Global Phylogeography of Marine Synechococcus in Coastal Areas Reveals Strong Community Shifts

Hugo Doré,† Jade Leconte,‡ Ulysse Guyet,§ Solène Breton,α Gregory K. Farrant,α David Demory,α⁎ Morgane Ratin,α Mark Hoebeke,β Erwan Corré,β Frances D. Pitt,γ Martin Ostrowski,ι‡ David J. Scanlan,ι Frédéric Partensky,ι Christophe Six,α Laurence Garczarekα

αSorbonne Université, CNRS, UMR 7144 Adaptation and Diversity in the Marine Environment (AD2M), Station Biologique de Roscoff (SBR), Roscoff, France
βCNRS, FR 2424, ABiMS Platform, Station Biologique, Roscoff, France
γUniversity of Warwick, School of Life Sciences, Coventry, United Kingdom

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 metagenomics data set 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 reveals several important characteristics and specificities of the coastal communities of Synechococcus worldwide.

IMPORTANCE Synechococcus is the second most abundant phytoplanktonic organism on Earth, and its wide genetic diversity allowed it to colonize all the oceans except for polar waters, with different clades colonizing distinct oceanic niches. In recent years, the use of global metagenomics data sets has greatly improved our knowledge of "who is where" by describing the distribution of Synechococcus clades or ecotypes in the open ocean. However, little is known about the global distribution of Synechococcus ecotypes in coastal areas, where Synechococcus is often the dominant phytoplanktonic organism. Here, we leverage the global Ocean Sampling Day metagenomics data set to describe Synechococcus community composition in coastal areas worldwide, revealing striking community shifts, in particular along the coasts of Europe. As temperature appears as an important driver of the community composition, we also characterize the thermal preferenda of 8 Synechococcus strains, bringing new insights into the adaptation to temperature of the dominant Synechococcus clades.

KEYWORDS marine cyanobacteria, Synechococcus, coastal areas, Ocean Sampling Day, temperature, niche partitioning, metagenomics

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 physicochemical 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 to 50°N latitudinal band, this organism preferentially thriving in oligotrophic areas, while 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 2 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 Herdman et al. (19) also encompasses 3 major Synechococcus/Cyanobium lineages, called sub-clusters (SC) 5.1 through 5.3 (9, 20). Although a number of phylogenetic studies based on individual markers have considered SC 5.2 and Cyanobium as being 2 distinct lineages (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 (27, 28), and in low salinity areas such as the Baltic Sea (29). 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, 30–32). 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, 33). Finally, SC 5.1, a lineage that rapidly diversified after the advent of the Prochlorococcus radiation (34, 35), 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 to 5.3 (12). From 10 to 15 phylogenetic clades have been defined within SC 5.1 depending on the phylogenetic marker (10, 30, 36) but studies of the global distribution patterns of Synechococcus populations in open ocean waters have shown that there are 5 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 (5, 12, 31, 32, 37, 38).

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’ (38–42). 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 (43), a similar change in membrane lipids (40, 44) and an increase of the photoprotection capacities using the orange carotenoid protein (OCP; 45). 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 (5, 12, 32, 37).

Although the variability of picocyanobacterial communities and the main physicochemical factors driving their composition are starting to be well understood in open ocean environments, the picture is much more fragmentary in coastal areas. Indeed, most coastal studies have concerned specific regions, such as the Baltic Sea, the Californian coast or estuarine waters (e.g., the Chesapeake Bay or coastal waters of Hong Kong; [21–23, 26, 28, 29, 45]) and/or a few long-term monitoring sites of coastal...
observatories (24, 46–50). Here, in order to get a more global view of the genetic diversity and biogeography of coastal populations of picocyanobacteria and to better understand how they vary between distinct coastal areas and differ from open ocean populations, we used metagenomic data from the Ocean Sampling Day (OSD) 2014 campaign (51), 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.

RESULTS AND DISCUSSION

Biogeography of coastal picocyanobacterial communities is influenced by seawater temperature. Most of the stations sampled during the OSD 2014 campaign (51) correspond to coastal areas with only 17 of 157 stations located over 11 nautical miles from the nearest coast. This data set 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. 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 largely outnumber Prochlorococcus in coastal areas (2, 7, 24, 52), was detected with sufficient coverage to perform reliable taxonomic assignment at the clade level in 102 out of the 150 OSD metagenomes, using a conservative lowest common ancestor approach for taxonomic assignment (see Materials and Methods). At most stations, the Synechococcus/Cyanobium community was dominated by 1 or 2 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 (6, 12, 15, 30, 32, 37), 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, which is in agreement with previous observations in open ocean waters (12, 30, 32, 37), as well as with the low temperatures of isolation sites of clade I and IV strains (39).

To further explore the role of temperature on the differential latitudinal distribution of members of clades I to IV, we characterized the thermal referenda of 8 strains belonging to these clades (Table 1, Fig. 2). While several strains belonging to clade I were previously shown to withstand colder temperatures than their tropical clade II counterparts (38–40, 53), growth optima and boundary limits for temperature were only available for 1 clade IV (38, 40, 42) and 2 closely related clade III strains (38, 40, 41, 54), and results were obtained in different light conditions, making them difficult to compare. Here, the direct comparison of clades I and IV strains, grown under the same conditions, showed quite similar thermal preferences. All tested strains of these 2 clades displayed an optimal temperature for growth of about 24°C according to our model fit (Fig. 2 and Table 2), and were 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 4 clades.

Besides the abundance of clades I and IV, coastal *Synechococcus* communities also exhibited some other specificities 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, 32, 37), 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 (9, 21, 36). 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 (Fig. 4B, cluster 5). Finally, clades V and VI, which were not distinguished from clade VII (and CRD1) in previous global surveys of *Synechococcus* distribution using the low-resolution 16S rRNA marker gene, were found to be locally abundant in the data set. While the V/VI/VII/CRD1 group was considered to be widely distributed in oceanic waters (15, 30, 55), 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, 49, 56).

A progressive latitudinal shift in *Synechococcus/Cyanobium* communities along the coast of Europe. Besides the above-mentioned 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 OSD 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, a clear progressive northward shift was observed toward the dominance of clade I in the North Sea (Fig. 1). Clustering of stations based on clade relative abundance indeed highlighted 2 groups of stations; the first one dominated by clade IV (Fig. 3, cluster 3) and the second one by clade I (Fig. 4B, cluster 4). Interestingly, clade I was found to dominate at stations that display a significantly lower salinity than those dominated by clades II or III (Fig. 3, 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.

Several potential reasons have been evoked to explain the variations in the clade I to clade IV ratio across space or time in coastal areas (46, 48, 57). These include differences in their respective adaptation to metal and/or nitrate concentrations (32, 46, 50), as well as transport and mixing of populations by advection, e.g., in the vicinity of the Svalbard island, where the Gulf Stream current brings clade IV populations in summer (4) or in 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 (58). Clade I was also suggested to be more coastal and opportunistic than clade IV (9, 46) but this hypothesis is not confirmed by this study since many coastal stations (cluster 3) are actually dominated by clade IV. Finally, this northward shift could also rely on differences occurring at a finer taxonomic level since several studies pointed out to the existence of several genotypes within clades I and/or IV, the relative abundance of which varies according to depth, latitude, phage interactions, season, or over the course of a bloom (4, 6, 27, 48, 50, 57, 59–61). In particular, coastal time series showed that different genotypes within clade I and/or IV follow distinct seasonal patterns, suggesting a

| TABLE 1 Information regarding the *Synechococcus* strains used in this study |
|---------------------------------|-----------------|--------|--------|---------|---------|---------|---------|---------|
| Strain name | CC9311 | ROS8604 | M16.1 | PROS-U-1 | RS9915 | A15-28 | MVIR-16-1 | MVIR-11-1 |
| RCC no. | 1086 | 2380 | 791 | 2369 | 2553 | 2556 | 2570 | 1695 |
| Clade | I (Ia) | I (Ib) | II (Ila) | II (Iib) | III (IIa) | III (IIb) | IV (Iva) | IV (Iva) |
| Pigment type | 3dA | 3a | 3a | 3dB | 3dB | 3c | 3d | 3a |
| Isolation site | California Current | English Channel | Gulf of Mexico | Moroccan upwelling | Red Sea - Gulf of Aqaba | Atlantic Ocean | Northern gyre | North Sea |
| Isolation latitude | 32° 0’ N | 48° 43’ N | 27° 42’ N | 30° 8’ N | 29° 28’ N | 31° 15’ N | 60° 19’ N | 56° 56’ N |
| Isolation longitude | 124° 31’ W | 3° 59’ W | 91° 18’ W | 10° 3’ W | 34° 55’ E | 20° 43’ W | 3° 29’ W | 3° 59’ E |
| Isolation date | 01/01/93 | 11/24/86 | 02/09/04 | 09/12/99 | 10/18/99 | 09/25/04 | 07/21/07 | 01/14/07 |
| Isolation depth (m) | 95 | 1 | 275 | 5 | 10 | 15 | 10 | 10 |
| Isolation temp. (°C) | 16.59 | 12.81 | 24.15 | 21.51 | 23.98 | 25.15 | 25.15 | 11.99 |
| Coast distance (km) | 458 | 0.5 | 156 | 35 | 3 | 369 | 78 | 201 |

aThe pigment type nomenclature is described in Humily et al. (2014).

bIsolation temperatures were retrieved from National Oceanic and Atmospheric Administration (NOAA) as described in Pittera et al. (2014).
differential adaptation of these genotypes to water temperature or other environmental parameters (48, 50, 60, 62).

Consistently, comparison of our experimental data with previous data acquired under the same light conditions (40) brings evidence that clade IV is comprised of distinct genotypes exhibiting different lower temperature boundary limits and, thus, potentially colonizing different thermal niches, as was also shown for clade I strains (39). Indeed, the 2 clade IV strains characterized here were sampled at high latitude (Table 1), and show a higher tolerance to cold temperatures than BL107, another clade

**TABLE 2** Parameters of growth versus temperature for 8 *Synechococcus* strains representative of the four most abundant clades

<table>
<thead>
<tr>
<th>Clade</th>
<th>Strain</th>
<th>$T_{opt}$ measured a</th>
<th>$T_{opt}$ model b</th>
<th>$T_{max}$ range d</th>
<th>$T_{max}$ measured e</th>
<th>$T_{max}$ model f</th>
<th>$T_{max}$ range g</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>CC9311</td>
<td>22</td>
<td>24.6</td>
<td>22.98–25.91</td>
<td>25</td>
<td>27.7</td>
<td>26.23–28.97</td>
</tr>
<tr>
<td></td>
<td>ROS8604</td>
<td>25</td>
<td>23.7</td>
<td>22.23–24.80</td>
<td>28</td>
<td>29.7</td>
<td>28.19–31.36</td>
</tr>
<tr>
<td>IV</td>
<td>MVIR-16-1</td>
<td>25</td>
<td>24.3</td>
<td>23.37–25.09</td>
<td>27</td>
<td>27.3</td>
<td>27.06–27.47</td>
</tr>
<tr>
<td></td>
<td>MVIR-11-1</td>
<td>22</td>
<td>23.8</td>
<td>22.42–25.05</td>
<td>27</td>
<td>27.3</td>
<td>26.93–27.63</td>
</tr>
<tr>
<td>III</td>
<td>RS9915</td>
<td>25</td>
<td>24.8</td>
<td>23.11–28.44</td>
<td>32</td>
<td>32.2</td>
<td>25.37–34.88</td>
</tr>
<tr>
<td></td>
<td>A15-28</td>
<td>25</td>
<td>23.4</td>
<td>22.97–23.72</td>
<td>32</td>
<td>33.9</td>
<td>32.79–34.82</td>
</tr>
<tr>
<td>II</td>
<td>M16.1</td>
<td>30</td>
<td>29.0</td>
<td>27.99–30.19</td>
<td>32</td>
<td>32.3</td>
<td>31.93–32.51</td>
</tr>
</tbody>
</table>

a$T_{opt}$, optimal temperature for growth.

b$T_{opt}$, measured values, see Fig. 3.

cValues estimated by a model of growth versus temperature fitted to the data shown in Fig. 3.

dConfidence intervals of model predictions (95%).

e$T_{max}$, maximal temperature for growth.

**FIG 2** Temperature preferenda of 8 marine *Synechococcus* strains. Growth rate as a function of temperature of acclimated growth. Two strains were chosen within each of the 4 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 ($n \geq 3$). The line represents the best fit of the Cardinal Temperature Model with Inflection (BR model; 86).
IV strain isolated in the Mediterranean Sea (40). Thus, the ecological drivers of clades I and IV distribution may be difficult to identify due to underlying differences within each clade, and a finer taxonomic resolution might be necessary to observe a significant effect of temperature on the distribution of these populations.

**FIG 3** Clusters of Ocean Sampling Day (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; two stations will cluster together if they have a similar composition in *Synechococcus* clades. The clustering dendrogram is available as Fig. S2. (A) The upper panel indicates water temperature. The lower panel shows the nine clusters of relative abundance profiles of *Synechococcus* clades. Categories 5.1 and Syn correspond to reads that could not be assigned to a clade but were assigned to the higher taxonomic levels of *Synechococcus* SC 5.1 or *Synechococcus* genus, respectively. (B) Geographical distribution of the nine clusters of OSD stations along the European coasts. A global map of cluster distribution is available as Fig. S3.
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 composition is quite similar to that previously described by Farrant et al. (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, 30). Most of the stations of the Adriatic Sea formed a distinct cluster (cluster 7), where the same clades were present but in different proportions, clade VI and SC 5.3 taking over clade III. Finally, stations OSD34 and OSD90, located on the Egyptian and Greek coasts, respectively, the only stations of the OSD data set comprising a high proportion of clade V or VIII, formed a cluster on their own (clusters 6 and 9, respectively). While these 4 clusters (clusters 1, 6, 7, and 9) are specific to the Mediterranean Sea, it is worth noting that 2 stations at the easternmost end of the Mediterranean Sea (OSD123 and OSD132) (Fig. 3 and Fig. S1) fell into cluster 2, dominated by clade II, and showed a clade composition very similar to the samples collected in the Red Sea (OSDS2 and OSDS3). While this observation could be due to similar environmental conditions in the Eastern Mediterranean Sea and the Red Sea, it also suggests that Israeli coastal areas may be influenced by waters entering the Mediterranean Sea via the Suez Canal, consistent with previous findings for *Synechococcus*, *Prochlorococcus*, as well as for many larger organisms (63, 64). Indeed, the water has been estimated to flow northward through the Canal until the end of June with up to 1250 m$^3$ s$^{-1}$, facilitating species migrations to the Mediterranean Sea (65).

Interestingly, the 3 specific clusters identified in the Mediterranean Sea displayed different temperature and salinity characteristics (Fig. 4A and B). The salinity range of stations in cluster 1 (dominated by clade III) was narrow (average salinity 37.90 psu, median 37.98 psu) and significantly higher than that of cluster 7 (dominated by clade VI and SC 5.3, average salinity 31.43 psu, median 32.77 psu), suggesting that clade VI and SC 5.3 are able to cope with lower salinities. Consistently, SC 5.3 was recently found to encompass members colonizing freshwater lakes (25, 33), while in the marine environment, this subcluster was reported both in strictly marine waters (12, 31) and in low salinity waters (66). Our study also brings new insights into the ecological niche occupied by clade VI, whose distribution was so far poorly known (30), and that appears to be restricted to coastal regions of intermediate salinity. All stations of the Adriatic Sea comprising cluster 6 were indeed sampled in the northwestern part of this area, where the influence of the Po River plume may be important (67). This distribution is consistent with previous observations of the closely related, and often co-occurring, clade V in low salinity surface
waters of the Adriatic Sea (49) and of both clades V and VI in the Pearl River Estuary (23). Laboratory experiments also showed that representative strains of these 2 clades can tolerate salinities as low as 15 psu (68). 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) (Fig. 1 and Fig. 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 (69).

Conclusion. The OSD data set 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 (5, 12, 32, 37, 70, 71). 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 that only a few physicochemical parameters were collected, this data set allowed us to considerably improve our knowledge of the distribution of Synechococcus/Cyanobium lineages in coastal areas, to gain insights into the realized environmental niches of the main ones, including some that were previously poorly known such as clade VI, as well as to reinforce hypotheses about thermal niche differentiation that were supported by laboratory experiments on a set of representative strains. Still, it is likely that the shifts observed here at the summer solstice would exhibit different latitudinal boundaries at other seasons since time series studies of Synechococcus community composition have revealed strong seasonal patterns, notably due the succession of different thermotypes (48, 50, 58, 60, 62). A continued effort toward global instantaneous surveys of microbial diversity in coastal areas over the long term and at different seasons would be invaluable to monitor the evolution of microbial communities in relation to global change.

MATERIALS AND METHODS

OSD metagenomics data. OSD 2014 is a global sampling campaign that took place on June 21st, 2014 and sampled 157 stations worldwide for metagenomes (Table 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 in (72) and at https://github.com/ocean-sampling-day/OSD2014. Data were downloaded from the EBI (see data availability below) 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 (73) before discarding reads shorter than 100 nucleotides. Contextual data were downloaded from PANGEA (https://doi.pangaea.de/10.1594/PANGAEA.854419) and the data used in this study are listed in Table S1: as the contextual data are very sparse for most parameters, we only used water temperature and salinity data that were available for a sufficient number of stations. A map of OSD stations used in this study is available as Fig. S1.

Taxonomic assignment of metagenomic reads. Because OSD metagenomes were not sequenced deeply enough to rely on a single high resolution marker gene for taxonomic assignment (12, 74), we used a Whole Genome Recruitment (WGR; [75, 76]) approach against a reference genome database of 863 publicly available complete genomes of aquatic bacteria (Table S2). The latter encompassed 141
The genomes of marine picocyanobacteria as well as 722 cyanobacterial or other aquatic microbial genomes, 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). This large number of outgroup genomes representative of the known diversity of the oceans was selected to minimize the risk of unspecific mapping on picocyanobacterial genomes.

BLASTN (v2.2.28+) (77, 78) was used to align metagenomic reads against this reference database. 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 percentage of identity of 30% (-perc_identity 30), a filter on e-value of 10^{-2} (-evalue 0.01), and by selecting the blastn algorithm (-task blastn) to allow for reads to map on multiple reference genomes. BLASTN results were then parsed using the Lowest Common Ancestor method (79). For each read, BLAST matches with over 80% identity, corresponding to a major discontinuity of the average nucleotide identity (ANI) values within the marine picocyanobacteria radiation (20), aligned over more than 90% of their length against a reference genome were kept if their BLAST score was within 5% of the best score. Then, the read was attributed to the lowest common ancestor of these matches (i.e., strain, clade, subcluster, or genus). Counts of reads assigned to the strain or subclade levels were ultimately aggregated by clade. Two additional categories were made for reads that could only be assigned to the level of Synechococcus subcluster 5.1 (SC 5.1 in Fig. 1 and 3) or even Synechococcus genus (Syn in Fig. 1 and 3). With such a method, potential remaining unspecifically mapped reads that would not have been filtered out by our first filter are highly unlikely to be assigned to the clade level. Moreover, this method allows to be very conservative on the taxonomic assignment of reads, and to avoid any misannotation of reads mapping to conserved regions of the genome.

Analysis of picocyanobacterial community composition. In order to account for the potential variation in genome length among clades, read counts were divided by the average genome length within each clade. To minimize the noise in recruitment data, we then removed from the data set stations with less than 600 recruited reads per million bp, corresponding to a genome coverage of ca. 16%, since each clade. To minimize the noise in recruitment data, we then removed from the data set stations with less than 600 recruited reads per million bp, corresponding to a genome coverage of ca. 16%, since reads are 242 bp long on average. Read counts at each station were further normalized by the total number of picocyanobacterial reads recruited at this station to assess relative abundances of taxa with respect to the cutoff used to delineate clusters are represented on Fig. S2. Figures were drawn in R v3.03 with package ggplot2 v1.0.1 (82).

Thermal preferenda of strains representative of the most abundant clades in situ. Two strains of the 4 most abundant Synechococcus clades in Fe-replete areas (clades I to IV) were selected from the Roscoff Culture Collection (Table 1) (http://roscoff-culture-collection.org/; [83]). Strains were grown in polystyrene flasks in PCR-S11 medium (84) supplemented with 1 mM sodium nitrate. The seawater was reconstituted using Red Sea Salts and distilled water. Cultures of the 8 strains were acclimated at least 2 weeks to a range of temperatures from 10°C to 33°C, within temperature-controlled chambers (Liebherr-Hausgeräte) and continuous light was provided by green/white/blue LEDs (Alpheus) at an irradiance of 20 μmol photons m^{-2} s^{-1}. After acclimation, cultures were split into 3 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) and stored at ～80°C until analysis (85). Cell concentration was determined using a flow cytometer (FACSCanto II, Becton, Dickinson) with laser emission set at 488 nm, and using distilled water as sheath fluid.

To estimate the maximum population growth rates, we used the following equation:

\[
\frac{dN}{dt} = \mu N
\]

Where \(N\) is the cell abundances (in cells mL\(^{-1}\)) and \(\mu\) is the maximum population growth rate (in days\(^{-1}\)). We estimated \(\mu\) as the coefficient of the linear regression model performed on log-transformed \(N\) (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; 86), which describes the maximal phytoplankton growth rate \(\mu_{\text{max}}\) as a function of temperature (T) as follows:

\[
\mu_{\text{max}} = \left\{ \begin{array}{ll}
0 & \text{for } T < T_{\text{min}} \\
\mu_{\text{opt}} \Phi(T) & \text{for } T_{\text{min}} < T < T_{\text{opt}} \\
0 & \text{for } T > T_{\text{opt}}
\end{array} \right.
\]

where:

November/December 2022 Volume 7 Issue 6 10.1128/msystems.00656-22
TEXT:

\[
\phi(T) = \frac{(T-T_{\text{max}})(T-T_{\text{min}})}{(T_{\text{opt}}-T_{\text{min}})(T_{\text{opt}}-T_{\text{max}})^2}
\]

\[T_{\text{max}}\] and \[T_{\text{min}}\] are the minimal and maximal growth temperatures and \[T_{\text{opt}}\] is the optimal growth temperature where \[\mu_{\text{max}} = \mu_{\text{opt}}\]. We estimated the cardinal growth temperatures \(T_{\text{max}}, T_{\text{opt}}, T_{\text{min}}\) and the optimal growth rate \(\mu_{\text{opt}}\) using the same procedure as in (86). Briefly, we used Synechococcus experimental growth rates obtained at different temperatures and fitted equations 1 and 2 by minimizing the Euclidian distance (fitting error) between model and data (Residual Sum of Squares) using the Scilab leastsq function. More information on the fitting procedure can be found in (86; section 2.2, parameter identification). Because of the shape of the growth response curve and the variability in the experimental data for low temperatures, our data did not allow to constrain \(T_{\text{min}}\), but this did not affect our estimation of other parameters (Table 2).

**Data availability.** 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). Contextual data collected at all OSD stations were retrieved from PANGAEA (https://doi.pangaea.de/10.1594/PANGAEA.854419; Ocean Sampling Day Consortium, 2015). All genomes of aquatic bacteria used as reference or outgroups in this study are publicly available and their database origin and accession numbers are listed in Table S2.

**SUPPLEMENTAL MATERIAL**

Supplemental material is available online only.

**FIG S1,** PDF file, 2.3 MB.

**FIG S2,** PDF file, 0.4 MB.

**FIG S3,** PDF file, 0.8 MB.

**TABLE S1,** XLSX file, 0.03 MB.

**TABLE S2,** XLSX file, 0.1 MB.

**ACKNOWLEDGMENTS**

We thank the OSD Consortium for sampling, sequencing, and making the data analyzed in this paper available, as well as the Roscoff Culture Collection (http://roscoff-culture-collection.org/) for providing Synechococcus strains used in this study.

Financial support for the OSD program was provided by the European Union program MicroB3 (UE-contract-287589) and authors were supported by the French “Agence Nationale de la Recherche” programs SAMOSA (ANR-13-ADAP-0010) and CINNAMON (ANR-17-CE02-0014-01), as well as the European program Assemble Plus (H2020-INFRAIA-1-2016-2017; grant no. 730984).

D.J.S. received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement No 883551).

We declare no competing interests.

**REFERENCES**


