

Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria

Gregory K. Farrant^{a,1,2}, Hugo Doré^{a,1}, Francisco M. Cornejo-Castillo^b, Frédéric Partensky^a, Morgane Ratin^a, Martin Ostrowski^c, Frances D. Pitt^d, Patrick Wincker^e, David J. Scanlan^d, Daniele Iudicone^f, Silvia G. Acinas^b, and Laurence Garczarek^{a,3}

^aSorbonne Universités, Université Paris 06, CNRS, UMR 7144, Station Biologique, CS 90074 Roscoff, France; ^bDepartment of Marine Biology and Oceanography, Institute of Marine Sciences (ICM), Consejo Superior de Investigaciones Científicas (CSIC), Barcelona ES-08003, Spain; ^cDepartment of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, NSW 2109, Australia; ^dSchool of Life Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom; ^eCommissariat à l’Energie Atomique et aux Energies Alternatives (CEA), Institut de Génomique, Genoscope, 91057 Evry, France; and ^fStazione Zoologica Anton Dohrn, 80121 Naples, Italy

Edited by David M. Karl, University of Hawaii, Honolulu, HI, and approved May 6, 2016 (received for review January 5, 2016)

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

molecular ecology | metagenomics | Tara Oceans | *Synechococcus* | *Prochlorococcus*

The ubiquitous marine picocyanobacteria *Prochlorococcus* and *Synechococcus* are major contributors to global chlorophyll biomass, together accounting for a quarter of global carbon fixation in marine ecosystems, a contribution predicted to further increase in the context of global change (1–3). Thus, determining how environmental conditions control their global distribution patterns, particularly at a fine taxonomic resolution (i.e., sufficient to identify lineages with distinct traits), is critical for understanding how these organisms populate the oceans and in turn contribute to global carbon cycling. The availability of numerous strains in culture and sequenced genomes make picocyanobacteria particularly well suited for cross-scale studies from genes to the global ocean (4). Physiological studies of a range of *Prochlorococcus* strains isolated from various depths and geographical regions notably revealed the occurrence of genetically distinct populations exhibiting different light or temperature growth optima and tolerance ranges (5, 6). These observations are congruent, on the one hand, with the well-known depth partitioning of genetically distinct *Prochlorococcus* populations in the ocean, with high light-adapted (hereafter HL) populations in the upper lit layer and low

light-adapted (hereafter LL) populations located further down the water column, and on the other hand, with the latitudinal partitioning between *Prochlorococcus* HLI and HLII clades that are adapted to temperate and tropical waters, respectively (5, 7, 8). For *Synechococcus*, although no clear depth partitioning (i.e., phototypes) has been observed so far, the occurrence of different “thermotypes” has been clearly demonstrated among strains isolated from different latitudes (9, 10). This latter finding agrees well with biogeographical patterns of the most abundant *Synechococcus* lineages, with members of clades I and IV restricted to cold and temperate waters, whereas clade II populations are mostly found in warm, (sub)tropical areas (11–13). Recently, several studies have shown that iron could also be an important parameter controlling the composition of picocyanobacterial community structure, as *Prochlorococcus* HLIII/IV ecotypes (14, 15) and *Synechococcus* clade CRD1 (16, 17) were shown to be dominant within high nutrient–low chlorophyll (HLNC) areas, where iron is limiting. Most of these studies considered members of the same clade—that is, *Prochlorococcus* clades HLI–VI and LLI–VI or *Synechococcus*

Significance

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

Author contributions: G.K.F. and L.G. designed research; G.K.F., H.D., and M.R. performed research; M.O., F.D.P., P.W., D.J.S., and S.G.A. contributed new reagents/analytic tools; M.O., F.D.P., and D.J.S. provided *petB* reference sequences; P.W. provided sequencing of the metagenomic dataset; G.K.F., H.D., F.M.C.-C., D.J., and L.G. analyzed data; and G.K.F., H.D., F.P., and L.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. KU377785-990, KU670814-6, KU705397-460, and KU937818-30).

¹G.K.F. and H.D. contributed equally to this work.

²Present address: Food Safety, Environment, and Genetics, Matis Ltd., 113 Reykjavik, Iceland.

³To whom correspondence should be addressed. Email: laurence.garczarek@sb-roscoff.fr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1524865113/-DCSupplemental.