Global phylogeography of marine *Synechococcus* in coastal areas reveals strikingly different communities than in the open ocean.

Hugo Doré[1,2], Jade Leconte[1], Ulysse Guyet[1,3], Solène Breton[1], Gregory K. Farrant[1], David Demory[1,4], Morgane Ratin[1], Mark Hoebeke[5], Erwan Corre[5], Frances D. Pitt[6,7], Martin Ostrowski[6,8], David J. Scanlan[6], Frédéric Partensky[1], Christophe Six[1] and Laurence Garczarek[1]

[1] Sorbonne Université, CNRS, UMR 7144 Adaptation and Diversity in the Marine Environment (AD2M), Station Biologique de Roscoff (SBR), Roscoff, France

[5] CNRS, FR 2424, ABiMS Platform, Station Biologique, CS 90074, Roscoff, France

[6] University of Warwick, School of Life Sciences, Coventry CV4 7AL, UK

[2] Current address: Department of Ecology, Evolution and Marine Biology; University of California, Santa Barbara, USA.

[3] Current address: Université de Bordeaux, CNRS, IBGC, UMR 5095, Bordeaux, 33000, France

[4] Current address: School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA

[7] Current address: Rosalind Franklin Laboratory, Leamington Spa, CV31 3RG, UK

[8] Current address: Climate Change Cluster, University of Technology, Broadway NSW 2007, Australia

Corresponding author: Laurence Garczarek (laurence.garczarek@sb-roscoff.fr)

Running Title: Phylogeography of *Synechococcus* in coastal areas

Key words: marine cyanobacteria, *Synechococcus*, coastal areas, Ocean Sampling Day, temperature, niche partitioning, metagenomics.

Competing Interests: The authors declare no competing interests.

Abstract

Marine *Synechococcus* comprise a numerically and ecologically prominent phytoplankton group, playing a major role in both carbon cycling and trophic networks in all oceanic regions except in the polar oceans. Despite their high abundance in coastal areas, our knowledge of *Synechococcus* communities in these environments is based on only a few local studies. Here, we use the global metagenome dataset of the Ocean Sampling Day (June 21st 2014) to get a snapshot of the taxonomic composition of coastal *Synechococcus* communities worldwide, by recruitment on a reference database of 141 picocyanobacterial genomes, representative of the whole *Prochlorococcus*, *Synechococcus* and *Cyanobium* diversity. This allowed us to unravel drastic community shifts over small to medium scale gradients of environmental factors, in particular along European coasts. The combined analysis of the phylogeography of natural populations and the thermophysiological characterization of eight strains, representative of the four major *Synechococcus* lineages (clades I to IV), also brought novel insights about the differential niche partitioning of clades I and IV, which most often co-dominate the *Synechococcus* community in cold and temperate coastal areas. Altogether, this study tackles the main differences between open-ocean and coastal communities worldwide.

Introduction

Better assessment of the spatial and temporal variability of the genetic diversity, structure and dynamics of marine phytoplankton communities is critical to predicting their future evolution in environments whose physico-chemical properties are continuously altered by the ongoing global change. The marine picocyanobacteria *Prochlorococcus* and *Synechococcus*, together accounting for about 25% of ocean net primary production [1], are key members of phytoplankton communities and constitute particularly relevant models to tackle this issue. *Prochlorococcus* distribution is restricted to the 45°S–50°N latitudinal band preferentially thriving in oligotrophic areas, whilst *Synechococcus* is present in all marine environments from the equator to subpolar waters but reaches its highest abundances in nutrient-rich areas [2–8]. The ability of these two genera to colonize a wide range of ecological niches is likely related to their large genetic diversity [9–13]. For *Prochlorococcus*, numerous environmental and laboratory studies have revealed the clear-cut niche partitioning between physiologically and genetically distinct ecotypes, with 'phototypes' [14], 'thermotypes' [3, 15, 16], and 'nutritypes' [12, 17, 18], occupying distinct light, thermal and nutrient (+Fe/- Fe) niches. Besides *Prochlorococcus*, 'Cluster 5' *sensu* [19] also encompasses three major *Synechococcus/Cyanobium* lineages, called sub-clusters (SC) 5.1 through 5.3 [10, 20]. Although a number of phylogenetic studies based on individual markers have considered SC 5.2 and *Cyanobium* as being two distinct lineages (see e.g. [21–23]), the delineation is unclear and it was recently proposed, based on comparative genomics, that all members of these lineages should be gathered into a single group (SC 5.2) named '*Cyanobium'*, even though the level of genomic diversity within this group is quite large [20, 24, 25]. SC 5.2 gathers freshwater and halotolerant representatives and thus in the marine environment, members of this group are only found in significant abundance in river-influenced coastal waters, such as the Chesapeake Bay [21, 22, 26] or the Pearl River estuary [23, 27], and in low salinity areas such as the Baltic

Sea [28]. SC 5.3 was long thought to contain only obligatory marine representatives and was shown to account for a significant fraction of the *Synechococcus* community in some specific marine areas, including the Mediterranean Sea and northwestern Atlantic Ocean [12, 29–31]. However, freshwater members of this group were recently discovered in the Tous reservoir (Spain) and were then found to be broadly distributed in temperate freshwater lakes [25, 32]. Finally, SC 5.1, a lineage that rapidly diversified after the advent of the *Prochlorococcus* radiation [33, 34], is by far the most widespread and abundant *Synechococcus* lineage in the open ocean environment, e.g. representing more than 93% of total *Tara* Oceans metagenomic reads assigned to SC 5.1-5.3 [12]. From 10 to 15 phylogenetic clades have been defined within SC 5.1 depending on the phylogenetic marker [11, 29, 35] but studies of the global distribution patterns of *Synechococcus* populations in open ocean waters have shown that there are five major clades *in situ* (I, II, III, IV and CRD1), with clades I and IV co-dominating *Synechococcus* communities in cold and temperate, nutrient-rich areas, while clades II, III and CRD1 preferentially thrive in warm waters [6, 12, 30, 31, 36]. Physiological measurement of temperature *preferenda* of strains belonging to clades I, II, III, IV and V isolated across different latitudes further confirmed the existence of warm (clades II, III, V) and cold (clades I and IV) 'thermotypes' [37–40]. Despite being phylogenetically distant, clades I and IV were further demonstrated to share a number of physiological adaptations to cold water, including a higher thermal sensitivity of phycobiliproteins [41], a similar change in membrane lipids [40, 42] and an increase of the photoprotection capacities using the orange carotenoid protein (OCP; [43]). Nutrients were also found to play a key role in structuring these populations, with clade II, the most abundant *Synechococcus* lineage in the ocean, dominating the *Synechococcus* community in N-poor areas, clade III in P-poor areas, while CRD1 is restricted to Fe-depleted waters [6, 12, 31, 36].

Although the variability of picocyanobacterial communities and the main physico-chemical factors driving their composition are starting to be well understood in open ocean environments, the picture is much more fragmentary in coastal areas, because only a few coastal sites have been studied to date [21, 22, 24, 27, 44–46]. To get a more global view of the genetic diversity and biogeography of coastal populations of picocyanobacteria, we used metagenomic data from the Ocean Sampling Day (OSD) 2014 campaign [47], encompassing 157 coastal samples collected all over the world at the summer solstice, employing the same protocol for collecting DNA samples and associated metadata. Using a whole genome recruitment (WGR) approach, we assessed the genetic diversity and the clade-level composition of *Synechococcus* communities in OSD samples. Given the previously recognized role of temperature in structuring *Synechococcus* communities, we then analyzed the distribution patterns of the different lineages in light of previously published and new comparative thermophysiological data on *Synechococcus* strains representative of the most abundant clades in the field. The excellent spatial resolution achieved in northern Atlantic and Mediterranean coastal waters allowed us to observe several spatial community shifts and to enlighten the roles of temperature and salinity as key drivers of coastal *Synechococcus* community composition.

Materials and methods

Ocean Sampling Day metagenomics data

OSD 2014 is a global sampling campaign that took place on June 21st, 2014 and sampled 157 stations worldwide for metagenomes (Dataset S1). The median distance to the nearest coast was 0.29 nautical miles (average: 6.3 nautical miles). Details about sampling methods can be found at https://repository.oceanbestpractices.org/handle/11329/616 [48]. Metagenomic data are available from the European Nucleotide Archive (http://www.ebi.ac.uk/ena/data/view/PRJEB8682) under the study accession number

PRJEB8682 (raw data) and from the European Bioinformatics Institute (EBI) Metagenomics portal under the project accession number ERP009703 (processed data). Data were downloaded from the EBI for 150 of the 157 stations for which a "processed reads without annotation" file was available, generated following the EBI analysis pipeline v2.0, available at https://www.ebi.ac.uk/metagenomics/pipelines/2.0. Briefly, Illumina MiSeq paired reads were merged using SeqPrep (https://github.com/jstjohn/seqprep) and trimmed for low quality ends, then sequences with more than 10% undetermined nucleotides were removed using Trimmomatic [49] before discarding reads shorter than 100 nucleotides. Contextual data collected at all OSD stations were retrieved from PANGAEA (https://doi.pangaea.de/10.1594/PANGAEA.854419; [50]) and those used in this study are listed in Dataset S1: only water temperature and salinity data were available in a sufficient number of stations to be used. A map of OSD stations used in this study is available as Fig. S1.

Taxonomic assignment of metagenomic reads

BLASTN v2.2.28+ [51, 52] was used to align metagenomic reads against a database of 863 complete genomes of aquatic bacteria (Dataset S2), gathering 141 genomes of marine picocyanobacteria and 722 outgroup genomes, the latter including 185 cyanobacterial genomes other than *Prochlorococcus* and marine *Synechococcus* listed in Cyanobase (http://genome.microbedb.jp/cyanobase/) as well as 537 genomes of other aquatic microbes downloaded from the proGenomes database (http://progenomes.embl.de/representatives.cgi). Only best-hit matches (option -max_target_seqs 1) with an e-value below 10^{-3} (-evalue 0.001) were kept, and reads matching outgroup genomes were discarded. Based on BLASTN results, reads aligning over more than 90% of their length on a picocyanobacterial genome were extracted from initial read files, and a second BLASTN was run against a database containing only marine picocyanobacterial genomes with default parameters except for a lower limit on

141 percentage of identity of 30% (-perc_identity 30), a filter on e-value of 10^{-2} (-evalue 0.01) and
142 by selecting the blastn algorithm (-task blastn). BLASTN results were then parsed using the
143 Lowest Common Ancestor method [53]. For each read, BLAST matches with over 80% ID
144 aligned over more than 90% of their length against a reference genome were kept if their
145 BLAST score was within 5% of the best score. Then, the read was attributed to the lowest
146 common ancestor of these matches (i.e., strain, clade, subcluster or genus). Counts of reads
147 assigned to the strain or clade levels were ultimately aggregated by clade. Two additional
148 categories were made for reads that could only be assigned to the level of *Synechococcus*
149 subcluster 5.1 (SC 5.1 in Figures 1 and 3) or even *Synechococcus* genus (*Syn* in Figs 1 and 3).
150
151 Analysis of picocyanobacterial community composition
152 In order to account for the potential variation in genome length among clades, read counts were
153 divided by the average genome length within each clade. To minimize the noise in recruitment
154 data, we then removed from the dataset stations with less than 600 recruited reads per million
155 bp, corresponding to a genome coverage of ca. 16%, since reads are 242 bp long on average.
156 Read counts at each station were further normalized by the total number of reads recruited at
157 this station to assess relative abundances of taxa. The R packages *cluster* v1.14.4 [54] and *vegan*
158 v2.2-1 [55] were used to cluster stations according to the Bray-Curtis distance. Figures were
159 drawn in R v3.03 with package *ggplot2* v1.0.1 [56].
160
161 Thermal *preferenda* of strains representative of the most abundant clades *in situ*
162 Two strains of each of the four most abundant *Synechococcus* clades in Fe-replete areas (clades
163 I to IV) were selected from the Roscoff Culture Collection (Table 1; http://roscoff-culture-
164 collection.org/; [57]). Strains were grown in polystyrene flasks in PCR-S11 medium [58]
165 supplemented with 1mM sodium nitrate. The seawater was reconstituted using Red Sea Salts

(Houston, TX, USA) and distilled water. Cultures of the eight strains were acclimated at least two weeks to a range of temperatures from 10°C to 33°C, within temperature-controlled chambers (Liebherr-Hausgeräte, Lienz, Austria) and continuous light was provided by green/white/blue LEDs (Alpheus, France) at an irradiance of 20 µmol photons m^{-2} s^{-1}. After acclimation, cultures were split into three biological replicates for each strain, and sampled once or twice a day until the stationary phase was reached.

For cell density measurements, aliquots of cultures were preserved with 0.25% glutaraldehyde grade II (Sigma Aldrich, St Louis, MO, USA) and stored at -80°C until analysis [59]. Cell concentration was determined using a flow cytometer (FACSCanto II, Becton Dickinson, San Jose, CA, USA) with laser emission set at 488 nm, and using distilled water as sheath fluid.

To estimate the maximum population growth rates, we considered that *Synechococcus* exponential growth followed:

$$\frac{dN}{dt} = \mu N$$

Where N is the cell abundances (in cell mL^{-1}) and μ is the maximum population growth rate (in $days^{-1}$). We estimated μ as the coefficient of the linear regression model performed on log-transform *N(t)* data during the exponential phase only.

To overcome the fact that discrete experimental measurements have a limited resolution, we estimated the cardinal growth parameters for each strain using the Cardinal Temperature Model with Inflection (BR model; [60]). This model helps describe the growth response of acclimated phytoplankton strains to temperature using four parameters (Table 2): the optimal temperature for growth (T_{opt}) at which the optimal growth rate (μ_{opt}) occurs, and the minimal and maximal temperatures for growth (T_{min} and T_{max}) at which $\mu = 0$. Our data did not allow to constrain T_{min}, but this does not affect our estimation of other parameters.

Results and Discussion

Biogeography of coastal picocyanobacterial communities is influenced by seawater temperature

Most of the stations sampled during the OSD 2014 campaign [47] correspond to coastal areas with only 17 of 157 stations located over 11 nautical miles from the nearest coast. This dataset displays a particularly good spatial resolution in some regions of the world ocean and notably along European and Eastern United States coasts, while only a few of the sampled sites were located in the southern hemisphere (7 out of 157; Fig. S1). Here, we used the 150 metagenomes obtained in the framework of this campaign, altogether totaling 41 Gbp (168.7 million reads), to assess the relative abundance of *Synechococcus/Cyanobium* and *Prochlorococcus* clades using a Whole Genome Recruitment (WGR) approach against a reference genome database encompassing 141 genomes of marine picocyanobacteria as well as 722 cyanobacterial or other aquatic microbial genomes, used as outgroups (Fig. 1). *Prochlorococcus* was only abundant at a few stations, likely due to the coastal localization of the sampling sites, and was therefore not included in subsequent analyses. By contrast, *Synechococcus/Cyanobium,* known to outnumber *Prochlorococcus* in coastal areas [2, 8, 24, 61], was detected with sufficient coverage to perform reliable taxonomic assignment at the clade level in 102 out of the 150 OSD metagenomes. At most stations, the *Synechococcus/Cyanobium* community was dominated by one or two taxa among SC 5.1 clades I-IV, SC 5.2 or SC 5.3 (Fig. 1). Consistent with previous studies on the picocyanobacterial distribution in open ocean waters [7, 12, 15, 29, 31, 36], clades I and IV dominated at latitudes above 35°N (except in the Mediterranean Sea) and clade II at latitudes below 35°N, while clade III was almost exclusively present and often dominant in the Mediterranean Sea. It is also worth noting that the co-occurrence of clades I and IV at the few stations beyond 35°S in the Southern hemisphere mirrored the profiles obtained at the same

latitude in the Northern hemisphere, in agreement with previous observations in open ocean waters [12, 29, 31, 36] as well as with the low temperatures of isolation sites of clade I and IV strains [38].

In order to further explore the role of temperature on the differential latitudinal distribution of members of clades I to IV, we characterized the thermal *preferenda* of eight strains belonging to these clades (Fig 2). While several strains belonging to clade I were previously shown to withstand colder temperatures than their tropical clade II counterparts [38, 40, 43], growth optima and boundary limits for temperature were only available for one clade IV [40, 62] and two closely related clade III strains [37, 40, 63] and results were obtained in different light conditions, making them difficult to compare. Here, the direct comparison of clades I and IV strains grown in the same conditions showed quite similar thermal preferences. All tested strains displayed an optimal temperature for growth of about 24°C according to our model fit (Fig. 2 and Table 2) and were all able to grow at the lowest tested temperature, 10°C, which is also the lowest temperature measured in the OSD 2014 stations where the *Synechococcus* community was analyzed. In comparison, clades II and III strains were not able to grow at temperatures of 13°C and below, thus confirming with several strains that clades I and IV are cold thermotypes, whereas clades II and III are warm thermotypes. Altogether, these results support the idea that differences in thermophysiology at least partially explain the latitudinal distribution of these four clades.

Besides the abundance of clades I and IV, coastal *Synechococcus* communities also exhibited some other specificities as compared to open ocean populations, notably the very low relative abundance of clade CRD1, which was shown to be prevalent in large regions of the open ocean that are limited by iron availability [12, 31, 36], as well as the dominance of SC 5.2 in the brackish Baltic sea and at stations along the Atlantic coast of North America, often co-occurring with a low proportion of clade VIII. The latter observation is most likely due to the influence

of riverine inputs at these OSD stations, these taxa being known to occur in estuarine areas and to contain strains growing over a large range of salinity [10, 21, 35]. This hypothesis was further confirmed by clustering stations according to the relative abundance profiles of *Synechococcus* clades (Fig. 3), which clearly separated stations dominated by subcluster 5.2 and showed that they had a lower salinity than most other stations (cluster 5, Fig. 4B). Finally, clades V and VI, which were not distinguished from clade VII (and CRD1) in previous global surveys of *Synechococcus* distribution using the 16S rRNA marker gene, were found to be locally abundant in the dataset. While the V/VI/VII/CRD1 group was considered to be widely distributed in oceanic waters [4, 15, 29], our analysis reveals the potential preference for coastal areas of the closely related clades V and VI. This result is consistent with the previous local observations of the occurrence of clade V- and VI-related sequences at some coastal sites in the Adriatic Sea and the Pearl River Estuary [23, 64, 65].

A progressive latitudinal shift in *Synechococcus/Cyanobium* communities along the coast of Europe

Besides the abovementioned specificities of coastal regions in terms of *Synechococcus/Cyanobium* community composition, we also observed changes in communities at a finer spatial scale along European coasts, where the sampling effort was the highest (see zoom in Fig. 1 and Fig. S1 for station numbers). While along the southern part of this latitudinal gradient from the Moroccan to French Atlantic coasts, *Synechococcus* communities were dominated by clade IV (e.g. OSD92), a progressive northward shift was observed towards the dominance of clade I in the North Sea (e.g., OSD164). Clustering of stations based on clade relative abundance indeed highlighted two groups of stations, the first one dominated by clade IV (cluster 3) and the second one by clade I (cluster 4; Fig. 3). Interestingly, clade I was found to dominate at stations that display a significantly lower salinity than those dominated by clades

II or III (clusters 1 and 2). These clade I-dominated stations also exhibited a significantly lower temperature (average 16.6°C, median 17°C) than all other clusters except cluster 3 dominated by clade IV (average temperature 19.1°C, median 19°C), the latter cluster of stations showing a significant difference in temperature only with cluster 2 (dominated by clade II). Thus, despite a clear latitudinal shift in the ratio of clade I to clade IV along the European coast, neither the difference in salinity nor the difference in seawater temperature seem to be sufficient to fully explain the observed changes.

While our thermal *preferenda* measurements confirmed that clades I and IV are both cold thermotypes, strains used for this study showed similar growth rates at cold temperatures (Fig 2). This supports the observation that temperature alone does not explain their differential distribution, with the caveat that only two strains per clade were analyzed. As previously suggested for clade I in a number of environmental studies [5, 7, 23, 66, 67], it is possible that clade I and potentially clade IV are comprised of distinct genotypes exhibiting different lower temperature boundary limits and colonizing different thermal niches. A previous study indeed showed variability in the minimal growth temperature of clade I strains in relation to their latitude of isolation [38], and comparison of our experimental data with previous data acquired under the same light conditions [40] brings evidence of such variability for clade IV. Indeed, the two clade IV strains characterized here were sampled at high latitude (Table 1) and show a higher tolerance to cold temperatures than BL107, another clade IV strain isolated in the Mediterranean Sea [40]. Thus, the ecological differences between clades I and IV are most probably difficult to identify due to underlying differences between genotypes at a finer taxonomic level, and a higher taxonomic resolution would be necessary if one wanted to observe a significant effect of temperature on the distribution of populations of these clades. Alternatively, it is possible that other parameters or combinations of parameters varying with latitude need to be considered to explain the shift in clades I and IV dominance. Notably, an

interaction between light and temperature on phytoplankton physiology has been described in a number of species [68], and these two parameters vary greatly with latitude.

Several other potential reasons have been previously evoked to explain the co-occurrence of clades I and IV and variations in their relative abundance *in situ*, including differences in their metal concentration requirements [31, 44] and transport and mixing of populations by advection. The latter hypothesis was notably suggested for *Synechococcus* populations from north-west of the Svalbard island (above 79°N), where the Gulf Stream current was proposed to be the source of clade IV populations in summer [5] and from the Korean Sea where the warm, oligotrophic Kuroshio Current was suggested to be responsible for the co-occurrence of clades I, II and IV populations [69]. Interestingly the only OSD sampling site close to Japan (OSD124, Fig. S1) was described by its sampler as a site where oceanic and coastal waters are sporadically interchanged, which could explain the unexpected profile of this station where clades I and II co-dominate (Fig. 1). Finally, other studies also suggested that clade I could be a more coastal and opportunistic clade than clade IV [10, 44], but this hypothesis does not seem to be confirmed by the present study since many coastal stations (cluster 3) are actually dominated by clade IV.

Local changes in *Synechococcus* communities in the Mediterranean Sea

Stations sampled in the Mediterranean Sea fell into several clusters based on their composition in *Synechococcus/Cyanobium* lineages. Most stations belonged to cluster 1, dominated by clade III with a low relative abundance of clades VI, WPC1 and SC 5.3 (Fig. 3). This clade composition is quite similar to that previously described [12] for open waters of the Mediterranean Sea, which was suggested to be related to specific features of this semi-enclosed sea and notably to its low phosphate concentration [12, 15, 29], a parameter that was not available in the OSD dataset. Most of the stations of the Adriatic Sea formed a distinct cluster

316 (cluster 7), where the same clades were present but in different proportions, clade VI and SC
317 5.3 taking over clade III. Finally, stations OSD34 and OSD90, located on the Egyptian and
318 Greek coasts, respectively, the only stations of the OSD dataset comprising a high proportion
319 of clade V or VIII, formed a cluster on their own (cluster 6 and 9). While these four clusters
320 (clusters 1, 6, 7 and 9) are specific to the Mediterranean Sea, it is worth noting that two stations
321 at the easternmost end of the Mediterranean Sea (OSD123 and OSD132, Figs. 3 and S1) fell
322 into cluster 2, dominated by clade II, and showed a clade composition very similar to the
323 samples collected in the Red Sea (OSD52 and OSD53). This suggests that Israeli coastal areas
324 are strongly influenced by waters entering the Mediterranean Sea via the Suez Canal, consistent
325 with previous findings for *Synechococcus, Prochlorococcus* as well as for larger organisms [70,
326 71].

327 Interestingly, the three specific clusters identified in the Mediterranean Sea displayed
328 different temperature and salinity characteristics (Fig. 4A-B). The salinity range of stations in
329 cluster 1 (dominated by clade III) was narrow (average salinity 37.90 psu, median 37.98 psu)
330 and significantly higher than that of cluster 7 (dominated by clade VI and SC 5.3, average
331 salinity 31.43 psu, median 32.77 psu), suggesting that clade VI and SC 5.3 are able to cope with
332 lower salinities. Consistently, SC 5.3 was recently found to encompass members colonizing
333 freshwater lakes [25, 32], while in the marine environment, this subcluster was reported both
334 in strictly marine waters [12, 30] and in low salinity waters [72]. Our study also brings new
335 insights into the ecological niche occupied by clade VI, whose distribution was so far poorly
336 known [29], and that appears to be restricted to coastal regions of intermediate salinity. All
337 stations of the Adriatic Sea comprising cluster 6 were indeed sampled in the northwestern part
338 of this area, where the influence of the Po River plume may be important [73]. This distribution
339 is consistent with previous observations of the closely-related and often co-occurring clade V
340 in low salinity surface waters of the Adriatic Sea [64] and of both clades V and VI in the Pearl

River Estuary [74]. Laboratory experiments also showed that representative strains of these two clades can tolerate salinities as low as 15 psu [75]. Still, we cannot exclude that besides low salinity, other local specificities linked to riverine input might also explain the predominance of SC 5.3 and clade VI in coastal areas of the Adriatic Sea.

A significant difference in water temperature was also found between cluster 1, dominated by clade III (average temperature 21.5°C, median 20.8°C) and cluster 2, dominated by clade II (average 26.5°C, median 27.1°C). This suggests that the shift observed at the easternmost part of the Mediterranean Sea from a dominance of clade III to a local dominance of clade II (stations OSD123 and OSD132, Figs. 1 and S1) might be related to a difference in water temperature. Interestingly, in contrast to clades I and IV that often co-occur, clades II and III seem to be nearly mutually exclusive, at least in the Mediterranean Sea, and the temperature limit above which clade II dominates seems to lie around 25°C (Fig. 3). In our experimental comparison of thermal *preferenda*, this corresponds to the temperature at which growth rates of clade II strains become higher than that of clade III strains, resulting in a higher optimal temperature of clade II compared to clade III strains (Table 2). Altogether, temperature and salinity appear as major factors driving the composition of *Synechococcus/Cyanobium* communities in coastal waters of the Mediterranean Sea, although other biotic and abiotic factors are most likely involved, notably the availability of phosphorus, a key limiting nutrient in this area [76].

Conclusion

The OSD dataset is unique, not only by providing an instantaneous snapshot of the microbial community composition but also because, by focusing on coastal areas, it nicely complements other recent global ocean surveys performed in the open ocean [6, 12, 31, 36, 77, 78]. In particular, the good spatial resolution of the sampling performed along the European coasts is well-adapted to observe shifts in communities and delineate their boundaries. Despite the fact

366 that only a few physico-chemical parameters were collected, this dataset allowed us to
367 considerably improve our knowledge of the distribution of *Synechococcus/Cyanobium* lineages
368 in coastal areas, to gain insights into the realized environmental niches of the main ones,
369 including some that were previously poorly known such as clade VI, as well as to reinforce
370 hypotheses about thermal niche differentiation that were supported by laboratory experiments
371 on a set of representative strains. A continued effort towards global instantaneous surveys of
372 microbial diversity in coastal areas over the long term and at different seasons would be
373 invaluable to monitor the evolution of microbial communities in relation to global change.
374

375 ## Acknowledgements

376 We thank the OSD Consortium for sampling, sequencing and making freely available the data
377 analyzed in this paper as well as the Roscoff Culture Collection (http://roscoff-culture-
378 collection.org/) for providing *Synechococcus* strains used in this study. Financial support for
379 the OSD program was provided by the European Union program MicroB3 (UE-contract-
380 287589) and authors were supported by the French "Agence Nationale de la Recherche"
381 programs SAMOSA (ANR-13-ADAP-0010) and CINNAMON (ANR-17-CE02-0014-01) as
382 well as the European program Assemble Plus (H2020-INFRAIA-1-2016-2017; grant no.
383 730984). DJS received funding from the European Research Council (ERC) under the European
384 Union's Horizon 2020 research and innovation programme (grant agreement No 883551).
385

386 ## Competing Interests

387 The authors declare no competing interests.
388

References

1. Flombaum P, Gallegos JL, Gordillo R a, Rincón J, Zabala LL, Jiao N, et al. Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc Natl Acad Sci U S A* 2013; **110**: 9824–9.
2. Partensky F, Blanchot J, Vaulot D. Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters: a review. *Bull Inst Océan* 1999; **19**: 457–475.
3. Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward EMS, Chisholm SW. Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* 2006; **311**: 1737–1740.
4. Scanlan DJ. Marine picocyanobacteria. In: Whitton BA (ed). *Ecology of Cyanobacteria II: Their Diversity in Space and Time*. 2012. Springer Netherlands, Dordrecht, pp 503–533.
5. Paulsen ML, Doré H, Garczarek L, Seuthe L, Müller O, Sandaa R-A, et al. *Synechococcus* in the Atlantic Gateway to the Arctic Ocean. *Front Mar Sci* 2016; **3**: 191.
6. Kent AG, Baer SE, Mouginot C, Huang JS, Larkin AA, Lomas MW, et al. Parallel phylogeography of *Prochlorococcus* and *Synechococcus*. *ISME J* 2019; **13**: 430–441.
7. Xia X, Cheung S, Endo H, Suzuki K, Liu H. Latitudinal and vertical variation of *Synechococcus* assemblage composition along 170° W transect from the south Pacific to the Arctic Ocean. *Microb Ecol* 2019; **77**: 333–342.
8. Visintini N, Martiny AC, Flombaum P. *Prochlorococcus*, *Synechococcus*, and picoeukaryotic phytoplankton abundances in the global ocean. *Limnol Oceanogr Lett* 2021; **6**: 207–215.

412 9. Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA, et al. Genome
413 divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation.
414 *Nature* 2003; **424**: 1042–1047.
415 10. Dufresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik BP, et al.
416 Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome*
417 *Biol* 2008; **9**: R90–R90.
418 11. Mazard S, Ostrowski M, Partensky F, Scanlan DJ. Multi-locus sequence analysis,
419 taxonomic resolution and biogeography of marine *Synechococcus*. *Environ Microbiol*
420 2012; **14**: 372–86.
421 12. Farrant GK, Doré H, Cornejo-Castillo FM, Partensky F, Ratin M, Ostrowski M, et al.
422 Delineating ecologically significant taxonomic units from global patterns of marine
423 picocyanobacteria. *Proc Natl Acad Sci* 2016; **113**: E3365–E3374.
424 13. Garcia CA, Hagstrom GI, Larkin AA, Ustick LJ, Levin SA, Lomas MW, et al. Linking
425 regional shifts in microbial genome adaptation with surface ocean biogeochemistry.
426 *Philos Trans R Soc B Biol Sci* 2020; **375**: 20190254.
427 14. Moore LR, Rocap G, Chisholm SW. Physiology and molecular phylogeny of coexisting
428 *Prochlorococcus* ecotypes. *Nature* 1998; **393**: 464–467.
429 15. Mella-Flores D, Mazard S, Humily F, Partensky F, Mahé F, Bariat L, et al. Is the
430 distribution of *Prochlorococcus* and *Synechococcus* ecotypes in the Mediterranean Sea
431 affected by global warming? *Biogeosciences* 2011; **8**: 2785–2804.
432 16. Chandler JW, Lin Y, Gainer PJ, Post AF, Johnson ZI, Zinser ER. Variable but persistent
433 coexistence of *Prochlorococcus* ecotypes along temperature gradients in the ocean's
434 surface mixed layer. *Environ Microbiol Rep* 2016; **8**: 272–284.

17. Rusch DB, Martiny AC, Dupont CL, Halpern AL, Venter JC. Characterization of *Prochlorococcus* clades from iron-depleted oceanic regions. *Proc Natl Acad Sci U S A* 2010; **107**: 16184–16189.

18. Malmstrom RR, Rodrigue S, Huang KH, Kelly L, Kern SE, Thompson A, et al. Ecology of uncultured *Prochlorococcus* clades revealed through single-cell genomics and biogeographic analysis. *ISME J* 2013; **7**: 184–198.

19. Herdman M, Castenholz RW, Waterbury JB, Rippka R. Form-genus XIII. *Synechococcus*. In: Boone D, Castenholz R (eds). *Bergey's Manual of Systematics of Archaea and Bacteria Volume 1*, 2nd Ed. 2001. Springer-Verlag, New York, pp 508–512.

20. Doré H, Farrant GK, Guyet U, Haguait J, Humily F, Ratin M, et al. Evolutionary mechanisms of long-term genome diversification associated with niche partitioning in marine picocyanobacteria. *Front Microbiol* 2020; **11**: 567431.

21. Chen F, Wang K, Kan J, Suzuki MT, Wommack KE. Diverse and unique picocyanobacteria in Chesapeake Bay, revealed by 16S-23S rRNA internal transcribed spacer sequences. *Appl Environ Microbiol* 2006; **72**: 2239–2243.

22. Cai H, Wang K, Huang S, Jiao N, Chen F. Distinct patterns of picocyanobacterial communities in winter and summer in the Chesapeake Bay. *Appl Environ Microbiol* 2010; **76**: 2955–2960.

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

24. Hunter-Cevera KR, Post AF, Peacock EE, Sosik HM. Diversity of *Synechococcus* at the Martha's Vineyard coastal observatory: Insights from culture isolations, clone libraries, and flow cytometry. *Microb Ecol* 2016; **71**: 276–289.

25. Cabello-Yeves PJ, Picazo A, Camacho A, Callieri C, Rosselli R, Roda-Garcia JJ, et al. Ecological and genomic features of two widespread freshwater picocyanobacteria. *Environ Microbiol* 2018; **20**: 3757–3771.

26. Chen F, Wang K, Kan J, Bachoon DS, Lu J, Lau S, et al. Phylogenetic diversity of *Synechococcus* in the Chesapeake Bay revealed by ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) large subunit gene (*rbcL*) sequences. *Aquat Microb Ecol* 2004; **36**: 153–164.

27. Xia X, Vidyarathna NK, Palenik B, Lee P, Liu H. Comparison of the seasonal variations of *Synechococcus* assemblage structures in estuarine waters and coastal waters of Hong Kong. *Appl Environ Microbiol* 2015; **81**: 7644–7655.

28. Haverkamp T, Acinas SG, Doeleman M, Stomp M, Huisman J, Stal LJ. Diversity and phylogeny of Baltic Sea picocyanobacteria inferred from their ITS and phycobiliprotein operons. *Environ Microbiol* 2008; **10**: 174–188.

29. Zwirglmaier K, Jardillier L, Ostrowski M, Mazard S, Garczarek L, Vaulot D, et al. Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ Microbiol* 2008; **10**: 147–161.

30. Huang S, Wilhelm SW, Harvey HR, Taylor K, Jiao N, Chen F. Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *ISME J* 2012; **6**: 285–97.

31. Sohm JA, Ahlgren NA, Thomson ZJ, Williams C, Moffett JW, Saito MA, et al. Co-occurring *Synechococcus* ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. *ISME J* 2015; **10**: 1–13.

32. Cabello-Yeves PJ, Haro-Moreno JM, Martin-Cuadrado AB, Ghai R, Picazo A, Camacho A, et al. Novel *Synechococcus* genomes reconstructed from freshwater reservoirs. *Front Microbiol* 2017; **8**: 1–13.

484 33. Urbach E, Robertson DL, Chisholm SW. Multiple evolutionary origins of
485 prochlorophytes within the cyanobacterial radiation. *Nature* 1992; **355**: 267–270.
486 34. Sánchez-Baracaldo P. Origin of marine planktonic cyanobacteria. *Sci Rep* 2015; **5**:
487 17418.
488 35. Fuller NJ, Marie D, Partensky F, Vaulot D, Post AF, Scanlan DJ. Clade-specific 16S
489 ribosomal DNA oligonucleotides reveal the predominance of a single marine
490 *Synechococcus* clade throughout a stratified water column in the Red Sea. *Appl Environ*
491 *Microbiol* 2003; **69**: 2430–2443.
492 36. Ahlgren NA, Belisle BS, Lee MD. Genomic mosaicism underlies the adaptation of
493 marine *Synechococcus* ecotypes to distinct oceanic iron niches. *Environ Microbiol* 2020;
494 **22**: 1801–1815.
495 37. Mackey KRM, Paytan A, Caldeira K, Grossman AR, Moran D, McIlvin M, et al. Effect
496 of temperature on photosynthesis and growth in marine *Synechococcus* spp. *Plant*
497 *Physiol* 2013; **163**: 815–829.
498 38. Pittera J, Humily F, Thorel M, Grulois D, Garczarek L, Six C. Connecting thermal
499 physiology and latitudinal niche partitioning in marine *Synechococcus*. *ISME J* 2014; **8**:
500 1221–1236.
501 39. Varkey D, Mazard S, Ostrowski M, Tetu SG, Haynes P, Paulsen IT. Effects of low
502 temperature on tropical and temperate isolates of marine *Synechococcus*. *ISME J* 2016;
503 **10**: 1252–1263.
504 40. Breton S, Jouhet J, Guyet U, Gros V, Pittera J, Demory D, et al. Unveiling membrane
505 thermoregulation strategies in marine picocyanobacteria. *New Phytol* 2020; **225**: 2396–
506 2410.
507 41. Pittera J, Partensky F, Six C. Adaptive thermostability of light-harvesting complexes in
508 marine picocyanobacteria. *ISME J* 2017; **11**: 112–124.

42. Pittera J, Jouhet J, Breton S, Garczarek L, Partensky F, Maréchal É, et al. Thermoacclimation and genome adaptation of the membrane lipidome in marine *Synechococcus*. *Environ Microbiol* 2018; **20**: 612–631.

43. Six C, Ratin M, Marie D, Corre E. Marine *Synechococcus* picocyanobacteria: Light utilization across latitudes. *Proc Natl Acad Sci U S A* 2021; **118**: e2111300118.

44. Tai V, Palenik B. Temporal variation of *Synechococcus* clades at a coastal Pacific Ocean monitoring site. *ISME J* 2009; **3**: 903–915.

45. Hunter-Cevera KR, Neubert MG, Solow AR, Olson RJ, Shalapyonok A, Sosik HM. Diel size distributions reveal seasonal growth dynamics of a coastal phytoplankter. *Proc Natl Acad Sci* 2014; **111**: 9852–9857.

46. Xia X, Liu H, Choi D, Noh JH. Variation of *Synechococcus* pigment genetic diversity along two turbidity gradients in the China seas. *Microb Ecol* 2018; **75**: 10–21.

47. Kopf A, Bicak M, Kottmann R, Schnetzer J, Kostadinov I, Lehmann K, et al. The ocean sampling day consortium. *GigaScience* 2015; **4**: 27–27.

48. ten Hoopen P, Cochrane G, MicroB3 Consortium. Ocean Sampling Day Handbook. 2016. https://repository.oceanbestpractices.org/handle/11329/616

49. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; **30**: 2114–2120.

50. Ocean Sampling Day Consortium P. Registry of samples and environmental context from the Ocean Sampling Day 2014. 2015. PANGAEA. http://datadiscoverystudio.org/geoportal/rest/metadata/item/6146e7124e4d49e5b6f26751a90a6e33/html

51. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; **215**: 403–410.

52. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics* 2009; **10**: 421–421.

53. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. *Genome Res* 2007; **17**: 377–386.

54. Maechler M, Peter R, Struyf A, Hubert M, Hornik K. Cluster: Cluster analysis basics and extensions. 2013.

55. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, et al. vegan: Community Ecology Package. 2015.

56. Wickham H. ggplot2: elegant graphics for data analysis. 2009. Springer New York.

57. Vaulot D, Le Gall F, Marie D, Guillou L, Partensky F. The Roscoff Culture Collection (RCC): a collection dedicated to marine picoplankton. *Nova Hedwig* 2004; **79**: 49–70.

58. Rippka R, Coursin T, Hess W, Lichtlé C, Scanlan DJ, Palinska KA, et al. *Prochlorococcus marinus* Chisholm *et al.* 1992 *subsp. pastoris subsp. nov.* strain PCC 9511, the first axenic chlorophyll *a2/b2*-containing cyanobacterium (Oxyphotobacteria). *Int J Syst Evol Microbiol* 2000; **50**: 1833–1847.

59. Marie D, Brussaard CPD, Partensky F, Vaulot D. Flow cytometric analysis of phytoplankton, bacteria and viruses. In: Robinson JP (ed). *Current Protocols in Cytometry*. 1999. John Wiley & Sons, New York, pp 1–15.

60. Bernard O, Rémond B. Validation of a simple model accounting for light and temperature effect on microalgal growth. *Bioresour Technol* 2012; **123**: 520–527.

61. Olson RJ, Chisholm SW, Zettler ER, Armbrust EV. Pigment, size and distribution of *Synechococcus* in the North Atlantic and Pacific oceans. *Limnol Oceanogr* 1990; **35**: 45–58.

556 62. Varkey D, Mazard S, Ostrowski M, Tetu SG, Haynes P, Paulsen IT. Effects of low
557 temperature on tropical and temperate isolates of marine *Synechococcus*. *ISME J* 2016;
558 **10**: 1252–1263.
559 63. Moore L, Goericke R, Chisholm S. Comparative physiology of *Synechococcus* and
560 *Prochlorococcus*: influence of light and temperature on growth, pigments, fluorescence
561 and absorptive properties. *Mar Ecol Prog Ser* 1995; **116**: 259–275.
562 64. Šilović T, Balagué V, Orlić S, Pedrós-Alió C. Picoplankton seasonal variation and
563 community structure in the northeast Adriatic coastal zone. *FEMS Microbiol Ecol* 2012;
564 **82**: 678–691.
565 65. Babic I, Petric I, Bosak S, Mihanovic H, Dupcic Radic I, Ljubesic Z. Distribution and
566 diversity of marine picocyanobacteria community: Targeting of *Prochlorococcus*
567 ecotypes in winter conditions (southern Adriatic Sea). *Mar Genomics* 2017; **36**: 3–11.
568 66. Tai V, Burton R, Palenik B. Temporal and spatial distributions of marine *Synechococcus*
569 in the Southern California Bight assessed by hybridization to bead-arrays. *Mar Ecol*
570 *Prog Ser* 2011; **426**: 133–147.
571 67. Choi DH, Noh JH, An SM, Choi YR, Lee H, Ra K, et al. Spatial distribution of cold-
572 adapted *Synechococcus* during spring in seas adjacent to Korea. *Algae* 2016; **31**: 231–
573 241.
574 68. Edwards KF, Thomas MK, Klausmeier CA, Litchman E. Phytoplankton growth and the
575 interaction of light and temperature: A synthesis at the species and community level.
576 *Limnol Oceanogr* 2016; **61**: 1232–1244.
577 69. Choi DH, Noh JH, Shim J. Seasonal changes in picocyanobacterial diversity as revealed
578 by pyrosequencing in temperate waters of the East China Sea and the East Sea. *Aquat*
579 *Microb Ecol* 2013; **71**: 75–90.

70. Feingersch R, Suzuki MT, Shmoish M, Sharon I, Sabehi G, Partensky F, et al. Microbial community genomics in eastern Mediterranean Sea surface waters. *ISME J* 2010; **4**: 78–87.

71. Katsanevakis S, Coll M, Piroddi C, Steenbeek J, Ben Rais Lasram F, Zenetos A, et al. Invading the Mediterranean Sea: biodiversity patterns shaped by human activities. *Front Mar Sci* 2014; **1**: 32.

72. Kim Y, Jeon J, Kwak MS, Kim GH, Koh IS, Rho M. Photosynthetic functions of *Synechococcus* in the ocean microbiomes of diverse salinity and seasons. *PLoS ONE* 2018; **13**: 4–8.

73. Lipizer M, Partescano E, Rabitti A, Giorgetti A, Crise A. Qualified temperature, salinity and dissolved oxygen climatologies in a changing Adriatic Sea. *Ocean Sci* 2014; **10**: 771–797.

74. Xia X, Guo W, Tan S, Liu H. *Synechococcus* assemblages across the salinity gradient in a salt wedge estuary. *Front Microbiol* 2017; **8**: 1254.

75. Marsan DW. Adaptive mechanisms of an estuarine *Synechococcus* based on genomics, transcriptomics, and proteomics. *ProQuest Dissertations and Theses*. 2016. University of Maryland, College Park.

76. Moutin T, Thingstad TF, Van Wambeke F, Marie D, Slawyk G, Raimbault P, et al. Does competition for nanomolar phosphate supply explain the predominance of the cyanobacterium *Synechococcus*? *Limnol Oceanogr* 2002; **47**: 1562–1567.

77. Pernice MC, Forn I, Gomes A, Lara E, Alonso-Sáez L, Arrieta JM, et al. Global abundance of planktonic heterotrophic protists in the deep ocean. *ISME J* 2015; **9**: 782–792.

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

Figure legends

Figure 1: Relative abundance of marine *Synechococcus* clades in OSD stations. Stations are located at the bottom of barplots of relative abundance. The insert shows a close-up version of Europe. Station numbers are shown in Supplementary Figure S1. Categories 5.1 and *Syn* correspond to reads that could not be assigned to a clade but were assigned to the level of *Synechococcus* SC 5.1 or *Synechococcus* genus, respectively.

Figure 2: Temperature preferenda of eight marine *Synechococcus* strains. Growth rate as a function of temperature of acclimated growth. Two strains were chosen within each of the four major clades I, II, III and IV (top to bottom). All cultures were grown at a light intensity of 20 µmol quanta m^{-2} s^{-1}. Error bars are standard deviation from the mean based on at least 3 replicates ($n \geq 3$).

Figure 3: Clusters of OSD stations based on relative abundance profiles of *Synechococcus* clades. OSD stations were clustered based on the relative abundance profiles of marine *Synechococcus* clades using Bray-Curtis distance. The upper panel indicates water temperature. Categories 5.1 and *Syn* correspond to reads that could not be assigned to a clade but were assigned to the level of *Synechococcus* SC 5.1 or *Synechococcus* genus, respectively.

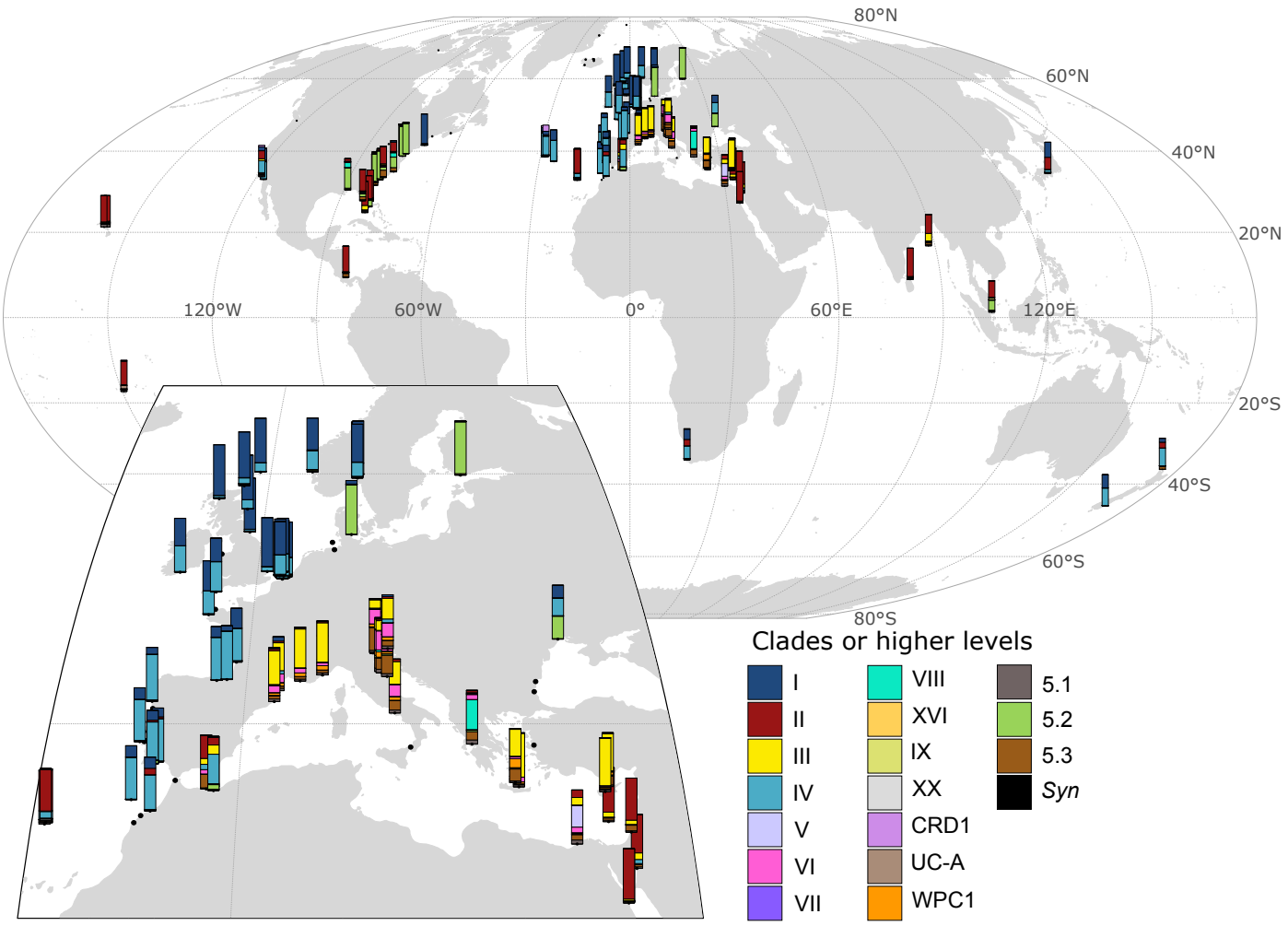
Figure 4: Violin plots showing the distribution of temperature and salinity for each cluster of OSD stations defined in Fig. 3. A. Temperature. B. Salinity. The black dot in each violin plot shows the median value. Different letters indicate significantly different distributions (Dunn test. adjusted p-value < 0.05). The same analysis considering distance to the nearest coast gave no significant result.

Supplementary information

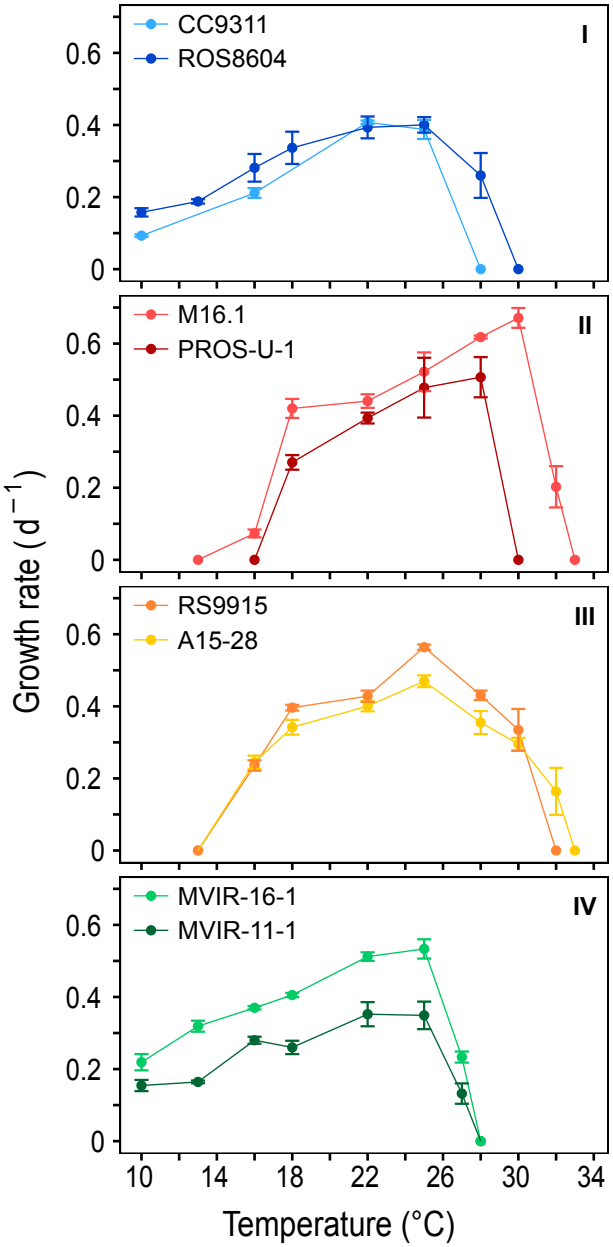
Supplementary Figure S1: **Map of OSD stations.** All OSD stations that were analyzed in this study are indicated by their number. The inset shows a close-up view of Europe.

Dataset S1: OSD samples used in this study. Characteristics and accession numbers of the OSD samples analyzed in this study and corresponding contextual data, as retrieved from PANGAEA (https://doi.pangaea.de/10.1594/PANGAEA.854419; [50]).

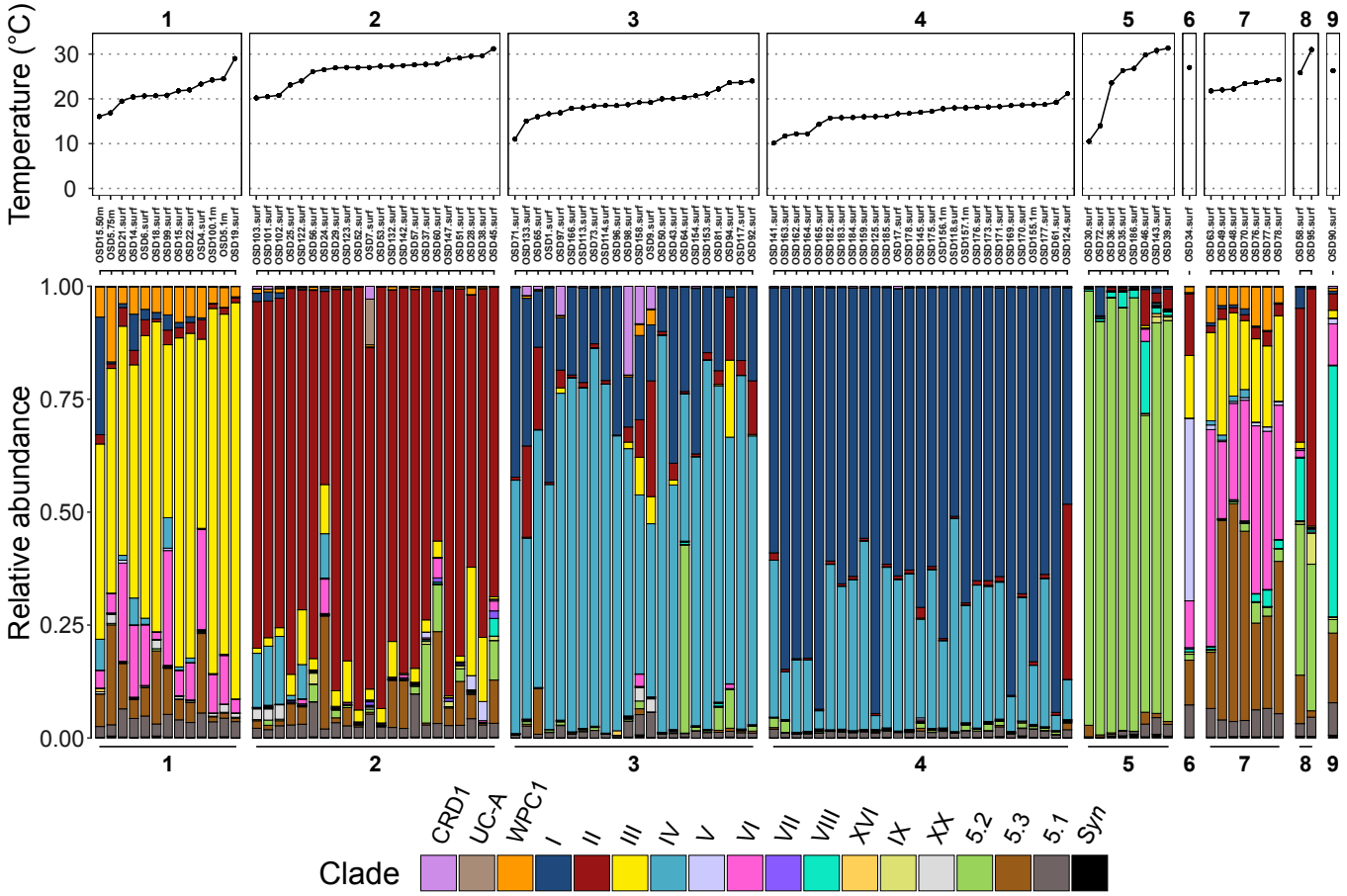
Dataset S2: **Summary data for the 863 complete genomes of aquatic bacteria used as reference in this study.** Genomes sequences were retrieved either from Cyanorak *v2.1* (www.sb-roscoff.fr/cyanorak), NCBI Genbank for additional *Synechococcus* whole genomes and for genomes other than marine *Synechococcus* and *Prochlorococcus* listed in Cyanobase (http://genome.microbedb.jp/cyanobase/), or proGenomes (http://progenomes.embl.de/index.cgi). The table includes subclade designation based on [11].



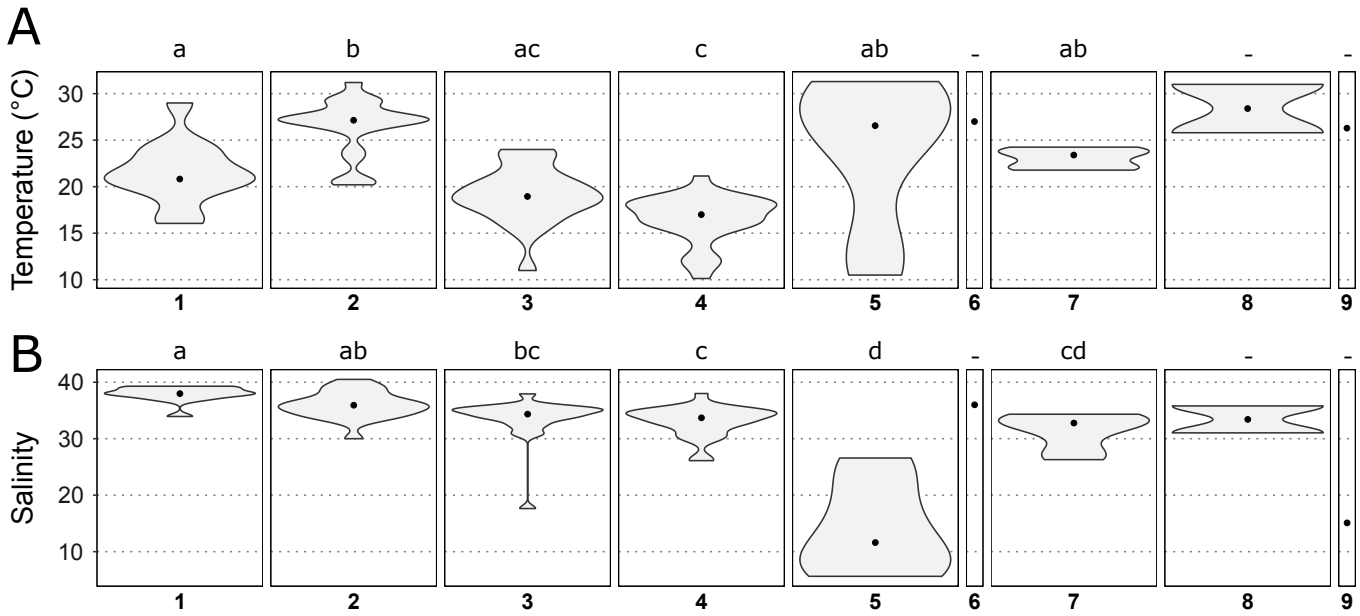
Doré et al. Figure 1



Doré et al. Figure 2



Doré et al. Figure 3



Doré et al. Figure 4