, although we cannot exclude
131 that the latter events may occur more often within members of a given ecotype.

132

133 **Distribution of flexible genes is tightly linked to environmental parameters and ESTUs**

134 In order to reduce the amount of data and better interpret the global distribution of
135 picocyanobacterial gene content, a correlation network of genes was built for each genus based
136 on relative abundance profiles of genes across *Tara* Oceans samples using Weighted Correlation
137 Network Analysis (WGCNA). This analysis emphasized four main modules of genes for
138 *Prochlorococcus* (Fig. S1A) and five for *Synechococcus* (Fig. S1B), each module being abundant in
139 a different set of stations. These modules were then associated with the available environmental
140 parameters (Figs. 2A-B) and to the relative abundance of *Prochlorococcus* or *Synechococcus* ESTUs
141 at each station (Figs. 2C-D). For instance, the *Prochlorococcus brown* module was strongly
142 correlated with nutrient concentrations, particularly nitrate and phosphate, and strongly anti-
143 correlated with Fe availability (Fig. 2A). This module thus corresponds to genes preferentially
144 found in Fe-limited high-nutrient low-chlorophyll (HNLC) areas. Indeed, the *brown* module
145 *eigengene* (Fig. S1A), i.e. the first principal component of gene abundances at the different
146 stations for this module, representative of the abundance profiles of genes for this module at the

147 different stations, showed the highest abundances at stations TARA-100 to 125, localized in the
148 South and North Pacific Ocean, as well as at TARA-052, a station located close to the northern
149 coast of Madagascar, likely influenced by the Indonesian throughflow originating from the tropical
150 Pacific Ocean [24, 37]. Furthermore, the correlation of the *Prochlorococcus brown* module with
151 the relative abundance of ESTUs at each station showed that it is also strongly associated with the
152 presence of HLIIIA and HLIVA (Fig. 2C), previously shown to constitute the dominant
153 *Prochlorococcus* ESTUs in low-Fe environments [24, 38, 39] but also the LLIB ESTU, found to
154 dominate the LLI population in these low-Fe areas [24]. Altogether, this example and analyses of
155 all other *Prochlorococcus* and *Synechococcus* modules (supplementary text) show that the
156 communities colonizing cold, Fe-, N-, and/or P-depleted niches possess specific gene repertoires
157 potentially involved in their adaptation to these particular environmental conditions.

158

159 **Identification of individual genes potentially involved in niche partitioning**

160 To identify genes relevant for adaptation to a specific set of environmental conditions and
161 enriched in specific ESTU assemblages, we selected the most representative genes from each
162 module (Dataset 5, Figs. 3 and S2). Most genes retrieved this way encode proteins of unknown or
163 hypothetical function (85.7% of 7,485 genes). However, among the genes with a functional
164 annotation (Dataset 6), a large fraction seems to have a function related to their realized
165 environmental niche (Figs. 3 and S2). For instance, many genes involved in the transport and
166 assimilation of nitrite and nitrate (*nirA*, *nirX*, *moaA-C*, *moaE*, *mobA*, *moeA*, *narB*, *M*, *nrtP*; [6]) as
167 well as cyanate, an organic form of nitrogen (*cynA*, *B*, *D*, *S*), are enriched in the *Prochlorococcus*
168 *blue* module, which is correlated with the HLIIA-D ESTU and to low inorganic N, P, and silica levels
169 and anti-correlated with Fe availability (Fig. 2A-C). This is consistent with previous studies showing
170 that while only a few *Prochlorococcus* strains in culture possess the *nirA* gene and even less the
171 *narB* gene, natural *Prochlorococcus* populations inhabiting N-poor areas do possess one or both
172 of these genes [40–42]. Similarly, numerous genes amongst the most representative of
173 *Prochlorococcus brown*, *red* and *turquoise* modules are related to adaptation of HLIIIA/IVA, HLIA
174 and LLIA ESTUs to Fe-limited, cold P-limited, and cold, mixed waters, respectively (Fig. 3).
175 Comparable results were obtained for *Synechococcus*, although the niche delineation was less
176 clear than for *Prochlorococcus* since genes within each module exhibited lower correlations with

177 the module *eigenvalue* (Fig. S2). These results therefore constitute a proof of concept that this
178 network analysis was able to retrieve niche-related genes from metagenomics data.

179

180 **Identification of eCAGs potentially involved in niche partitioning**

181 In order to better understand the function of niche-related genes, notably of the numerous
182 unknowns, we then integrated global distribution data with gene synteny in reference genomes
183 using a network approach (Datasets 7 and 8). This led us to identify clusters of adjacent genes in
184 reference genomes, and thus potentially involved in the same metabolic pathway (Figs. 4, S3 and
185 S4; Dataset 6). These clusters were defined within each module and thus encompass genes with
186 similar distribution and abundance *in situ*. Hereafter, these environmental clusters of adjacent
187 genes will be called “eCAGs”.

188

189 *eCAGs related to nitrogen metabolism*

190 The well-known nitrate/nitrite gene cluster involved in uptake and assimilation of inorganic
191 forms of N (see above), which is present in most *Synechococcus* genomes (Dataset 6), was
192 expectedly not restricted to a particular niche in natural *Synechococcus* populations, as shown by
193 its quasi-absence from WGCNA modules. In *Prochlorococcus*, this cluster is separated into two
194 eCAGs enriched in low-N areas (Fig. S5A-B), most genes being included in Pro-eCAG_002, present
195 in only 13 out of 118 *Prochlorococcus* genomes, while *nirA* and *nirX* form an independent eCAG
196 (Pro-eCAG_001) due to their presence in many more genomes. The quasi-core *ureA-G/urtB-E*
197 genomic region was also found to form a *Prochlorococcus* eCAG (Pro-eCAG_003) that was
198 impoverished in low-Fe compared to other regions (Fig. S5C-D), in agreement with its presence in
199 only two out of six HLIII/IV genomes. We also uncovered several other *Prochlorococcus* and
200 *Synechococcus* eCAGs that seem to be involved in the transport and/or assimilation of more
201 unusual and/or complex forms of nitrogen, which might either be degraded into elementary N
202 molecules or possibly directly used by cells for e.g. the biosynthesis of proteins or DNA. Indeed,
203 we detected in both genera an eCAG (Pro-eCAG_004 and Syn-eCAG_001 ; Figs. S6A-B, Dataset 6)
204 that encompasses *speB2*, an ortholog of *Synechocystis* PCC 6803 *sl1077*, previously annotated as
205 encoding an agmatinase [29, 43] and which was recently characterized as a guanidase that
206 degrades guanidine rather than agmatine to urea and ammonium [44]. *E. coli* produces guanidine

207 under nutrient-poor conditions, suggesting that guanidine metabolism is biologically significant
208 and potentially prevalent in natural environments [44, 45]. Furthermore, the *ykkC* riboswitch
209 candidate, which was shown to specifically sense guanidine and to control the expression of a
210 variety of genes involved in either guanidine metabolism or nitrate, sulfate, or bicarbonate
211 transport, is located immediately upstream of this eCAG in *Synechococcus* reference genomes, all
212 genes of this cluster being predicted by RegPrecise 3.0 to be regulated by this riboswitch (Fig. S6C;
213 [45, 46]). The presence of *hypA* and *B* homologs within this eCAG furthermore suggests that, in
214 the presence of guanidine, these homologs could be involved in the insertion of Ni_2^+ , or another
215 metal cofactor, in the active site of guanidinase. The next three genes of this eCAG, which encode
216 an ABC transporter similar to the TauABC taurine transporter in *E. coli* (Fig. S6C), could be involved
217 in guanidine transport in low-N areas. Of note, the presence in most *Synechococcus/Cyanobium*
218 genomes possessing this eCAG of a gene encoding a putative Rieske Fe-sulfur protein
219 (CK_00002251) downstream of this gene cluster, seems to constitute a specificity compared to
220 the homologous gene cluster in *Synechocystis* sp. PCC 6803. The presence of this Fe-S protein
221 suggests that Fe is used as a cofactor in this system and might explain why this gene cluster is
222 absent from picocyanobacteria thriving in low-Fe areas, while it is present in a large proportion of
223 the population in most other oceanic areas (Figs. S6A-B).

224 Another example of the use of organic N forms concerns compounds containing a cyano
225 radical ($C\equiv N$). The cyanate transporter genes (*cynABD*) were indeed found in a *Prochlorococcus*
226 eCAG (Pro-eCAG_005, also including the conserved hypothetical gene CK_00055128; Fig. S7A-B).
227 While only a small proportion of the *Prochlorococcus* community possesses this eCAG in warm,
228 Fe-replete waters, it is absent from other oceanic areas in accordance with its low frequency in
229 *Prochlorococcus* genomes (present in only two HLI and five HLII genomes). In *Synechococcus* these
230 genes were not included in a module, and thus are not in an eCAG (Dataset 6; Fig. S7C), but seem
231 widely distributed despite their presence in only a few *Synechococcus* genomes (mostly in clade
232 III strains; [6, 47, 48]). Interestingly, we also uncovered a 7-gene eCAG (Pro-eCAG_006 and Syn-
233 eCAG_002), encompassing a putative nitrilase gene (*nitC*), which also suggests that most
234 *Synechococcus* cells and a more variable fraction of the *Prochlorococcus* population could use
235 nitriles or cyanides in warm, Fe-replete waters and more particularly in low-N areas such as the
236 Indian Ocean (Fig. 5A-B). The whole operon (*nitHBCDEFG*; Fig. 5C), called Nit1C, was shown to be

237 upregulated in the presence of cyanide and to trigger an increase in the rate of ammonia
238 accumulation in the heterotrophic bacterium *Pseudomonas fluorescens* [49], suggesting that like
239 cyanate, cyanide could constitute an alternative nitrogen source in marine picocyanobacteria as
240 well. However, given the potential toxicity of these C≡N-containing compounds [50], we cannot
241 exclude that these eCAGs could also be devoted to cell detoxification [45, 47]. Such an example
242 of detoxification has been described for arsenate and chromate that, as analogs of phosphate and
243 sulfate respectively, are toxic to marine phytoplankton and must be actively exported out of the
244 cells [51, 52].

245 We detected the presence of an eCAG encompassing *asnB*, *pyrB2*, and *pydC* (Pro-eCAG_007,
246 Syn-eCAG_003, Fig. S8), which could contribute to an alternative pyrimidine biosynthesis pathway
247 and thus provide another way for cells to recycle complex nitrogen forms. While this eCAG is
248 found in only one fifth of HLII genomes and in quite specific locations for *Prochlorococcus*, notably
249 in the Red Sea, it is found in most *Synechococcus* cells in warm, Fe-replete, N and P-depleted
250 niches, consistent with its phyletic pattern showing its absence only from most clade I, IV, CRD1,
251 and EnvB genomes (Fig. S8; Dataset 6). More generally, most N-uptake and assimilation genes in
252 both genera were specifically absent from Fe-depleted areas, including the *nirA/narB* eCAG for
253 *Prochlorococcus*, as mentioned by Kent et al. [36] as well as guanidinase and nitrilase eCAGs. In
254 contrast, picocyanobacterial populations present in low-Fe areas possess, in addition to the core
255 ammonium transporter *amt1*, a second transporter *amt2*, also present in cold areas for
256 *Synechococcus* (Fig. S9). Additionally, *Prochlorococcus* populations thriving in HNLC areas also
257 possess two amino acid-related eCAGs that are present in most *Synechococcus* genomes, the first
258 one involved in polar amino acid N-II transport (Pro-eCAG_008; *natF-G-H-bgtA*; [53]; Fig. S10A-B)
259 and the second one (*leuDh-soxA-CK_00001744*, Pro-eCAG_009, Fig. S10C-D) that notably
260 encompasses a leucine dehydrogenase, able to produce ammonium from branched-chain amino
261 acids. This highlights the profound difference in N acquisition mechanisms between HNLC regions
262 and Fe-replete, N-depleted areas: the primary nitrogen sources for picocyanobacterial
263 populations dwelling in HNLC areas seem to be ammonium and amino acids, while N acquisition
264 mechanisms are more diverse in N-limited, Fe-replete regions.

265

266 eCAGs related to phosphorus metabolism

267 Adaptation to P depletion has been well documented in marine picocyanobacteria showing that
268 while in P-replete waters *Prochlorococcus* and *Synechococcus* essentially rely on inorganic
269 phosphate acquired by core transporters (PstSABC), strains isolated from low-P regions and
270 natural populations thriving in these areas additionally contain a number of accessory genes
271 related to P metabolism, located in specific genomic islands [6, 14, 30–32, 54]. Here, we indeed
272 found in *Prochlorococcus* an eCAG containing the *phoBR* operon (Pro-eCAG_010) that encodes a
273 two-component system response regulator, as well as an eCAG including the alkaline phosphatase
274 *phoA* (Pro-eCAG_011), both present in virtually the whole *Prochlorococcus* population from the
275 Mediterranean Sea, the Gulf of Mexico and the Western North Atlantic Ocean, which are known
276 to be P-limited [30, 55] (Fig. S11A-B). By comparison, in *Synechococcus*, we only identified the
277 *phoBR* eCAG (Syn-eCAG_005, Fig. S11C) that is systematically present in warm waters whatever
278 the limiting nutrient, in agreement with its phyletic pattern in reference genomes showing its
279 specific absence from cold thermotypes (clades I and IV, Dataset 6). Furthermore, although our
280 analysis did not retrieve them within eCAGs due to the variability of gene content and synteny in
281 this genomic region, even within each genus, several other P-related genes were enriched in low-
282 P areas but partially differed between *Prochlorococcus* and *Synechococcus* (Figs. 3, S2, S11 and
283 Dataset 6). While the genes putatively encoding a chromate transporter (ChrA) and an arsenate
284 efflux pump ArsB were present in both genera in different proportions, a putative transcriptional
285 phosphate regulator related to PtrA (CK_00056804; [56]) was specific to *Prochlorococcus*.
286 *Synechococcus* in contrast harbors a large variety of alkaline phosphatases (PhoX, CK_00005263
287 and CK_00040198) as well as the phosphate transporter SphX (Fig. S11).

288 Phosphonates, i.e. reduced organophosphorus compounds containing C–P bonds that
289 represent up to 25% of the high-molecular-weight dissolved organic P pool in the open ocean,
290 constitute an alternative P form for marine picocyanobacteria [57]. We indeed identified, in
291 addition to the core phosphonate ABC transporter (*phnD1-C1-E1*), a second previously
292 unreported putative phosphonate transporter *phnC2-D2-E2-E3* (Pro-eCAG_012; Fig. 6A). Most of
293 the *Prochlorococcus* population in strongly P-limited areas of the ocean harbored these genes,
294 while they were absent from other areas, consistent with their presence in only a few
295 *Prochlorococcus* and no *Synechococcus* genomes. Furthermore, as previously described [58–60],

296 we found a *Prochlorococcus* eCAG encompassing the *phnYZ* operon involved in C-P bond cleavage,
297 the putative phosphite dehydrogenase *ptxD*, and the phosphite and methylphosphonate
298 transporter *ptxABC* (Pro-eCAG_0013, Dataset 6, and Fig. 6B, [60–62]). Compared to these
299 previous studies that mainly reported the presence of these genes in *Prochlorococcus* cells from
300 the North Atlantic Ocean, here we show that they actually occur in a much larger geographic area,
301 including the Mediterranean Sea, the Gulf of Mexico, and the ALOHA station (TARA_132) in the
302 North Pacific, even though they were present in a fairly low fraction of *Prochlorococcus* cells.
303 These genes occurred in an even larger proportion of the *Synechococcus* population, although not
304 found in an eCAG for this genus (Fig. S12, Dataset 6). *Synechococcus* cells from the Mediterranean
305 Sea, a P-limited area dominated by clade III [24], seem to lack *phnYZ*, in agreement with the
306 phyletic pattern of these genes in reference genomes, showing the absence of this two-gene
307 operon in the sole clade III strain that possesses the *ptxABDC* gene cluster. In contrast, the
308 presence of the complete gene set (*ptxABDC-phnYZ*) in the North Atlantic, at the entrance of the
309 Mediterranean Sea, and in several clade II reference genomes rather suggests that it is primarily
310 attributable to this clade. Altogether, our data indicate that part of the natural populations of
311 both *Prochlorococcus* and *Synechococcus* would be able to assimilate phosphonate and phosphite
312 as alternative P-sources in low-P areas using the *ptxABDC-phnYZ* operon. Yet, the fact that no
313 picocyanobacterial genome except *P. marinus* RS01 (Fig. 6C) possesses both *phnC2-D2-E2-E3* and
314 *phnYZ*, suggests that the phosphonate taken up by the *phnC2-D2-E2-E3* transporter could be
315 incorporated into cell surface phosphonoglycoproteins that may act to mitigate cell mortality by
316 grazing and viral lysis, as recently suggested [63].

317

318 *eCAGs related to iron metabolism*

319 As for macronutrients, it has been hypothesized that the survival of marine picocyanobacteria in
320 low-Fe regions was made possible through several strategies, including the loss of genes encoding
321 proteins that contain Fe as a cofactor, the replacement of Fe by another metal cofactor, and the
322 acquisition of genes involved in Fe uptake and storage [14, 15, 36, 39, 64]. Accordingly, several
323 eCAGs encompassing genes encoding proteins interacting with Fe were found in modules anti-
324 correlated to HNLC regions in both genera. These include three subunits of the (photo)respiratory
325 complex succinate dehydrogenase (*SdhABC*, Pro-eCAG_014, Syn-eCAG_006, Fig. S13; [65]) and

326 Fe-containing proteins encoded in most abovementioned eCAGs involved in N or P metabolism,
327 such as the guanidinase (Fig. S6), the NitC1 (Fig. 5), the *pyrB2* (Fig. S8), the phosphonate (Figs. 6
328 and S12), and the urea and inorganic nitrogen eCAGs (Fig. S5). Most *Synechococcus* cells thriving
329 in Fe-replete areas also possess the *sodT/sodX* eCAG (Syn-eCAG_007, Fig. S14A-B) involved in
330 nickel transport and maturation of the Ni-superoxide dismutase (SodN), these three genes being
331 in contrast core in *Prochlorococcus*. Additionally, *Synechococcus* from Fe-replete areas, notably
332 from the Mediterranean Sea and the Indian Ocean, specifically possess two eCAGs (Syn-eCAG_008
333 and 009; Fig. S14C-D), involved in the biosynthesis of a polysaccharide capsule that appear to be
334 most similar to the *E. coli* groups 2 and 3 *kps* loci [66]. These extracellular structures, known to
335 provide protection against biotic or abiotic stress, were recently shown in *Klebsiella* to provide a
336 clear fitness advantage in nutrient-poor conditions since they were associated with increased
337 growth rates and population yields [67]. However, while these authors suggested that capsules
338 may play a role in Fe uptake, the significant reduction in the relative abundance of *kps* genes in
339 low-Fe compared to Fe-replete areas (t-test p value <0.05 for all genes of the Syn-eCAG_008 and
340 009; Fig. S14C) and their absence in CRD1 strains (Dataset 6) rather suggests that these capsules
341 may be too energy-consuming for some picocyanobacteria thriving in this particular niche, while
342 they may have a more meaningful and previously overlooked role in their adaptation to low-P and
343 low-N niches.

344 Several eCAGs were in contrast enriched in populations dwelling in HNLC environments,
345 dominated by *Prochlorococcus* HLIIIA/HLIVA/LLIB and *Synechococcus* CRD1A/EnvBA ESTUs (Fig.
346 2). The vast majority of *Prochlorococcus* cells thriving in low-Fe regions possess an eCAG
347 encompassing the *ctaC2-D2-E2* operon, also found in 85% of all *Synechococcus* reference
348 genomes, including all CRD1 (Fig. 7; Dataset 6). This eCAG encodes the alternative respiratory
349 terminal oxidase ARTO, a protein complex that has been suggested to be part of a minimal
350 respiratory chain in the cytoplasmic membrane, potentially providing an additional electron sink
351 under Fe-deprived conditions [68, 69]. Furthermore, a *Synechocystis* mutant in which the *ctaD2*
352 and *ctaE2* genes were inactivated was found to display markedly impaired Fe reduction and
353 uptake rates as compared to wild-type cells, suggesting that ARTO is involved in the reduction of
354 Fe(III) to Fe(II) prior to its transport through the plasma membrane via the Fe(II) transporter FeoB

355 [70]. Thus, the presence of the ARTO system appears to represent a major and previously
356 unreported adaptation for *Prochlorococcus* populations thriving in low-Fe areas.

357 Both *Prochlorococcus* and *Synechococcus* thriving in low-Fe waters also possess eCAGs
358 encoding the TonB-dependent siderophore uptake operon *fecDCAB-tonB-exbBD* (Pro-eCAG_015
359 and Syn-eCAG_013-014, Dataset 6). This gene cluster, which is found in a few picocyanobacterial
360 genomes and was previously shown to be anti-correlated with dissolved Fe concentration [14, 15,
361 64], is indeed systematically present in a significant part of the *Prochlorococcus* and
362 *Synechococcus* population in low-Fe areas (Fig. S15). However, it is also present in a small fraction
363 of the populations thriving in the Indian Ocean, consistent with its occurrence in two
364 *Prochlorococcus* HLII and one *Synechococcus* clade II genomes (Dataset 6). Finally, a large
365 proportion of the *Synechococcus* populations in HNLC areas possess i) a large eCAG involved in
366 glycine betaine synthesis and transport (Syn-eCAG_010, Fig. S15A-B; [6, 71, 72]), almost absent
367 from low-N areas, ii) an eCAG encompassing a flavodoxin and a thioredoxin reductase (Syn-
368 eCAG_011, Fig. S16C-D), mostly absent from low-P areas, and iii) the *nfeD-floT1-floT2* eCAG (Syn-
369 eCAG_012, Fig. S17A-B) involved in the production of lipid rafts, potentially affecting cell shape
370 and motility [6, 73].

371

372 eCAGs enriched in cold waters

373 Besides genes involved in nutrient acquisition and metabolism, several *Prochlorococcus* and
374 *Synechococcus* eCAGs were found to be correlated with low temperature. A closer examination
375 of *Prochlorococcus* eCAGs however, shows that their occurrence is not directly related to
376 temperature adaptation but mainly explained by the prevalence at high latitude of either i) the
377 HLIA ESTU (Fig. 2A, C and Fig. 4), the *red* module encompassing most of the above-mentioned
378 eCAGs involved in P-uptake and assimilation pathways, or ii) the LLIA ESTU, present in surface
379 waters at vertically-mixed stations, the *turquoise* module mainly gathering *Prochlorococcus* LL-
380 specific genes, such as Pro-eCAG_017, involved in phycoerythrin-III biosynthesis (*ppeA*, *cpeFTZY*,
381 *unk13*). As concerns *Synechococcus*, although a fairly high number of eCAGs were identified in the
382 *tan* module associated with ESTUs IA and IVA-C (Fig. 2B, D and Fig. S4), only very few are
383 conserved in more than two reference strains and/or have a characterized function (Dataset 6).

384 Among these, at least one eCAG is clearly related to adaptation to cold waters, the orange
385 carotenoid protein (OCP) operon (*ocp-crtW-frp*; Syn-eCAG_016). Indeed, this operon is involved
386 in a photoprotective process [74] that provides cells with the ability to deal with oxidative stress
387 under cold temperatures [75]. Accordingly, our data shows that *Synechococcus* populations
388 colonizing mixed waters at high latitudes or in upwelling areas all possess this eCAG (Fig. S18),
389 highlighting the importance of this photoprotection system in *Synechococcus* populations
390 colonizing cold and temperate areas. *Synechococcus* populations thriving in cold waters also
391 appear to be enriched in eCAGs involved in gene regulation such as transcription factors involved
392 in the regulation of the CA4-A form of the type IV chromatic acclimation process (*fciA-B*; Syn-
393 eCAG_017), consistent with the predominance of *Synechococcus* CA4-A cells in temperate or cold
394 environments [76–78](Dataset 6). Altogether, the fairly low number of eCAGs strongly associated
395 with temperature supports the hypothesis that adaptation to cold temperature is not mediated
396 by evolution of gene content but rather of protein sequences [5, 6, 36, 79].

397

398 **Conclusions**

399 Our analysis of *Prochlorococcus* and *Synechococcus* gene distributions at the global scale using
400 the deeply sequenced metagenomes collected along the *Tara* Oceans expedition transect
401 revealed that each picocyanobacterial community has a specific gene repertoire, with different
402 sets of accessory genes being highly correlated with distinct ESTUs and physicochemical
403 parameters. As previously suggested for *Prochlorococcus* [36], this strong correlation between
404 taxonomy and gene content strengthens the idea that, in both genera, genome evolution mainly
405 occurs by vertical transmission and selective gene retention, and that lateral gene transfers
406 between ecotypes are fairly scarce. By combining information about gene synteny in 256
407 reference genomes with the distribution and abundance of these genes in the field, we further
408 managed to delineate suites of adjacent genes likely involved in the same metabolic pathways
409 that may have a crucial role in adaptation to specific niches. These analyses confirmed previous
410 observations about the niche partitioning of individual genes and a few genomic regions involved
411 in nutrient uptake and assimilation [14, 15, 25, 31, 33, 36, 40, 42]. Most importantly, even though
412 our network approach likely only revealed the lower boundary of the number of eCAGs actually
413 present in different niches, due to the incompleteness of some reference genomes, this approach
414 highlighted that some previously detected individual genes are part of larger genomic regions and

415 unveiled several novel genomic regions. Although we cannot exclude that some genes enriched
416 in a specific niche are not adaptive *per se* but either hitchhiked along with an adaptive gene [80]
417 or occurred from passive transport of ecotype populations outside their niche [79, 81], it is
418 reasonable to assume that many eCAGs identified in the present study could confer cells a fitness
419 benefit in particular niches and were thus retained by natural selection, or in contrast have been
420 counter-selected in certain environments (such as eCAGs that are absent from low-Fe
421 environments). This study revealed the potential importance of the uptake and assimilation of
422 organic forms of nutrients, which might either be directly used by cells e.g. for the biosynthesis of
423 proteins or DNA, or be degraded into inorganic N and/or P forms. Consequently, many eCAGs
424 potentially involved in the uptake and assimilation of complex compounds, such as guanidine,
425 C≡N-containing compounds, or pyrimidine were present in both N- and P-depleted waters, and
426 might constitute an advantage in areas of the world ocean co-limited in these nutrients [30], while
427 they were absent from HNLC areas. Our data also suggests that adaptation to Fe-limitation relies
428 on specific adaptation mechanisms including the reduction of Fe(III) to Fe(II) using ARTO, Fe
429 scavenging using siderophores, as well as reduction of the Fe quota and of energy-consuming
430 mechanisms, such as polysaccharide capsule biosynthesis. Altogether, this study provides novel
431 insights into the genetic basis of niche partitioning in key members of the phytoplankton
432 community. A future challenge will clearly be biochemically characterizing the function of
433 adaptive genes in these eCAGs (Datasets 5 and 6), which are sometimes present in only a few or
434 even a single cultured strain but which can occur in a large part or even the whole *Prochlorococcus*
435 and/or *Synechococcus* population occupying a specific niche *in situ*.

436

437

438 **Materials and Methods**

439 *Tara* Oceans metagenomic reads from surface waters corresponding to the bacterial size-fraction
440 [24, 82] were recruited against 256 reference genomes using MMseqs2 Release 11-e1a1c (76) and
441 then mapped to an extended database, including 722 outgroup cyanobacterial genomes (Dataset
442 9; Supplementary Methods). Picocyanobacterial reads were then taxonomically assigned to either
443 *Prochlorococcus* or *Synechococcus* and functionally assigned to a cluster of likely orthologous
444 genes (CLOGs) as defined in the information system Cyanorak v2.1 [19]. After normalization by
445 gene and read length, filtration steps included the selection of i) samples containing more CLOGs

446 than the average number of genes in a *Synechococcus* or *Prochlorococcus* HL genome, ii) CLOGs
447 with a gene coverage higher than 1 in at least 2 of the selected samples and iii) non-core genes
448 [6].

449 *Tara* Oceans stations were clustered using Ward's minimum variance [83] based on Bray-
450 Curtis similarities between the relative abundance of either CLOG or picocyanobacterial ESTUs as
451 defined based on the *petB* marker gene [24]. CLOG abundance profiles were also used to perform
452 co-occurrence analyses by weighted genes correlation network analysis (WGCNA, [84, 85]) to
453 delineate modules of CLOGs that share a similar distribution pattern. The *eigengene* of each
454 module was then correlated to environmental parameters, retrieved from PANGAEA
455 (www.pangaea.de/), and to the relative abundance of *petB* ESTUs. Furthermore, the most
456 representative genes of each module were identified as those most correlated to the *eigengene*.
457 Finally, we then defined eCAGs within each module by searching adjacent genes (less than 6 genes
458 apart in 80% of the genomes possessing them) in the 256 reference genomes (Datasets 7 and 8)
459 that were used to build a network of the corresponding CLOGs (node) according to the graph
460 embedder (GEM) or the Fruchterman-Reingold layout algorithms implemented in the R package
461 igraph [86].

462 A detailed version of these methods is available in Supplementary Methods.

463

464 **Contributions**

465 HD and LG conceived and supervised the project. MR, MO, GKF, DJS, FP, LG, and PW participated
466 in the generation of genomes and/or metagenomes used in this study. UG, HD, GKF, and JL
467 performed the bioinformatics analyses. LBG, MH, JS, GL, and EC developed the Cyanorak
468 information system and FP, HD, UG, JL, MF, MO, DJS, and LG curated the genome database. UG,
469 BA, and DE performed the network and statistical analyses. HD, UG, FP, DJS, DE, and LG wrote the
470 manuscript.

471

472 **Acknowledgments**

473

474 This work was supported by the French “Agence Nationale de la Recherche” Programs SAMOSA
475 (ANR-13-ADAP-0010), CINNAMON (ANR-17-CE02-0014-01), EFFICACY (ANR-19-CE02-0019), and
476 France Génomique (ANR-10-INBS-09) as well as the European Union program Assemble+ (Horizon
477 2020, under grant agreement number 287589). We acknowledge Christophe Six for his help with
478 cloning some of the *Synechococcus* strains used in this study and Francisco M. Cornejo-Castillo for
479 helpful discussions. We also thank the support and commitment of the *Tara* Oceans coordinators
480 and consortium, Agnès b. and E. Bourgois, the Veolia Environment Foundation, Région Bretagne,
481 Lorient Agglomération, World Courier, Illumina, the EDF Foundation, FRB, the Prince Albert II de
482 Monaco Foundation, the *Tara* schooner, and its captains and crew. *Tara* Oceans would not exist
483 without continuous support from 23 institutes (<http://oceans.taraexpeditions.org>).

484

485 **References**

- 486 1. Field CB, Behrenfeld MJ, Randerson JT, & Falkowski PG. Primary production of the
487 biosphere: Integrating terrestrial and oceanic components. *Science* 1998; **281**: 237–240.
- 488 2. Litchman E, de Tezanos Pinto P, Edwards KF, Klausmeier CA, Kremer CT, Thomas
489 MK. Global biogeochemical impacts of phytoplankton: a trait-based perspective. *J Ecol* 2015;
490 **103**: 1384–1396.
- 491 3. Abonyi A, Horváth Z, Ptasnik R. Functional richness outperforms taxonomic richness in
492 predicting ecosystem functioning in natural phytoplankton communities. *Freshwater Biol* 2018;
493 **63**: 178–186.
- 494 4. Ye L, Chang C-W, Matsuzaki SS, Takamura N, Widdicombe CE, Hsieh C. Functional
495 diversity promotes phytoplankton resource use efficiency. *J Ecol* 2019; **107**: 2353–2363.
- 496 5. Kettler GC, Martiny AC, Huang K, Zucker J, Coleman ML, Rodrigue S, et al. Patterns
497 and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet* 2007; **3**:
498 e231.
- 499 6. Doré H, Farrant GK, Guyet U, Haguait J, Humily F, Ratin M, et al. Evolutionary
500 mechanisms of long-term genome diversification associated with niche partitioning in marine
501 picocyanobacteria. *Front Microbiol* 2020; **11**: 567431.
- 502 7. Nef C, Madoui M-A, Pelletier É, Bowler C. Whole-genome scanning reveals
503 environmental selection mechanisms that shape diversity in populations of the epipelagic diatom
504 *Chaetoceros*. *PLoS Biol* 2022; **20**: e3001893.
- 505 8. Read BA, Kegel J, Klute MJ, Kuo A, Lefebvre SC, Maumus F, et al. Pan genome of the
506 phytoplankton *Emiliania* underpins its global distribution. *Nature* 2013; **499**: 209–213.
- 507 9. Piganeau G, Grimsley N, Moreau H. Genome diversity in the smallest marine
508 photosynthetic eukaryotes. *Res Microbiol* 2011; **162**: 570–577.

- 509 10. Delmont TO, Eren AM. Linking pangenomes and metagenomes: the *Prochlorococcus*
510 metapangenome. *PeerJ* 2018; **6**: e4320.
- 511 11. Tully BJ, Graham ED, Heidelberg JF. The reconstruction of 2,631 draft metagenome-
512 assembled genomes from the global oceans. *Sci Data* 2018; **5**: 170203.
- 513 12. Massana R, López-Escardó D. Metagenome assembled genomes are for eukaryotes too.
514 *Cell Genomics* 2022; **2**: 100130.
- 515 13. Duncan A, Barry K, Daum C, Eloë-Fadrosch E, Roux S, Schmidt K, et al. Metagenome-
516 assembled genomes of phytoplankton microbiomes from the Arctic and Atlantic Oceans.
517 *Microbiome* 2022; **10**: 67.
- 518 14. Garcia CA, Hagstrom GI, Larkin AA, Ustick LJ, Levin SA, Lomas MW, et al. Linking
519 regional shifts in microbial genome adaptation with surface ocean biogeochemistry. *Phil Trans*
520 *Roy Soc B Biol Sci* 2020; **375**: 20190254.
- 521 15. Ahlgren NA, Belisle BS, Lee MD. Genomic mosaicism underlies the adaptation of
522 marine *Synechococcus* ecotypes to distinct oceanic iron niches. *Environ Microbiol* 2020; **22**:
523 1801–1815.
- 524 16. Garcia SL, Stevens SLR, Crary B, Martinez-Garcia M, Stepanauskas R, Woyke T, et al.
525 Contrasting patterns of genome-level diversity across distinct co-occurring bacterial populations.
526 *ISME J* 2018; **12**: 742–755.
- 527 17. Berube PM, Biller SJ, Hackl T, Hogle SL, Satinsky BM, Becker JW, et al. Single cell
528 genomes of *Prochlorococcus*, *Synechococcus*, and sympatric microbes from diverse marine
529 environments. *Sci Data* 2018; **5**: 180154.
- 530 18. Biller SJ, Berube PM, Dooley K, Williams M, Satinsky BM, Hackl T, et al. Marine
531 microbial metagenomes sampled across space and time. *Sci Data* 2018; **5**: 180176.
- 532 19. Garczarek L, Guyet U, Doré H, Farrant GK, Hoebeker M, Brillet-Guéguen L, et al.
533 Cyanorak v2.1: a scalable information system dedicated to the visualization and expert curation
534 of marine and brackish picocyanobacteria genomes. *Nucl Acids Res* 2021; **49**: D667–D676.
- 535 20. Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A, et al. Single-
536 cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science*
537 2014; **344**: 416–420.
- 538 21. Flombaum P, Gallegos JL, Gordillo R a, Rincón J, Zabala LL, Jiao N, et al. Present and
539 future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*.
540 *Proc Natl Acad Sci USA* 2013; **110**: 9824–9.
- 541 22. Visintini N, Martiny AC, Flombaum P. *Prochlorococcus*, *Synechococcus*, and
542 picoeukaryotic phytoplankton abundances in the global ocean. *Limnol Oceanogr* 2021; **6**: 207–
543 215.
- 544 23. Paulsen ML, Doré H, Garczarek L, Seuthe L, Müller O, Sandaa R-A, et al.
545 *Synechococcus* in the Atlantic gateway to the Arctic Ocean. *Front Mar Sci* 2016; **3**: 191.
- 546 24. Farrant GK, Doré H, Cornejo-Castillo FM, Partensky F, Ratin M, Ostrowski M, et al.

- 547 Delineating ecologically significant taxonomic units from global patterns of marine
548 picocyanobacteria. *Proc Natl Acad Sci USA* 2016; **113**: E3365–E3374.
- 549 25. Sohm JA, Ahlgren NA, Thomson ZJ, Williams C, Moffett JW, Saito MA, et al. Co-
550 occurring *Synechococcus* ecotypes occupy four major oceanic regimes defined by temperature,
551 macronutrients and iron. *ISME J* 2016; **10**: 333–345.
- 552 26. Ahlgren NA, Rocap G. Diversity and distribution of marine *Synechococcus*: Multiple
553 gene phylogenies for consensus classification and development of qPCR assays for sensitive
554 measurement of clades in the ocean. *Front Microbiol* 2012; **3**: 213–213.
- 555 27. Biller SJ, Berube PM, Lindell D, Chisholm SW. *Prochlorococcus*: The structure and
556 function of collective diversity. *Nature Rev Microbiol* 2015; **13**: 13–27.
- 557 28. Huang S, Wilhelm SW, Harvey HR, Taylor K, Jiao N, Chen F. Novel lineages of
558 *Prochlorococcus* and *Synechococcus* in the global oceans. *ISME J* 2012; **6**: 285–297.
- 559 29. Kent AG, Baer SE, Mouginito C, Huang JS, Larkin AA, Lomas MW, et al. Parallel
560 phylogeography of *Prochlorococcus* and *Synechococcus*. *ISME J* 2019; **13**: 430–441.
- 561 30. Ustick LJ, Larkin AA, Garcia CA, Garcia NS, Brock ML, Lee JA, et al. Metagenomic
562 analysis reveals global-scale patterns of ocean nutrient limitation. *Science* 2021; **372**: 287–291.
- 563 31. Martiny AC, Coleman ML, Chisholm SW. Phosphate acquisition genes in
564 *Prochlorococcus* ecotypes: Evidence for genome-wide adaptation. *Proc Natl Acad Sci USA* 2006;
565 **103**: 12552–12557.
- 566 32. Martiny AC, Huang Y, Li W. Occurrence of phosphate acquisition genes in
567 *Prochlorococcus* cells from different ocean regions. *Environ Microbiol* 2009; **11**: 1340–1347.
- 568 33. Kashtan N, Roggensack SE, Berta-Thompson JW, Grinberg M, Stepanauskas R,
569 Chisholm SW. Fundamental differences in diversity and genomic population structure between
570 Atlantic and Pacific *Prochlorococcus*. *ISME J* 2017; **11**: 1997–2011.
- 571 34. Zheng Y, Roberts RJ, Kasif S. Genomic functional annotation using co-evolution profiles
572 of gene clusters. *Genome Biol* 2002; **3**: research0060.1.
- 573 35. Overbeek R, Fonstein M, D’Souza M, Pusch GD, Maltsev N. The use of gene clusters to
574 infer functional coupling. *Proc Natl Acad Sci USA* 1999; **96**: 2896–2901.
- 575 36. Kent AG, Dupont CL, Yooseph S, Martiny AC. Global biogeography of
576 *Prochlorococcus* genome diversity in the surface ocean. *ISME J* 2016; **10**: 1856–1865.
- 577 37. Song Q, Gordon AL, Visbeck M. Spreading of the Indonesian throughflow in the Indian
578 Ocean. *J Phys Oceanogr* 2004; **34**: 772–792.
- 579 38. West NJ, Lebaron P, Strutton PG, Suzuki MT. A novel clade of *Prochlorococcus* found
580 in high nutrient low chlorophyll waters in the South and Equatorial Pacific Ocean. *ISME J* 2011;
581 **5**: 933–944.
- 582 39. Rusch DB, Martiny AC, Dupont CL, Halpern AL, Venter JC. Characterization of
583 *Prochlorococcus* clades from iron-depleted oceanic regions. *Proc Natl Acad Sci USA* 2010; **107**:

- 584 16184–16189.
- 585 40. Martiny AC, Kathuria S, Berube PM. Widespread metabolic potential for nitrite and
586 nitrate assimilation among *Prochlorococcus* ecotypes. *Proc Natl Acad Sci USA* 2009; **106**:
587 10787–10792.
- 588 41. Berube PM, Rasmussen A, Braakman R, Stepanauskas R, Chisholm SW. Emergence of
589 trait variability through the lens of nitrogen assimilation in *Prochlorococcus*. *eLife* 2019; **8**:
590 e41043–e41043.
- 591 42. Berube PM, Biller SJ, Kent AG, Berta-Thompson JW, Roggensack SE, Roache-Johnson
592 KH, et al. Physiology and evolution of nitrate acquisition in *Prochlorococcus*. *ISME J* 2015.
- 593 43. Burnat M, Li B, Kim SH, Michael AJ, Flores E. Homospermidine biosynthesis in the
594 cyanobacterium *Anabaena* requires a deoxyhypusine synthase homologue and is essential for
595 normal diazotrophic growth. *Mol Microbiol* 2018; **109**: 763–780.
- 596 44. Wang B, Xu Y, Wang X, Yuan JS, Johnson CH, Young JD, et al. A guanidine-degrading
597 enzyme controls genomic stability of ethylene-producing cyanobacteria. *Nat Commun* 2021; **12**:
598 5150.
- 599 45. Nelson JW, Atilho RM, Sherlock ME, Stockbridge RB, Breaker RR. Metabolism of free
600 guanidine in Bacteria is regulated by a widespread riboswitch class. *Mol Cell* 2017; **65**: 220–230.
- 601 46. Novichkov PS, Kazakov AE, Ravcheev DA, Leyn SA, Kovaleva GY, Sutormin RA, et al.
602 RegPrecise 3.0 – A resource for genome-scale exploration of transcriptional regulation in
603 bacteria. *BMC Genomics* 2013; **14**: 745.
- 604 47. Kamennaya NA, Post AF. Characterization of cyanate metabolism in marine
605 *Synechococcus* and *Prochlorococcus* spp. *Appl Environ Microbiol* 2011; **77**: 291–301.
- 606 48. Kamennaya NA, Post AF. Distribution and expression of the cyanate acquisition
607 potential among cyanobacterial populations in oligotrophic marine waters. *Limnol Oceanogr*
608 2013; **58**: 1959–1971.
- 609 49. Jones LB, Ghosh P, Lee J-H, Chou C-N, Kunz DAY 2018. Linkage of the Nit1C gene
610 cluster to bacterial cyanide assimilation as a nitrogen source. *Microbiol* 2018; **164**: 956–968.
- 611 50. Pablo F, Stauber JL, Buckney RT. Toxicity of cyanide and cyanide complexes to the
612 marine diatom *Nitzschia closterium*. *Water Res* 1997; **31**: 2435–2442.
- 613 51. Saunders JK, Rocap G. Genomic potential for arsenic efflux and methylation varies
614 among global *Prochlorococcus* populations. *ISME J* 2016; **10**: 197–209.
- 615 52. Riedel GF. Influence of salinity and sulfate on the toxicity of chromium(vi) to the
616 estuarine diatom *Thalassiosira Pseudonana*. *J Phycol* 1984; **20**: 496–500.
- 617 53. Pernil R, Picossi S, Mariscal V, Herrero A, Flores E. ABC-type amino acid uptake
618 transporters Bgt and N-II of *Anabaena* sp. strain PCC 7120 share an ATPase subunit and are
619 expressed in vegetative cells and heterocysts. *Mol Microbiol* 2008; **67**: 1067–1080.
- 620 54. Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, Delong EF, et al.

- 621 Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* 2006; **311**: 1768–
622 1770.
- 623 55. Moore CM, Mills MM, Arrigo KR, Berman-Frank I, Bopp L, Boyd PW, et al. Processes
624 and patterns of oceanic nutrient limitation. *Nature Geosci* 2013; **6**: 701–710.
- 625 56. Tetu SG, Brahmsha B, Johnson DA, Tai V, Phillippy K, Palenik B, et al. Microarray
626 analysis of phosphate regulation in the marine cyanobacterium *Synechococcus* sp. WH8102.
627 *ISME J* 2009; **3**: 835–849.
- 628 57. Clark LL, Ingall ED, Benner R. Marine phosphorus is selectively remineralized. *Nature*
629 1998; **393**: 426–426.
- 630 58. Feingersch R, Philosof A, Mejuch T, Glaser F, Alalouf O, Shoham Y, et al. Potential for
631 phosphite and phosphonate utilization by *Prochlorococcus*. *ISME J* 2012; **6**: 827–834.
- 632 59. Martinez A, Tyson GW, DeLong EF. Widespread known and novel phosphonate
633 utilization pathways in marine bacteria revealed by functional screening and metagenomic
634 analyses. *Environ Microbiol* 2010; **12**: 222–238.
- 635 60. Sosa OA, Casey JR, Karl DM. Methylphosphonate oxidation in *Prochlorococcus* strain
636 MIT9301 supports phosphate acquisition, formate excretion, and carbon assimilation into purines.
637 *Appl Environ Microbiol* 2019; **85**: e00289-19.
- 638 61. Martínez A, Osburne MS, Sharma AK, DeLong EF, Chisholm SW. Phosphite utilization
639 by the marine picocyanobacterium *Prochlorococcus* MIT9301. *Environ Microbiol* 2012; **14**:
640 1363–1377.
- 641 62. McSorley FR, Wyatt PB, Martinez A, DeLong EF, Hove-Jensen B, Zechel DL. PhnY and
642 PhnZ comprise a new oxidative pathway for enzymatic cleavage of a carbon–phosphorus bond. *J*
643 *Am Chem Soc* 2012; **134**: 8364–8367.
- 644 63. Acker M, Hogle SL, Berube PM, Hackl T, Coe A, Stepanauskas R, et al. Phosphonate
645 production by marine microbes: Exploring new sources and potential function. *Proc Natl Acad*
646 *Sci USA* 2022; **119**: e2113386119.
- 647 64. Malmstrom RR, Rodrigue S, Huang KH, Kelly L, Kern SE, Thompson A, et al. Ecology
648 of uncultured *Prochlorococcus* clades revealed through single-cell genomics and biogeographic
649 analysis. *ISME J* 2013; **7**: 184–198.
- 650 65. Cooley JW, Vermaas WFJ. Succinate dehydrogenase and other respiratory pathways in
651 thylakoid membranes of *Synechocystis* sp. strain PCC 6803: capacity comparisons and
652 physiological function. *J Bacteriol* 2001; **183**: 4251–4258.
- 653 66. Whitfield C. Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*.
654 *Annu Rev Biochem* 2006; **75**: 39–68.
- 655 67. Buffet A, Rocha EPC, Rendueles O. Nutrient conditions are primary drivers of bacterial
656 capsule maintenance in *Klebsiella*. *Proc Roy Soc B Biol Sci* 2021; **288**: 20202876.
- 657 68. Lea-Smith DJ, Ross N, Zori M, Bendall DS, Dennis JS, Scott SA, et al. Thylakoid
658 terminal oxidases are essential for the cyanobacterium *Synechocystis* sp. PCC 6803 to survive

659 rapidly changing light intensities. *Plant Physiol* 2013; **162**: 484–495.

660 69. Lea-Smith DJ, Bombelli P, Vasudevan R, Howe CJ. Photosynthetic, respiratory and
661 extracellular electron transport pathways in cyanobacteria. *Biochim Biophys Acta Bioenerget*
662 2016; **1857**: 247–255.

663 70. Kranzler C, Lis H, Finkel OM, Schmetterer G, Shaked Y, Keren N. Coordinated
664 transporter activity shapes high-affinity iron acquisition in cyanobacteria. *ISME J* 2014; **8**: 409–
665 417.

666 71. Scanlan DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR, et al.
667 Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol Rev* 2009; **73**: 249–299.

668 72. Ford BA, Ranjit P, Mabbutt BC, Paulsen IT, Shah BS. ProX from marine *Synechococcus*
669 spp. show a sole preference for glycine-betaine with differential affinity between ecotypes.
670 *Environ Microbiol* 2022; **24**: 6071–6085.

671 73. Dempwolff F, Wischhusen HM, Specht M, Graumann PL. The deletion of bacterial
672 dynamin and flotillin genes results in pleiotrophic effects on cell division, cell growth and in cell
673 shape maintenance. *BMC Microbiol* 2012; **12**: 298.

674 74. Boulay C, Wilson A, D’Haene S, Kirilovsky D. Identification of a protein required for
675 recovery of full antenna capacity in OCP-related photoprotective mechanism in cyanobacteria.
676 *Proc Natl Acad Sci USA* 2010; **107**: 11620–11625.

677 75. Six C, Ratin M, Marie D, Corre E. Marine *Synechococcus* picocyanobacteria: Light
678 utilization across latitudes. *Proc Natl Acad Sci USA* 2021; **118**.

679 76. Xia X, Partensky F, Garczarek L, Suzuki K, Guo C, Yan Cheung S, et al.
680 Phylogeography and pigment type diversity of *Synechococcus* cyanobacteria in surface waters of
681 the northwestern Pacific Ocean. *Environ Microbiol* 2017; **19**: 142–158.

682 77. Grébert T, Doré H, Partensky F, Farrant GK, Boss ES, Picheral M, et al. Light color
683 acclimation is a key process in the global ocean distribution of *Synechococcus* cyanobacteria.
684 *Proc Natl Acad Sci USA* 2018; **115**: E2010–E2019.

685 78. Sanfilippo JE, Nguyen AA, Karty JA, Shukla A, Schluchter WM, Garczarek L, et al.
686 Self-regulating genomic island encoding tandem regulators confers chromatic acclimation to
687 marine *Synechococcus*. *Proc Natl Acad Sci USA* 2016; **113**: 6077–6082.

688 79. Larkin AA, Martiny AC. Microdiversity shapes the traits, niche space, and biogeography
689 of microbial taxa: The ecological function of microdiversity. *Environ Microbiol Rep* 2017; **9**: 55–
690 70.

691 80. Barton NH. Genetic hitchhiking. *Philosophical Transactions of the Royal Society of*
692 *London Series B: Biological Sciences* 2000; **355**: 1553–1562.

693 81. Wiedenbeck J, Cohan FM. Origins of bacterial diversity through horizontal genetic
694 transfer and adaptation to new ecological niches. *FEMS Microbiol Rev* 2011; **35**: 957–976.

695 82. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, et al. Structure
696 and function of the global ocean microbiome. *Science* 2015; **348**: 1261359–1261359.

- 697 83. Szmrecsanyi B. Grammatical variation in British English dialects: A study in corpus-
698 based dialectometry. 2012. Cambridge University Press, Cambridge.
- 699 84. Zhang B, Horvath S. A general framework for weighted gene co-expression network
700 analysis. *Stat Appl Genet Mol Biol* 2005; **4**: Article17.
- 701 85. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network
702 analysis. *BMC Bioinfo* 2008; **9**: 559.
- 703 86. Csardi G, Nepusz T. The igraph software package for complex network research.
704 *InterJournal, Complex Systems*. 2006. 1695
- 705
- 706

707 **Figure Legends**

708

709 **Figure 1. Comparison of clustering based on relative abundance profiles of ecologically**
710 **significant taxonomic units (ESTUs) and of flexible genes for both picocyanobacteria.** A.
711 *Prochlorococcus*. B. *Synechococcus*. Leaves of the trees correspond to stations along the *Tara*
712 Oceans transect that are colored according to the code shown at the bottom of the trees, which
713 corresponds to ESTU assemblages as determined previously [24] by clustering stations exhibiting
714 similar ESTU relative abundance profiles shown here on the right of each tree (for global
715 distribution maps of ESTU assemblages, see Fig. 3B and 4B in [24]). ESTUs were colored according
716 to the palette below each panel. Dotted lines in dendrograms indicate discrepancies between tree
717 topologies. Accessory genes correspond to all genes except those defined as large-core genes in
718 a previous study [6]. Of note, due to a slightly different clustering method (cf. materials and
719 methods), assemblage 7 (dark grey stations in 1B), which was discriminated from assemblage 6 in
720 the Farrant et al. (2016) now clusters with this assemblage. Abbreviations: IO, Indian Ocean; MS,
721 Mediterranean Sea; NAO, North Atlantic Ocean; NPO, North Pacific Ocean; RS, Red Sea; SAO,
722 South Atlantic Ocean; SO, Southern Ocean.

723

724 **Figure 2. Correlation of picocyanobacterial module eigengenes to physico-chemical parameters**
725 **and ESTU abundance.** A, B. Correlation of module eigengenes to physico-chemical parameters for
726 *Prochlorococcus* (A) and *Synechococcus* (B). C, D. Correlation of module eigengenes to relative
727 abundance profiles of ESTUs *sensu* [4]. Pearson (A, B) and Spearman (C, D) correlation coefficients
728 (r and ρ , respectively) are indicated by the color scale. Each module is identified by a specific
729 color and the number between brackets specifies the number of genes in each module. The
730 *eigengene* is representative of the relative abundance of genes of a given module at each *Tara*
731 Oceans station. Non-significant correlations (Student asymptotic p value > 0.01) are marked by a
732 cross. Φ_{sat} : index of iron limitation derived from satellite data. PAR30: satellite-derived
733 photosynthetically available radiation at the surface, averaged on 30 days. DCM: depth of the
734 deep chlorophyll maximum.

735

736 **Figure 3. Violin plots highlighting the most representative genes of each *Prochlorococcus***
737 **module.** For each module, each gene is represented as a dot positioned according to its

738 correlation with the eigengene for each module, the most representative genes being localized
739 on top of each violin plot. Genes mentioned in the text and/or in Dataset 6 have been colored
740 according to the color of the corresponding module, indicated by a colored bar above each
741 module. The text above violin plots indicates the most significant environmental parameter(s)
742 and/or ESTU(s) for each module, as derived from Fig. 2.

743

744 **Figure 4. Delineation of *Prochlorococcus* eCAGs, defined as a set of genes that are both adjacent**
745 **in reference genomes and share a similar *in situ* distribution.** Nodes correspond to individual
746 genes with their gene name (or significant numbers of the CK number, e.g. 1234 for CK_00001234)
747 and are colored according to their WGCNA module. A link between two nodes indicates that these
748 two genes are less than five genes apart in at least one genome. The bottom insert shows the
749 most significant environmental parameter(s) and/or ESTU(s) for each module, as derived from Fig.
750 2.

751

752 **Figure 5. Global distribution map of the eCAG involved in nitrile or cyanide transport and**
753 **assimilation.** (A) *Prochlorococcus* (Pro-eCAG_006) and (B) *Synechococcus* Syn-eCAG_002. (C) The
754 genomic region in *Prochlorococcus marinus* MIT9301. The size of the circle is proportional to
755 relative abundance of each genus as estimated based on the single-copy core gene *petB* and this
756 gene was also used to estimate the relative abundance of other genes in the population. Black
757 dots represent *Tara* Oceans stations for which *Prochlorococcus* or *Synechococcus* read abundance
758 was too low to reach the threshold limit.

759

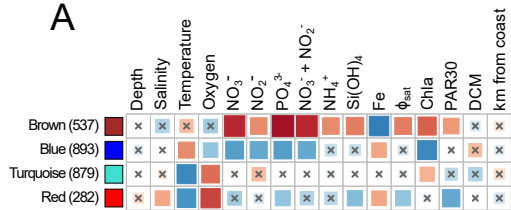
760 **Figure 6. Global distribution map of eCAGs putatively involved in phosphonate and phosphite**
761 **transport and assimilation.** *Prochlorococcus* (A) Pro-eCAG_012 putatively involved in
762 phosphonate transport, (B) Pro-eCAG_013, involved in phosphonate/phosphite uptake and
763 assimilation and phosphonate C-P bond cleavage, (C) The genomic region encompassing both
764 *phnC2-D2-E2-E3* and *ptxABDC-phnYZ* specific to *P. marinus* RS01. The size of the circle is
765 proportional to relative abundance of *Prochlorococcus* as estimated based on the single-copy core
766 gene *petB* and this gene was also used to estimate the relative abundance of other genes in the

767 population. Black dots represent *Tara* Oceans stations for which *Prochlorococcus* read abundance
768 was too low to reach the threshold limit.

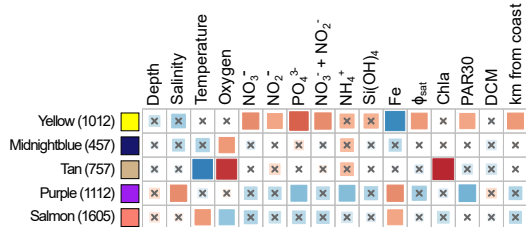
769

770 **Figure 7. Global distribution map of the *Prochlorococcus* eCAGs involved in the biosynthesis of**
771 **an alternative respiratory terminal oxidase (ARTO). (A) *Prochlorococcus* Pro-eCAG_016, (B)**
772 ***Synechococcus* Syn-eCAG_015. The size of the circle is proportional to relative abundance of each**
773 **genus as estimated based on the single-copy core gene *petB* and this gene was also used to**
774 **estimate the relative abundance of other genes in the population. Black dots represent *Tara***
775 **Oceans stations for which *Prochlorococcus* or *Synechococcus* read abundance was too low to**
776 **reach the threshold limit.**

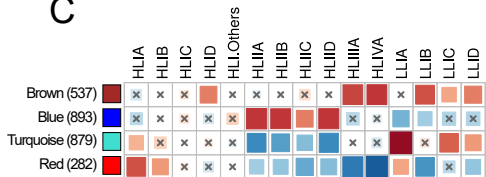
A



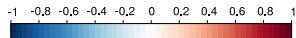
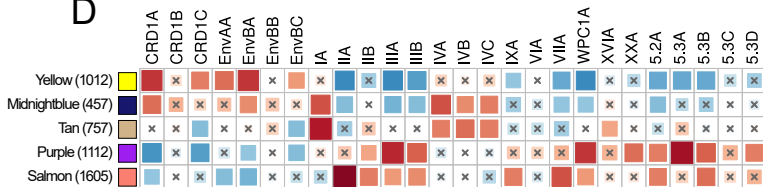
B

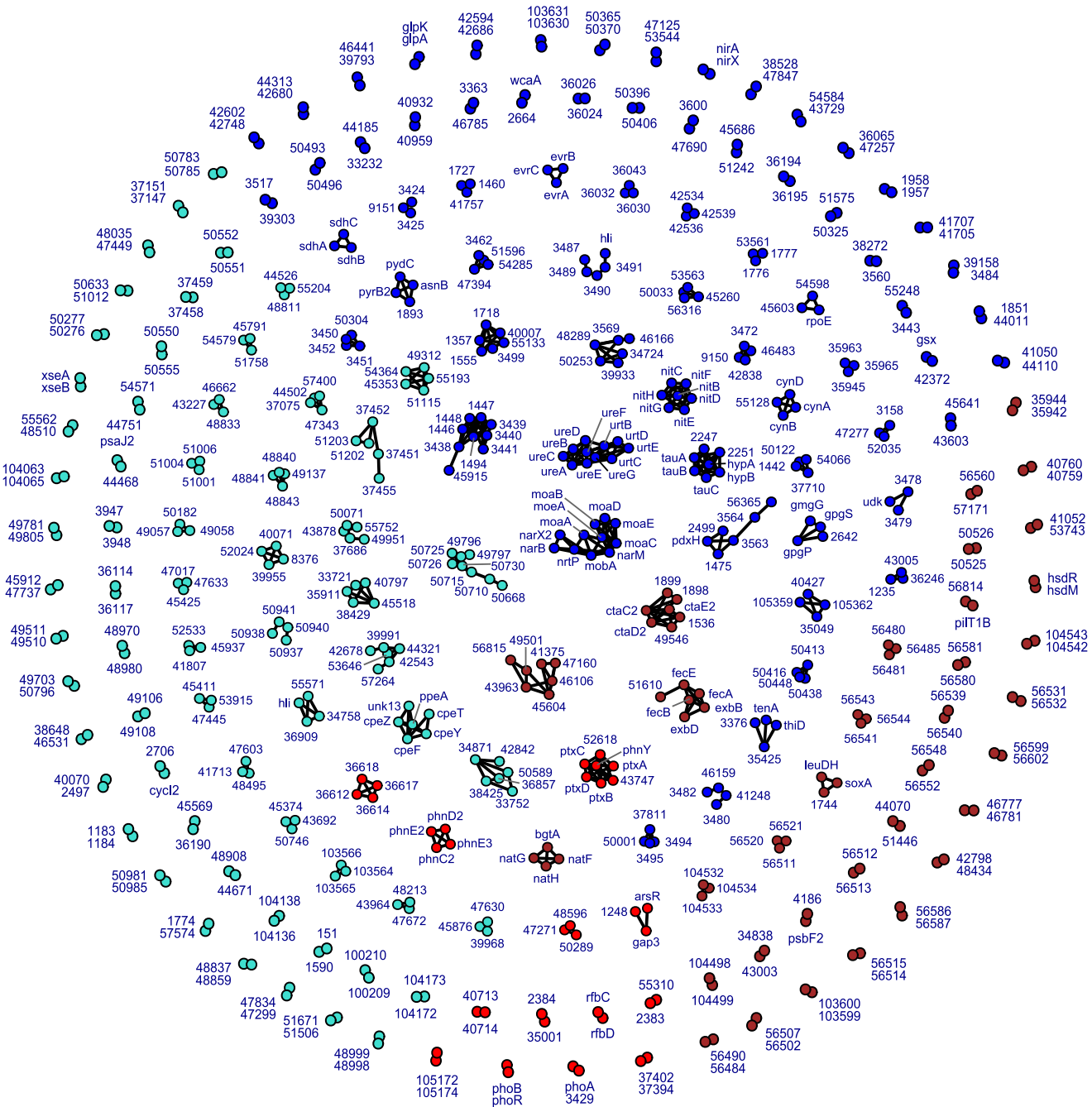


C



D





Modules



Env. niches

-N, +Fe

-Fe

-P, cold, +Fe

Cold, mixed water

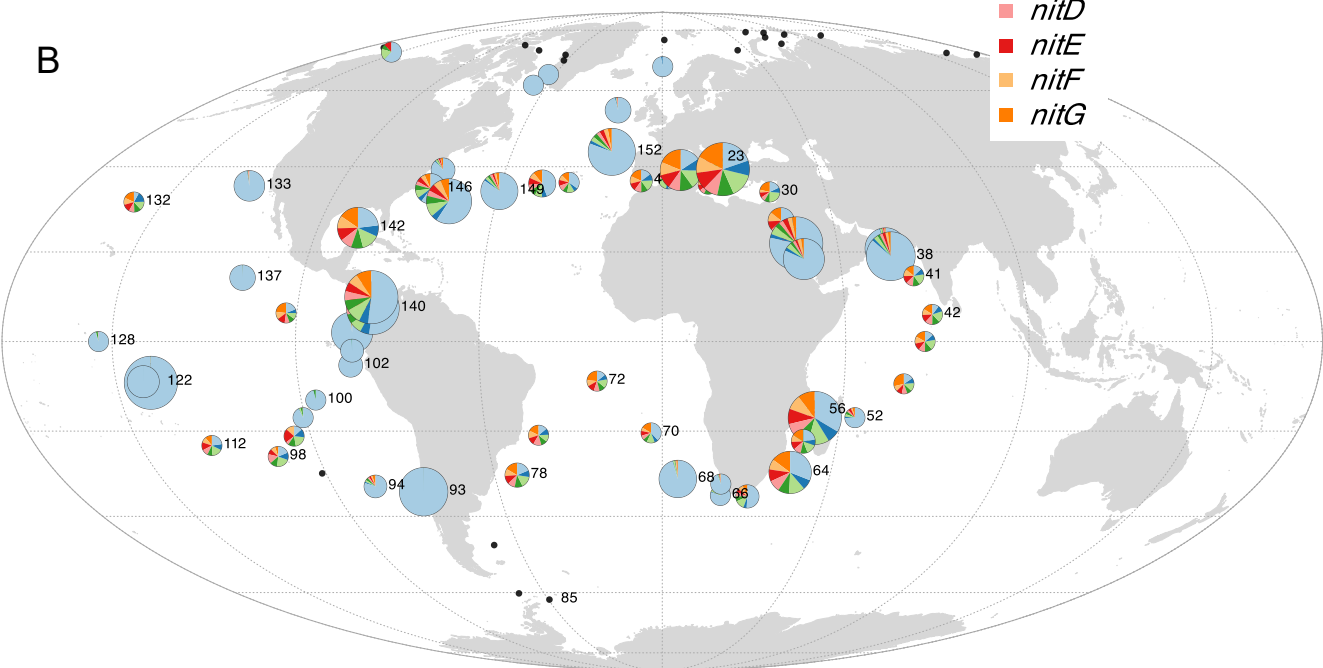
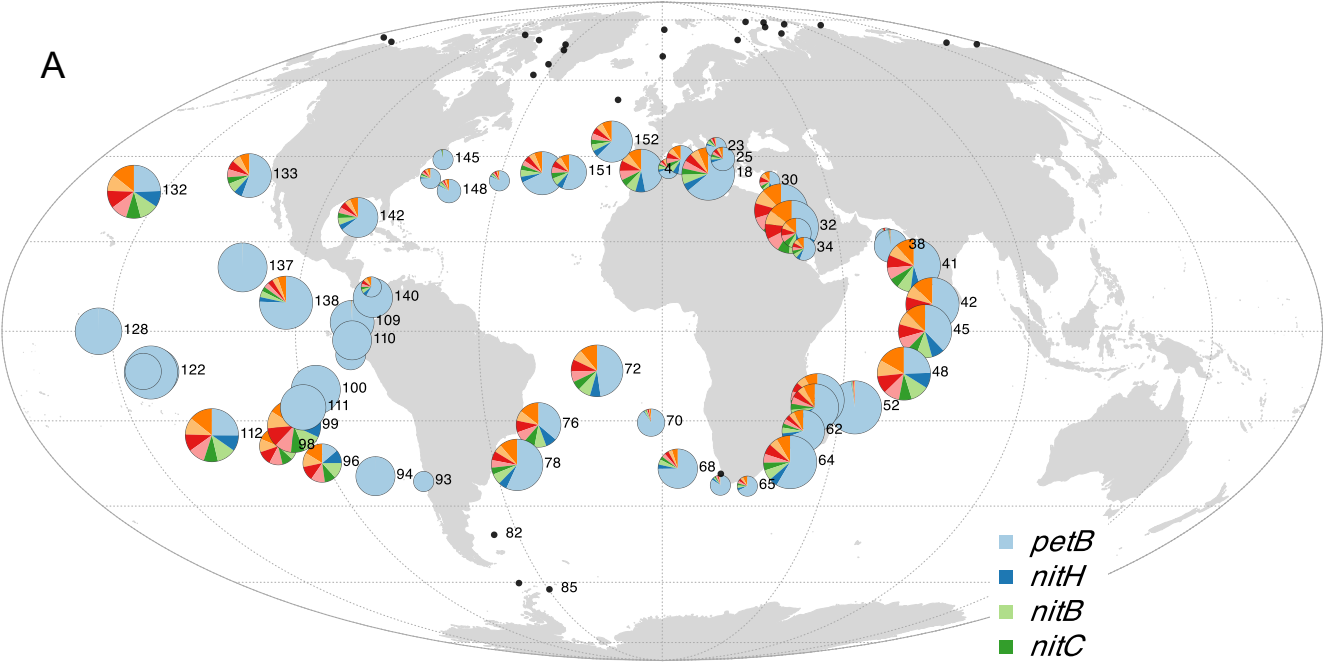
Major ESTUs

HLIA-D

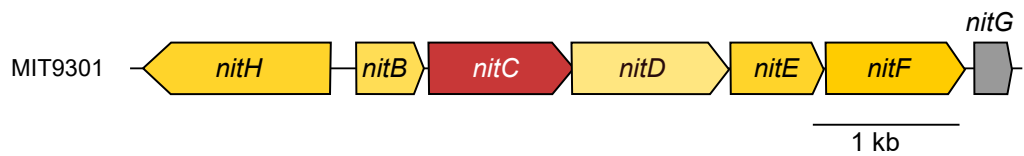
HLIIIA/IVA, LLIB

HLIA

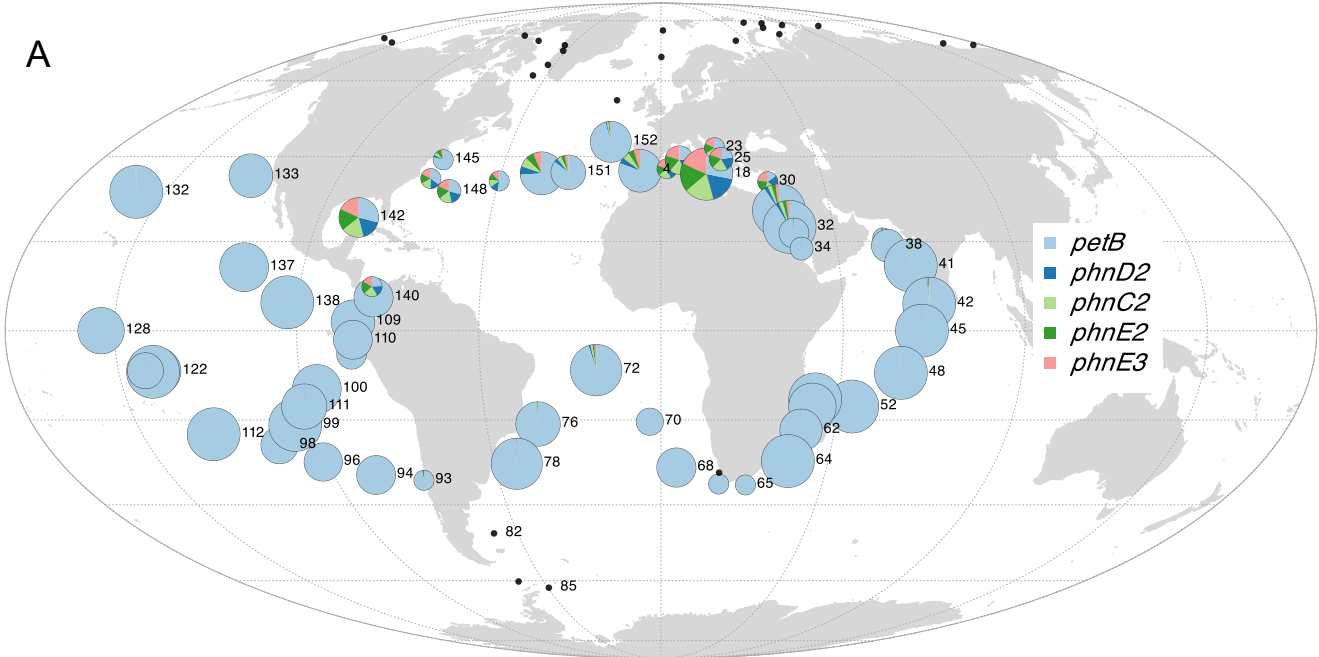
LLIA



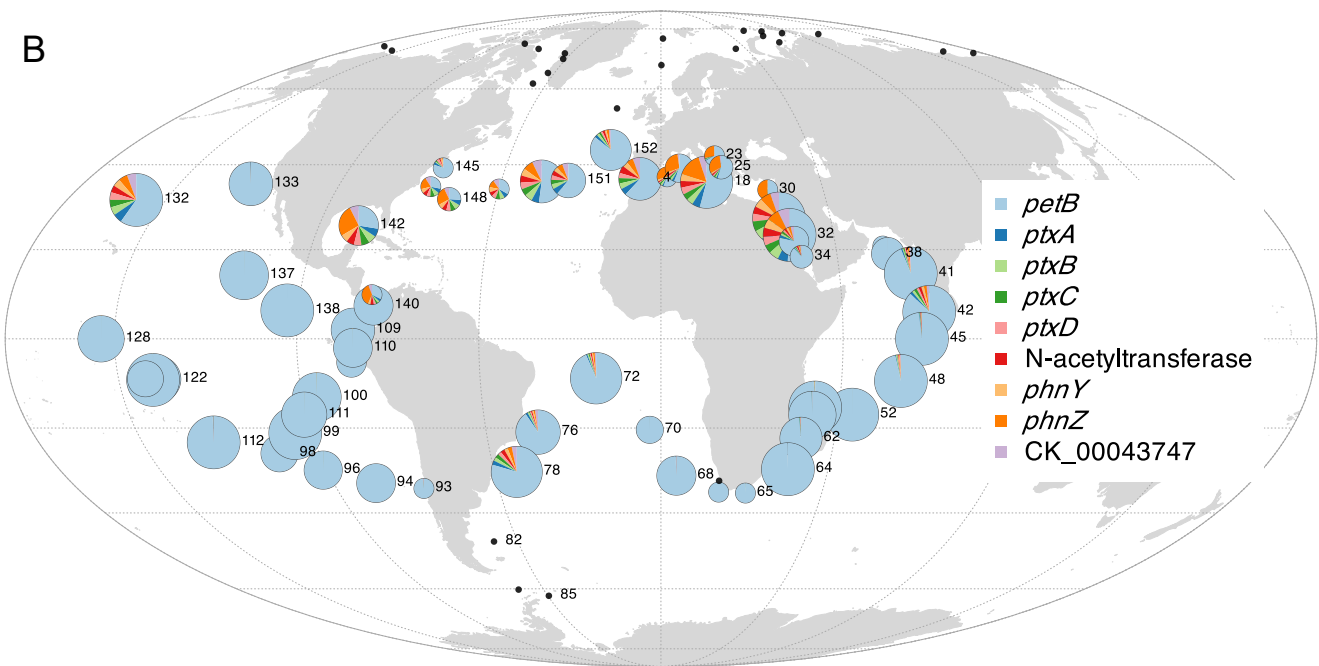
C



A



B



C

