Adaptation à la niche écologique chez deux représentants majeurs du phytoplancton marin, Synechococcus et Prochlorococcus : des gènes à l'écosystème

Hugo Doré

THESE DE DOCTORAT DE
L'UNIVERSITE PIERRE ET MARIE CURIE

Spécialité Ecologie Microbienne
Ecole doctorale « Sciences de la Nature et de l'Homme : évolution et écologie » (ED 227)

Présentée par

Hugo DORÉ

En vue de l'obtention du grade de

DOCTEUR de L'UNIVERSITE PIERRE ET MARIE CURIE

Adaptation à la niche écologique chez deux représentants majeurs du phytoplancton marin, *Synechococcus* et *Prochlorococcus* : des gènes à l'écosystème

Soutenue le 14 décembre 2017, devant le jury composé de :

Dr. GARCIA-FERNANDEZ José Manuel.....Rapporteur
Université de Cordoue

Dr. LINDELL Debbie.....Rapporteur
Technion - Israel Institute of Technology

Dr. MAIGNIEN Loïs.....Examinateur
Université de Bretagne Occidentale

Dr. OBERNOSTERER Ingrid.....Examinateur
CNRS, Université Pierre et Marie Curie

Dr. DAUBIN Vincent.....Examinateur
Université Claude Bernard Lyon 1

Pr. THIEBAUT Eric.....Examinateur
Université Pierre et Marie Curie

Dr. GARCZAREK Laurence.....Directrice de thèse
CNRS, Université Pierre et Marie Curie

Station Biologique de Roscoff
UMR 7144 Adaptation et Diversité en Milieu Marin, Equipe Procaryotes Phototrophes Marins

A mes grand-parents

RESUMÉ

Les picocyanobactéries marines *Prochlorococcus* et *Synechococcus* sont les organismes photosynthétiques les plus abondants sur la planète et sont présentes dans tous les océans à l'exception de l'océan Austral, un succès écologique sans doute rendu possible par leur grande diversité génétique. Au cours de cette thèse, j'ai cherché à mieux comprendre les liens entre diversité génétique et adaptation à la niche écologique chez ces deux genres. Tout d'abord, l'étude de la répartition des populations de picocyanobactéries à l'échelle mondiale à l'aide d'un marqueur taxonomique très résolutif m'a permis de définir des unités taxonomiques écologiquement significatives pour les taxons dominants, d'améliorer la délimitation de leurs niches écologiques et d'identifier les principaux facteurs abiotiques influençant leur répartition *in situ*. La deuxième partie de ce travail de thèse a visé à identifier les bases génétiques de l'adaptation des picocyanobactéries marines à des niches écologiques distinctes. D'un point de vue évolutif, l'étude comparative de 81 génomes de ces organismes a révélé le rôle combiné des gains et pertes de gènes et des substitutions d'acides aminés dans la diversification de ces deux genres. L'analyse de la répartition de l'ensemble des gènes connus de picocyanobactéries marines dans l'océan mondial a également permis de montrer que chaque communauté, adaptée à des conditions environnementales données, possède un répertoire de gènes distinct. Enfin, le dernier volet de cette thèse a consisté en la caractérisation physiologique d'une souche modèle de *Synechococcus* et de 4 souches représentatives des écotypes dominant *in situ* soumises à des stress lumineux et thermiques afin de mieux comprendre les bases moléculaires de l'adaptation et la variabilité écotypique de la réponse au stress chez *Synechococcus*, notamment par une approche de transcriptomique à grande échelle. Les résultats obtenus au cours de cette thèse ont donc permis non seulement d'améliorer notre connaissance des niches écologiques occupées par les picocyanobactéries marines et du niveau taxonomique à considérer pour différencier les écotypes au sein de ces deux genres, mais également de mieux comprendre les mécanismes leur ayant permis de s'adapter à ces niches variées et de coloniser presque tous les océans.

ABSTRACT

The marine picocyanobacteria *Synechococcus* and *Prochlorococcus* are the two most abundant photosynthetic organisms on the planet and are present in all oceans, except the Antarctic Ocean, an ecological success likely related to their wide genetic diversity. During this PhD thesis, I explored the links between genetic diversity and niche adaptation in these two genera. First, the analysis of the distribution of picocyanobacterial populations at the global scale using a high-resolution taxonomic marker, allowed me to define ecologically significant taxonomic units for all major taxa, to improve the delineation of their ecological niches and to identify the main abiotic factors influencing their *in situ* distribution. The second part of this work aimed at identifying the genetic bases of adaptation of marine picocyanobacteria to distinct niches. From an evolutionary point of view, the comparative analysis of 81 genomes of these organisms revealed the combined role of gene gains and losses and of substitutions in protein sequences in the diversification of both genera. Analysis of the distribution of all known picocyanobacterial genes in the global ocean also allowed to show that each community, adapted to specific environmental conditions, possesses a distinct gene repertoire. Finally, the last part of this work has consisted in the physiological characterization of one model *Synechococcus* strain and four other strains representative of the predominating ecotypes in the field, which were submitted to light and thermal stresses in order to better understand the molecular bases of niche adaptation and the ecotypic variability of the stress response in *Synechococcus*, notably through an extensive transcriptomic study. Altogether, results obtained during this PhD have allowed not only to improve our knowledge of the ecological niches occupied by marine picocyanobacteria and of the taxonomical level to consider to differentiate ecotypes within both genera, but also to better understand the mechanisms that allowed them to adapt to these various niches and to colonize almost all the oceans.

Remerciements

The grass was greener

The light was brighter

With friends surrounded

Pink Floyd – High Hopes

En premier lieu, je souhaite remercier les membres de mon jury d'avoir accepté de lire et d'évaluer mon travail. Merci donc aux deux rapporteurs de cette thèse, Debbie Lindell et Jose-Manuel Garcia Fernandez, ainsi qu'aux quatre examinateurs, Vincent Daubin, Ingrid Obernosterer, Loïs Maignien et Eric Thiébaut pour le temps qu'ils consacreront à la lecture de ce manuscrit.

Je remercie également ceux qui ont effectué le suivi de ce travail pendant les trois dernières années, les membres de mon comité de thèse : Daniele Iudicone, Samuel Chaffron et Frédérique Viard. Merci pour votre curiosité et votre intérêt pour mes recherches ainsi que pour vos remarques judicieuses.

Le travail présenté dans les pages qui vont suivre n'aurait pas pu voir le jour sans le financement qui m'a été accordé pour effectuer ma thèse, financement de l'UPMC obtenu par le biais d'un contrat doctoral spécifique de l'ENS Lyon. Je remercie donc ces deux entités, en particulier l'ENS Lyon puisque je ne serai pas en train d'écrire ces mots sans la formation qui m'y a été délivrée, et parce que les années d'école ont été particulièrement épanouissantes.

Cette thèse n'aurait pas non plus vu le jour sans la curiosité de ceux qui l'ont imaginée et préparée par des années de travail acharné sur leurs bestioles préférées. Merci à Laurence de m'avoir proposé de continuer le travail après un bref passage en Master. Merci également d'avoir su écouter mes idées (parfois loufoques) et les prendre en compte, et d'avoir insisté pour imposer les tiennes quand ça en valait la peine (d'ailleurs oui, je ferai ces workflows… un jour !). J'ai encore beaucoup à apprendre (heureusement), mais j'aurai bien progressé grâce à toi ! Je ne peux pas remercier Laurence sans remercier Fred, puisque vous passez tellement de temps ensemble que tous vos travaux sont mêlés. Merci Fred pour tes bons mots et ton sens de la perfection (mais pas pour les « il suffit juste de » !). Vous faites une sacrée paire !

Pour faire fructifier les bonnes idées, il faut un bon cadre de vie et un bon cadre de travail. Si j'ai accepté de m'exiler au fin fond du Finistère Nord, c'est parce que la Station Biologique offre ce cadre. Merci donc à ceux qui la font vivre et lui donne de l'énergie. Merci également aux services communs qui nous facilitent le travail, notamment à Nathalie pour sa gestion. Au niveau du labo, un énorme merci à Céline qui nous sauve de toutes les situations administratives périlleuses en un rien de temps, et qui en plus nous fait rire.

Pendant les deux premières années de ma thèse j'ai eu la chance de pouvoir profiter d'un contrat de monitorat, et je remercie donc tous les enseignants qui ont partagé leurs heures de cours, ainsi que Stéphane pour m'avoir aiguillé vers eux.

Le travail de recherche est avant tout un travail d'équipe, et de nombreuses personnes ont participé à l'élaboration des résultats présentés ici. Un merci général à tous ces collaborateurs, qui sont intervenus de près ou de loin. Je remercie en particulier Greg pour tout ce qui touche à la bioinfo, ainsi que la plateforme ABiMS, en particulier Mark pour son investissement dans Cyanorak et pour sa malice, mais également les autres membres de la plateforme pour avoir installé les dizaines d'outils que j'ai voulu tester, et pour avoir fermé les yeux quand je trustais tous les cœurs disponibles. J'en profite pour remercier Ulysse pour sa participation tout aussi bioinformatique. Une pensée aussi pour Damien, qui a des idées folles et qui sait les suivre. Côté manip, merci à Théo d'être venu faire son stage avec moi ainsi qu'à Garance pour avoir tous les deux commencé à tester nos nouvelles hypothèses. Enfin merci à toute la troupe des stress SAMOSA (‘on a oublié le shift back !’), Momo, An, Juju, Théophile, Solène, Christophe, Fred, Laurence, pour les moments intenses de ces grosses manips dont on ne garde que les bons souvenirs.

La thèse, comme chacun sait, n'est pas que professionnelle, c'est une tranche de vie. Côté « associatif », merci à l'ex-bureau de l'AJC avec qui j'ai pu participer à mettre un peu de vie dans la Station, ainsi qu'à l'équipe JJC 2016 (Ambre et Laure) pour l'organisation sans faille (mais pas sans effort) de cette journée. Merci aussi à ceux qui ont partagé mon bureau (Laure, Greg, Ulysse, Adriana), à quatre dans un si petit espace, ça forge des liens !

Côté détente, il y a des routines dont on ne veut pas se défaire, et parmi elles le mercredi au Ty Pierre. Merci à toute son équipe pour son accueil, en particulier Goul et Marie. Le groupe des habitués a beaucoup évolué au cours du temps, mais il y en a qui sont restés fidèles au poste depuis le premier jour : Marine et Alexis. Merci à tous les deux pour le Ty Pierre et les soirées chez vous, en bref pour nous changer du boulot. Une pensée aussi pour tous ceux qui ont fréquenté ces « mercredi » (je suis sûr d'en oublier), Théophile, Juju, Tristan, Doriane, David, Valérian pour les premiers, puis Ambre, Laure, Eloïse, Camille, Ulysse, Martin et encore plus récemment Axelle, Charles et Laura. Je pense aussi à tous ceux qui sont passés plus ou moins longtemps à ces rendez-vous, notamment les stagiaires de la première année et les membres de la coloc de choc (Marine, David, Garf et Thibault) pour tous les bons moments qu'on a pu passer.

Merci à tous ceux que j'ai vus tous les jours et qui nous font oublier qu'on est au travail : les membres du groupe « phytopk ». Merci pour les pauses cafés mémorables et les tonnes de gâteaux (surtout depuis qu'Emilie est arrivée), à Martin, Emilie, Sarah, Estelle, Christian, Laure, Ulysse, Margot, Solène, Adriana (là je ne peux vraiment pas faire toute la liste !)… et aussi aux membres passés du groupe : Valérian, Roseline, David, Delphine D., Delphine S., Klervi, Tristan, Juju, vous nous manquez ! Merci Domi, Flo, Fabienne, Pris pour votre joie de vivre et pour l'accueil dans votre bureau quand je « trainais » dans les couloirs. Enfin un simple merci ne suffirait pas pour Momo, une mention très spéciale pour toi ! Il suffit que tu sois là pour être sûr que tout ira bien.

Puisque je parlais de pause : merci aux rois du baby, Juju, Tristan et Théophile pour nos parties mémorables. Le jonglage a fini par remplacer le baby, et merci Margot pour ces nouvelles pauses.

Comment ne pas évoquer les anciens roscovites de la Team Rosko International, une valeur sûre, et ces soirées à jouer dans une odeur de tourbe. Doriane, Tristan, Juju, Théophile, merci à tous les quatre ! Théophile on se suit depuis un moment (8 ans !), merci pour ton regard critique et tes arguments irréfutables (« Non. »), et même presque pour ton indécision chronique.

Quelle belle transition pour parler de ceux qui étaient là avant l'ère roscovite et qui ont su rester en contact : une pensée émue pour les anciens, Marc et Lise, Thibault et Mathilde, Théophile, Michel, Nina, Félix, Camille. Merci pour les voyages, et merci d'être là quelque part.

Enfin il y a ceux qui 'm'ont connu et soutenu avant tous les autres', merci à toute ma famille, savoir que vous êtes là me suffit pour vivre tranquille, même en rédigeant ma thèse (c'est dire !).

Et bien sûr, la vie est plus belle à deux. Merci à Laura, avec qui j'ai partagé ces trois dernières années et qui m'a supporté et chouchouté jusqu'aux derniers instants de rédaction. Tu es ma motivation, merci d'être là, et merci pour ce que je lis dans tes yeux.

MERCI A TOUS

TABLE DES MATIERES

LISTE DES FIGURES

INTRODUCTION

« Prendre des p'tits bouts d'trucs et puis
les assembler ensemble »
Stupéflip – La Menuiserie

1. Importance écologique du phytoplancton marin et des picocyanobactéries marines

1.1 Caractéristiques générales du phytoplancton marin

1.1.1 Plancton : quelques définitions

Qu'est-ce que le plancton ? Un enfant de 6 ans répondrait « c'est ce que mangent les baleines ». Cette réponse n'est pas fondamentalement fausse, mais un peu réductive : le terme plancton, du grec πλαγκτός « errant, instable », désigne l'ensemble des organismes pélagiques[1] se déplaçant passivement du fait de leur incapacité à lutter contre le courant (Brierley, 2017). Le plancton regroupe une grande diversité d'organismes de taille, de forme et de mode de vie variables, des virus de quelques dizaines de nanomètres aux méduses, en passant par les bactéries, les microalgues, les petits invertébrés, les larves de poissons…

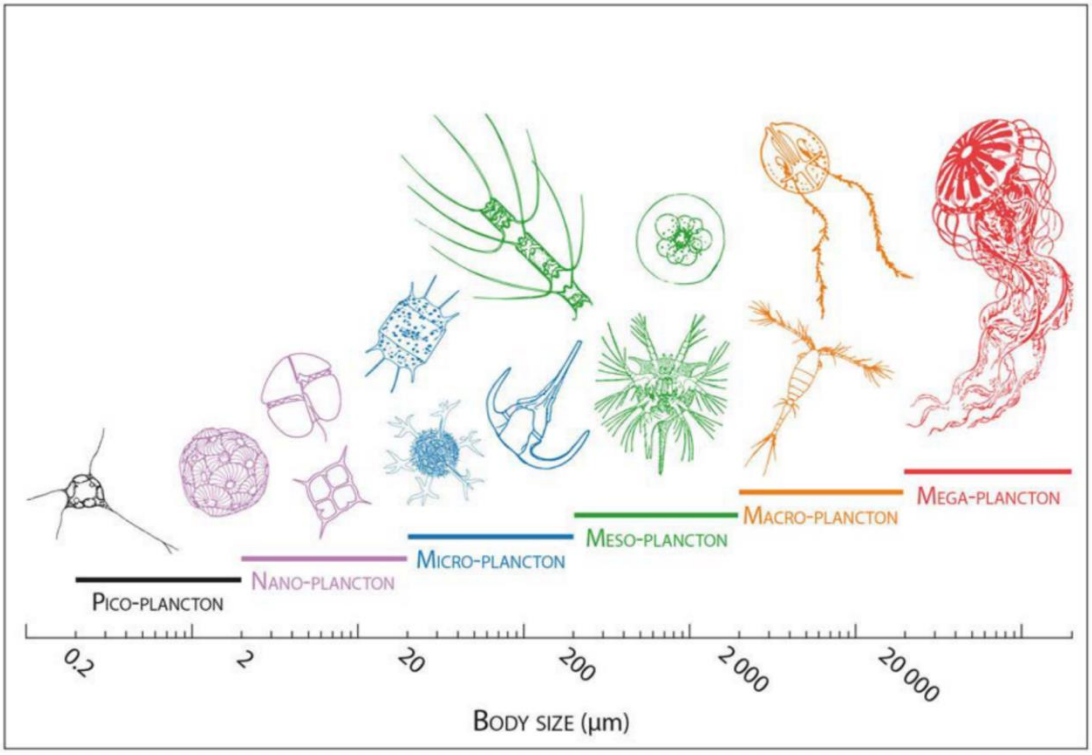


Figure 1 Classification du plancton en différentes classes de taille. Figure issue de (Biard, 2015).

Le plancton est subdivisé en catégories sur la base de différents critères, notamment un critère de taille (Figure 1). Les bactéries libres unicellulaires sont par exemple trouvées essentiellement dans la classe du picoplancton (0,2-2 µm), et pour les plus grosses dans le nanoplancton (2-20 µm). Le deuxième critère de distinction est le mode vie, qui a permis historiquement de distinguer le plancton « végétal » (phytoplancton, du grec φυτόν, plante) du plancton animal (zooplancton, du

[1] L'adjectif pélagique désigne les organismes vivant dans la colonne d'eau, par opposition aux organismes benthiques qui vivent sur ou dans le fond marin.

grec ζῷον, animal), auxquels s'ajoutent à présent le bactérioplancton qui regroupe les bactéries et le virioplancton qui désigne les virus planctoniques (Sieburth *et al.*, 1978).

Le terme de phytoplancton désigne donc tous les organismes planctoniques ayant un mode de vie végétal, c'est-à-dire similaire au mode de vie des plantes terrestres dans le sens où la photosynthèse oxygénique en est la base.

1.1.2 Photosynthèse oxygénique et besoins en nutriments

La photosynthèse consiste à utiliser l'énergie lumineuse pour réduire le carbone minéral et former de la matière organique. La photosynthèse oxygénique en est la forme la plus répandue, et se décompose en deux phases la phase claire dans laquelle l'énergie lumineuse est utilisée pour former des molécules dont l'énergie (chimique) et le pouvoir réducteur sont directement utilisables par la cellule et la phase sombre où ces molécules sont utilisées pour fixer le carbone minéral (CO_2) au sein de molécules organiques (Figure 2).

La première phase, phase claire, se déroule dans la membrane des thylakoïdes des chloroplastes (chez les Eucaryotes) ou dans les membranes thylakoïdales pour les cyanobactéries Figure 2A. Lors de cette première phase, l'énergie des photons est captée par les pigments de l'antenne collectrice en contact avec le photosystème II, et transmise jusqu'au centre réactionnel de ce photosystème jusqu'à une molécule de chlorophylle qu'elle excite. La chlorophylle sous forme excitée est fortement réductrice, et cède un électron à une autre protéine du photosystème. Après cette oxydation, elle récupère un électron fourni par son apoprotéine et issu de la dissociation de l'eau par le complexe d'évolution de l'oxygène, également situé au sein du photosystème II. Cette première étape constitue la photodissociation de l'eau, son bilan étant l'oxydation de l'eau en dioxygène sous l'effet de la lumière. L'électron cédé par la chlorophylle suit ensuite une chaine de transport au sein de la membrane, représentée sur la Figure 2A. Les différentes étapes le long de cette chaine entrainent la formation d'un gradient de protons entre les deux côtés de la membrane, gradient dont l'énergie est dissipée au niveau de l'ATP synthase pour former des molécules d'ATP, sorte de « monnaie énergétique » de la cellule, utilisée dans de très nombreux processus cellulaires.

Le potentiel rédox des électrons est diminué au niveau du photosystème I en utilisant une nouvelle fois l'énergie lumineuse, et ils finissent leur course en étant utilisés pour réduire le $NADP^+$ en $NADPH,H^+$, dont le pouvoir réducteur sera aussi utilisé lors de nombreux processus cellulaires.

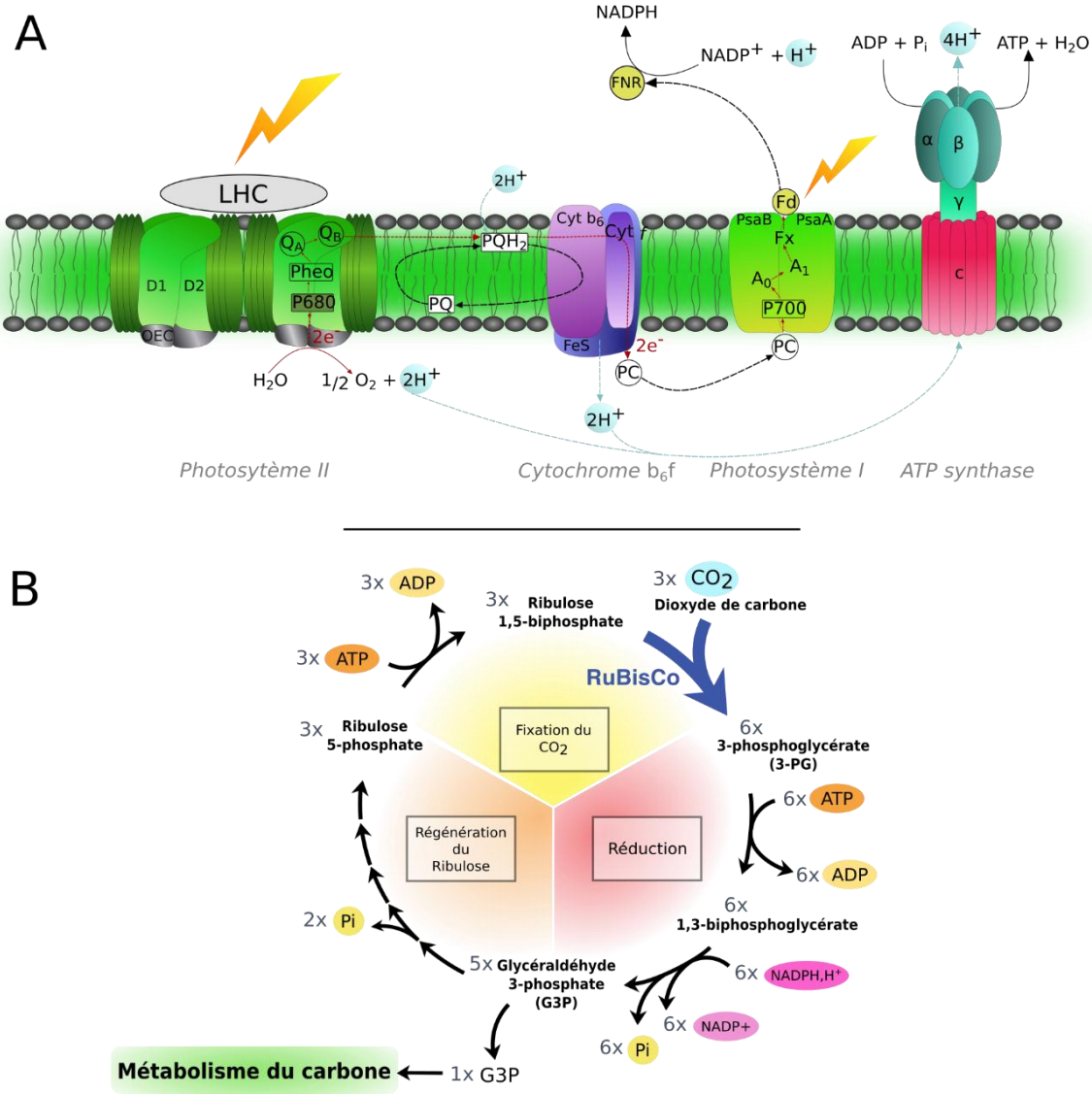


Figure 2 Fonctionnement général de la photosynthèse. A. Phase claire de la photosynthèse, durant laquelle l'énergie lumineuse est utilisée pour former de l'ATP et des pouvoirs réducteurs ($NADPH,H^+$) utilisés dans la phase sombre. Schéma réalisé par Justine Pittera (Pittera, 2015). B. Phase sombre, au cours de laquelle est fixé le CO_2 au sein de molécules à 3 carbones qui entrent dans le métabolisme cellulaire du carbone (cycle de Calvin-Benson). LHC : Light Harvesting Complex. OEC : Oxygen Evolving Complex. PQ : Plastoquinone. PC : Plastocyanine.

Dans la deuxième phase, dite phase sombre, ATP et $NADPH,H^+$ sont utilisés pour fixer le carbone minéral grâce à une enzyme essentielle, la RubisCO (Ribulose-1,5-bisphosphate carboxylase/oxygenase), au cours du cycle de Calvin-Benson (Figure 2B). Les molécules à 3 carbones sortant de ce cycle sont ensuite intégrées au métabolisme cellulaire du carbone, qui entre dans la composition de toutes les biomolécules (ADN, ARN, protéines, lipides, etc…).

La caractéristique principale du phytoplancton est donc l'utilisation de la photosynthèse oxygénique, et cela a une implication importante pour son écologie. En effet, puisqu'il a besoin de lumière pour croître, il ne peut être viable que dans des zones où l'intensité lumineuse est suffisamment intense. On considère en général que cela correspond à la zone située entre la surface et la profondeur où parvient 1% de l'intensité lumineuse de surface (zone euphotique). La

profondeur de cette limite dépend de la transparence de l'eau, qui est influencée par différents facteurs, notamment le degré de stratification verticale de la colonne d'eau, la concentration en particules en suspension et la productivité biologique. Un certain nombre d'organismes phytoplanctoniques sont cependant mixotrophes par rapport au carbone (c.-à-d. capables d'utiliser des sources de carbone organique), ce qui leur permet de subvenir à leurs besoins en carbone lorsque la photosynthèse n'y suffit plus (un exemple extrême étant la survie pendant la nuit polaire) (Stoecker *et al.*, 2017).

En dehors du carbone, les organismes phytoplanctoniques sont dépendants pour leurs besoins de nutriments, qu'ils doivent trouver dans leur environnement sous forme minérale ou organique. Leur disponibilité varie à différentes échelles spatio-temporelles ; à titre d'exemple, elle explique en partie l'occurrence de *blooms* printaniers de phytoplancton dans les régions tempérées, où le mélange hivernal lié au vent et à la houle enrichit l'eau en nutriments (Figure 3).

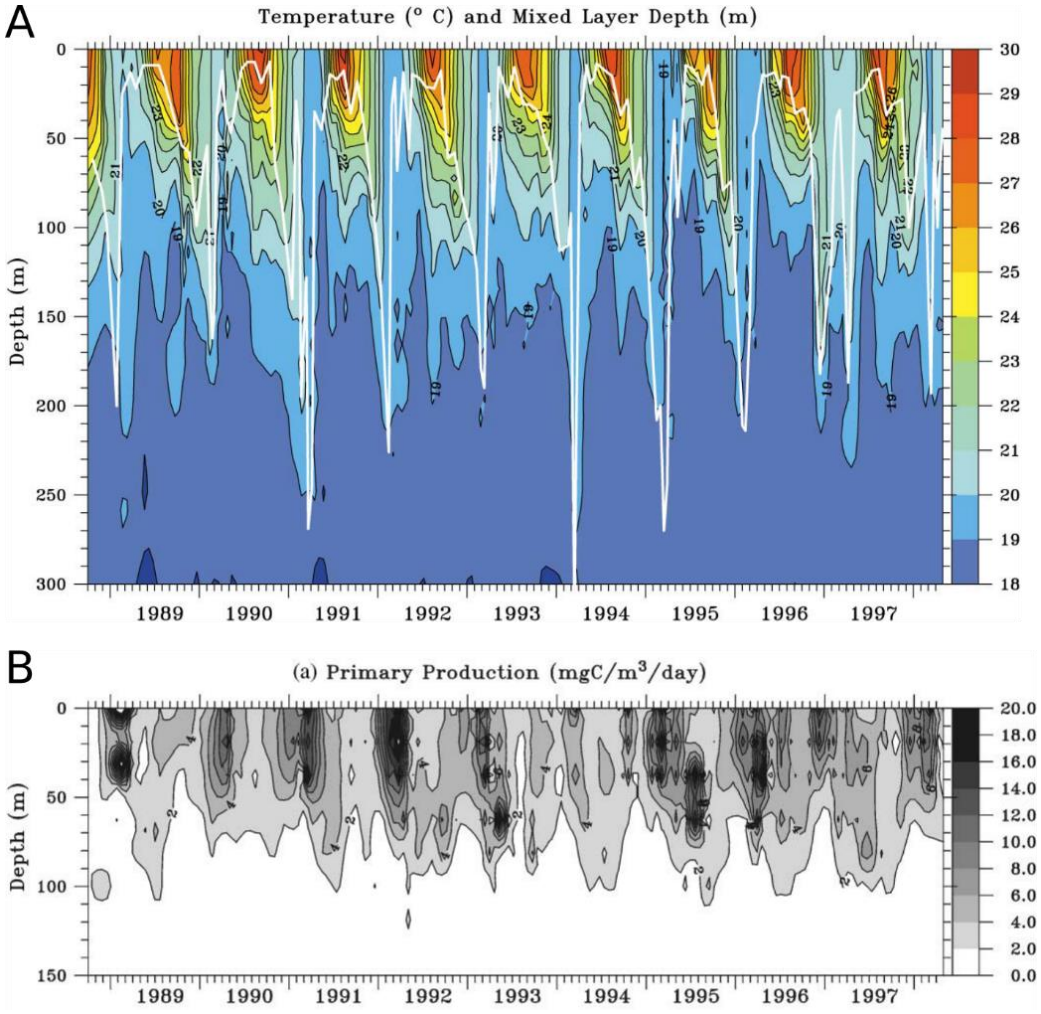


Figure 3 Variations temporelles de de la température et de la production primaire à la station BATS (Atlantique Nord). A. Variations temporelles de la température de l'eau (échelle de couleur) et de la profondeur de la couche de mélange (ligne blanche). Le mélange hivernal est bien visible. B. Variation de la production primaire (niveaux de gris) sur la même période temporelle. Les *blooms* phytoplanctoniques sont visibles par la forte production primaire (couleur noire). Source : (Steinberg *et al.*, 2001).

Au niveau spatial, les variations de la disponibilité en nutriments influent fortement sur l'abondance du phytoplancton, comme le révèle par exemple la concentration en chlorophylle *a* (Chl *a*) estimée à partir de données satellitaires (Figure 4). Ainsi, certaines régions, notamment les gyres des océans Pacifique et Atlantique sont particulièrement oligotrophes (c.-à.d présentant une concentration en Chl *a* en surfaceparticulièrement faible). Selon les régions, différents éléments nutritifs deviennent limitants pour la croissance du phytoplancton, mais il s'agit en général de l'azote, du phosphate ou de métaux traces comme le fer (Figure 4C).

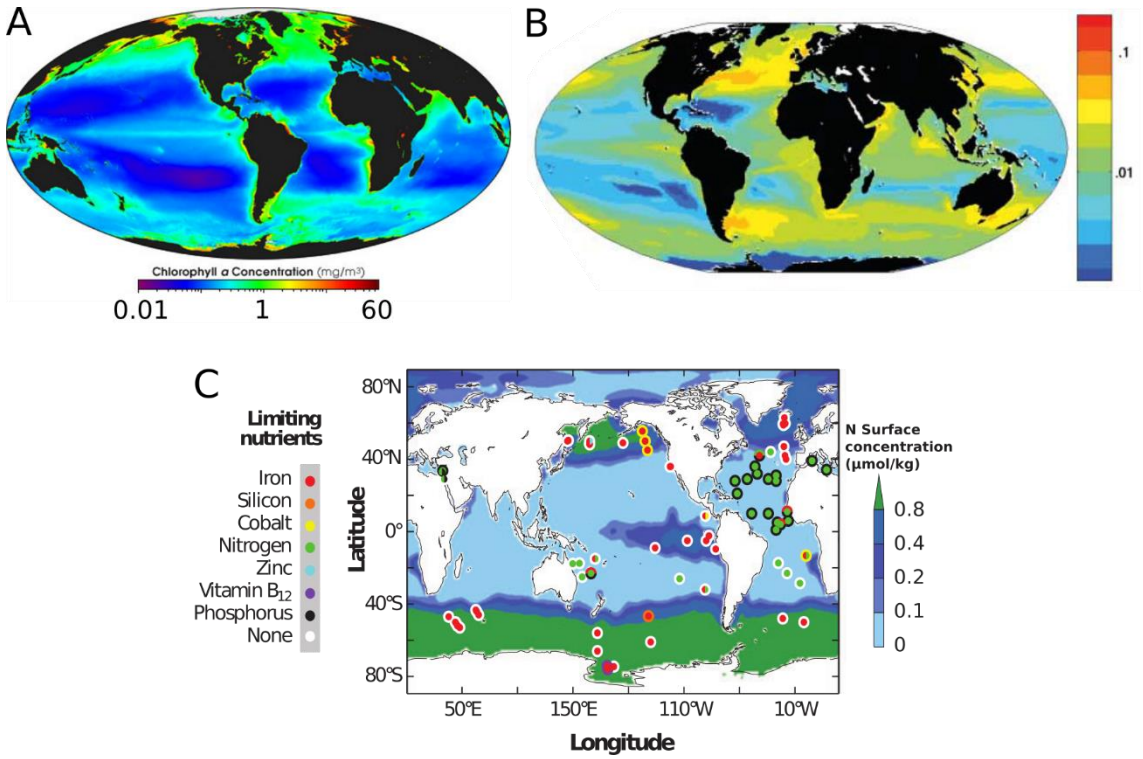


Figure 4 Facteurs limitant la croissance du phytoplankton dans les oceans. Limitation of phytoplankton growth in the ocean. A. Concentration en chlorophylle a dans les océans (mg/m^3), montrant les vastes gyres oligostrophes des oceans Atlantique et Pacifique. Source : SeaWiFS/NASA (earthobservatory.nasa.gov). B. Biomasse phytoplanctonique estimée à partir d'un modèle numérique et moyennée sur les 50 premiers mètres. L'échelle numérique est en mM Phosphate. Modifié à partir de (Follows *et al.*, 2007). C. Limitation du phytoplancton par la concentration en nutriments. Le fond de carte correspond à la concentration en azote en surface. La couleur de chaque disque indique quel élément est limitant, et plusieurs couleurs sont indiquées en cas de co-limitation. La couleur du contour des disques correspond à la seconde limitation, quand elle a été mesurée. Figure modifiée à partir de (Palenik, 2015; Moore *et al.*, 2013).

1.1.3 Diversité du phytoplancton marin

Derrière le terme de phytoplancton que nous avons utilisé jusqu'ici se cache une immense diversité d'organismes de tailles, de formes et de modes de vie différents. Les Eubactéries ne sont représentées que par les cyanobactéries, dont certaines sont capables de fixer l'azote atmosphérique (N_2), souvent en formant des filaments montrant un certain degré de spécialisation cellulaire, alors que d'autres sont unicellulaires. De nombreuses lignées phytoplanctoniques ont en revanche évolué

chez les Eucaryotes, certaines par l'endosymbiose secondaire d'une microalgue elle-même phytoplanctonique. Pour ne donner que quelques exemples emblématiques de cette diversité, les Straménopiles comprennent des membres se protégeant d'un frustule de silice aux ornements complexes (les diatomées), quand d'autre en sont dépourvus (Figure 5A,B); au sein des Haptophytes, les Coccolithophoridés ont quant à eux un test (ou coquille) calcaire aux formes variables (Figure 5C,E), alors que les cellules d'autres groupes sont nues (Worden *et al.*, 2015). Au-delà de cette diversité morphologique évidente, lorsqu'on dispose d'un microscope, existe une grande diversité génétique, qui commence seulement à être appréhendée (de Vargas *et al.*, 2015).

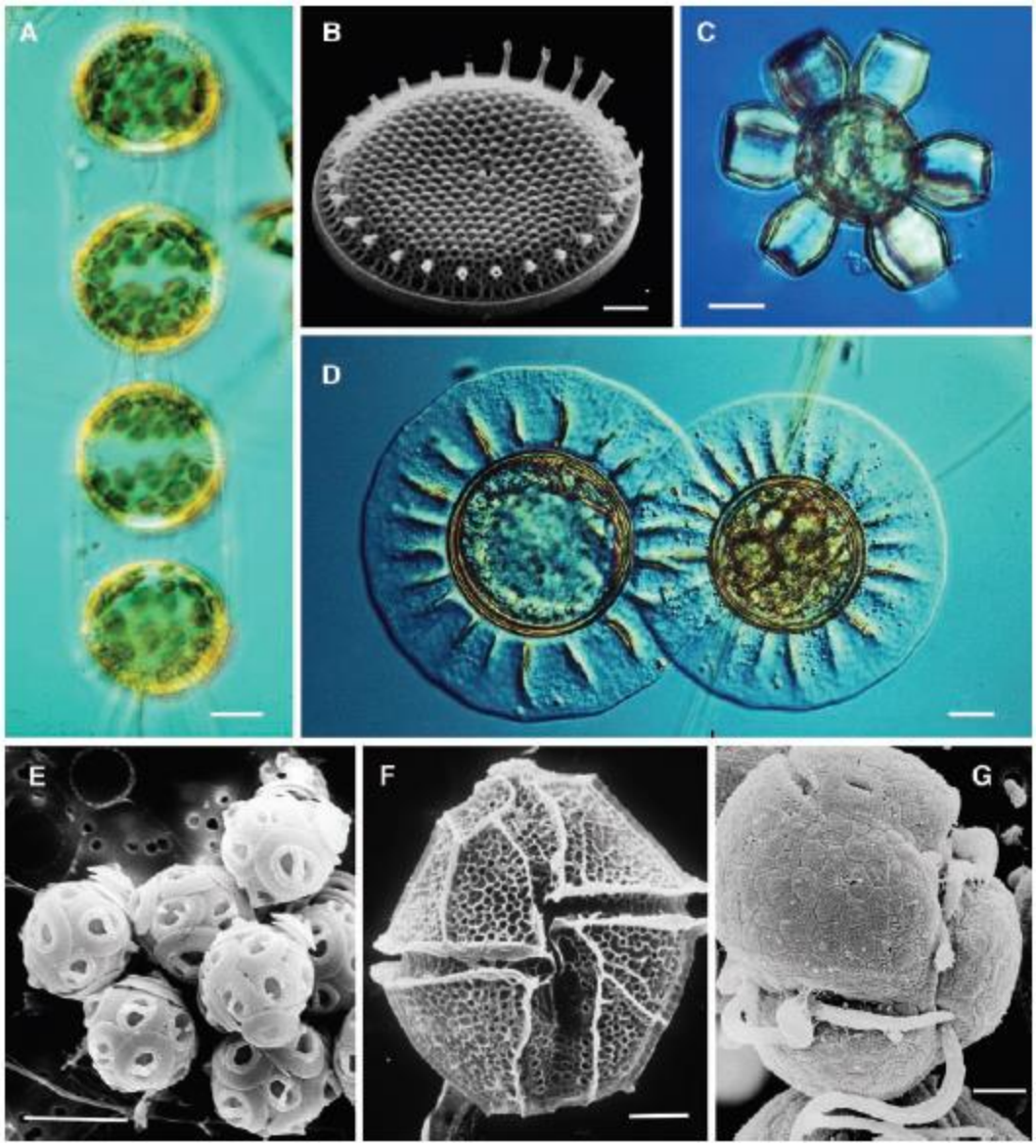


Figure 5 Echantillon de la diversité des organismes phytoplanctoniques. A. La diatomée Stephanopyxis nipponica. B. Valve d'un frutule de la diatomée *Thalassiosira pacifica*. C. Le Coccolithophoridé *Scyphospahaera apsteinii*. D. Pterosperma moebii (Prasinophycées). E. Un amas de coccosphères du Coccolithophoridé *Gephyrocapsa oceanica*. F. Le dinoflagellé *Lingulodinium polyedra*. G. Le dinoflagellé *Gyrodinium galatheanum*. Echelles : (A, C, E, F) 10 µm; (B et G) 2 µm; (D) 25 µm. [Photographes: (A, C, D, F) F. J. R. Taylor; (B) E. Simons; (E) G.Hallegraeff; (G) G. Gaines]. Figure issue de (Falkowski, 2004).

Un seul échantillon d'eau de mer peut contenir une grande diversité d'organismes planctoniques, diversité qui a été source d'émerveillement, mais également de questionnement scientifique. Elle a par exemple amené Hutchinson à énoncer ce qu'il appelle le « paradoxe du plancton » : du fait de la relative homogénéité de l'échantillon, cette diversité est en opposition avec le « principe d'exclusion compétitive »[2]. Hutchinson a avancé plusieurs hypothèses pour expliquer ce paradoxe, notamment la variabilité temporelle des conditions environnementales, la taille des populations de micro-organismes planctoniques et les pressions de prédation qui s'exercent sur elles, hypothèses qui sont encore explorées aujourd'hui (Hutchinson, 1961).

1.2 Le phytoplancton, acteur essentiel des grands cycles biogéochimiques

1.2.1 Le phytoplancton au cours des temps géologiques

Les variations de la biochimie des enveloppes fluides de la Terre au cours des temps géologiques, et en particulier avant le Phanérozoïque, sont assez mal connues, mais certains éléments font consensus. L'océan et l'atmosphère archéens (avant 2,5 milliards d'années (Ga)) étaient fortement réducteurs et jusqu'au début du Protérozoïque (il y a 2,5 Ga), le dioxygène en était quasi-absent (Lenton and Daines, 2017). La photosynthèse oxygénique, apparue au cours de l'Archéen chez les ancêtres des cyanobactéries actuelles, s'est renforcée au début du Protérozoïque pour des raisons encore incertaines mais potentiellement liées à l'acquisition de la multicellularité par les cyanobactéries (Schirrmeister *et al.*, 2015). Ce renforcement a induit une oxydation d'abord rapide (*Great Oxydation Event*, entre 2,3 et 1,8 Ga), puis plus progressive de l'atmosphère et des océans, qui n'ont probablement été totalement oxydés qu'à la fin du Protérozoïque il y a 500 millions d'années (Ma) (Lenton and Daines, 2017).

Ce n'est qu'après la « grande oxydation » que sont apparues les formes phytoplanctoniques, puisque les cyanobactéries planctoniques sont estimées être apparues dans la deuxième moitié du Protérozoïque (Sánchez-Baracaldo, 2015), et c'est probablement à la même époque que sont apparus les premiers eucaryotes photosynthétiques par endosymbiose. Les cyanobactéries et les microalgues de la lignée verte semblent avoir dominé les communautés phytoplanctoniques jusqu'au milieu du Mésozoïque, où ont eu lieu des changements majeurs dont l'origine est encore incomprise. A cette époque se sont développés les organismes planctoniques calcifiants (Coccolithophoridés et Foraminifères) et les Dinoflagellés, suivis plus récemment par les Diatomées (apparues vers la fin du Crétacé) (Figure 6), groupes qui, avec les cyanobactéries, dominent actuellement les océans (Falkowski, 2004; Hull, 2017). Ces chagements ont eu des effets majeurs sur la chimie des océans et sur le climat, notamment du fait de la séquestration de CO_2 sous forme de carbonate de calcium, qui a rendu les océans beaucoup plus résilients face aux

[2] Cette théorie postule que si deux organismes sont en compétition pour l'occupation d'une même niche, soit l'un deux est amené à disparaître, soit ils se différencient pour occuper des niches distinctes (Hardin, 1960).

variations de la concentration atmosphérique en CO_2 (Hull, 2017). L'ensemble de l'écosystème marin a également été impacté, le développement de ces groupes phytoplanctoniques ayant augmenté la productivité primaire des océans et donc favorisé la diversification des organismes qui en dépendent.

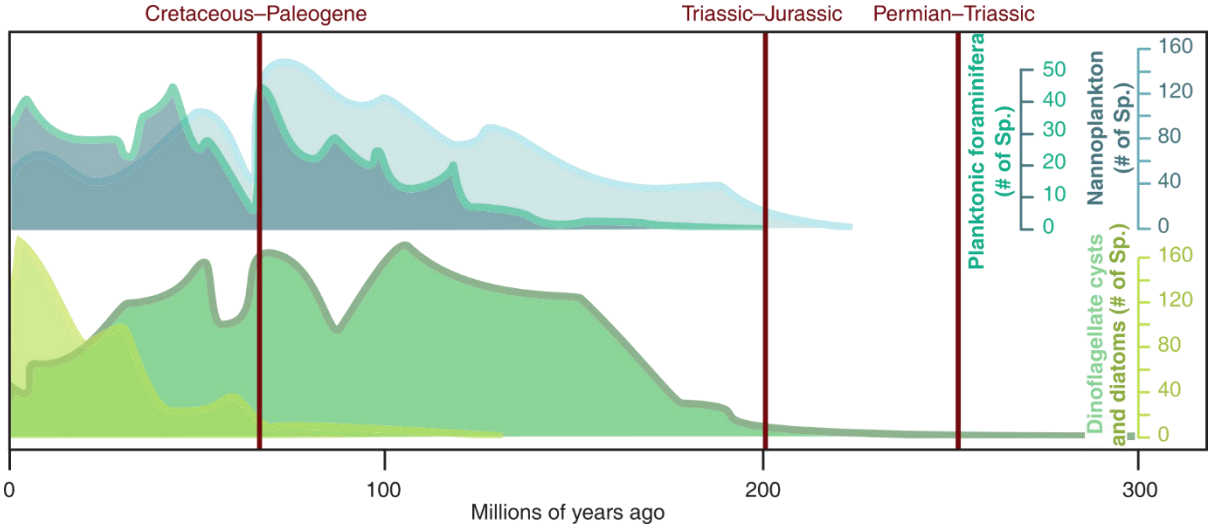


Figure 6 Evolution de la diversité de quelques groupes phytoplanctoniques majeurs au cours des 200 derniers millions d'années. Le nombre d'espèces de Foraminifères planctoniques (en bleu foncé), de nanoplancton calcifiant (dont les Coccolithophoridés, bleu clair), de Dinoflagellés (en vert foncé) et de Diatomées (en vert clair), évalué à partir des enregistrements fossiles. Les principales extinctions de masse sont indiquées par des traits rouges. Figure issue de (Hull, 2017).

1.2.2 L'influence du phytoplancton sur les cycles biogéochimiques

Par son activité photosynthétique, le phytoplancton est responsable de la production primaire de l'ensemble de l'écosystème pélagique : il produit la matière organique consommée par les autres compartiments de l'écosystème. Du fait de la grande surface océanique de notre planète, et de l'abondance du phytoplancton, les organismes phytoplanctoniques fixent ainsi collectivement $48{,}5.10^{12}$ kg de CO_2 par an, soit 46% de la production primaire nette globale (Field, 1998). Cette matière organique est ensuite transmise aux niveaux trophiques supérieurs lorsque le phytoplancton est consommé par le zooplancton, ou par des protistes hétérotrophes ou mixotrophes qui seront eux-mêmes potentiellement consommés par le zooplancton, qui à son tour nourrira des Métazoaires de plus grosse taille (Figure 7). A chaque maillon de cette chaine, une partie de la matière organique est utilisée pour la respiration qui libère le CO_2 fixé par la photosynthèse. Une grande partie du carbone fixé ne suit cependant pas cette chaîne : les organismes phytoplanctoniques en libèrent une partie dans l'eau au cours de leur vie et lorsqu'ils meurent, souvent du fait de parasites ou de virus. Cette matière organique (sous forme dissoute ou particulaire) est utilisée comme source d'énergie par les bactéries hétérotrophes ou mixotrophes, qui la « recyclent » en libérant des formes minérales (CO_2, azote et phosphate inorganiques), qui seront à nouveau disponibles pour le phytoplancton. Cette étape est désignée sous le terme de « boucle microbienne » (Azam *et al.*, 1983). Enfin une partie de la matière organique, qualifiée de récalcitrante, n'est pas recyclée sur place, coule sous forme d'agrégats et quitte la zone euphotique. Une grande partie sera dégradée au cours de sa chute

et nourrira les écosystèmes profonds, mais cette matière (dont le carbone) reste piégée pour une période très longue (plusieurs milliers d'années) dans les océans. C'est en ce sens que l'écosystème planctonique constitue une « pompe biologique » de carbone, une partie du carbone fixé par le phytoplancton étant stocké dans les océans. Si ce carbone atteint le fond de l'océan, il peut ensuite entrer dans des processus de diagénèse pour former des roches sédimentaires, auquel cas il sera stocké à très long terme (Figure 7). Ce modèle général, et en particulier la place que prend chaque compartiment, varie selon les conditions environnementales, en fonction du lieu considéré et de la saison.

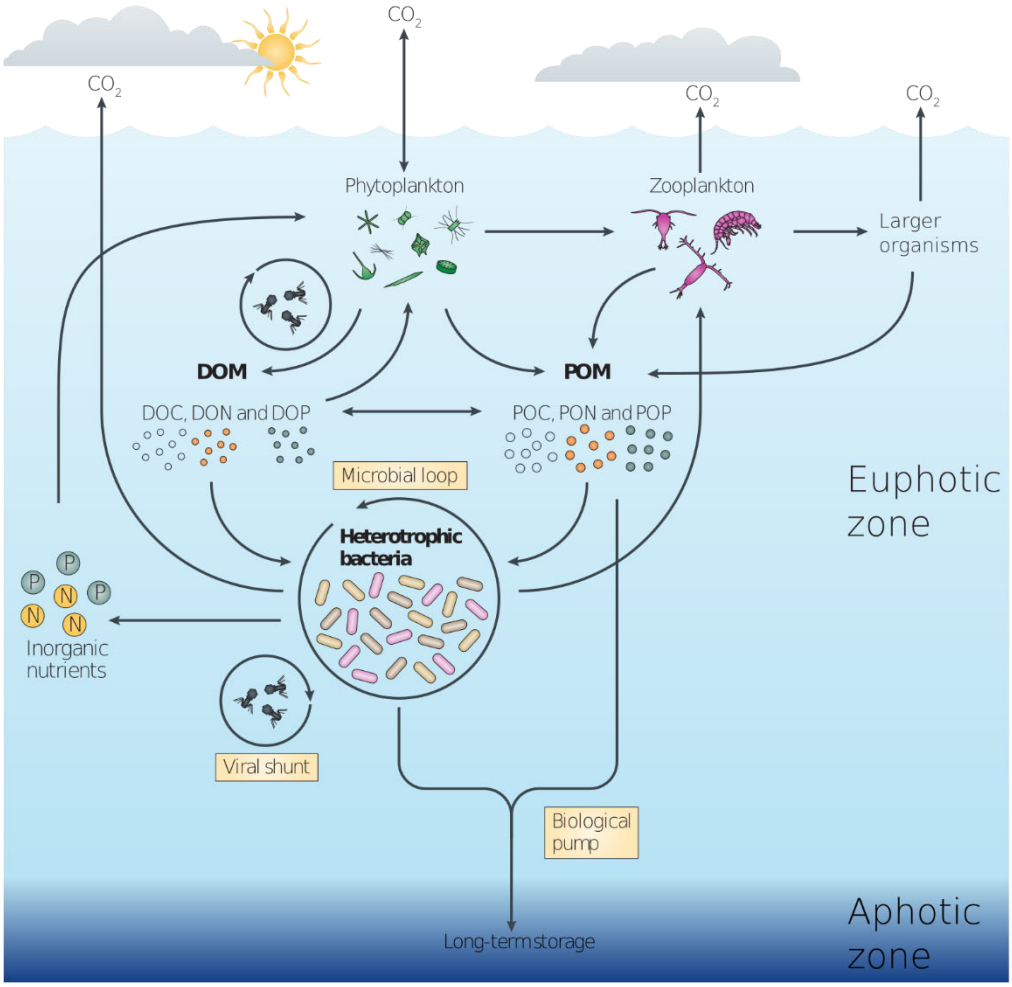


Figure 7 Cycle de la matière organique au sein de l'écosystème planctonique. DOM : Matière organique dissoute (DOC, DON, DOP : Carbone, Azote et Phosphate organiques dissouts). POM : Matière organique particulaire (POC, PON, POP : Carbone, Azote et Phosphate organiques particulaires). Figure modifiée à partir de (Buchan *et al.*, 2014).

Cette description montre à quel point les cycles élémentaires au sein de l'écosystème planctonique reposent sur les interactions biotiques entre ses membres, que ce soit sous forme de symbiose, de prédation, de parasitisme ou de commensalisme[3]. Au sein de cet écosystème, le

[3] Le commensalisme est un type d'interaction biologique dans lequel l'organisme commensal bénificie de l'interaction sans pour autant que cela représente un coût pour son hôte.

phytoplancton est un acteur majeur du cycle du carbone, mais aussi du cycle de l'azote (certaines cyanobactéries étant capables de fixer le N_2 atmosphérique) et du soufre (certains organismes du phytoplancton produisant des espèces soufrées qui diffusent dans l'atmosphère). Il relie également ces cycles à ceux d'un certain nombre d'éléments moins abondants, notamment de métaux comme le fer qui est essentiel à la photosynthèse (Tagliabue *et al.*, 2017).

1.3 Les picocyanobactéries marines et leur place dans l'écosystème phytoplanctonique

1.3.1 La découverte des picocyanobactéries marines

Les protistes du nanoplancton et des classes de taille supérieures sont reconnus dans leur diversité depuis la fin du XIXe siècle (leurs formes ont été rendues célèbres dessins d'Ernst Haeckel (Haeckel, 1904)), mais les connaissances sur les procaryotes planctoniques n'ont pu se développer qu'avec les progrès techniques de la deuxième moitié du XXe siècle. Alors qu'en milieu océanique, les seules cyanobactéries connues appartenaient au genre *Trichodesmium* dont les membres forment des colonies multicellulaires capables de fixer l'azote atmosphérique, l'utilisation de la microscopie à épifluorescence a permis de mettre en évidence l'existence de cyanobactéries planctoniques unicellulaires à la fois abondantes et ubiquistes, faciles à identifier grâce à leur auto-fluorescence jaune-orange liée à la présence de phycoérythrine, un pigment spécifique de l'appareil pigmentaire des cyanobactéries et des algues rouges (voir ci-dessous) (Waterbury *et al.*, 1979). Cette première description associait des données de comptage *in situ* obtenues en microscopie à épifluorescence à des électronographies permettant de classer les bactéries observées dans le genre *Synechococcus*, contenant déjà des souches isolées en eau douce, sur la base de critères morphologiques (Figure 8).

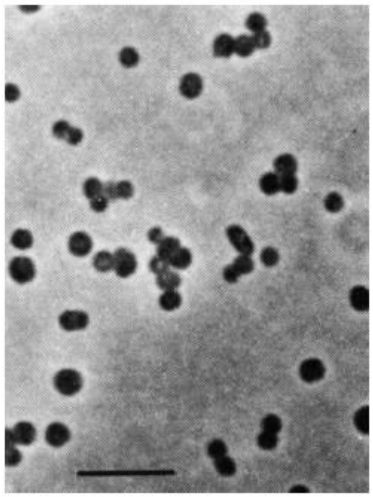


Figure 8 Figure accompagnant l'article de Waterbury caractérisant pour la première fois les *Synechococcus* marins. Figure issue de (Waterbury *et al.*, 1979).

Quelques années plus tard, l'application de la cytométrie en flux aux échantillons d'eau de mer prélevés en milieu océanique dans différentes régions a permis de mettre en évidence l'existence et l'abondance d'un autre genre de picocyanobactérie se distinguant de *Synechococcus* par sa taille plus petite et son absence de fluorescence orange indiquant l'absence de phycoérythrine (Chisholm *et al.*, 1988). Sa pigmentation particulière, en particulier la présence de divnyl-chlorophylle *a* et *b* (dv-chl *a* et dv-chl *b*) et l'absence de (monovinyl-) chlorophylle *a*, lui a valu le nom de *Prochlorococcus* (Chisholm *et al.*, 1992). D'abord décrit à partir d'échantillons prélevés en profondeur, où il abonde en particulier à la base de la zone euphotique, l'abondance en surface de cellules aux caractéristiques proches a ensuite été démontrée (Olson *et al.*, 1990). Les caractéristiques morphologiques révélées par microscopie se sont avérées correspondre à celles de cyanobactéries déjà observée 10 ans plutôt, mais qui n'avaient pas été caractérisées (Johnson and Sieburth, 1979).

1.3.2 Caractéristiques morphologiques et appareils pigmentaires de *Synechococcus* et *Prochlorococcus*

La comparaison des morphologies de *Synechococcus* et *Prochlorococcus* à partir de clichés de microscopie électronique (Figure 9) révèle un certain nombre de similitudes et de différences. Les cellules de *Synechococcus* sont rondes à ovales, avec une taille autour de 0,9 µm, tandis que *Prochlorococcus* présente deux formes de cellules : certaines ovales de taille assez similaire à *Synechococcus* et retrouvées essentiellement en profondeur, d'autres sphériques et plus petites (0,5 µm) plus abondantes en surface (Johnson and Sieburth, 1979; Partensky *et al.*, 1999a). Comme toutes les cyanobactéries, ce sont des bactéries à Gram-, présentant donc une membrane externe en plus de la membrane plasmique, les deux étant séparées par une couche de peptidoglycane. La surface externe et la structure de la paroi sont variables selon les souches, et l'épaisseur de cette paroi serait particulièrement réduite chez *Prochlorococcus* (Waterbury *et al.*, 1986; Ting *et al.*, 2007). Aucune des souches cultivées de picocyanobactéries marines ne possède de flagelle, mais certaines souches de *Synechococcus* présentent une paroi extracellulaire spécifique qui leur confère une motilité originale, qualifiée de « grouillante » (*swarming*) (Waterbury *et al.*, 1985; Brahamsha, 1999).

Aux deux membranes cellulaires s'ajoutent celles des thylakoïdes, formant un réseau de membrane intracellulaire, lieu des réactions de la phase claire de la photosynthèse. Le cytosol des cellules de *Synechococcus* et *Prochlorococcus* contient des carboxysomes où se trouvent concentrées la RuBisCO et l'anhydrase carbonique permettant l'assimilation du carbone, et peut contenir des granules de glycogène servant au stockage du carbone, selon les conditions de culture. Enfin certaines souches de *Synechococcus* produisent des inclusions cytoplasmiques de polyphosphate (servant au stockage du phosphate) en fonction des conditions environnementales (Waterbury *et al.*, 1986; Mazard *et al.*, 2012b).

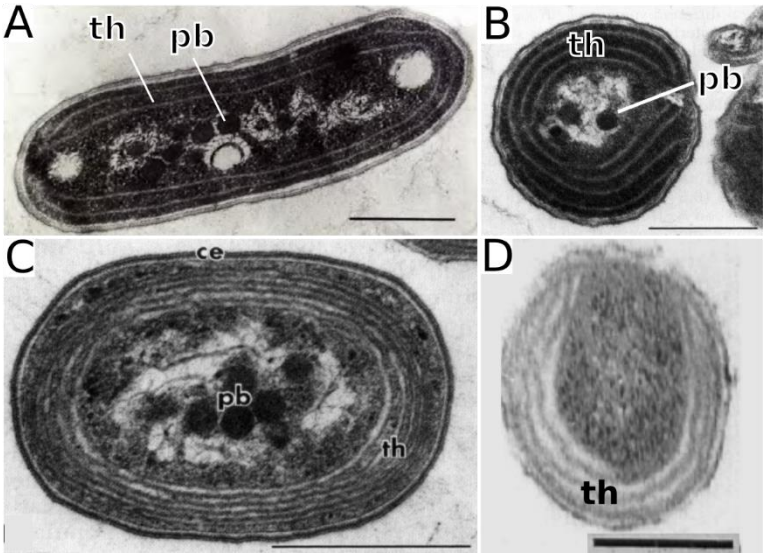


Figure 9 Electronographies de *Synechococcus* et *Prochlorococcus.* A. *Synechococcus* WH7804. Echelle : 0,5 µm. Source : (Waterbury *et al.*, 1986). B. *Synechococcus* non caractérisé issu d'un échantillon d'eau de mer. Echelle : 0,5 µm. Source : (Johnson and Sieburth, 1979). C. *Prochlorococcus* non caractérisé issu d'un échantillon d'eau de mer prélevé à 100m Echelle : 0,5 µm. Source : (Johnson and Sieburth, 1979). D. *Prochlorococcus* MED4. Echelle : 0,2 µm. Source : (Lichtlé *et al.*, 1995). th : thylakoïdes. pb : polyhedral bodies correspondant aux carboxysomes. ce : enveloppe cellulaire.

Outre la différence dans la taille des cellules, la différence ultrastructurale majeure entre *Synechococcus* et *Prochlorococcus* est la disposition des membranes thylakoïdales au sein du cytoplasme. Chez *Synechococcus*, ces membranes forment des cercles concentriques largement espacés les uns des autres (Figure 9A,B), alors que chez *Prochlorococcus,* elles sont accolées les unes aux autres et à la membrane plasmique (Figure 9C,D) (Johnson and Sieburth, 1979; Waterbury *et al.*, 1986; Chisholm *et al.*, 1988). Ces membranes sont parfois interrompues au niveau d'un pôle de la cellule. Chez *Prochlorococcus* différentes dispositions existent, certaines cellules (notamment celles isolées en profondeur) ayant plus de couches de membranes thylakoïdales qui soit couvrent le périmètre de la cellule, soit sont interrompues à un pôle, formant un « fer à cheval » (Lichtlé *et al.*, 1995). D'autres cellules (souvent isolées en surface) possèdent moins de membranes thylakoïdales et la configuration « fer à cheval » peut être poussée jusqu'à n'avoir qu'une moitié de la cellule contenant des thylakoïdes (Figure 9D) (Lichtlé *et al.*, 1995).

Cette différence entre les deux genres dans l'espacement des thylakoïdes est due à une composition et une structure radicalement différentes de leurs appareils pigmentaires. En effet, *Synechococcus* possède, comme la plupart des cyanobactéries, de gros complexes protéiques localisés entre les membranes thylakoïdales (ce qui explique leur espacement) et spécialisés dans la récolte d'énergie lumineuse, les phycobilisomes. Au contraire, *Prochlorococcus* ne possède pas de phycobilisomes mais des protéines collectrices de lumière intramembranaires (Pcb) liant des molécules de chlorophylle (Figure 10).

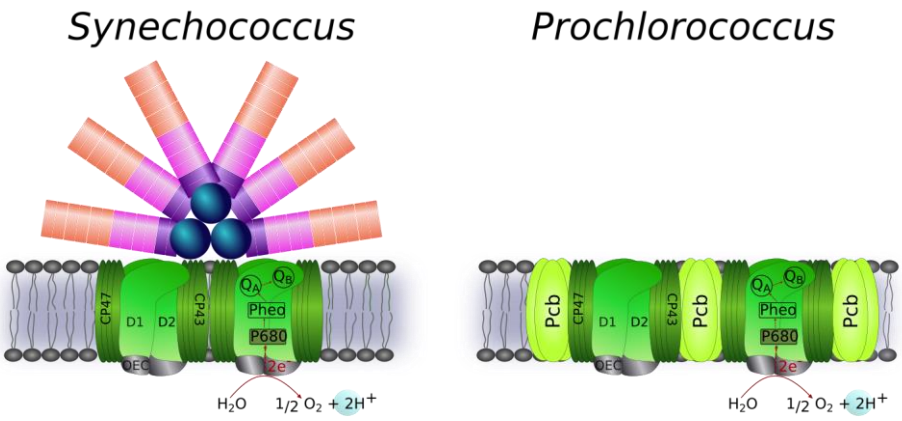


Figure 10 Comparaison des systèmes de collection de lumière de *Prochlorococcus* et *Synechococcus*. *Synechococcus* (à gauche) collecte la lumière grâce aux phycobilisomes, complexes multiprotéiques cytosoliques accolés à la membrane. Les antennes de *Prochlorococcus* (à droite) sont quant à elles constituées de la protéine Pcb, protéine intramembranaire liant les molécules de dv-chl *a* et de (dv-) chl *b*. Les systèmes de collection de lumière sont ici représentés au contact du photosystème II. Figure modifiée à partie de la version originale créée pour *Synechococcus* par Justine Pittera (Pittera, 2015).

Les phycobilisomes sont des complexes multiprotéiques constitués d'un cœur et de six bras rayonnant à partir du cœur. Ils sont composés de phycobiliprotéines (PBPs), protéines associées à des chromophores (les phycobilines), reliées entre elles par des protéines de liaison (linker). Les cultures de *Synechococcus* peuvent avoir des couleurs variant du vert à l'orange en passant par le rose selon les souches (Figure 11), et ces différentes pigmentations sont dues à la composition des phycobilisomes en PBPs. Il existe 4 types de PBPs chez *Synechococcus*, chacune liant des combinaisons de chromophores différentes ce qui leur confère des propriétés d'absorption différentes, et la composition en PBPs permet de définir 3 types pigmentaires principaux (Six *et al.*, 2007c). Le cœur du phycobilisome est toujours composé d'allophycocyanine (APC), qui lie la phycocyanobiline (PCB), chromophore de couleur bleue. Les différents types pigmentaires se différencient par la composition en PBPs des bras du phycobilisome (Figure 11A). Dans le type pigmentaire 1, les bras sont constitués uniquement de phycocyanine (PC) liant uniquement la PCB (couleur bleue). Dans le type pigmentaire 2, les bras contiennent à la fois de la PC (liant ici à la fois de la PCB et de la phycoérythrobiline (PEB) de couleur rouge) et de la phycoérythrine I (PEI) liant de la PEB (de couleur rouge). Enfin, dans le type pigmentaire 3 les bras contiennent de la PC (liant ici de la PCB, de la PEB et/ou de la phycourobiline (PUB) de couleur orange), de la PEI et de la phycoérythrine II (PEII) qui lient à la fois la PEB et la PUB (Six *et al.*, 2007c).

Du fait de cette composition, les souches de type pigmentaire 1 sont colorées en vert, celles de type pigmentaire 2 en rose et celles de type pigmentaire 3 ont une couleur orangée (Figure 11B). Des sous-types sont également différenciés au sein du type pigmentaire 3 en fonction du ratio entre le contenu en PUB et le contenu en PEB. Chez les souches de type 3a ce ratio est faible, il est intermédiaire chez les souches 3b et élevé chez les souches 3c (Six *et al.*, 2007c). Enfin le type 3d est capable de modifier sa composition pigmentaire en fonction de la couleur de la lumière utilisée pour le cultiver : il ressemble aux 3b quand il croît sous une lumière blanche ou verte, et aux 3c sous une lumière bleue (Palenik, 2001; Humily *et al.*, 2013). Chez ces deux derniers types pigmentaires (3c et 3d), la PEI fixe à la fois de la PEB et de la PUB, ce qui permet d'augmenter la proportion de PUB.

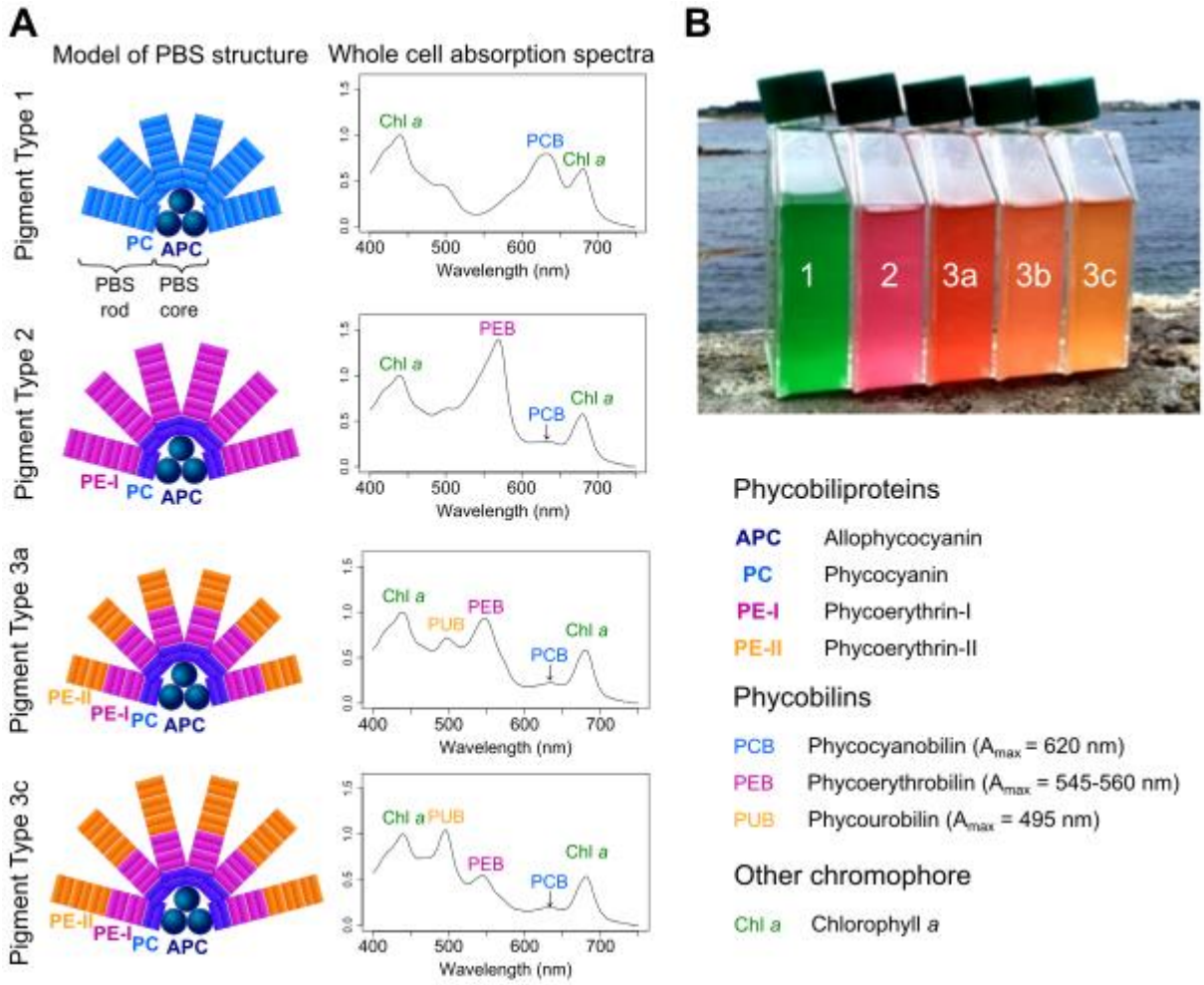


Figure 11 Caractéristiques des types pigmentaires de *Synechococcus*. A. Structure des phycobilisomes (à gauche) et spectres d'absorption (à droite) des types pigmentaires 1, 2, 3a et 3c. L'absorbtion est exprimée en unités arbitraires. Figure réalisée par Théophile Grébert (Grébert *et al., in revision*). B. Photographie des différents types pigmentaires de *Synechococcus*. Photo : Florian Humily.

Ces différences de pigmentation impliquent que les différents types pigmentaires absorbent préférentiellement certaines longueurs d'ondes : les populations de type 1 absorbent préférentiellement la lumière rouge, celles de type 2 et 3a la lumière verte, et celles de type 3b et 3c la lumière bleue, tandis que les 3d s'acclimatent à la longueur d'onde de la lumière ambiante (Six *et al.*, 2007c). L'étude de la répartition des types pigmentaires, en utilisant la cytométrie en flux qui permet de distinguer les populations « *high*-PUB » (types 3b, 3c, et potentiellement 3d) et « *low*-PUB » (types 2, 3a et potentiellement 3d), a révélé qu'en surface les premières dominent fortement au large dans les régions oligotrophes, alors que les secondes dominent plus près des côtes dans les régions eutrophes et mésotrophes (Olson *et al.*, 1988; Olson and Chisholm, 1990; Lantoine and Neveux, 1997; Michelle Wood *et al.*, 1998). Par ailleurs dans les régions eutrophes et mésotrophes, la proportion de cellules « *high*-PUB » augmente avec la profondeur, une tendance qui n'est pas toujours observée dans les régions oligotrophes où celles-ci dominent dans toute la colonne d'eau (Olson and Chisholm, 1990; Lantoine and Neveux, 1997). Les types pigmentaires se répartissent donc sur un gradient horizontal côte-large en surface et, selon les régions, sur un

gradient vertical surface-profondeur. De ce point de vue, les types pigmentaires représentent une adaptation à la diversité des niches de qualité de lumière disponibles dans l'environnement : dans les zones eutrophes et mésotrophes, plus turbides et riches en Chl a, la lumière bleue est fortement absorbée à la surface et les types pigmentaires utilisant au mieux les lumières rouge (type 1) et verte (type 2) sont favorisés (Stomp *et al.*, 2007). Au contraire dans les zones oligotrophes, la lumière bleue n'est que peu absorbée en surface et est largement disponible, favorisant les spécialistes de la lumière bleue (« *high*–PUB », types 3b et 3c). Enfin, les longueurs d'ondes bleues sont celles qui sont les moins absorbées par les molécules d'eau et qui pénètrent le plus profond dans la colonne d'eau, expliquant la dominance des « *high*-PUB » à ces profondeurs (Stomp *et al.*, 2007). La cytométrie en flux ne permet cependant pas de distinguer les types 3a des types 3d acclimatés à la lumière verte, ou les types 3c des types 3d acclimatés à la lumière bleue, et il est donc impossible à partir de ces données de déterminer si la répartition observée est due à des populations de différents types pigmentaires adaptés à des niches différentes, à l'acclimatation des populations capables de modifier leur appareil pigmentaire, ou aux deux processus à la fois.

Chez *Prochlorococcus* la diversité pigmentaire est bien moindre, mais la cytométrie en flux a malgré tout permis de distinguer deux types de populations : les populations « *dim* » qui présentent une faible fluorescence et sont essentiellement retrouvées en surface, et les populations « *bright* », fortement fluorescentes et retrouvées principalement en profondeur (Campbell and Vaulot, 1993; Partensky *et al.*, 1996). Les souches isolées en surface ont également un ratio chl *b*/chl *a* plus faible que les souches isolées en profondeur, ce qui leur a valu les dénominations de « low *b*/*a* » et « high *b*/*a* » (Partensky *et al.*, 1993; Moore *et al.*, 1995). Ceci correspond à une adaptation à l'intensité lumineuse disponible, les populations en profondeur absorbant plus efficacement la lumière. L'utilisation de dv-chl *a* et *b*, dont le spectre d'absorption est décalé vers le bleu par rapport aux versions monovinyl, les rend très efficace pour utiliser les longueurs d'ondes bleues qui sont abondantes dans les régions oligotrophes, et qui sont les seules disponibles au bas de la zone euphotique (Goericke and Repeta, 1992; Moore *et al.*, 1995).

1.3.3 La place des picocyanobactéries marines dans l'écosystème planctonique

Les picocyanobactéries marines sont retrouvées dans la plupart des régions océaniques, avec des différences de répartition entre *Synechococcus* et *Prochlorococcus* en fonction de quelques paramètres clés. Des deux genres, *Synechococcus* est celui qui présente la répartition la plus large (Figure 12A): il est retrouvé dans tous les océans à l'exception des régions polaires où la température de l'eau est inférieure à 0°C. Son abondance en surface varie de 500 à $3{,}6.10^6$ cellules/mL, un maximum observé dans la région d'*upwelling* du dôme du Costa Rica où la croissance du phytoplancton eucaryote est limitée par les conditions particulières de cette zone (Partensky *et al.*, 1999a; Saito *et al.*, 2005; Flombaum *et al.*, 2013). De façon générale, *Synechococcus* est plus abondant dans les régions riches en nutriments comme les *upwellings* ou les zones côtières, mais il est également toujours présent dans les régions oligotrophes, à une concentration proche de 1000 cellules/mL. Dans les zones subtropicales oligotrophes où la

température reste élevée toute l'année, il est cependant plus abondant en hiver quand le mélange de la colonne d'eau l'enrichit en nutriments (Partensky *et al.*, 1999a). En comparaison, la zone de répartition de *Prochlorococcus* est plus réduite (Figure 12B) : son abondance diminue fortement au-delà de 45°N ou S, et il est rarement observé à plus de 50°N ou S, ce qui correspond à une limite de température entre 10 et 13°C (Johnson *et al.*, 2006; Flombaum *et al.*, 2013). Lorsqu'il co-occurre avec *Synechococcus*, il est généralement plus abondant que ce dernier, notamment dans les régions oligotrophes où son abondance atteint 10^5 cellules/mL, soit deux ordres de grandeur de plus que *Synechococcus*. Ce patron de dominance s'inverse en revanche dans les régions enrichies comme les *upwellings* et au-delà de 40°N ou S, et à l'inverse de *Synechococcus*, la croissance de *Prochlorococcus* est limitée lors des épisodes de mélange de la colonne d'eau (Lindell and Post, 1995).

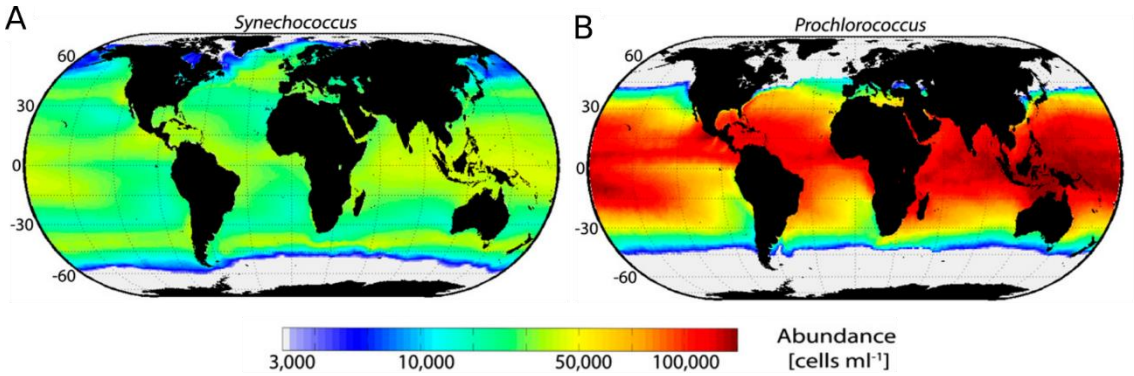


Figure 12 Abondance annuelle moyenne de *Synechococcus* et *Prochlorococcus* dans l'océan mondial. L'abondance calculée à partir d'un modèle de niche est indiquée en cellules/mL. A. *Synechococcus*. B. *Prochlorococus*. Source : (Flombaum *et al.*, 2013).

La différence entre les deux genres est également très marquée si l'on s'intéresse à leur répartition verticale dans la colonne d'eau. En effet l'abondance de *Synechococcus* est généralement maximale en surface ou au niveau de la nitracline (zone où le gradient vertical de concentration en nitrates et le plus important), mais elle diminue rapidement avec la profondeur et il n'est observé au bas de la zone euphotique qu'à très faible concentration (≤150 cellules/mL). Au contraire, *Prochlorococcus* est à la fois abondant en surface et en profondeur, où il est en général le membre dominant du phytoplancton au bas voire sous la limite de la zone euphotique (Partensky *et al.*, 1999a; Flombaum *et al.*, 2013). Sa capacité à croître à ces profondeurs serait liée non seulement à sa pigmentation particulière évoquée ci-dessus, qui lui permet d'absorber efficacement les longueurs d'ondes bleues, mais également à sa petite taille qui limite la diffusion de la lumière par la cellule et maximise la quantité de lumière absorbée par la cellule (Morel *et al.*, 1993).

Les caractéristiques de la répartition des picocyanobactéries marines évoquées ci-dessus varient en fonction des saisons, en particulier les limites latitudinales (en lien avec les variations de température), et la répartition verticale (en lien avec le mélange de la colonne d'eau) (Flombaum *et al.*, 2013). L'abondance de *Synechococcus* augmente par exemple fortement au printemps dans les régions tempérées (sur la côte Nord-Américaine, ce *bloom* intervient quand la température de l'eau dépasse les 6°C), alors que l'augmentation d'abondance de *Prochlorococcus* est plus progressive et intervient plus tardivement.

En faisant abstraction de ces variations, c'est-à-dire en faisant la moyenne sur l'année sur l'ensemble des océans, les picocyanobactéries marines représentent 10% du picoplancton sur les 200 premiers mètres de la colonne d'eau. De par leur abondance, elles sont responsables d'environ 25% de la production primaire nette dans les océans. *Prochlorococcus* est globalement plus abondant en termes de nombre de cellules, estimé à 2.9×10^{27} cellules contre 7.0×10^{26} cellules pour *Synechococcus* (Flombaum *et al.*, 2013), notamment du fait de sa dominance dans les larges étendues oligotrophes et en profondeur. Cependant, étant donnée leur différence de taille, *Synechococcus* fixe deux fois plus de carbone (8 Gt/an, 16,7% de la production primaire nette des océans) que *Prochlorococcus* (4 Gt/an, 8,5%). Cette masse de carbone suit les différentes voies du réseau trophique décrit dans la partie 1.2.2: une partie est libérée au cours de la vie de la cellule sous forme de matière organique dissoute ou de vésicules extracellulaires (Biller *et al.*, 2014b), tandis qu'une partie des cellules est lysée par les cyanophages et libère de la matière organique dissoute ou particulaire, et qu'une autre partie est consommée par des organismes « brouteurs » (notamment des flagellés hétérotrophes et des Ciliés, ainsi que par certains membres mixotrophes du phytoplancton) (Christaki *et al.*, 1999; Guillou *et al.*, 2001; Frias-Lopez *et al.*, 2009). En effet, malgré un taux de croissance *in situ* correspondant en moyenne à une division par jour, la concentration cellulaire varie peu au cours du temps, ce qui suggère un taux de mortalité compensant le taux de croissance (Partensky *et al.*, 1999a; Hunter-Cevera *et al.*, 2014). Enfin, *Prochlorococcus* et *Synechococccus* sont tous les deux capables d'utiliser certaines sources organiques de nutriments (voir partie 3.2 de cette introduction), et présentent donc un certain degré de mixotrophie, ce qui leur ajoute encore un degré de connectivité au sein du réseau trophique. Ces quelques données estimées à partir des observations effectuées *in situ* et sur des cultures en laboratoire donnent une idée du rôle crucial joué par les picocyanobactéries marines dans l'écosystème planctonique au niveau mondial.

L'ubiquité des picocyanobactéries marines, et leur capacité à occuper des milieux aux caractéristiques très différentes (tant au niveau de la température de l'eau que dans la disponibilité en nutriments et en lumière) suggèrent l'existence d'une diversité intra-genre constituant des populations adaptées à des niches écologiques [4]différentes et/ou une grande plasticité phénotypique de ces organismes, deux aspects qui seront explorés dans les parties suivantes de cette introduction.

[4] La niche écologique peut avoir différentes définitions. Dans cette thèse, nous utilisons ce terme pour désigner l'ensemble des conditions biotiques et abiotiques permettant la survie d'un organisme. Pour un paramètre environnemental donné, la limite de la niche est donc la limite à partir de laquelle l'organisme ne peut plus maintenir sa population à un effectif stable.

2. Diversité génétique et répartition géographique des picocyanobactéries marines

2.1 Diversité génétique des picocyanobactéries marines

Comme nous l'avons vu dans la partie précédente, la diversité des picocyanobactéries marines a d'abord été reconnue à travers des caractéristiques phénotypiques mesurables dans l'environnement et/ou sur des cultures en laboratoire, notamment leur contenu pigmentaire et la fluorescence associée. Cependant dès leur découverte, cette diversité a également été reliée à des caractéristiques génétiques, porte d'accès de l'histoire évolutive de ce groupe. Ainsi dans les années 1980, 3 genres ont été définis sur la base du %GC (pourcentage de bases G+C dans le génome) pour classer taxonomiquement le vaste groupe nommé *Synechococcus* qui rassemblait les « petites cyanobactéries unicellulaires de forme cylindrique à ovoïde, se reproduisant par fission binaire transverse sur un seul plan, et ne possédant pas de gaine », à la fois d'eau douce et marines : *Cyanobacterium* (39-43 %GC), *Synechococcus* (47-56 %GC) et *Cyanobium* (66-71 %GC) (Rippka *et al.*, 1979). Cependant les 27 premières souches marines de *Synechococcus* marines se sont avérées avoir entre 54.9 et 62.4 %GC, et cette classification a rapidement été abandonnée (Waterbury *et al.*, 1986). La classification linnéenne n'ayant pas été modifiée depuis lors, le genre *Synechococcus* est resté largement polyphylétique, regroupant des organismes marins et d'eau douce éloignés phylogénétiquement – mais le groupe des picocyanobactéries marines comprend encore quelques souches nommées *Cyanobium* sp. Si *Prochlorococcus* a quant à lui été nommé ainsi à cause de sa ressemblance pigmentaire avec les prochlorophytes du genre *Prochloron* (Chisholm *et al.*, 1988), le séquençage de marqueurs génétiques a rapidement démontré sa proximité avec les *Synechococcus* marins. Bien que le genre *Synechococcus* soit polyphylétique, la radiation des picocyanobactéries marines forme quant à elle un groupe monophylétique bien défini au sein des Cyanobactéries (Shih *et al.*, 2013), et le genre *Prochlorococcus* y forme un sous-groupe également monophylétique Figure 13.

Bien que la taxonomie linnéenne n'ait pas évolué, l'utilisation de marqueurs génétiques variés tels que l'ADNr 16S (Moore *et al.*, 1998; Fuller *et al.*, 2003), *rpoC1* (gène de la sous-unité γ de l'ARN polymérase, (Palenik *et al.*, 1997; Toledo and Palenik, 1997; Ferris and Palenik, 1998)), *ntcA* (codant pour un régulateur du métabolisme de l'azote (Lindell and Post, 2001; Penno *et al.*, 2006)), *petB-petD* (codant pour des sous-unités du cytochrome $b_{6/f}$ (Urbach *et al.*, 1998; Mazard *et al.*, 2012a)) ou l'ITS (*internal transcribed spacer* de l'ADNr 16S-23S (Rocap *et al.*, 2002; Ahlgren and Rocap, 2006)) a progressivement permis de distinguer différents clades au sein des genres *Synechococcus* et *Prochlorococcus*. Ces clades, découverts au fur et à mesure d'amplifications *in situ* et sur des clones isolés en culture, sont robustes à l'utilisation de différents marqueurs bien que les relations entre clades, c.-à-d. les bipartitions profondes de l'arbre phylogénétique, soient mal résolues. La classification actuelle comprend près de 20 clades chez *Synechococcus* et 11 clades chez *Prochlorococcus* (Huang *et al.*, 2012) (l'arbre phylogénétique représenté sur la Figure 13, basé sur le gène de l'ARNr 16S, montre certains de ces clades). Les *Synechococcus* marins sont divisés en trois « sous-clusters » (subcluters 5.1, 5.2 et 5.3 (Dufresne *et al.*, 2008), Figure 13), le

subcluster 5.1 regroupant la majorité des souches décrites et semblant être le plus diversifié sur la base des séquences aujourd'hui disponibles.

Figure 13 Arbre phylogénétique des picocyanobactéries marines basé que le gène 16S de l'ARN ribosomal. Arbre en Neighbor-Joining issu de (Scanlan et al., 2009) et modifié par J. Pittera (Pittera, 2015). Les ronds blancs indiquent des valeurs de bootstrap entre 70 et 90%, et les ronds noires des valeurs de bootstrap supérieures à 90%. Seule une partie des clades décrits est présentée. Noter que la résolution du marqueur utilisé (gène de l'ARNr 16S) ne permet pas de différencier les clades V, VI, VII et CRD1.

De nouveaux clades sont cependant encore régulièrement décrits dans des régions du monde jusqu'ici peu séquencées, notamment au sein des sub-clusters 5.2 et 5.3, Huang *et al.* décrivant par exemple 6 clades au sein du subcluster 5.3 (Chen *et al.*, 2006; Huang *et al.*, 2012). Certains nouveaux clades ont été décrits à partir de séquences environnementales d'un marqueur donné, et ne sont donc pas représentés dans les phylogénies basées sur d'autres marqueurs (p. ex. au sein du subcluster 5.1 les clades EnvA et EnvB décrits à partir du marqueur *petB*, ou le clade CRD2 décrit à partir de l'ITS). Enfin certains marqueurs, notamment le gène de l'ARNr 16S, sont trop conservés pour pouvoir différencier une partie des clades qui sont identifiés à l'aide d'autres marqueurs. A l'inverse, certains gènes moins conservés offrent une très bonne résolution taxonomique : l'utilisation du gène *petB* a ainsi permis de mettre en évidence des sous-clades au sein de la plupart des clades du subcluster 5.1, signes de l'existence d'une microdiversité au sein de ces ensembles taxonomiques (Mazard *et al.*, 2012a).

Chez *Prochlorococccocus*, les populations retrouvées en surface forment une lignée monophylétique subdivisée en six clades (HLI-VI) et celles retrouvées en profondeur forment plusieurs lignées (LLI-LLVI) dont la diversité est encore mal connue (Huang *et al.*, 2012; Biller *et al.*, 2014a). Le gène marqueur le plus souvent utilisé pour distinguer ces clades est l'ITS, la séquence de l'ARNr 16S étant très conservée (p. ex. plus de 99% d'identité au sein du clade HL) (Rocap *et al.*, 2002). De manière similaire à *Synechococcus*, l'utilisation d'une résolution taxonomique plus élevée (notamment en différenciant les variants d'ITS au nucléotide près) souligne l'existence d'une microdiversité importante au sein des clades présentés sur la Figure 13 (Kashtan *et al.*, 2014; Larkin *et al.*, 2016; Needham *et al.*, 2017).

2.2 La répartition biogéographique des picocyanobactéries marines, signe d'une adaptation à la niche écologique

2.2.1 La biogéographie : la science de la répartition des espèces

Si l'utilisation de la cytométrie en flux a permis de mettre en évidence une répartition différentielle des populations de picocyanobactéries marines aux propriétés pigmentaires distinctes dans les océans et le long de la colonne d'eau, ces populations se sont avérées être également distinctes d'un point de vue génétique : les populations de *Prochlorococcus* « high b/a » et « low b/a » ont rapidement été identifiées comme formant deux groupes phylogénétiques distincts sur la base de l'ARNr 16S (Moore *et al.*, 1998), correspondant respectivement aux groupes LL (« low-light ») et HL (« high light »). Les populations de surface (HL) se différencient donc d'un point de vue évolutif des populations de profondeur (LL). L'étude physiologique de ces deux groupes (développée dans la partie 3) a également révélé que les souches LL (de profondeur) était adaptées à croître à des intensités lumineuses moindre par rapport aux souches HL, en accord avec la répartition des populations dans la colonne d'eau (Moore *et al.*, 1998; West and Scanlan, 1999).

En termes d'adaptation à l'intensité lumineuse, le clade HL correspond donc à un écotype [5]au sens strict, puisqu'il forme un groupe monophylétique possédant des adaptations physiologiques lui permettant d'occuper une niche environnementale particulière. Cet exemple illustre l'idée que les clades déterminés uniquement à partir de la séquence de gènes marqueurs forment des entités écologiques adaptées à des conditions environnementales distinctes et présentant par conséquent une répartition géographique différentielle. C'est l'idée de base sous-jacente aux études biogéographiques, qui ont démarré par l'étude de la répartition des espèces de plantes terrestres. Mais si, de manière similaire, les botanistes du XVIIIe et XIXe siècle s'apprêtaient à trouver les mêmes espèces de plantes dans les régions présentant un climat et des caractéristiques pédologiques similaires, ils se sont rapidement rendu compte que l'inventaire floristique d'une région dépendait au moins tout autant, voire plus, de son histoire que de ses caractéristiques physico-chimiques – ce qui s'explique notamment, à la lumière de nos connaissances actuelles, par des processus démographiques (migrations, dispersion limitée…) et par l'isolement géographique permettant des spéciations allopatriques[6].

Qu'en est-il donc des picocyanobactéries marines ? Contrairement aux plantes, les bactéries semblent moins limitées dans leurs capacités de dispersion, en particulier dans un milieu relativement ouvert comme l'océan. Dans ce contexte, il est tentant de penser que les « espèces » bactériennes sont ubiquistes, mais qu'une espèce donnée sera spécifiquement favorisée dans certaines régions du fait de la sélection par les paramètres environnementaux (biotiques et abiotiques) – une idée résumée dans la formule de Baas Becking « everything is everywhere, *but*, the environment selects » (Becking, 1934; De Wit and Bouvier, 2006).

2.2.2 Biogéographie des clades de *Prochlorococcus*

L'utilisation de gènes marqueurs pour traquer les populations de picocyanobactéries *in situ* a permis d'apporter quelques réponses aux questions mentionnées précédemment. Outre la répartition différentielle des clades LL et HL, les populations au sein du clade HL se répartissent dans l'océan selon quelques gradients physico-chimiques : le clade HLI domine entre 30°S et 45°S et entre 35°N et 45°N, soit aux limites Nord et Sud de la répartition de *Prochlorococcus* dans des eaux relativement froides et peu stratifiées, alors que le clade HLII domine la majorité des régions situées entre ces deux latitudes, dans des eaux chaudes et stratifiées (Bouman *et al.*, 2006; Johnson *et al.*, 2006; Zinser *et al.*, 2007; Zwirglmaier *et al.*, 2007, 2008) (Figure 14). Cette répartition correspond à un gradient de température de l'eau et de stabilité de la colonne d'eau, et est reliée à des différences physiologiques mesurées sur des souches représentatives de ces deux clades

[5] Le terme d'écotype a différentes acceptations. Au sens large, il peut désigner un groupe d'organismes partageant les mêmes préférences écologiques. Au sens strict, utilisé ici il s'agit d'une population se distinguant des autres à la fois par ses caractéristiques génétiques et par sa niche écologique, et qui présente donc des caractéristiques physiologiques mesurables la différenciant des autres populations (Turesson, 1922).

[6] Le terme de spéciation allopatrique désigne la formation de deux espèces à partir d'une seule espèce du fait d'une divergence due à une séparation physique et/ou géographique.

(Figure 14): les souches HLI sont adaptées à des températures plus faibles que les souches HLII, ce qui explique la répartition de leurs populations respectives le long d'un gradient latitudinal (Johnson *et al.*, 2006; Zinser *et al.*, 2007).

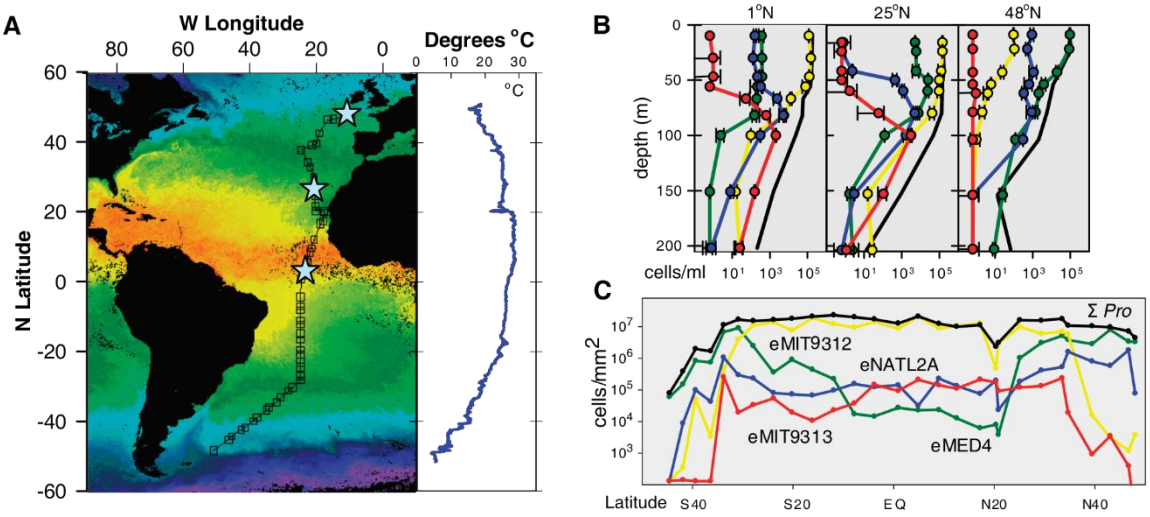


Figure 14 Répartition latitudinale des écotypes de *Prochlorococcus*. A. Transect AMT13 et profil de température de l'eau en surface. B. Profils verticaux d'abondance des écotypes de *Prochlorococcus* aux points indiqués par des étoiles sur la carte. C. Abondance intégrée sur la colonne d'eau des différents écotypes de *Prochlorococcus* le long du transect. Rouge : LLIV, bleu : LLI, jaune : HLII, vert : HLI. Figures modifiées à partir de (Johnson *et al.*, 2006).

Les clades HLIII et HLIV (d'abord connus sous le nom HNLC1 et HNLC2) ont quant à eux été découverts plus récemment à la fois par l'analyse de métagénomes des océans Pacifique et Indien et par l'amplification de séquences *in situ* dans le Pacifique équatorial (Rusch *et al.*, 2010; West *et al.*, 2011; Huang *et al.*, 2012). Ces deux clades dominent dans une zone latitudinale habituellement dominée par le clade HLII, c.-à-d. dans les eaux intertropicales stratifiées, mais sont restreints aux régions regroupées sous le sigle HNLC (High Nutrient Low Chlorophyll[7]) qui sont relativement riches en macronutriments mais où les concentrations en fer sont très faibles et limitent la croissance du phytoplancton (Martin *et al.*, 1994; Behrenfeld and Kolber, 1999; Behrenfeld *et al.*, 2008). Les clades HLV et HLVI ont pour l'instant été très peu décrits, du fait de leur découverte récente et de leur rareté, mais le premier (HLV) semble être co-occurrent des clades HLIII et HLIV dans les régions HNLC, bien qu'en faible abondance, et le clade HLVI semble être présent plus en profondeur que les autres populations HL (Huang *et al.*, 2012). La répartition de ces deux clades reste cependant à étudier plus en détails.

La biogéographie des clades LL, qu'on ne retrouve pas ou peu en surface, est bien moins connue. Le clade LLI est retrouvé à des profondeurs intermédiaires entre celles dominées par les clades HL et celles dominées par les autres LL, et supporte – contrairement à la plupart des clades LL – un passage de courte durée dans les eaux de surface (Zinser *et al.*, 2007; Zwirglmaier *et al.*,

[7] Les régions HNLC sont des régions où malgré une concentration en nutriments relativement élevée, la productivité primaire, estimée à partir de la concentration en chlorophylle, est faible.

2007; Malmstrom *et al.*, 2010). A l'opposé, le clade LLIV est retrouvé à la limite inférieure de la zone euphotique et ne survit pas à l'exposition aux intensités lumineuses rencontrées en surface (Rocap *et al.*, 2002; Malmstrom *et al.*, 2010). Les clades LLV et LLVI ont quant à eux été découverts dans une zone limitée en oxygène présentant des conditions de suboxie particulières, et n'ont pour l'instant pas été décrits ailleurs (Lavin *et al.*, 2010).

De façon générale, pour les clades les mieux caractérisés, il semble donc que quelques paramètres environnementaux soient structurants pour la répartition horizontale et verticale des populations de *Prochlorococcus* : l'intensité lumineuse, la température et la disponibilité en fer. Ces paramètres influent sur la biogéographie à différents niveaux de la phylogénie : les différences d'intensité lumineuse sont liées aux branches profondes de l'arbre de *Prochlorococcus* différenciant les HL des LL, tandis que la disponibilité en fer et la température ont influé sur la différenciation des clades HL plus récemment au cours de l'évolution (Figure 15). D'autres paramètres abiotiques ne permettent pas d'expliquer la répartition des clades, notamment la concentration en macronutriments (azote et phosphate) (Martiny *et al.*, 2009c; Biller *et al.*, 2014a). Nous verrons plus loin dans cette introduction que différentes populations au sein de chaque clade possèdent des adaptations différentes pour l'utilisation de ces nutriments, ce qui explique le manque de lien avec la phylogénie.

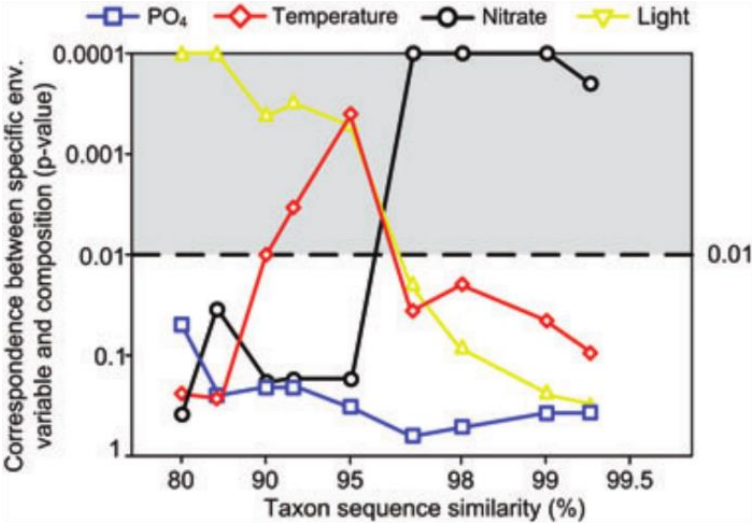


Figure 15 Relation entre le niveau taxonomique observé et l'influence des variables environnementales sur la structure des communautés de *Prochlorococcus*. L'abscisse correspond au pourcentage de similarité (ITS) utilisé pour définir les taxons de *Prochlorococcus,* et l'ordonnée à la significativité de l'association entre une variable environnementale et la composition taxonomique de la communauté (p-value d'une analyse canonique de correspondance partielle). La zone grisée correspond à une forte significativité (p-value < 0.01). Par exemple, la lumière influe sur la composition des communautés de *Prochlorococcus* quand les taxons sont définis sur la base d'un pourcentage de similarité entre 80 et 95%. Figure issue de (Martiny *et al.*, 2009c).

La vision d'une relation entre clades et paramètres environnementaux chez *Prochlorococcus* s'est complexifiée ces dernières années avec l'étude de la diversité présente au sein de chacun des clades – rendue possible par l'utilisation de marqueurs hautement résolutifs et l'amélioration des

techniques de séquençage. Cette microdiversité au sein des écotypes précédemment décrits a principalement été étudiée au sein des clades HLI et HLII, et semble présenter une répartition différentielle dans le temps et dans l'espace, bien que l'effet des différents paramètres environnementaux (biotiques et abiotiques) soit encore mal compris (Kashtan *et al.*, 2014; Larkin *et al.*, 2016; Kashtan *et al.*, 2017).

2.2.3 Biogéographie des clades de *Synechococcus*

Chez *Synechococcus*, les études biogéographiques ont également montré une répartition différentielle des clades, notamment des clades les mieux détectés par les méthodes d'amplification. Seulement quelques études se sont intéressées à la répartition des clades de *Synechococcus* à l'échelle du bassin océanique ou à l'échelle globale (Zwirglmaier *et al.*, 2007, 2008; Mazard *et al.*, 2012a; Huang *et al.*, 2012; Sohm *et al.*, 2015), et leurs résultats sont complétés par des études plus localisées. Le sub-cluster 5.2 semble n'être retrouvé que dans des eaux peu salées, au niveau d'estuaires et/ou dans les régions arctiques (Chen *et al.*, 2006; Huang *et al.*, 2012; Larsson *et al.*, 2014). La répartition des populations du subcluster 5.3 est assez peu connue, mais des études récentes suggèrent qu'il est relativement ubiquiste (Huang *et al.*, 2012; Sohm *et al.*, 2015), et que différents clades peuvent être distingués au sein du subcluster dont certains sont retrouvés préférentiellement en profondeur (Huang *et al.*, 2012).

Le subcluster le plus étudié est de loin le subcluster 5.1. Les travaux le concernant ont montré que les clades I et IV dominent dans les eaux tempérées et froides, mésotrophes à eutrophes notamment les régions côtières ou les régions d'upwelling (Zwirglmaier *et al.*, 2007, 2008; Tai and Palenik, 2009; Tai *et al.*, 2011; Ahlgren and Rocap, 2012; Mazard *et al.*, 2012a; Huang *et al.*, 2012; Sohm *et al.*, 2015), le clade II étant son équivalent tropical, dominant dans les zones oligotrophes à mésotrophes plus chaudes (Figure 16) (Zwirglmaier *et al.*, 2008). Le clade III semble quant à lui restreint aux régions oligotrophes (Zwirglmaier *et al.*, 2008; Mazard *et al.*, 2012a; Sohm *et al.*, 2015), et peut dominer la communauté de *Synechococcus* en Méditerranée (Mella-Flores *et al.*, 2011) et en mer Rouge (Post *et al.*, 2011) (Figure 16). Les clades CRD1 et CRD2 ont été identifiés dans le dôme du Costa Rica, où *Synechococcus* atteint les plus fortes abondances mesurées *in situ* (jusqu'à $3{,}6.10^6$ cellules/mL). La répartition géographique de ce clade est restée floue jusqu'à son évaluation très récente, qui a mis en évidence sa dominance sur les autres clades de *Synechococcus* non seulement dans la région du dôme du Costa Rica (Ahlgren *et al.*, 2014) mais également dans certaines régions limitées en fer des océans Atlantique et Pacifique (Sohm *et al.*, 2015) (Figure 15).

En dehors de ces cinq clades dominants, la biogéographie des clades restants de *Synechococcus* est encore mal connue. Les clades V, VI et VII ont d'abord été considérés comme ubiquistes mais relativement peu abondants (Zwirglmaier *et al.*, 2008; Huang *et al.*, 2012), mais les amorces utilisées pour détecter ces clades n'étaient pas spécifiques et amplifiaient probablement en partie des séquences du clade CRD1. Leur répartition géographique est donc peu connue. Le clade VIII est présent dans les eaux peu salées, par exemple dans les zones estuariennes, et les souches de ce clade qui ont été isolées sont capables de croître dans un milieu d'eau douce (Fuller *et al.*, 2003). Les clades XV et XVI sont peu abondants, mais ont été trouvés essentiellement dans

les régions de transition entre les zones dominées par les clades I et IV, et les zones dominées par le clade II (Sohm *et al.*, 2015). Certaines études suggèrent cependant que ces clades sont retrouvés légèrement plus en profondeur que les autres, au moins dans la région du dôme du Costa Rica (Ahlgren *et al.*, 2014; Guttierrez-Rodriguez *et al.*, 2014; Sohm *et al.*, 2015). De nombreux clades ne sont décrits qu'à partir d'un seul marqueur et à une localisation précise, et en l'absence de souche de référence il est impossible d'établir les liens entre les clades décrits à partir de marqueurs différents (Ahlgren and Rocap, 2012). Si la description de ces « clade-mystères » montre la diversité de *Synechococcus*, leur écologie (et parfois même leur place dans la phylogénie du genre) est encore floue.

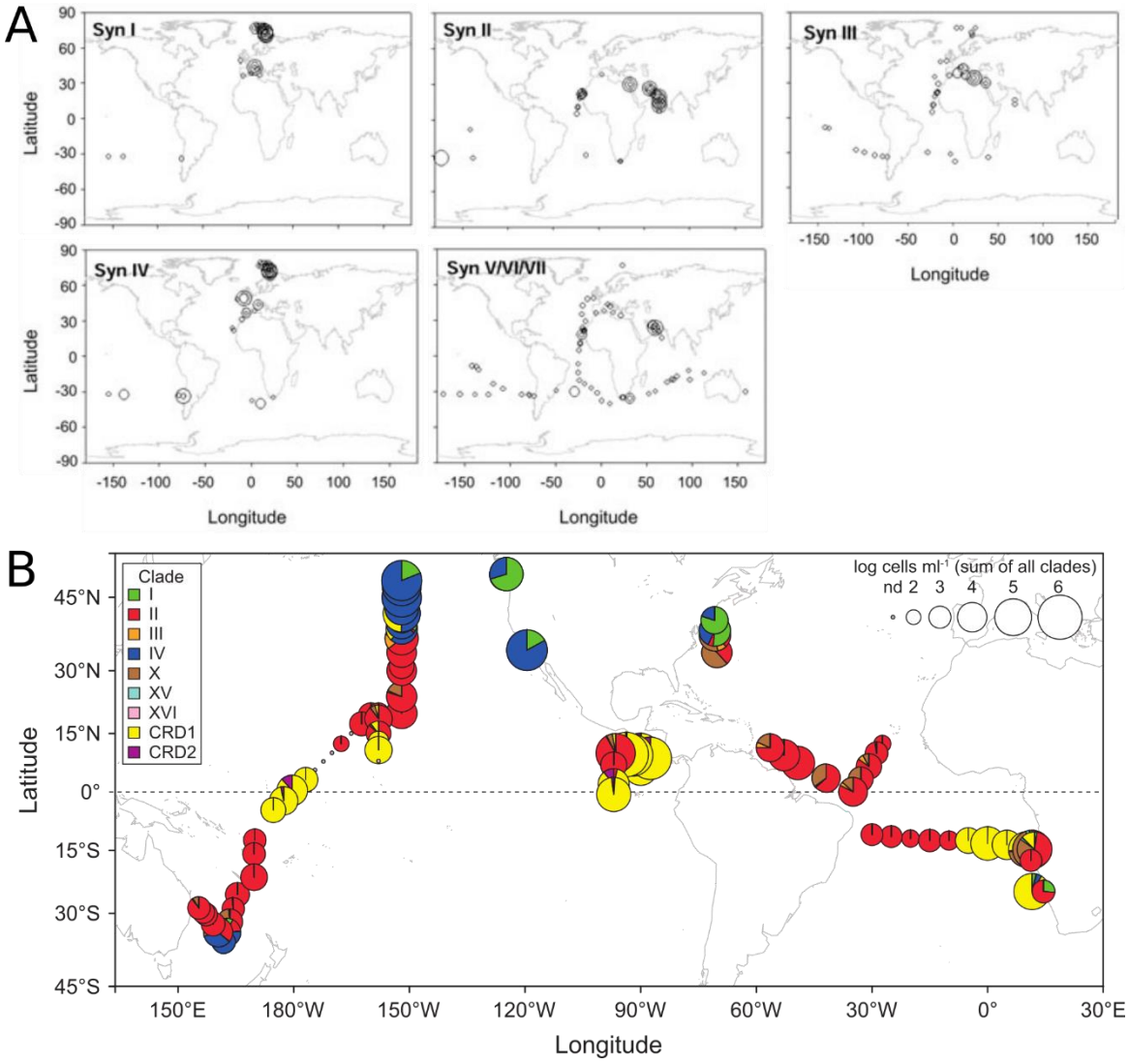


Figure 16 Répartition des clades de *Synechococcus* à l'échelle mondiale, en surface. A. Abondance relative des principaux clades de *Synechococcus* estimée par hybridation *dot blot* sur le gène marqueur 16S. Noter que la méthode utilisée ne permet pas de distinguer les clades V, VI, VII et CRD1. La taille des cercles est proportionnelle à l'abondance relative du clade. Source : (Zwirglmaier *et al.*, 2008). B. Abondance relative des principaux clades de *Synechococcus* estimée par qPCR sur le marqueur 16S-23S ITS. Le clade X correspond au subcluster 5.3. La taille des cercles représente la concentration totale en *Synechococcus* en cellules/mL (échelle logarithmique). Source : (Sohm *et al.*, 2015).

De manière générale, les principaux paramètres qui semblent influer sur la composition de communautés de *Synechococcus* sont donc la température, la concentration en macronutriments (phosphate et azote), la disponibilité en fer (et peut être en d'autres métaux traces) ainsi que, dans certains cas, la salinité. Contrairement à ce qui est observé pour *Prochlorococcus,* l'intensité lumineuse ne semble pas structurer les populations de *Synechococcus*. L'influence de ces paramètres sur la répartition des clades est cependant bien moins comprise chez *Synechococcus*, potentiellement du fait de l'existence d'une diversité au sein des clades établis. En effet de manière similaire à ce que nous avons décrits pour *Prochlorococcus,* l'utilisation de marqueurs résolutifs a permis de montrer l'existence d'une microdiversité (voir ci-dessus) dont la biogéographie reste à étudier en détails mais semble être différente selon les sous-clades (Tai *et al.*, 2011; Mazard *et al.*, 2012a). Si la plus grande diversité génétique observée chez *Synechococcus* explique peut-être pourquoi son étendue géographique est plus vaste, nous sommes donc encore loin de comprendre les facteurs qui gouvernent la répartition de ses communautés.

2.2.4 Au-delà de l'espace : variations temporelles des communautés de picocyanobactéries

A la répartition géographique s'ajoute une variation temporelle des communautés de *Prochlorococcus* et *Synechococcus*. En effet, la composition des communautés suit les variations saisonnières des conditions environnementales, en accord avec les préférences écologiques décrites ci-dessus. La composition en écotypes de *Prochlorococcus* et leur répartition verticale dans la colonne d'eau est stable tout au long de l'année au large d'Hawaï, où la saisonnalité est faible et où la colonne d'eau est stratifiée toute l'année. Au contraire, l'abondance des différents écotypes et leur répartition dans la colonne d'eau varie en fonction de la saison au large des Bermudes, où la colonne d'eau est profondément mélangée en hiver (ce qui a pour effet de diminuer la température de l'eau et d'augmenter la concentration en nutriments) : en surface par exemple, l'écotype HLI atteint son pic d'abondance au printemps, suite à l'événement de mélange hivernal, quand les eaux sont encore froides, alors que le pic d'HLII est atteint durant l'été. Lors du mélange hivernal, le clade LLI peut se retrouver en forte abondance en surface, alors qu'il se trouve à une profondeur intermédiaire en été (Malmstrom *et al.*, 2010). Ces patrons de variations annuelles sont très reproductibles d'une année à l'autre (Zwirglmaier *et al.*, 2007; Malmstrom *et al.*, 2010), ce qui souligne la dépendance entre la composition en écotypes et les conditions environnementales.

Chez *Synechococcus*, les variations saisonnières de l'abondance des différents clades semblent également reliées à la disponibilité en nutriments et éventuellement à la température de l'eau, en lien avec le degré de stratification de la colonne d'eau (Fuller *et al.*, 2006; Tai and Palenik, 2009; Tai *et al.*, 2011; Post *et al.*, 2011; Choi *et al.*, 2013), et les sous-clades semblent également présenter des variations saisonnières dans leur abondance (Tai *et al.*, 2011). La variabilité saisonnière n'ayant été étudiée qu'au niveau de certains sites localisés, l'intégration de la répartition spatiale et de la répartition temporelle est encore loin d'être comprise, d'autant plus que le rôle joué par le déplacement des masses d'eaux (courants) est rarement pris en compte.

2.3 Explorer la diversité des picocyanobactéries marines : à la recherche des bases génétiques de l'adaptation

2.3.1 20 ans de génomique bactérienne

Le premier génome bactérien entièrement séquencé a été publié en 1995 (Fleischmann *et al.*, 1995) et le nombre de génomes disponibles a depuis augmenté de manière exponentielle, poussé par les progrès des techniques de séquençage et des méthodes d'analyse bio-informatique (Loman and Pallen, 2015). Les premiers génomes ont d'abord surpris par la quantité de gènes non caractérisés qu'ils comprenaient, y compris pour des organismes ayant pourtant servi de modèle pendant plusieurs dizaines d'années, comme *Escherichia coli* ou *Bacillus subtilis* (Blattner *et al.*, 1997; Kunst *et al.*, 1997). Les premières comparaisons génomiques intra-spécifiques ont ensuite révélé la grande variabilité du contenu en gènes des génomes bactériens, qualifiés de génomes « mosaïques », différentes souches d'une même espèce ne partageant qu'une partie de leurs gènes et possédant une part non négligeable de gènes spécifiques à chaque souche(Welch *et al.*, 2002). Ce constat, étendu à l'ensemble des bactéries, a amené à définir deux catégories de gènes en fonction de leur état de conservation au sein d'un jeu de génomes donné : les gènes core sont ceux partagés par l'ensemble des souches, tandis que les gènes accessoires ne sont présents que dans une partie des souches (Figure 17) (Lan and Reeves, 2000). Parmi ces gènes accessoires, une partie n'est retrouvée que dans un génome – on qualifie alors ces gènes d' « uniques ». Etant donné que chaque souche du jeu de données peut posséder des gènes uniques, le répertoire génétique d'une espèce ou d'un taxon ne se limite pas à celui d'une souche. Il s'étend au contraire à l'ensemble des gènes trouvés dans au moins une souche et est qualifié de « pangénome » (Figure 17) (Tettelin *et al.*, 2005). Ce terme de pangénome peut désigner soit l'ensemble des gènes présents dans au moins un génome, soit l'ensemble des gènes potentiellement disponibles pour le taxon étudié, y compris dans les populations non séquencées.

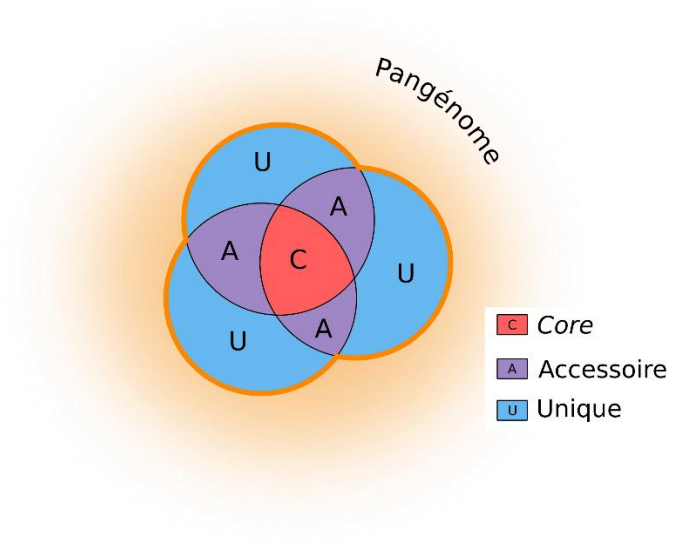


Figure 17 Définition des gènes core, accessoires et uniques. Chaque cercle représente un génome. Les gènes *core* (C) sont trouvés dans tous les génomes, les gènes accessoires (A) sont trouvés dans au moins deux génomes mais absents dans au moins un génome. Enfin, les gènes présents dans une seule souche sont

qualifiés d'uniques (U). Le pangénome correspond soit à l'ensemble des gènes trouvés dans au moins une souche séquencée (trait orange), soit l'ensemble des gènes trouvés dans le taxon considéré in situ (halo orange).

L'étendue de la diversité du contenu en gènes des espèces bactériennes a accentué, sinon déclenché, un débat virulent au sujet de l'existence même d'espèces prokaryotes – en tant qu'entités naturelles répondant à un concept défini et reconnaissable selon des critères objectifs. Le débat n'est pas tranché et ne pourra l'être avant d'avoir choisi un concept définissant l'espèce, mais les faits invoqués par les différentes parties méritent d'être évoqués. La remise en cause de l'existence d'espèces trouve son origine dans la reconnaissance du rôle des transferts horizontaux de gènes (THG ; voir l'encart Transferts horizontaux de gènes et îlots génomiques

), c.-à-d. de l'acquisition de gènes provenant d'organismes tiers, dans l'évolution des bactéries. La diversité du contenu en gènes des espèces n'est qu'un argument parmi d'autres prouvant l'importance des THG chez les prokaryotes, qui ont été mis en avant du fait de leur rôle dans l'acquisition et la dissémination intra-et inter-spécifique de gènes de virulence et de gènes de résistance aux antibiotiques chez des bactéries pathogènes.

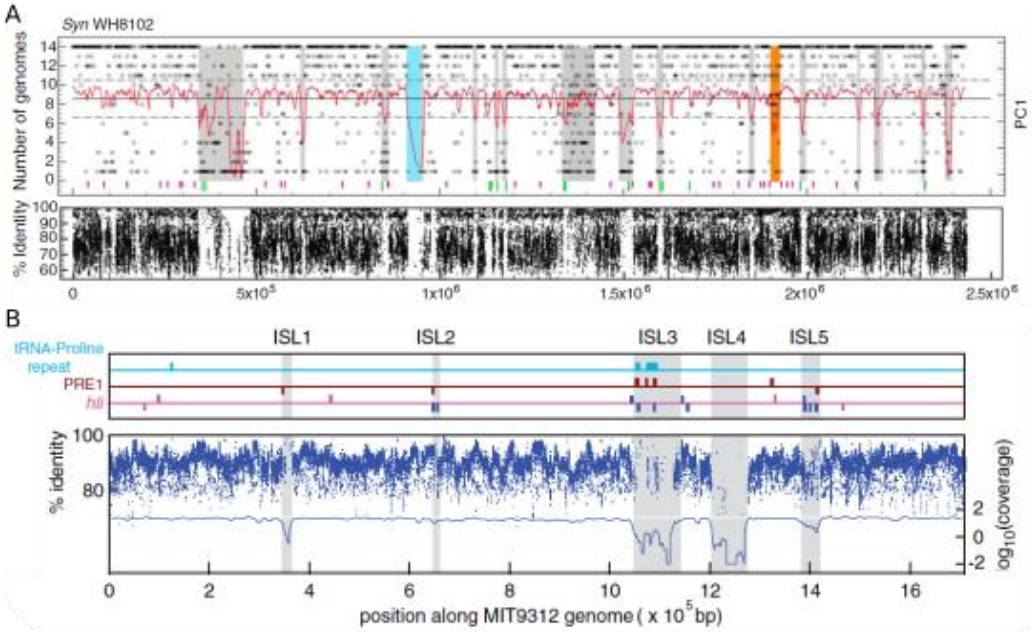
Transferts horizontaux de gènes et îlots génomiques

Le terme de transfert horizontal, par opposition à la transmission verticale, désigne le transfert d'un gène entre deux organismes qui ne sont pas parent et enfant. Chez les bactéries, ces gènes sont soit portés par des plasmides, soit intégrés dans le génome par recombinaison (homologue ou hétérologue). Le fragment d'ADN étranger peut pénétrer dans la cellule soit par transduction, c.-à-d. suite à l'infection par un bacteriophage, soit par conjugaison avec une autre bactérie, soit par transformation naturelle, c.-à-d. la récupération d'ADN libre présent dans l'environnement. Les fragments génomiques transférés (et maintenus dans les génomes) sont en général constitués de plusieurs gènes, formant souvent un ou plusieurs opérons, et conférant une ou des fonctions particulières à la cellule – par exemple, la résistance à un antibiotique.

Du fait du mécanisme de recombinaison, l'intégration dans le génome se fait préférentiellement dans certaines régions entourées de gènes conservés (par exemple des ARNt, ou des séquences d'insertion d'éléments transposables). Ce regroupement forme des îlots génomiques où se concentrent les gènes acquis par THG et qui ont plusieurs caractéristiques liées à leur origine :

- La composition en nucléotides des îlots est plus proche de celle de la cellule d'origine que de la cellule hôte, donc leur composition en bases est différente de celle du reste du génome ;
- Ils sont peu conservés entre génomes proches, et forment donc des trous dans la synténie ;
- La phylogénie des gènes d'îlots est différente de la phylogénie verticale de l'hôte ;
- Les gènes d'îlots sont peu représentés dans les génomes du taxon, et sont donc peu recrutés dans les jeux de données de métagénomique.

Ces caractéristiques permettent d'identifier les îlots génomiques (la figure présente deux exemples d'îlots identifiés chez *Synechococcus* et *Prochlorococcus*). Elles peuvent être plus ou moins visibles en fonction de la distance phylogénétique séparant l'organisme d'origine et l'organisme cible, et en fonction du temps de résidence des gènes dans le génome : plus les gènes résident longtemps dans le génome, plus leur composition en nucléotides s'homogénéisera avec celle du génome, et s'ils se répandent dans la population, ils seront plus fortement représentés dans les données de métagénomique.



Ilots génomiques de picocyanobactéries marines. A. Ilots génomiques détectés chez *Synechococcus* WH8102 par la déviation de la signature tétranucléotidique. La conservation des gènes au sein des génomes de picocyanobactéries et leur représentation dans les métagénomes du Global Ocean Sampling sont aussi représentées. Figure issue de (Dufresne *et al.*, 2008) B. Ilots détectés chez *Prochlorococcus* MIT9312 par identification de « trous » dans la synténie par comparaison avec la souche MED4. La présence d'éléments répétés et la représentation dans les métagénomes sont également représentées. Figure issue de (Coleman *et al.*, 2006). Noter les différents degrés de conservation des différents îlots.

Les gènes transférés horizontalement ne suivent en général pas la phylogénie verticale de l'organisme étudié (l'incohérence entre la phylogénie de référence de l'organisme et celle du gène est d'ailleurs une des méthodes utilisées pour détecter les THG), ce qui remet en question la possibilité d'établir une phylogénie verticale. Certains auteurs considèrent donc que la fréquence des THG est telle que les relations entre bactéries ne peuvent pas être représentées par un arbre dichotomique, mais devraient plutôt être représentées sous forme de réseau dans lequel il est difficile de définir des entités génétiques distinctes (Doolittle, 1999; Gogarten *et al.*, 2002; Doolittle, 2012).

Si le nombre élevé de transferts horizontaux n'est généralement pas remis en cause (voir p. ex. (Ochman *et al.*, 2005)), le taux de transfert dépend des organismes et de leur mode de vie (quelques exemples sont décrits dans la revue (Loman and Pallen, 2015)). Bien que du fait des THG, on ne puisse pas définir avec certitude la phylogénie verticale d'un organisme sur la base d'un seul gène, les études basées sur l'utilisation de plusieurs gènes concaténés ont révélé l'existence de clusters génétiques distincts (voir p. ex.(Ochman *et al.*, 2005; Lerat *et al.*, 2005; Hanage *et al.*, 2006). (Daubin *et al.*, 2003) ont notamment montré que la plupart des gènes *core* sont en accord avec une phylogénie verticale unique, et les arbres phylogénétiques basés sur l'alignement de l'ensemble de ces gènes donnent donc probablement la meilleure estimation de la phylogénie verticale. De façon plus générale, l'identité nucléotidique moyenne des gènes (ou des fragments de génomes) pouvant être alignés entre deux génomes (ANI pour *Average Nucleotide Identity*) semble permettre de définir des entités génétiques distinctes : les espèces définies historiquement (souvent sur la base de critères physiologiques) ont en général une ANI supérieure à 94% d'identité (Konstantinidis and Tiedje, 2005a) et ce seuil semble correspondre également au critère d'hybridation de l'ADN utilisé pour la définition d'espèces (Goris *et al.*, 2007) (Figure 18).

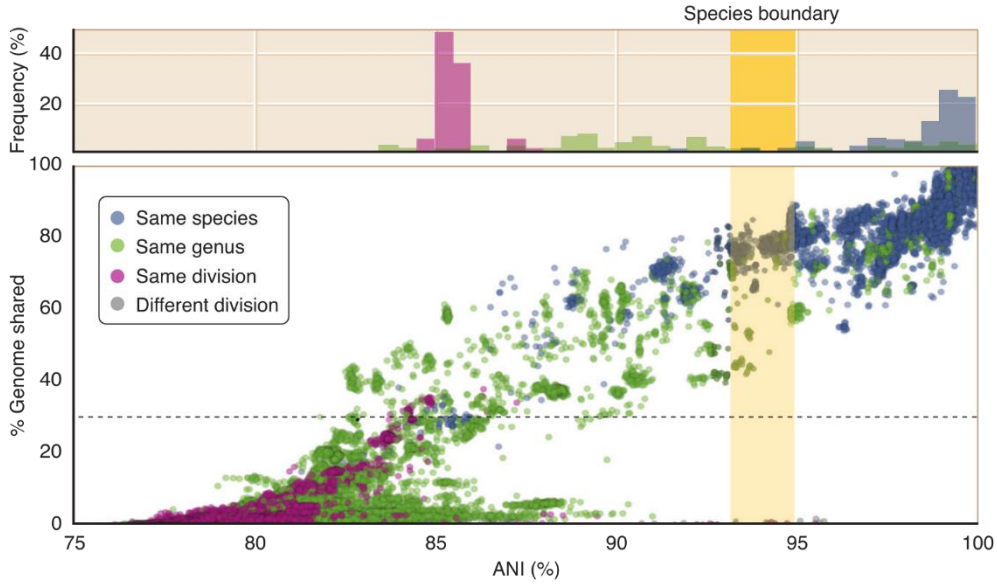


Figure 18 Définition d'espèce sur la base de l'identité nucléotidique moyenne (ANI). Les paires de génomes présentant plus de 94% d'ANI font en général partie de la même espèce décrite. Au contraire sous ce seuil, les génomes appartiennent en général à des espèces distinctes. Source : (Rodriguez-r and Konstantinidis, 2014)

La plupart, si ce n'est l'ensemble, de ces considérations sont fondées essentiellement sur l'étude de génomes de bactéries présentant un intérêt particulier pour l'homme, soit parce qu'elles sont commensales et/ou pathogènes, soit parce qu'elles présentent un intérêt industriel (*Lactobacillus*, pathogènes de plantes comme certains *Pseudomonas*…). Au sein de ces genres de bactéries, plusieurs dizaines de génomes ont été complètement séquencés, et jusqu'à plusieurs milliers si sont pris en compte les génomes « partiels » (c.-à-d. dont l'assemblage n'est pas complet) (Figure 19). L'essor des technologies de séquençage à haut-débit et la diminution des coûts associés a en effet permis de démocratiser l'accès au séquençage, souvent au détriment de la qualité du génome (Loman and Pallen, 2015).

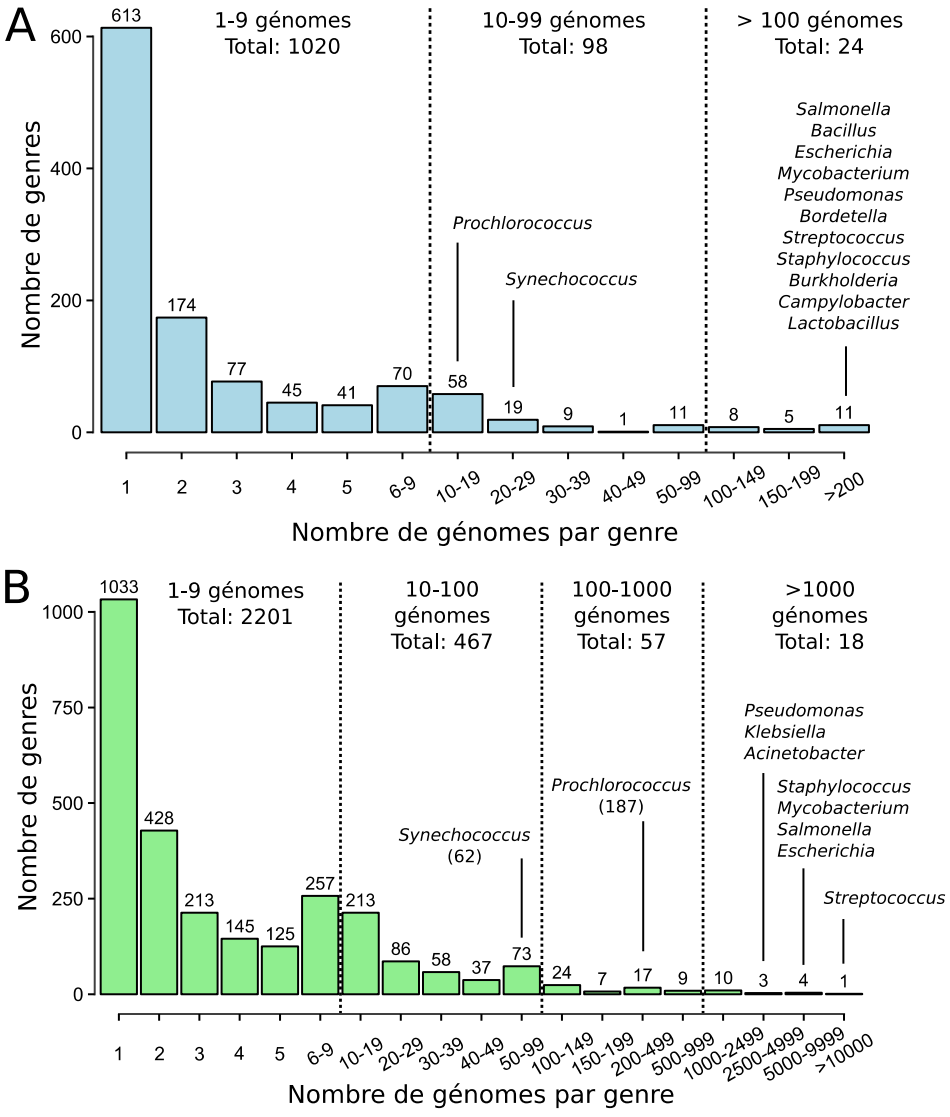


Figure 19 Répartition des génomes séquencés dans la base de données NCBI. Pour chaque catégorie de nombre de génomes séquencés, le nombre de genres de bactéries est indiqué. A. Seuls les génomes complets ou chromosomes complets sont considérés (total : 9625 génomes). B. Sans restriction de qualité (total : 104663 génomes). Les genres les plus séquencés sont indiqués, ainsi que la position de *Synechococcus* et *Prochlorococcus*. Noter que la classification par genre ne distingue pas les *Synechococcus* marins des *Synechococcus* d'eau douce. Données téléchargées sur le site du NCBI le 24/08/2017.

Si les genres de bactéries purement environnementaux sont moins bien représentés dans les bases de données, ils ont néanmoins profité de cette démocratisation, et les bactéries marines (dont les picocyanobactéries) figurent en bonne place : les dernières études de comparaison de génomes complets de picocyanobactéries portaient sur 12 génomes de *Prochlorococcus* (Kettler *et al.*, 2007) et 11 génomes de *Synechococcus* (Dufresne *et al.*, 2008). Depuis ces travaux, le nombre de génomes a encore augmenté pour atteindre 43 génomes de *Prochlorococcus* (auxquels s'ajoutent de nombreux génomes partiels issus du séquençage de cellules uniques (Kashtan *et al.*, 2014)) et 54 génomes de *Synechococcus* marins dont une partie ne sont pas encore publié et ont été analysés au cours de cette thèse. Les picocyanobactéries marines constituent donc des modèles de choix pour étudier les questions touchant à l'évolution en milieu naturel et à l'adaptation à l'environnement, processus qui suivent vraisemblablement des modalités différentes de celles observées chez les organismes commensaux ou pathogènes.

2.3.2 Caractéristiques générales des génomes de picocyanobactéries marines

La comparaison des génomes de *Prochlorococcus* entre eux et avec les génomes de *Synechococcus* a révélé la réduction drastique du génome de *Prochlorococcus* au cours de l'évolution. Cette réduction progressive de la taille du génome et du nombre de gènes a atteint différents stades selon les clades : les génomes du clade LLIV, phylogénétiquement le plus proche des *Synechococcus*, a une taille équivalente à ceux de *Synechococcus*, ce qui prête à penser que l'ancêtre commun à tous les *Prochlorococcus* possédait un génome de taille similaire (Rocap *et al.*, 2003; Dufresne *et al.*, 2005). Le génome des clades LLII/III et LLI est plus petit, et la réduction est maximale pour les souches des clades HLI et HLII (Figure 20). Cette diminution de la taille du génome s'est faite en parallèle d'une diminution du %GC, qui s'explique d'un point de vue mécanistique par la perte de gènes de réparation de l'ADN empêchant les transversions d'une paire de bases G:C vers une paire A:T (Rocap *et al.*, 2003; Dufresne *et al.*, 2005). De telles réductions génomiques ont également été observées chez des pathogènes obligatoires et des bactéries endosymbiontes, le cas le plus extrême étant celui des organelles des cellules Eucaryotes (Ochman, 2001). Chez ces organismes très fortement dépendants de leur hôte, la réduction génomique a été expliquée par la faible taille efficace de la population qui augmente l'effet de la dérive, et par une sélection purifiante relâchée du fait de la stabilité du milieu de vie (au sein de l'organisme hôte). Ces deux arguments ne peuvent pas être appliqués à *Prochlorococcus* dont la taille de population est extrêmement large, et dont la taille efficace de population est probablement également très élevée (Dufresne *et al.*, 2005; Flombaum *et al.*, 2013), et sur le génome duquel la sélection purifiante ne semble pas être relâchée (Dufresne *et al.*, 2005). Au contraire, la réduction génomique de *Prochlorococcus* s'expliquerait par l'avantage évolutif que représente un petit génome dans les régions océaniques oligotrophes : un petit génome riche en bases A et T nécessite moins d'azote et de phosphate, ce qui peut s'avérer avantageux dans les régions pauvres en nutriments. Par ailleurs les codons comprenant plus de A et de T codent pour des acides aminés plus pauvres en azote, ce qui permet de réduire d'autant plus la quantité d'azote nécessaire aux cellules (Dufresne *et al.*, 2005). Cette optimisation de

l'utilisation des nutriments serait encore plus poussée, puisque *Prochlorococcus* serait capable de réduire la taille de certaines de ses protéines en cas de limitation en azote, en utilisant des sites d'initiation de la transcription alternatifs (Read *et al.*, 2017). Par ailleurs le fait de posséder un petit génome permet de réduire la taille des cellules et d'augmenter leur rapport surface/volume, ce qui favorise l'assimilation des nutriments. Ces arguments sont surtout valables pour les clades HL, qui dominent dans un environnement pauvre en nutriment, mais moins pour les clades LL, qui dominent en profondeur où la concentration en nutriments est plus élevée. Dans leur cas, l'hypothèse que la diminution de la taille permet d'augmenter l'efficacité de la photosynthèse a été avancée (Dufresne *et al.*, 2005). De façon remarquable, une autre bactérie planctonique marine présente la même réduction génomique : SAR11, la bactérie hétérotrophe la plus abondante dans les eaux de surface (dont les caractéristiques sont détaillées dans la revue récente de Giovannoni (Giovannoni, 2017)). Les cellules de SAR11 sont retrouvées dans les mêmes environnements que *Prochlorococcus* et ont également optimisé leurs besoins en nutriments, et les cas de réduction génomique pourraient être relativement répandus dans l'environnement (Swan *et al.*, 2013; Giovannoni *et al.*, 2014). A la sélection pour l'utilisation optimale des nutriments s'ajoute une deuxième explication, plus récente, pour expliquer la perte de gènes chez ces organismes. (Morris *et al.*, 2012) ont exposé une théorie, dite de la « Reine Noire » (*Black Queen Hypothesis*[8], BQH), selon laquelle un des moyens pour un organisme d'optimiser sa consommation en nutriments serait d'arrêter de produire certains composés, et d'utiliser à la place ceux produits par les bactéries voisines et directement présents dans l'environnement. Selon cette théorie, les différentes populations bactériennes sont donc largement interdépendantes au niveau métabolique. Si elle ne permet pas d'expliquer l'ensemble des pertes de gènes (Giovannoni *et al.*, 2014), elle constitue une explication élégante pour expliquer certaines pertes (Morris *et al.*, 2012), et ses implications sont actuellement explorées notamment à travers les interactions entre *Prochlorococcus* et SAR11 (Braakman *et al.*, 2017).

[8] Le nom de cette théorie fait écho avec la théorie de la « Reine Rouge » (van Valen, 1973), qui désigne la coévolution de deux organismes par une forme de « course aux armements », par exemple entre un parasite et son hôte. Le nom de « Reine Noire » a quant à lui été choisi en référence au jeu de cartes « la dame de pique », dans laquelle la stratégie consiste à éviter cette carte (Morris *et al.*, 2012).

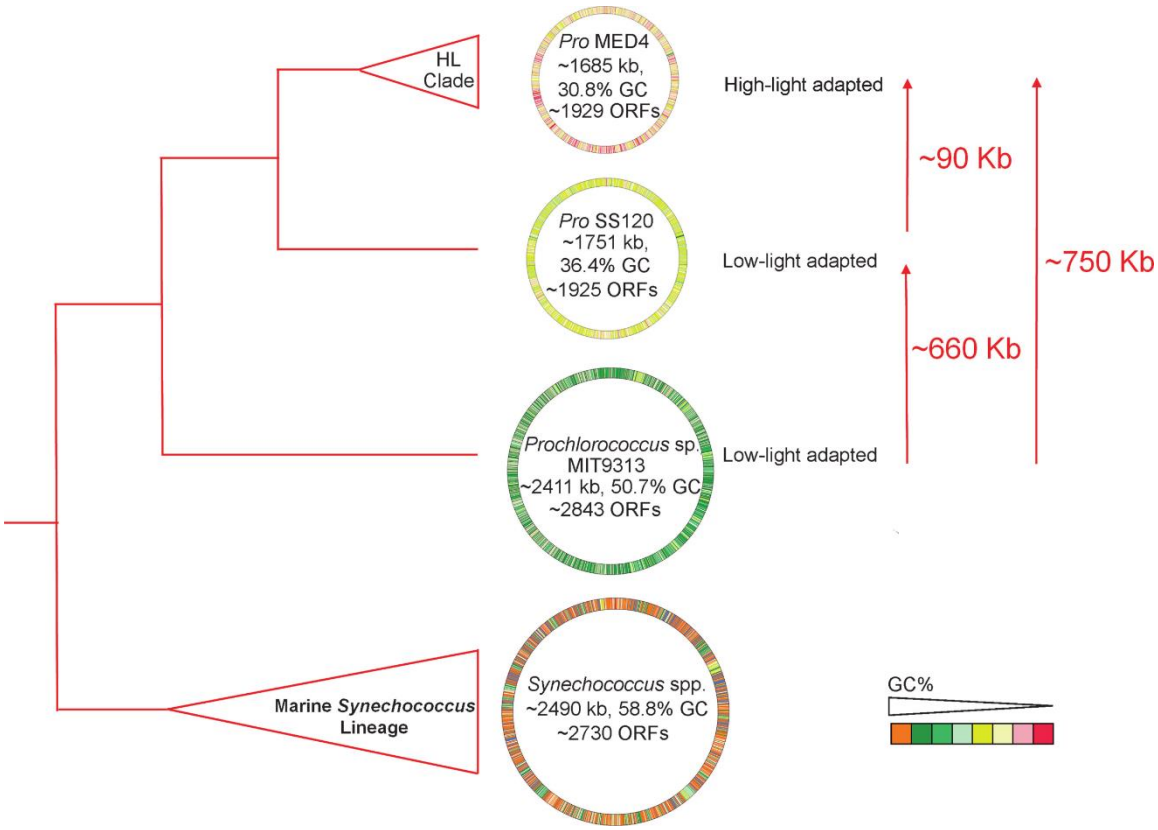


Figure 20 Représentation schématique de la réduction génomique observée chez *Prochlorococcus.* La réduction de la taille du génome et la diminution du pourcentage des bases G et C sont représentées. Figure issue de (Scanlan, 2012).

En dehors de ces considérations spécifiques à *Prochlorococcus,* les deux genres de picocyanobactéries marines partagent un certain nombre de caractéristiques génomiques. Leur génomes sont petits (par rapport à la moyenne des bactéries), et contiennent peu de paralogues (c.-à-d. de gènes présents en plusieurs copies suite à des duplications), mais possèdent un ensemble commun de gènes nécessaires à leur mode de vie autotrophe (Scanlan *et al.*, 2009). La phylogénie basée sur la concaténation des gènes *core* (communs à tous les génomes de picocyanobactéries séquencés) est en accord avec celle basée sur des gènes marqueurs (voir ci-dessus), et est souvent mieux résolue, bien que restreinte aux souches dont le génome est séquencé (Dufresne *et al.*, 2008).

De façon similaire aux autres procaryotes, le contenu en gènes des génomes de *Prochlorococcus* et *Synechococcus* est très variable entre souches, ce qui implique de nombreux gains et pertes de gènes au cours de l'évolution (Kettler *et al.*, 2007). Les gènes accessoires et/ou uniques ainsi que les gènes récemment gagnés sont souvent regroupés dans certaines régions du génome, qualifiées d' « îlots génomiques » (Kettler *et al.*, 2007; Dufresne *et al.*, 2008; voir l'encart Transferts horizontaux de gènes et îlots génomiques

). Ces îlots ont été initialement identifiés en comparant la synténie (c.-à-d. l'ordre des gènes dans le génome) entre souches de *Prochlorococcus* (Coleman *et al.*, 2006), une méthode qui s'applique plus difficilement à *Synechococcus*, chez qui ils ont été identifiées par leur composition spécifique en nucléotides (Dufresne *et al.*, 2008). Ces régions recrutent peu de *reads* issus de l'environnement, ce qui confirme la variabilité de leur contenu en gènes. Le fait que les différents

clades de picocyanobactéries présentent des préférences écologiques distinctes et la variabilité du contenu en gènes ont amené à chercher à relier le phénotype (mesuré au laboratoire, ou par le biais de la niche environnementale occupée) au génotype des différentes souches.

2.3.3 Variabilité génétique et adaptation à l'environnement : quelques exemples

Le meilleur moyen de vérifier le lien entre un gène et un phénotype est de muter le gène et d'observer le phénotype obtenu. Cependant la mutagénèse n'est pas maîtrisée chez *Prochlorococcus,* et bien qu'une cinquantaine de mutants aient été obtenus chez *Synechococcus*, 40 d'entre eux concernent une seule souche (WH8102, clade III) et son application est loin d'être routinière. La disponibilité de génomes séquencés permet cependant de progresser dans l'étude des gènes de picocyanobactéries, à travers des études de génomique fonctionnelle (par exemple en étudiant l'expression de l'ensemble des gènes du génome dans des conditions spécifiques) ou par comparaison génomique entre des souches dont l'écologie est relativement bien connue. Nous donnerons ici quelques exemples de cas où la comparaison des génomes de *Synechococcus* et *Prochlorococcus* a permis de mettre en évidence le lien entre l'adaptation aux conditions environnementales et certains gènes ou régions génomiques.

Chez *Prochlorococcus,* les génomes de souche HL contiennent moins de copies du gène codant pour la protéine de l'antenne collectrice de lumière Pcb que les souches LL, mais plus de gènes encodant les protéines HLIP (High Light Inducible Protein) qui pourraient intervenir dans la dissipation de l'excès d'énergie lumineuse (voir aussi partie 3.1 (He *et al.*, 2001; Havaux *et al.*, 2003)), et un ou plusieurs gènes reliés aux photolyases, qui pourraient intervenir dans la réparation des dommages de l'ADN liés à l'exposition aux UV (Rocap *et al.*, 2003; Coleman and Chisholm, 2007; Partensky and Garczarek, 2010) (Figure 21). Les souches du clade LLI, dont la position est intermédiaire dans la colonne d'eau, présentent quant à elles un contenu en gène intermédiaire. Ainsi le contenu en gène explique au moins en partie la différenciation écologique des populations HL et LL. A l'inverse, les gènes d'assimilation des macronutriments sont répartis dans les génomes de façon relativement indépendante de la phylogénie (Figure 21), et leur présence dans une souche semble plutôt reliée à son lieu d'isolement (Martiny *et al.*, 2006; Coleman and Chisholm, 2007; Martiny *et al.*, 2009b). Leur répartition dans les génomes suggère une acquisition par transfert horizontal, d'autant plus que certains de ces gènes sont retrouvés dans des îlots génomiques (Coleman *et al.*, 2006; Kettler *et al.*, 2007), et il semble donc qu'ils soient impliqués dans une adaptation récente à un environnement local. Les îlots génomiques décrits chez *Prochlorococcus* contiennent par ailleurs certains gènes de réponse au stress, ainsi que de nombreux gènes impliqués dans la modification de la surface cellulaire, dont le rôle probable est de protéger les cellules contre le broutage et les infections virales (Rocap *et al.*, 2003; Coleman *et al.*, 2006; Coleman and Chisholm, 2007; Kettler *et al.*, 2007), ce qui montre l'importance de THG dans ce domaine. Les prédateurs sont en effet considérés comme une force évolutive majeure des génomes de

Prochlorococcus, maintenant la diversité à la fois par une course aux armements[9] entre prédateurs et proie et par sélection fréquence-dépendante (Avrani *et al.*, 2011).

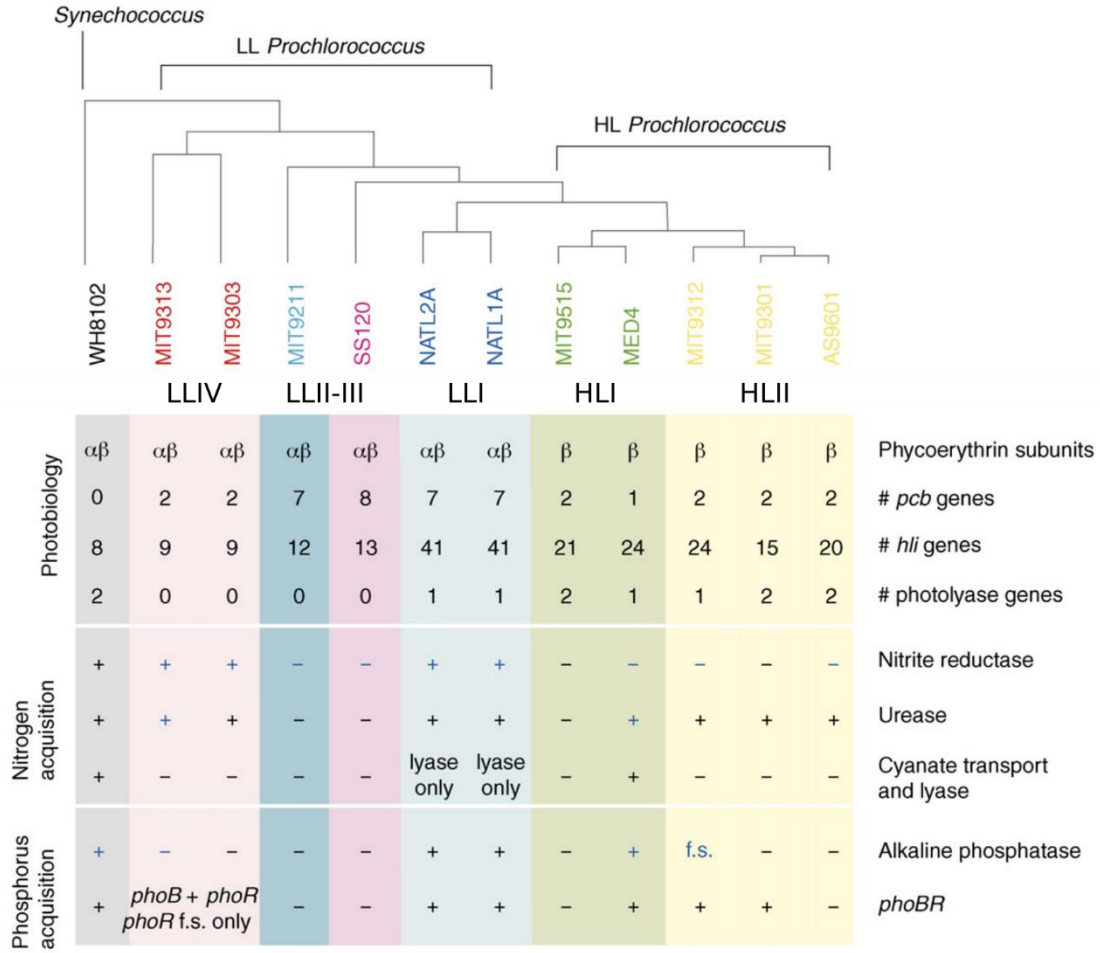


Figure 21 Répartition au sein des souches de *Prochlorococcus* de gènes liés à l'adaptation aux conditions d'intensité lumineuse et de concentration en nutriments. Les gènes *pcb* sont impliqués dans l'absotrption de l'énergie lumineuse, et les gènes *hli* et de photolyase protègent les cellules des dommages causés par l'excès d'énergie lumineuse et d'UV. Les gènes indiqués dans les deux cadres inférieurs sont impliqués dans l'utilisation de différentes sources d'azote et de phosphore. Source : (Coleman and Chisholm, 2007).

Chez *Synechococcus*, la comparaison génomique a révélé des différences plus globales entre souches isolées au large et souches d'origine côtière : les premières, vivant dans un milieu relativement stable, semblent posséder moins de gènes impliqués dans la régulation (systèmes à deux composants et facteurs sigma), de façon assez similaire à ce qui est observé chez *Prochlorococcus* (Palenik *et al.*, 2003; Rocap *et al.*, 2003; Palenik *et al.*, 2006; Dufresne *et al.*, 2008; Scanlan *et al.*, 2009). Cette perte de systèmes de régulation qui pourrait être compensée au moins en partie par le développement de la régulation par les petits ARN (Steglich *et al.*, 2008; Gierga *et al.*, 2012; Kopf and Hess, 2015). Les souches côtières semblent également posséder plus

[9] La course aux armements désigne la coévolution de deux organismes au cours de laquelle de nouvelles « armes » sont constamment sélectionnées chez un parasite pour contrer les « défenses » sélectionnées en retour chez leurs hôtes. Voir aussi la note précédente.

de gènes de transport et de stockage de métaux (Fe, Cu), en lien avec un plus grand nombre de métalloenzymes, ce qui pourrait être relié à la plus forte variabilité de la disponibilité de ces éléments en milieu côtier (Palenik *et al.*, 2006; Dufresne *et al.*, 2008). Enfin les souches issues de zones peu salées (clade VIII et subcluster 5.2) semblent partager un grand nombre de gènes malgré la distance phylogénétique qui les sépare, mais la plupart de ces gènes ne sont pas caractérisés.

Comme *Prochlorococcus, Synechococcus* possède un certain nombre d'îlots génomiques, dont les gènes sont impliqués (pour ceux dont la fonction est connue) essentiellement dans la modification de la surface cellulaire, ainsi que dans le transport de certains nutriments (Palenik *et al.*, 2003, 2006; Dufresne *et al.*, 2008). Ces îlots présentent des degrés de conservation variables, ce qui suggère des origines plus ou moins anciennes (Dufresne *et al.*, 2008). L'exemple le mieux documenté est un îlot très conservé chez *Synechcococcus* au sein duquel se trouvent une grande partie de gènes codant pour les protéines impliquées dans la synthèse des bras des phycobilisomes et sa régulation (Six *et al.*, 2007c; Dufresne *et al.*, 2008). La région est de taille variable, et son contenu en gène explique l'existence des différents types pigmentaires décrits précédemment (Figure 22). Un autre îlot, beaucoup plus petit, et existant sous deux formes différentes, est responsable de l'acclimatation chromatique (et donc du dernier type pigmentaire, 3d) (Humily *et al.*, 2013). Cet exemple illustre bien le lien entre phénotype, génotype et écologie, les différents types pigmentaires se répartissant en fonction de la qualité (couleur) de la lumière trouvée dans l'environnement.

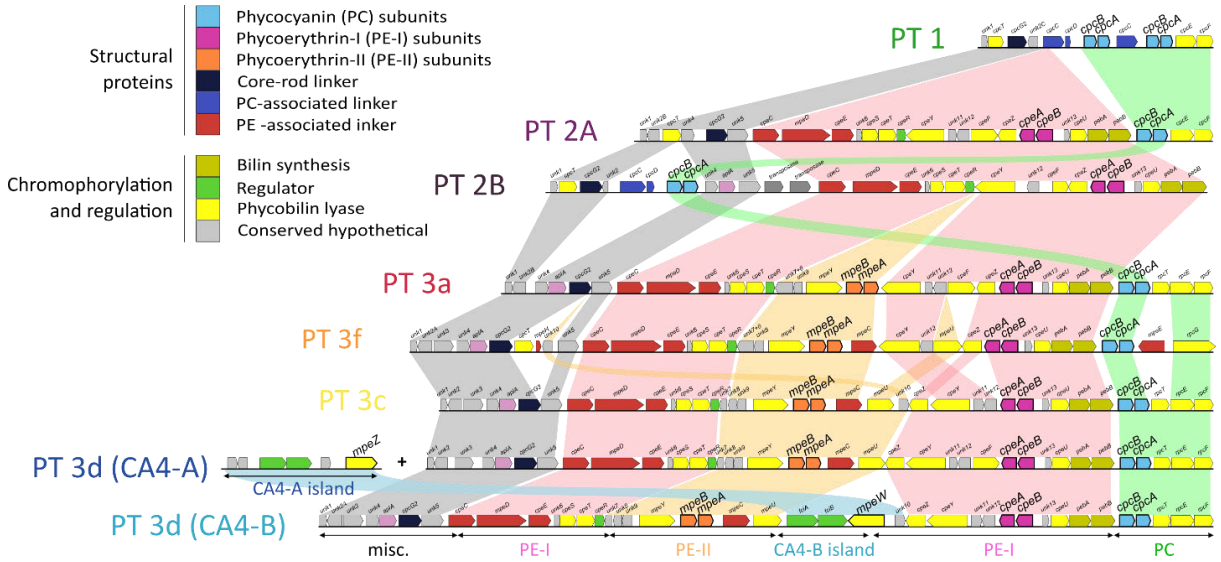


Figure 22 Organisation génomique de la région codant pour les bras des phycobilisomes chez *Synechococcus*. Cette région correspond à un îlot génomique. La région s'est complexifiée du type pigmentaire 1 (PT1) au type pigmentaire 3d (PT 3d) par l'acquisition de blocs de gènes au sein de l'îlot. Les rubans de couleurs permettent de suivre la synténie des différents blocs de gènes. Noter la présence chez les types pigmentaires 3d d'un îlot supplémentaire qui peut prendre deux formes, CA4-A et CA4-B. L'îlot CA4-B est inséré dans la région phycobilisome, alors que l'îlot CA4-A est en général localisé ailleurs dans le génome. Le type pigmentaire 2B correspond à une organisation particulière retrouvée uniquement dans des séquences environnementales (Larsson *et al.*, 2014), et le type 3f présente une organisation le différenciant du type 3c bien qu'aucune différence phénotypique n'ait pour l'instant été identifiée. PT : type pigmentaire. Figure réalisée par Théophile Grébert.

En plus de ces îlots génomiques, *Synechococccus* présente la particularité de posséder des « gènes géants » qui montrent également les caractéristiques de THG, dont la fonction est inconnue

en-dehors d'un d'entre eux, *swmB*, impliqué dans la motilité très particulière de certaines souches du clade III (Dufresne *et al.*, 2008).

Pour *Synechococcus* comme pour *Prochlorococcus,* les phages[10] ont joué, et jouent probablement encore, un rôle important dans la structuration des génomes. Outre la pression sélective évoquée ci-dessus, ils servent de vecteurs à la transmission de gènes, comme le révèlent à la fois la présence de gènes d'origine phagique dans les génomes de picocyanobactéries (Palenik *et al.*, 2003; Rocap *et al.*, 2003; Lindell *et al.*, 2004; Dufresne *et al.*, 2008) et la présence de gènes de picocyanobactéries dans les génomes de phage (Lindell *et al.*, 2004; Sullivan *et al.*, 2005). En plus des phages, une étude récente a révélé que les picocyanobactéries marines produisent des vésicules extracellulaires qui contiennent entre autres de l'ADN, et dont l'un des rôles pourrait être de favoriser les THG (Biller *et al.*, 2014b).

Si ces quelques exemples donnent une première idée des bases génétiques de l'adaptation à des niches différentes, notre compréhension de cette adaptation est encore incomplète. Le croisement des données de génomique (qui servent de référence) avec les données de métagénomique est utile dans ce contexte, parce qu'il donne accès à la répartition des gènes dans les populations naturelles, y compris celles non cultivées. Cela a par exemple permis de confirmer que les populations de *Prochlorococcus* possédant le plus de gènes d'assimilation du phosphate se trouvent dans les régions où celui-ci est faiblement disponible (Martiny *et al.*, 2006, 2009b), de cibler les populations de *Prochlorococcus* HL capables d'assimiler le nitrate afin de les isoler (Berube *et al.*, 2014), ou de mettre en évidence l'importance de la mixotrophie chez les picocyanobactéries marines (Yelton *et al.*, 2016). Plus récemment, les données de métagénomique ont été utilisées pour relier directement les gènes aux conditions environnementales, s'affranchissant ainsi de la relation entre gènes et populations (Kent *et al.*, 2016). Ces approches n'en sont qu'à leurs débuts, mais sont prometteuses dans l'objectif de relier le génotype des différentes populations à la niche qu'elles occupent dans l'environnement.

Si l'étude du contenu en gène est importante pour mieux cerner l'écologie des picocyanobactéries, il ne faut pas négliger le fait que chacun de ces gènes n'est qu'un élément du métabolisme cellulaire, et que leur expression est intégrée au niveau de l'organisme. Il est donc tout aussi important d'analyser la régulation de ces gènes pour comprendre comment les cellules font face aux variations des conditions environnementales, et dans quelles circonstances elles peuvent survivre. Ce sont ces aspects physiologiques qui seront évoqués dans la troisième et dernière partie de cette introduction.

[10] Le terme de phage est utilisé pour désigner les bactériophages, c.-à-d. les virus infectant des bactéries. Le terme plus spécifique de cyanophage est employé pour désigner les phages infectant des cyanobactéries.

3. Les principales adaptations physiologiques des picocyanobactéries marines à leur environnement

3.1 La photosynthèse, processus au cœur de la vie des picocyanobactéries

3.1.1 Les préférences des picocyanobactéries en termes de quantité de lumière

La très large répartition des picocyanobactéries marines, tant verticale qu'horizontale, que nous avons détaillée dans la partie précédente, a d'emblée interpellé les océanographes. Il a en effet été noté dès leur découverte que les picocyanobactéries marines étaient capables de se maintenir à des concentrations cellulaires élevées dans l'ensemble de la zone euphotique, c'est-à-dire dès lors que l'intensité lumineuse disponible pour la photosynthèse est suffisante (Waterbury *et al.*, 1986). Dès 1987, Kana et Glibert (Kana and Glibert, 1987) ont étudié en détail la photo-physiologie d'une souche océanique de *Synechococcus*, WH7803, soumise à des intensités lumineuses de 30 à 2000 µmol photons.m^{-2}.s^{-1} en lumière continue. Contrairement aux études qui l'ont précédée, celle-ci montre qu'en acclimatant progressivement la souche à des intensités lumineuses croissantes (en multipliant par deux ou trois l'intensité entre les paliers d'acclimatation et en laissant 3 à 4 jours d'acclimatation à chaque palier), WH7803 est capable de maintenir son taux de croissance à une valeur quasi-maximale de 1,87 $jour^{-1}$ même à 2000 µmol photons.m-2.s-1 (Figure 23A). La courbe exprimant le taux de croissance en fonction de l'intensité lumineuse sature cependant à partir d'une intensité lumineuse (Isat) de 200 µmol photons.m^{-2}.s^{-1}. Des résultats similaires ont été obtenus sur la souche WH8103 (en alternance jour/nuit 14h/10h) et la souche WH8102 (en lumière continue), bien que ces études se limitent à une intensité lumineuse maximale de 450 et 650 µmol photons.m^{-2}.s^{-1}, respectivement (Moore *et al.*, 1995; Six *et al.*, 2004), et que celle menée sur WH8102 mentionne un échec de la culture à 1000 µmol photons.m^{-2}.s^{-1}. Dans ces trois cas, l'intensité lumineuse minimale nécessaire pour observer une croissance cellulaire est très faible, entre 2 et 10 µmol photons.m^{-2}.s^{-1} (Six *et al.*, 2004). Les *Synechococcus* marins étudiés montrent donc une capacité d'acclimatation à une très large gamme d'intensités lumineuses, bien que le faible nombre de souches étudiées jusqu'ici limite notre connaissance de la variabilité de cette capacité d'acclimatation entre souches et de son lien éventuel avec la diversité pigmentaire et/ou génétique du genre.

Chez *Prochlorococcus,* les premières études de photo-physiologie ont montré des différences entre souches avant même la mise en évidence de l'existence de populations de profondeurs différant des populations de surface (Partensky *et al.*, 1993). Parmi les premières souches isolées, les souches SS120 et MED4 ont révélé des différences dans leurs capacités d'acclimatation à l'intensité lumineuse : SS120 est capable de croître à une intensité lumineuse plus faible que MED4, et voit son taux de croissance diminuer à une intensité lumineuse optimale pour la croissance de MED4 (Moore *et al.*, 1995) (Figure 23B). Ces différences et leur lien avec la

profondeur d'isolement ont été confirmés par l'isolement de plusieurs souches low-light et high-light différenciées sur la base de leur pigmentation et l'étude de leurs capacités d'acclimatation (Figure 23C) (Moore *et al.*, 1998; Moore and Chisholm, 1999).

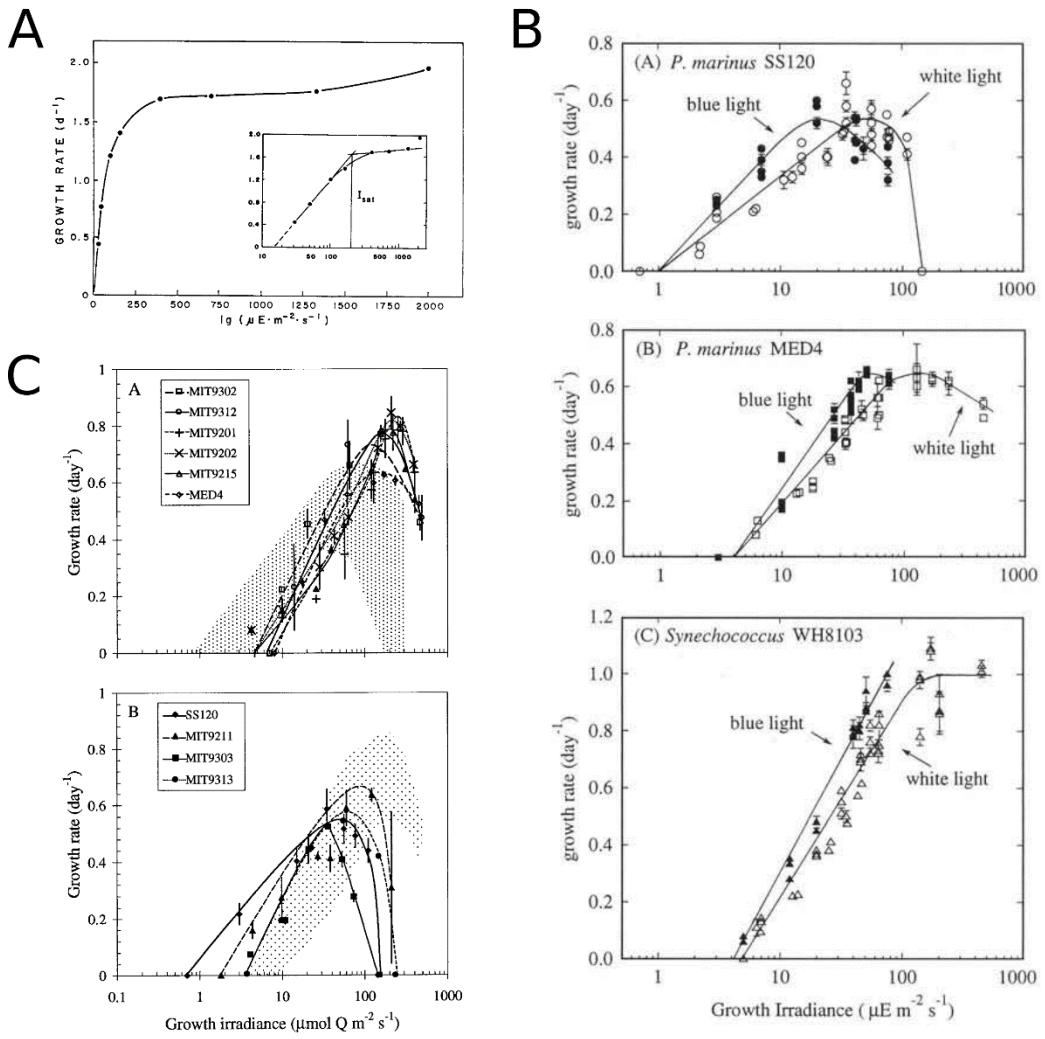


Figure 23 Taux de croissance en fonction de l'intensité lumineuse pour différentes souches de *Synechococcus* et *Prochlorococcus*. A. Figure issue de (Waterbury et al., 1986) pour WH7803. B. Figure issue de (Moore et al., 1995), pour les souches SS120 (*Prochlorococcus* LLII, panneau supérieur), MED4 (*Prochlorococcus* HLI, panneau central) et *Synechococcus* WH8103, panneau inférieur. C. Figure issue de (Moore and Chisholm, 1999), pour différentes souches de *Prochlorococcus* HL (panneau supérieur) et LL (panneau inférieur). Ces différents résultats montrent les capacités variables de photoacclimatation chez les picocyanobactéries marines.

Synechococcus et *Prochlorococcus* sont donc capables de s'acclimater à l'intensité lumineuse ambiante, mais cette capacité est plus limitée chez *Prochlorococcus* qui supporte une gamme plus faible d'intensités lumineuses. Chez ce dernier, la colonisation de l'ensemble de la zone euphotique s'est faite par la spécialisation de certaines populations dans des gammes d'intensité lumineuse différentes (photoadaptation), tandis que *Synechococcus* montre une capacité d'acclimatation plus large, sans qu'une spécialisation de certaines populations aux eaux de surface ou aux eaux plus profondes n'ait été clairement mise en évidence (voir la partie 1.3.2 concernant les préférences en termes de qualité de lumière pour ce genre).

3.1.2 Effets de l'excès de lumière sur la physiologie des picocyanobactéries marines : la photoinhibition

La diminution du taux de croissance au-delà d'une certaine intensité lumineuse chez *Prochlorococcus* (et sa stabilisation chez *Synechococcus*), visible sur la Figure 23, met en évidence l'existence d'une photo-inhibition chez ces organismes, c.-à.-d. d'une limitation de l'efficacité de la photosynthèse par une forte intensité lumineuse. L'inhibition de la photosynthèse par la lumière elle-même est connue depuis longtemps chez l'ensemble des organismes effectuant la photosynthèse oxygénique (plantes, algues vertes, cyanobactéries…) et un certain nombre de mécanismes ont été avancés pour expliquer ce phénomène, que nous résumerons ici rapidement.

La photoinhibition est due principalement à la dégradation du centre réactionnel du photosystème II (RCII) du fait de différents facteurs (Latifi *et al.*, 2009) (Figure 24) : i) la dégradation du complexe manganèse (complexe d'oxydation de l'eau, COE) au sein du centre réactionnel, notamment du fait du faible pH au sein du lumen, qui empêche la réduction de la chlorophylle excitée par la lumière. Cette chlorophylle excitée est très oxydante et peut soit dégrader directement les composants du RCII à proximité et en particulier la protéine D1, soit entrainer la production du radical $^1O_2$, une espèce réactive de l'oxygène (ROS pour Reactive Oxygen Species) ; ii) Une différence d'efficacité entre les PSII et le PSI, ou une limitation de l'oxydation du NADPH en raison d'un ralentissement de la phase obscure de la photosynthèse (cycle de Calvin) entrainant un excès de PSI ou de quinones à l'état réduit, qui réduisent alors le dioxygène en $O_2^{\bullet -}$, transformé ensuite en H_2O_2 et $OH^{\bullet}$, espèces réactives de l'oxygène. Dans les deux cas, les ROS ainsi formées seraient aussi potentiellement responsables de la dégradation du RCII.

La dégradation du RCII le rendant inactif, il est rapidement réparé par le remplacement de la protéine D1 par une D1 nouvellement synthétisée. Cependant, les ROS peuvent également limiter ce remplacement en inhibant la synthèse de D1, ce qui aggrave le phénomène. La photoinhibition devient apparente lorsque le taux de réparation est inférieur au taux de dégradation. (Nishiyama *et al.*, 2006; Murata *et al.*, 2007; Tyystjärvi, 2008).

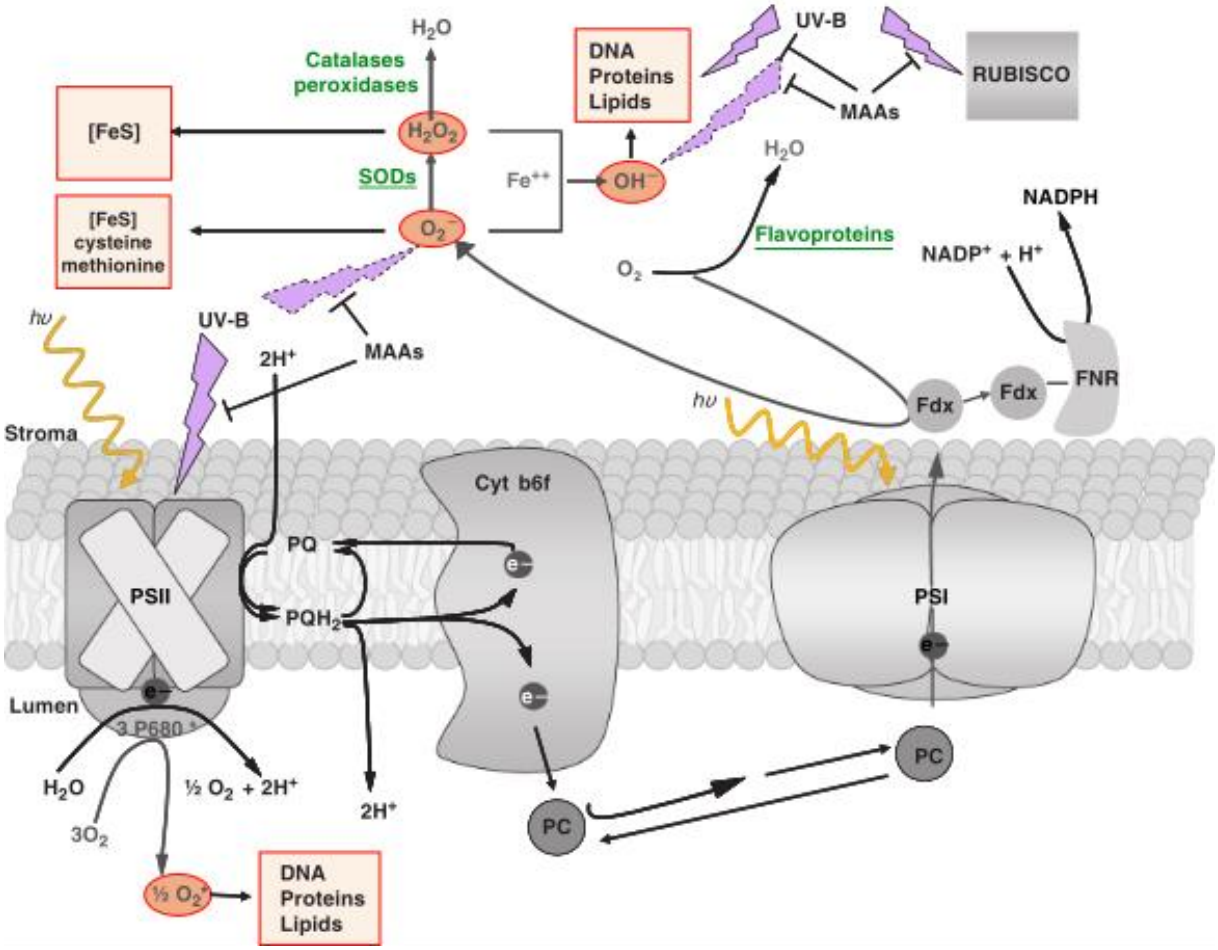


Figure 24 Production d'espèces réactives de l'oxygène (ROS) au niveau de la chaîne de transport d'électrons photosynthétique. Les ROS sont indiquées par un rond rouge, et leurs cibles potentielles sont encadrées en rouge. Les enzymes de détoxification sont indiquées en vert. Les ROS sont produites par réaction de l'oxygène produit issu de la photodissociation de l'eau avec les électrons de la chaine de transport photosynthétique au niveau du PSI et du PSII. Leurs cibles sont nombreuses dans la cellule (protéines Fe-S, acide aminés soufrés, ADN, protéines, lipides…) et leur effet est donc fortement délétère. Figure colorisée à partir de (Latifi *et al.*, 2009).

Les effets d'un excès de lumière sur les picocyanobactéries marines sont en accord avec ce modèle de photoinhibition développé sur d'autres organismes. En effet, les cultures acclimatées à une faible intensité lumineuse et soumises brutalement à une forte intensité lumineuse (stress lumineux) montrent une forte diminution de l'efficacité photosynthétique du PSII, que ce soit chez *Prochlorococcus* ou chez *Synechococcus* (Garczarek *et al.*, 2008; Six *et al.*, 2007a) (Figure 25). De façon intéressante cette dégradation de l'efficacité photosynthétique du PSII dépend à la fois de la souche utilisée et de l'état d'acclimatation des cultures avant le stress, les cultures acclimatées à une intensité lumineuse intermédiaire ou élevée résistant mieux au stress lumineux (Garczarek *et al.*, 2008). *Prochlorococcus* semble plus sensible que *Synechococcus*, en particulier les souches LL qui seraient incapables de retrouver une activité normale du PSII même après un rétablissement de conditions de faible intensité lumineuse, contrairement aux souches HL et aux souches de *Synechococcus* (Six *et al.*, 2007a) (Figure 25). De façon similaire à ce qui est observé au cours de ces stress lumineux artificiels, les cultures acclimatées à un cycle jour/nuit mimant les variations

lumineuses quotidiennes subies dans l'environnement montrent une diminution de l'efficacité du PSII en relation avec l'augmentation de l'intensité lumineuse dans la journée, et ce de façon plus prononcée chez *Prochlorococcus* (Mella-Flores *et al.*, 2012).

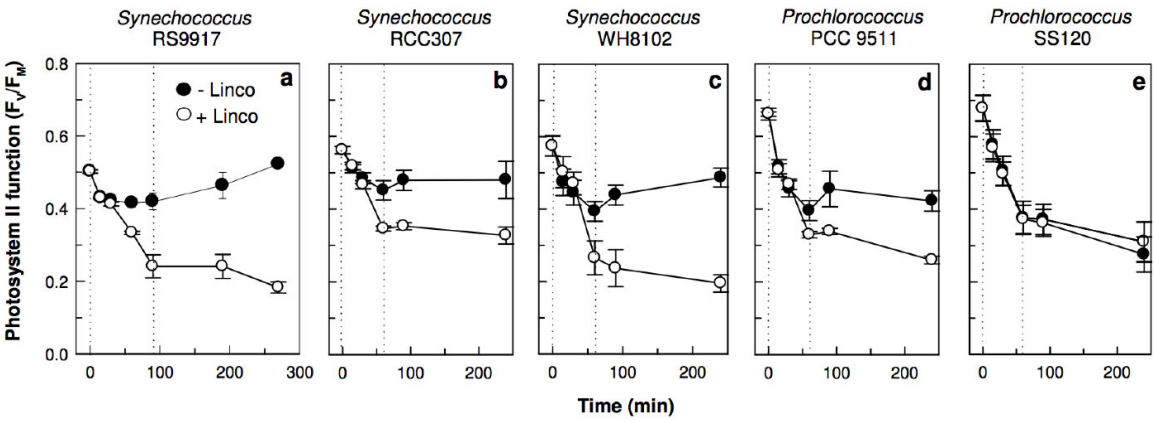


Figure 25 Effet d'un stress de forte intensité lumineuse sur différentes souches de *Prochlorococcus* et *Synechococcus*. Les cultures ont été soumises pendant la période délimitée par les pointillés à une intensité lumineuse 10 fois supérieure à celle utilisée pour leur acclimatation, puis l'intensité lumineuse initiale a été rétablie pour tester la capacité de récupération des cellules après le stress. Les cultures ont été traitées (+ Linco, en blanc) ou non (- Linco, en noir) avec de la lincomycine, inhibiteur de synthèse protéique, au début de l'expérience. La lincomycine empêchant la néosynthèse de protéine D1, les PSII endommagés ne peuvent pas être réparés au cours du traitement + Linco. La différence entre les deux courbes révèle donc la capacité des cellules à réparer leurs PSII endommagés par la synthèse de nouvelles protéines. Figure issue de (Six *et al.*, 2007a).

Dans les deux cas (stress artificiel ou cycle jour/nuit), les taux de dégradation et de réparation de la protéine D1 reflètent bien les variations observées : la quantité cellulaire de D1 augmente fortement lors d'un stress lumineux sur des cultures acclimatées en faible lumière ainsi qu'au court de la journée en cycle jour/nuit. La néo-synthèse de protéine D1 est en général déclenchée rapidement lors du changement de condition, et plus cette réparation est efficace, plus les cellules sont capables de rétablir une physiologie normale après le stress (Six *et al.*, 2007a). De la même façon, les cultures acclimatées en forte lumière montrent constitutivement un taux de réparation élevé de la protéine D1 (Blot *et al.*, 2011), ce qui constitue une des stratégies d'acclimatation aux fortes intensités lumineuses et explique pourquoi elles résistent mieux au stress lumineux.

Comme nous l'avons expliqué plus haut, la dégradation de la protéine D1 est liée à la formation de ROS à proximité des membranes thylakoïdales. Les cellules cultivées à une forte intensité lumineuse produisent plus de ROS, et sont donc plus dépendantes de la réparation de D1. Ceci explique qu'elles soient plus sensibles à l'ajout d'agents oxydants, qui ont un effet inhibiteur sur la réparation de D1 (Blot *et al.*, 2011). De la même façon, lors d'un cycle jour/nuit, l'effet délétère des agents oxydants sur la physiologie cellulaire est plus élevé à midi (Mella-Flores *et al.*, 2012). Enfin, aux effets de l'exposition au spectre visible de la lumière s'ajoutent, en milieu naturel, ceux des radiations UV. L'ajout d'une composante UV au stress lumineux accentue l'effet du stress en diminuant l'efficacité du PSII, et en augmentant le taux de dégradation de la protéine D1 ainsi que la réponse des cellules au stress oxydant (Garczarek *et al.*, 2008; Blot *et al.*, 2011).

3.1.3 Les mécanismes de protection face à l'excès d'énergie lumineuse

L'excès d'énergie lumineuse a donc des effets fortement délétères sur les picocyanobactéries, et celles-ci possèdent un certain nombre de mécanismes de protection pour y faire face. La protéine D1 servant de fusible au niveau du PSII, il n'est pas étonnant que l'un de ces mécanismes soit l'augmentation de son taux de synthèse – quand cela est possible. Si le taux de synthèse est déjà à son maximum et que la dégradation de la protéine D1 ne peut pas être compensée, le stress cellulaire est d'autant plus fort. Outre l'augmentation du taux de synthèse de D1, *Synechococcus* répond également au stress lumineux (forte lumière et/ou UV) en changeant l'isoforme de cette protéine. En effet, *Synechococccus* possède plusieurs copies du gène *psbA* encodant la protéine D1, une copie codant pour l'isoforme D1:1 exprimée en conditions de faible intensité lumineuse, et une ou plusieurs copies codant pour l'isoforme D1:2, exprimée en conditions de forte intensité lumineuse (que ce soit lors d'un stress ou en condition acclimatée) (Garczarek *et al.*, 2008). Il semblerait que l'isoforme D1:1 confère une plus grande efficacité au PSII, alors que l'isoforme D1:2 permet de limiter la production de ROS et donc la dégradation du PSII. Le fait de posséder les deux isoformes permet donc à la fois une forte efficacité photosynthétique quand l'intensité lumineuse est faible, et une meilleure résistance à la photoinhibition quand elle est élevée.

Si la réparation des photosystèmes II par le renouvellement de la protéine D1 est essentielle pour maintenir une activité photosynthétique en cas de photoinhibition, d'autres mécanismes sont également mis en œuvre pour s'acclimater aux variations de l'intensité lumineuse ambiante en évitant la dégradation du PSII. Un premier ensemble de mécanismes permettent de limiter la quantité d'énergie lumineuse captée par la cellule.

Tout d'abord, *Synechococcus* diminue fortement la surface cellulaire de thylakoïdes lorsque l'intensité lumineuse augmente, ce qui entraine la diminution du nombre de phycobilisomes et donc de la quantité d'énergie lumineuse captée (Kana and Glibert, 1987). Ensuite, *Synechococcus* et *Prochlorococcus* diminuent leur concentration cellulaire en pigments photosynthétiques (Figure 26) : les cellules exposées à une forte intensité lumineuse voient leur concentration en chlorophylle *a* et β-carotène (chez *Synechococcus*) ou dv-chlorophylles *a* et *b* et α-carotène (chez *Prochlorococcus*) diminuer (Kana and Glibert, 1987; Moore *et al.*, 1995; Six *et al.*, 2004). Chez *Synechococcus*, la structure des phycobilisomes est également modifiée avec un raccourcissement des bras des phycobilisomes qui perdent la PEII disposée à leur extrémité distale. Ces changements peuvent être interprétés comme une acclimatation pour limiter la quantité d'énergie lumineuse captée par le système photosynthétique. Du point de vue inverse, l'augmentation de la concentration en pigments photosynthétiques quand l'intensité lumineuse diminue peut être interprétée comme une photo-acclimatation visant à maximiser la captation d'énergie lumineuse - une augmentation de la quantité de PE par cellule avec la profondeur a d'ailleurs été observée *in situ* (Wyman, 1992). La concentration en zeaxanthine, autre pigment accessoire, reste quant à elle constante chez ces deux organismes ce qui questionne son rôle de pigment photoprotecteur. Elle pourrait cependant être importante pour la détoxification des ROS produites par le PSII (Six *et al.*, 2004).

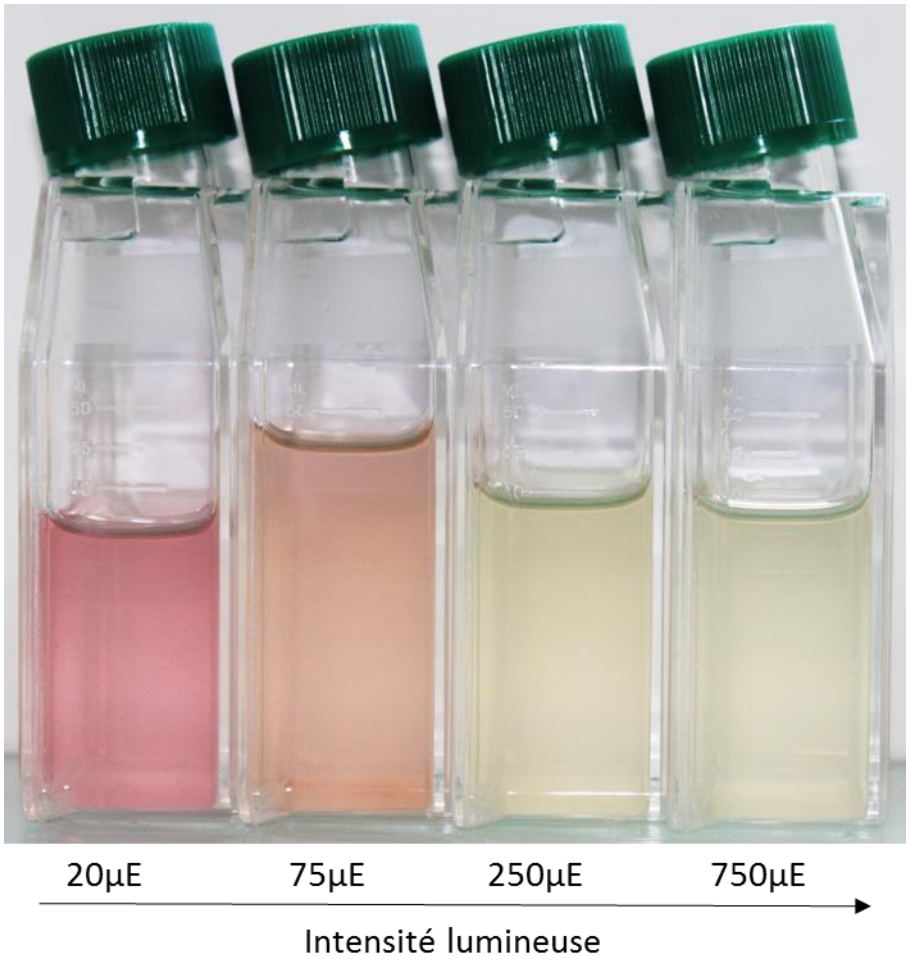


Figure 26 Effet de l'intensité lumineuse sur la pigmentation de *Synechococcus*. Cultures de la souche WH7803 (type pigmentaire 3a) acclimatées à des intensités de lumière croissantes (exprimées en $\mu E.m^{-2}.s^{-1}$). Le changement de couleur est dû au raccourcissement des bras des phycobilisomes, et la diminution de l'intensité de la coloration à la diminution du nombre de phycobilisomes. Photo : H. Doré.

Enfin, en cas de déséquilibre entre l'activité du PSII et du PSI, l'énergie lumineuse captée peut être redistribuée entre les deux photosystèmes par un processus appelé « transition d'état ». Les longueurs d'ondes d'absorption des deux photosystèmes étant différentes, il arrive que l'un des deux photosystèmes fonctionne plus efficacement que l'autre ce qui bloque la chaine de transport d'électrons à l'état réduit (si le PSII est plus efficace) ou à l'état oxydé (si le PSI est plus efficace). Les transitions d'état permettent une meilleure répartition de l'énergie lumineuse captée par un mécanisme qui n'a pas été étudié chez les picocyanobactéries marines mais chez d'autres cyanobactéries possédant des phycobilisomes, reflétant probablement assez bien ce qui se passe chez *Synechococcus*. Le mécanisme précis est encore débattu, mais implique probablement la translocation des phycobilisomes depuis le PSII vers le PSI (et inversement), voire peut-être des photosystèmes eux-mêmes (Figure 27). Elles ont lieu sur une échelle de temps de quelques secondes (ce qui plaide en faveur d'un mouvement des phycobilisomes, l'encombrement de la membrane thylakoïdale limitant la fluidité membranaire), et sont régulées par l'état d'oxydoréduction de l'environnement des thylakoïdes : la réduction de la plastoquinone va par exemple entrainer une transition d'état vers « l'état 2 », où les PBSs transmettent l'énergie au PSI.

Ces transitions d'état permettent également de contrôler l'équilibre entre la voie « classique » du transport d'électrons (aboutissant à la réduction du NADP en NADPH) et les voies cycliques, qui maintiennent la force proton-motrice génératrice d'ATP sans produire de pouvoir réducteur. En effet à l'état 2, l'activité supérieure du PSI par rapport au PSII favorise les voies cycliques du transport d'électrons (Mullineaux, 2014).

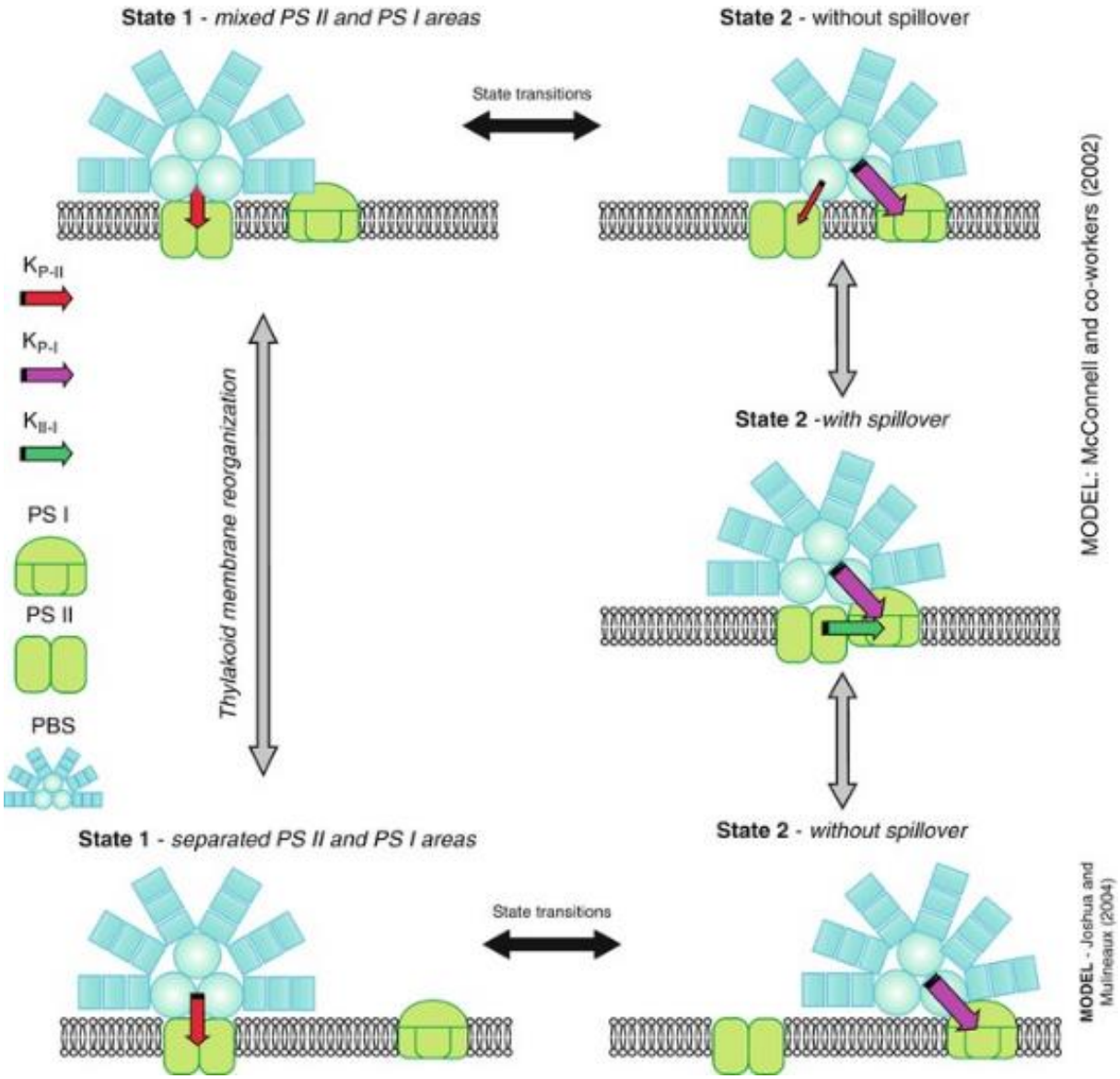


Figure 27 Modèles de fonctionnement des transitions d'état. L'énergie captée par les phycobilisomes (PBS) est majoritairement transmise au PSII à l'état 1, et au PSI à l'état 2. Le *spillover* désigne la transmission d'énergie directe entre les photosystèmes. Les flèches colorées représentes le taux de transfert d'énergie des PBS au PSII (K_{P-II}, en rouge), des PBS au PSI (K_{P-I}, en violet) et entre les deux photosystèmes (K_{II-I}, en vert). Figure issue de (Kirilovsky *et al.*, 2014).

Un deuxième ensemble de mécanismes, regroupés sous le terme NPQf (pour Non-Photochemical Quenching of fluorescence) permet de dissiper l'énergie lumineuse captée par les phycobilisomes ou les photosystèmes avant qu'elle soit transmise à la chaine de transfert d'électrons. Un premier mécanisme est utilisé pour dissiper l'énergie lumineuse reçue sous forme

de chaleur avant même qu'elle atteigne le centre réactionnel, et est donc très efficace pour limiter la dégradation des centres réactionnels. Chez *Synechococcus*, cette dissipation est effectuée grâce à une protéine associée à un pigment caroténoïde, l'OCP (pour *Orange Carotenoid Protein*). L'OCP dans sa version la plus courante est activée par la lumière par un changement de conformation, et la forme active se place au niveau du cœur d'allophycocyanine du PBS, où elle dissipe l'énergie issue de la lumière sous forme de chaleur. L'OCP est donc à la fois le senseur de la lumière et l'effecteur de la dissipation d'énergie. Le retour sous la forme inactive de l'OCP est catalysé par la protéine FRP (*Fluorescence Recovery Protein*) (voir la revue récente de Kerfeld *et al.*, 2017). Tous les génomes de *Synechococcus* séquencés jusqu'ici possèdent les gènes codant pour l'OCP et la protéine associée FRP, mais ces gènes sont absents des génomes de *Prochlorococcus* (qui ne possède pas de phycobilisome) (Scanlan *et al.*, 2009; Kerfeld *et al.*, 2017). La protéine IsiA (Iron-stress induced), impliquée dans la réponse au stress en fer (voir partie 3.2.4 ci-dessous) est également induite en conditions de forte lumière et pourrait avoir une activité de NPQ (Havaux *et al.*, 2005; Bailey *et al.*, 2005). Le gène codant pour cette protéine est également absent de tous les génomes de *Prochlorococcus,* mais ces derniers sont malgré tout capable de dissiper l'excès d'énergie lumineuse reçue par ses photosystèmes, bien que le mécanisme ne soit pas connu (Bailey *et al.*, 2005; Six *et al.*, 2007a). Une famille de protéines connue sous le nom de HliP (High-Light Inducible Protein) et présente aussi bien chez les plantes que chez les cyanobactéries a également connu une forte expansion au sein des génomes de *Prochlorococcus.* Il a été suggéré que ces protéines pourraient avoir une activité de NPQ en réponse à une forte intensité lumineuse chez *Synechocystis* PCC6803 (He *et al.*, 2001; Havaux *et al.*, 2003), et un certain nombre d'entre elles sont surexprimées en condition de stress lumineux chez *Prochlorococcus* (Steglich *et al.*, 2006), et deux d'entre elles chez *Synechococcus* (Berg *et al.*, 2011). Le mécanisme précis par lequel a lieu la dissipation de l'énergie lumineuse n'est cependant pas encore élucidé.

Les mécanismes détaillés ci-dessus sont cependant insuffisants pour empêcher la formation d'espèces réactives de l'oxygène du fait de la proximité entre l'O_2 issu de l'oxydation de l'eau et les électrons qui sont transférés dans la chaine de transport photosynthétiques. Les ROS, très réactives, peuvent réagir avec de nombreux composants cellulaires en oxydant les lipides, les protéines et les acides nucléiques – ce qui les rend très toxiques pour la cellule. Plusieurs mécanismes permettent de limiter la concentration cellulaire en ROS soit en captant les électrons en excès pour éviter leur formation, soit en les réduisant rapidement après leur formation.

Chez *Synechococcus* sp. WH8102, les électrons peuvent être captés au niveau de la plastoquinone par une oxydase, probablement la *plastoquinone terminal oxydase* PTOX (Bailey *et al.*, 2008), ce qui permettrait d'éviter l'accumulation de plastoquinone sous forme réduite. Un orthologue du gène *ptox* est présent chez certaines souches de *Synechococcus* et dans la plupart des souches de *Prochlorococcus* HL et LLI (Scanlan *et al.*, 2009). Le gène correspondant est par ailleurs surexprimé en condition de forte intensité lumineuse (Steglich *et al.*, 2006) et au cours de la journée lors d'un cycle jour/nuit (Mella-Flores *et al.*, 2012) chez *Prochlorococcus,* ce qui tend à confirmer son rôle dans la régulation de l'état de réduction de la chaine d'oxydoréduction quand la lumière est en excès.

Les électrons en excès peuvent également être utilisés pour produire de l'eau en réduisant l'oxygène, processus connu sous le nom de réaction de Mehler. Contrairement aux plantes dans lesquelles cette réaction passe par la production d'$O_2^{\bullet -}$ au niveau du PSI, chez *Synechocystis* PCC6803 elle est réalisée grâce à deux flavoprotéines qui réduisent directement l'O_2 en eau grâce aux électrons fournis par le PSI (Helman *et al.*, 2003) (Figure 28). Toutes les picocyanobactéries marines séquencées possèdent des orthologues de ces deux protéines (Scanlan *et al.*, 2009), et le gène codant pour l'une d'entre elle est surexprimé chez *Prochlorococcus* MED4 en conditions de forte intensité lumineuse (Steglich *et al.*, 2006).

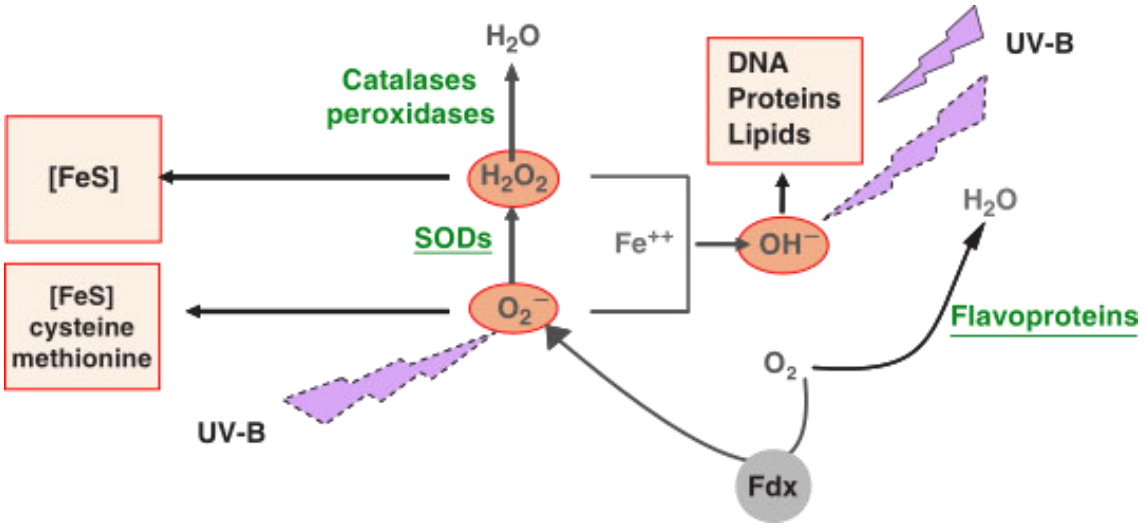


Figure 28 Exemples de mécanismes de détoxification en cas de stress oxydant. Les enzymes permettant d'éviter la formation de ROS ou de les éliminer sont indiquées en vert. Fdx : Ferrédoxine. Figure modifiée à partir de (Latifi *et al.*, 2009).

Les picocyanobactéries marines possèdent également un certain nombre de gènes codant pour des protéines qui permettent d'éliminer les ROS, dont la distribution phylétique est variable (Scanlan *et al.*, 2009). Ainsi la plupart des *Synechococcus* possèdent une catalase-peroxydase (éliminant le peroxyde d'hydrogène H_2O_2) qui est en revanche absente chez *Prochlorococcus.* Toutes possèdent au moins un gène codant pour une superoxyde-dismutase (éliminant l'ion superoxyde) parmi la variété des superoxyde dismutases associées à différents ions métalliques (Ni, Fe, Cu/Zn or Mn). Elles possèdent également plusieurs gènes codant pour des peroxiredoxines et thioredoxines qui peuvent éliminer différents types d'hydroxyperoxides (Scanlan *et al.*, 2009). La plupart de ces gènes sont surexprimés lors de l'augmentation lumineuse au cours d'un cycle jour/nuit chez les souches testées (Mella-Flores *et al.*, 2012). Enfin, l'*orange carotenoid protein* et les *high-light inducible proteins* évoquée plus haut jouent peut être également un rôle dans l'élimination de l'oxygène singulet produit au niveau du centre réactionnel du PSII (Kerfeld *et al.*, 2017; He *et al.*, 2001).

Les picocyanobactéries marines possèdent donc tout un ensemble de mécanismes de protection contre les effets délétères de la lumière, contrepartie nécessaire à son utilisation comme

source d'énergie pour assimiler le carbone minéral. Nous allons à présent nous intéresser à l'utilisation des principaux autres nutriments nécessaires à la production de biomasse.

3.2 L'acquisition des nutriments et son intégration au niveau cellulaire

De façon similaire à ce qui a été développé pour le phytoplancton en général en première partie de cette introduction, les picocyanobactéries marines, quand elles sont en surface où l'intensité lumineuse est suffisante, sont la plupart du temps limitées dans leur croissance par la concentration en quelques nutriments clés. Selon les régions de l'océan et les périodes de l'année, elles peuvent ainsi être limitées par la concentration en azote (Hunt *et al.*, 2013; Van Mooy and Devol, 2008; Shilova *et al.*, 2017), par la concentration en fer (Mann and Chisholm, 2000; Shilova *et al.*, 2017) et de façon épisodique par la concentration en phosphate (Fuller *et al.*, 2005; Saito *et al.*, 2014). Nous détaillerons dans cette partie les modes d'acquisition des différents nutriments ainsi que les réponses physiologiques déclenchées par les carences en nutriments et leurs liens avec les autres processus cellulaires.

3.2.1 Diversité des sources d'azote et de phosphate utilisées par les picocyanobactéries marines

L'azote (N) et le phosphore (P, sous forme de phosphate) sont des éléments essentiels du fait de leur abondance dans la constitution de nombres de molécules organiques. Au sein des cellules, l'azote est présent essentiellement dans les protéines, qui constituent sa forme principale d'utilisation (74 à 84% de l'azote chez une souche de *Synechococcus* d'eau douce), et dans les acides nucléiques (ADN et ARN, 1,8 à 2,8%) (Geider and La Roche, 2002; Lourenço *et al.*, 1998). L'azote entre également dans la composition des pigments et de différents osmolytes, ainsi que dans celle des composants de la paroi bactérienne (N-acétylglucosamine et N-acétylmuramique). Dans les cellules phytoplanctoniques, le phosphate est quant à lui retrouvé essentiellement dans les acides nucléiques (ADN, ARN) et les phospholipides membranaires, ainsi que sous forme de NTP (nucléotides tri-phosphates), et éventuellement sous forme de granules de polyphosphate servant de réserve (Geider and La Roche, 2002). Les lipopolysaccharides de la paroi externe des bactéries contiennent également en général du phosphate. L'assimilation de ces deux éléments est donc essentielle pour permettre la croissance cellulaire (augmentation de la biomasse cellulaire, notamment protéique) et la division (réplication de l'ADN).

Synechococcus et *Prochlorococcus* sont capables d'utiliser une grande variété de composés azotés. Chez *Synechococcus*, toutes les souches testées sont capables de croître avec comme seule source d'azote l'ammonium (NH_4^+) ou le nitrite (NO_2^-), et toutes sauf quatre (MITS9220, RS9913, RS9914 et RS9917) peuvent utiliser le nitrate (Waterbury *et al.*, 1986; Moore *et al.*, 2002; Fuller *et al.*, 2003). En revanche, aucune souche n'est capable de fixer le diazote, contrairement à d'autres cyanobactéries marines dont l'emblématique *Trichodesmium* (Waterbury *et al.*, 1986). Une partie des souches peut également utiliser des formes organiques d'azote, comme l'urée (NH_2-CO-NH_2)

ou le cyanate (OCN^-) (Waterbury *et al.*, 1986; Moore *et al.*, 2002; Kamennaya *et al.*, 2008). Enfin *Synechococcus* est capable d'utiliser les acides-aminés libres dans l'environnement, et leur acquisition est stimulée par la lumière, mais leur utilisation comme seule source d'azote n'a pas été testée (Paerl, 1991; Zubkov *et al.*, 2003; Michelou *et al.*, 2007; Mary *et al.*, 2007). Parmi toutes ces sources d'azote, les sources privilégiées (c.-à-d. soutenant la meilleure croissance des cultures) semblent être l'ammonium et l'urée (Moore *et al.*, 2002). Chez *Prochlorococcus,* les premières souches testées en culture ont également révélé leur capacité à croître en présence d'ammonium ou d'urée comme seule source d'azote, avec une préférence pour l'ammonium (Rippka *et al.*, 2000; Palinska *et al.*, 2000; Moore *et al.*, 2002). En revanche, seules quelques souches LL pouvaient utiliser le nitrite, et aucune ne pouvait utiliser le nitrate. Ce n'est que récemment que des souches capables d'utiliser le nitrate (deux souches LL et une souche HL) ont été isolées et leur génome séquencé (Berube *et al.*, 2014), après que leur existence eut été révélée dans l'environnement par différentes méthodes (Casey *et al.*, 2007; Martiny *et al.*, 2009a). Enfin certaines souches sont capables d'utiliser le cyanate (Kamennaya *et al.*, 2008), et comme chez *Synechococcus* l'absorption d'acides aminés est répandue et stimulée par la lumière (Zubkov *et al.*, 2003; Michelou *et al.*, 2007; Mary *et al.*, 2007).

De la même façon, les picocyanobactéries marines sont capables d'utiliser une grande variété de sources de phosphate, dont la gamme est variable selon les souches. Toutes les souches étudiées sont capables d'assimiler le phosphate inorganique (Pi) et la grande majorité peuvent utiliser des polyphosphates et l'ATP en les hydrolysant (Rippka *et al.*, 2000; Moore *et al.*, 2005; Mazard *et al.*, 2012b), ainsi que l'ADN libre (Mazard *et al.*, 2012b). L'utilisation d'autres sources de phosphate organique dépend des souches, certaines pouvant croître en utilisant du β-glycérophosphate, du glucose-6-phosphate ou de l'AMP cyclique. Ces capacités sont relativement bien reliées au contenu en gènes des différentes souches (Scanlan *et al.*, 2009).

3.2.2 Les modes d'acquisition des macronutriments

Les modalités du transport et de l'assimilation des macronutriments chez les picocyanobactéries sont relativement bien connues, car elles sont bien conservées avec les autres Cyanobactéries et les Eubactéries en général.

L'incorporation de l'azote dans les molécules organiques de la cellule se fait essentiellement (à l'exception des acides aminés) par l'association de l'ammonium au 2-oxoglutarate (composé carboné se trouvant au croisement des métabolismes du carbone et de l'azote) pour former un acide-aminé glutamate, une étape catalysée par le complexe GS-GOGAT (Glutamine Synthase - Glutamate OxoGlutarate AminoTransferase) (Figure 29). Ce complexe utilise la ferrédoxine réduite lors de la photosynthèse comme donneur d'électrons.

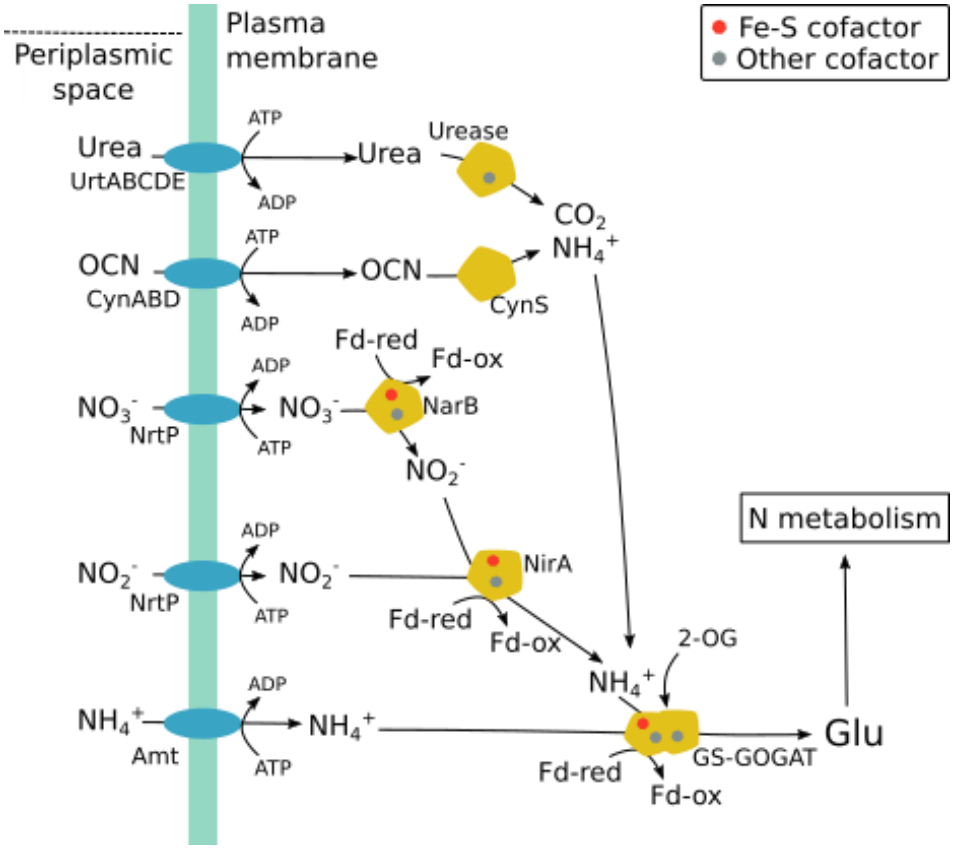


Figure 29 Mécanismes d'assimilation de l'azote. Différents transporteurs transmembranaires permettent l'absorption de l'azote sous forme minérale ou organique. Les formes oxydées sont réduites en ammonium NH4+ en bénéficiant des électrons fournis par la ferrédoxine (Fd). L'ammonium est ensuite intégré au métabolisme de l'azote par la formation de glutamate (Glu).

L'incorporation de l'azote nécessitant que celui-ci soit sous forme d'ammonium, cela implique que les autres formes d'azote captées par la cellule soient transformées avant leur assimilation (Figure 29). Le nitrate (NO_3^-) et le nitrite (NO_2^-) sont absorbés chez les picocyanobactéries marines par le même transporteur transmembranaire, NrtP (Bird and Wyman, 2003; Scanlan *et al.*, 2009). Le nitrate est réduit en nitrite par la nitrate réductase NarB, et le nitrite (issu de nitrate ou directement absorbé) est réduit en ammonium par la nitrite réductase NirA. Ces deux enzymes NarB et NirA utilisent également comme donneur d'électrons la ferrédoxine réduite lors de la photosynthèse (ou la flavodoxine) (Flores *et al.*, 2005). Contrairement aux mécanismes homologues retrouvés chez les plantes et la plupart des bactéries, ces enzymes ne sont donc pas dépendantes du NADPH, puisque la Ferrédoxine est directement utilisée comme donneur d'électrons. Cela les rend en revanche très dépendantes de l'activité photosynthétique.

L'ensemble de ces enzymes (GS-GOGAT, NarB, NirA) utilisent des cofacteurs pour la transmission des électrons, notamment des protéines fer-soufre (Fe-S) et NarB utilise également comme cofacteur la molydoptérine (cofacteur contenant du Molybdène et dont les gènes de synthèse sont retrouvés chez toutes les souches possédant *narB*) (Flores *et al.*, 2005). Leur fonctionnement dépend donc de la présence de micronutriments, en particulier de métaux.

Les sources organiques d'azote (en dehors des acides aminés) sont également transformées pour être assimilées sous forme d'ammonium NH_4^+ (Figure 29). L'urée est absorbée par les cellules grâce à une perméase de type ABC, encodée par le complexe de gènes *urtA/B/C/D/E* qui forme un opéron (Valladares *et al.*, 2002; Scanlan *et al.*, 2009). Elle est ensuite dégradée par l'uréase, grosse métallo-enzyme nickel-dépendante encodée par les gènes *ureA/B/C* (pour la partie catalytique) et *ureD/E/F/G* (Collier *et al.*, 1999) pour donner une molécule de CO_2 et deux $NH4^+$. Le cyanate utilisé par les cellules peut quant à lui avoir deux origines : soit une origine externe, auquel cas il est absorbé par un transporteur encodé par les gènes *cynA/B/D*, soit une origine interne comme produit du cycle de l'urée. Ce cyanate étant toxique pour les cellules, il est dégradé par la cyanase (CynS) en NH_4^+ et CO_2. Cette décomposition peut également se faire spontanément dans le milieu (c.-à-d. sans intervention de l'enzyme), ce qui explique que certaines souches ne possédant pas le transporteur puissent malgré tout pousser sur le cyanate comme source d'azote (Kamennaya *et al.*, 2008). Enfin les acide-aminés présents dans l'environnement sont absorbés grâce à des transporteurs transmembranaires qui n'ont pas été caractérisés en détail (Scanlan *et al.*, 2009), et suivent ensuite directement les voies du métabolisme des acide-aminés.

En ce qui concerne le phosphate, son incorporation dans le métabolisme cellulaire se fait également sous une forme privilégiée, le phosphate inorganique (Pi). Celui-ci est capté dans l'espace périplasmique par la protéine PstS (Scanlan *et al.*, 1993; Ikeya *et al.*, 1997) puis est transféré dans le cytoplasme par le transporteur PstABC (Moore *et al.*, 2005) (Figure 30). Les gènes codant pour ces protéines sont présents dans l'ensemble des souches, avec un nombre variable de copies de *pstS* (Moore *et al.*, 2005; Scanlan *et al.*, 2009). Les molécules organiques contenant du phosphate sont quant à elles dégradées par une ou des phosphatase(s), qui sont probablement périplasmiques (le Pi étant ensuite assimilé par la cellule) et dont le nombre peut aller jusqu'à cinq dans un même génome (Moore *et al.*, 2005; Scanlan *et al.*, 2009). La plus commune est la phosphatase alcaline (encodée par le gène *phoA*) mais d'autres ont par exemple une activité 5'-nucleotidase permettant de récupérer le phosphate d'acides nucléiques. Tous les génomes possèdent également les gènes d'acquisition des phosphonates (composés comportant une liaison carbone-phosphore), même si la voie de dégradation n'a pas été formellement décrite (Moore *et al.*, 2005; Scanlan *et al.*, 2009; Feingersch *et al.*, 2012). Enfin, tous les génomes possèdent les gènes de dégradation des polyphosphates (*ppK* et *ppX*) bien que la formation de granules de polyphosphate en milieu riche en phosphate n'ait été observée que chez certaines souches de *Synechococcus* (Waterbury *et al.*, 1986; Mazard *et al.*, 2012b). Les picocyanobactéries marines sont particulièrement bien armées pour faire face à la limitation en phosphate, ce qui explique qu'on les retrouve en abondance dans les régions où la disponibilité en phosphate est très faible, de l'ordre du nM (Moutin *et al.*, 2002; Fuller *et al.*, 2005; Zubkov *et al.*, 2007; Talarmin *et al.*, 2015).

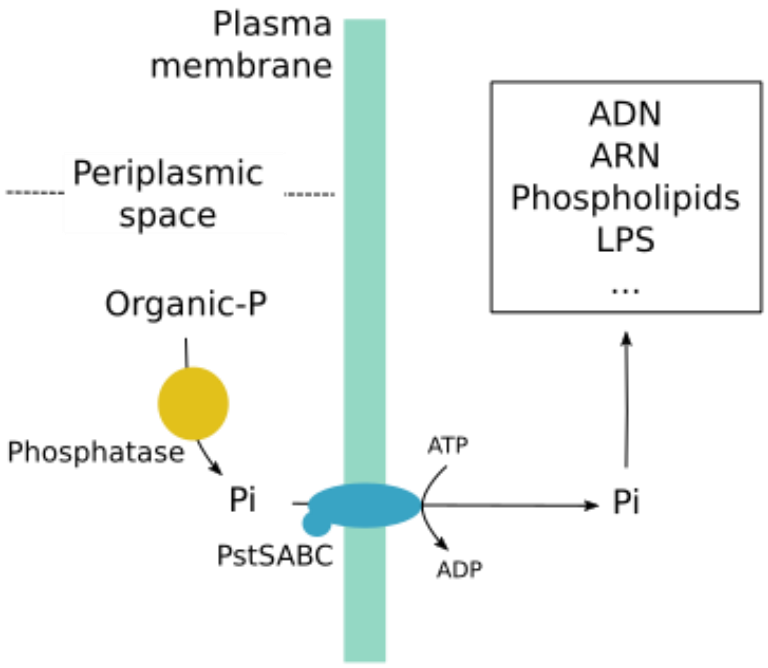


Figure 30 Modes d'acquisition du phosphate. Les formes organiques sont hydrolysées dans l'espace périplasmique, et le phosphate entre dans la cellule sous forme de phosphate inorganique (Pi). LPS : lipopolysaccharides.

Comme évoqué dans la partie 2.3.3, la présence dans les génomes de *Prochlorococcus* des gènes d'assimilation des macronutriments N et P sous différentes formes ne semble pas être reliée à la phylogénie verticale du genre, mais plutôt aux caractéristiques environnementales du lieu d'isolement des souches. Chez *Synechococcus*, malgré un nombre croissant de souches caractérisées, la présence d'un seul génome séquencé par clade (Scanlan *et al.*, 2009) n'a pour l'instant pas permis d'identifier le lien éventuel entre les capacités d'assimilation des macronutriments et les clades et leur écologie.

3.2.3 Les effets physiologiques de la carence en macronutriments

Les picocyanobactéries marines étant abondantes dans les régions oligotrophes, où les concentrations en macronutriments sont faibles, elles possèdent un certain nombre de mécanismes permettant de faire face à une carence en ces éléments.

Etant donnée l'abondance de l'azote dans les macromolécules, la suppression de toute forme d'azote du milieu de culture a un effet délétère rapide sur la croissance des cellules. En l'absence de source d'azote, *Synechococcus* et *Prochlorococcus* interrompent leur croissance (Lindell and Post, 2001; Steglich *et al.*, 2001; Lindell *et al.*, 2002; Tolonen *et al.*, 2006), et plusieurs études montrent chez *Prochlorococcus* une diminution de l'efficacité photosynthétique du photosystème II (Steglich *et al.*, 2001; Lindell *et al.*, 2002; Tolonen *et al.*, 2006) ainsi qu'une augmentation du rapport carbone/azote (C/N) (Lindell *et al.*, 2002), et il est probable que ces résultats s'appliquent

également à *Synechococcus*. Une étude plus détaillée chez *Prochlorococcus* SS120 montre également la diminution du taux de réoxydation de la quinone au niveau du photosystème II et la diminution du contenu cellulaire en protéine D1, ce qui montre que la privation d'azote est susceptible d'augmenter la photoinhibition et de créer un stress oxydant (Steglich *et al.*, 2001). Chez *Synechococcus,* la dégradation de la phycoérythrine a également été observée en cas de privation d'azote, révélant que les phycobiliprotéines (qui constituent une part importante du contenu en azote) peuvent servir de molécules de stockage (Wyman *et al.*, 1985). Bien qu'une phycoérytrhine non photosynthétique soit présente chez la plupart des souches de *Prochlorococcus,* celle-ci n'est pas dégradée en conditions pauvres en azote, et ne joue donc pas le rôle de molécule de stockage (en tout cas dans la seule souche étudiée) (Steglich *et al.*, 2001). Ces différents aspects révèlent les liens étroits qui existent entre le métabolisme de l'azote et la photosynthèse (donc le métabolisme du carbone) chez les picocyanobactéries marines.

Les picocyanobactéries marines montrent une très grande capacité d'acclimatation à une faible disponibilité du phosphate. Bien que la limitation en phosphate retarde la réplication et donc la division cellulaire (Vaulot *et al.*, 1996), les cultures sont capables de survivre bien plus longtemps à une privation totale de phosphate qu'à une privation d'azote (Rippka *et al.*, 2000; Lindell and Post, 2001). Cette capacité de survie est liée à la possibilité qu'ont les picocyanobactéries marines de réduire drastiquement leur quota intracellulaire en phosphate. Les rapports N:P et C:P mesurés en conditions riches en phosphate sont autour de 20 et 140 pour *Synechococcus* et autour de 20 et 190 pour *Prochlorococcus,* légèrement plus élevées que le ratio de Redfield (16 et 106 respectivement) correspondant à la moyenne des organismes phytoplanctoniques (Heldal *et al.*, 2003; Bertilsson *et al.*, 2003). Leurs besoins en phosphate sont donc plus faibles que ceux de la moyenne du phytoplancton, mais le plus impressionnant est leur capacité à réduire leur quota de P en conditions limitées : les ratios C:P et N:P sont alors multipliés par 4, ce qui signifie que les besoins en phosphate sont divisés par 4 (Bertilsson *et al.*, 2003; Fu *et al.*, 2006). Le quota mesuré est alors si faible que plus de 50% du phosphate intracellulaire est contenu dans l'ADN du chromosome, ce qui laisse moins de la moitié du phosphate pour les autres composants cellulaires dont l'ARN (Bertilsson *et al.*, 2003). La concentration limite à partir de laquelle cette réduction est observée semble être autour de 1 µM pour *Synechococcus* sp. WH7803, et elle s'accompagne d'une diminution du taux de croissance (Fu *et al.*, 2006). Le mécanisme principal de cette réduction du quota en P est le remplacement des phospholipides membranaires par des lipides ne contenant pas de phosphate, la tête polaire comprenant soit du galactose (monogalactosyl-diacyl-glycerol MGDG et digalactosyl-diacyl-glycerol DGDG) soit du sulfate (sulfoquinovosyl-diacyl-glycerol SQDG). Chez *Synechococcus* et *Prochlorococcus,* la majorité des lipides membranaires est sous forme de SQDG y compris en conditions riches en phosphate, et seulement 2 à 10% des lipides est sous forme de phospholipide (essentiellement du phosphatidylglycerol PG) (Cuhel and Waterbury, 1984; Van Mooy *et al.*, 2006). La proportion de sulfolipides augmente fortement en cas de limitation : le rapport SQDG/PG est multiplié par 1,5 chez *Prochlorococcus* et par 6 à 20 chez *Synechococcus* (Van Mooy *et al.*, 2009). Le remplacement des lipides n'explique pas à lui seule les faibles besoins en P des picocyanobactéries marines. Elles possèdent également des lipopolysaccharides (LPS) à leur surface qui sont dépourvus de P, sans que l'on sache si cela

correspond à une adaptation spécifique ou simplement à une forme primitive de LPS (Snyder *et al.*, 2009). Enfin, une étude de modélisation métabolique récente suggère que d'autres processus cellulaires sont impactés, notamment par l'élimination chez *Prochlorococcus* de réactions métaboliques impliquant l'utilisation de phosphate, et par la réduction de la production d'ARN (Casey *et al.*, 2016). Ce dernier point est appuyé par l'étude de l'expression des gènes en cas de limitation en phosphate chez *Synechococcus* sp. WH8102, qui montre une répression de plusieurs gènes impliqués dans la traduction (Tetu *et al.*, 2009).

En plus de ces modifications physiologiques, la limitation en nutriments déclenche la mise en place de mécanismes d'utilisation de sources alternatives d'azote et de phosphate. La réponse à une privation d'azote et à un changement de source d'azote est finement régulée au niveau génétique, notamment par la protéine NtcA, un régulateur transcriptionnel conservé qui a de nombreuses cibles dans le génome (Vega-palas *et al.*, 1990). Le gène codant pour cette protéine, *ntcA*, est retrouvé dans tous les génomes de *Synechococcus* et *Prochlorococcus* (Lindell and Padan, 1998; Lindell *et al.*, 2002; Scanlan *et al.*, 2009). L'ammonium étant la source privilégie d'azote, les gènes codant pour l'import et la transformation des autres sources sont très peu exprimés en présence d'ammonium : l'expression de *ntcA* est réprimée et les cellules n'utilisent ni le nitrite, ni le nitrate si l'ammonium est également présent. A l'inverse, l'expression de *ntcA* est activée en présence d'une autre source d'azote si l'ammonium est absent (Lindell and Padan, 1998; Lindell and Post, 2001; Lindell *et al.*, 2002; Wyman and Bird, 2007), et elle est maximale en l'absence de source d'azote (Lindell and Post, 2001; Tolonen *et al.*, 2006). Cette remarquable dépendance de l'expression de *ntcA* à la disponibilité en azote en fait un marqueur de choix pour évaluer le stress en azote *in situ*. Il a ainsi été utilisé pour évaluer le stress en mer rouge chez *Synechococcus*, révélant que les populations présentes n'étaient pas en manque d'ammonium (expression basale de *ntcA*) (Lindell and Post, 2001; Lindell *et al.*, 2005b).

Outre l'expression des gènes de transport et de transformation de l'azote et du régulateur *ntcA*, la privation d'azote entraine de nombreuses modifications transcriptionnelles : les gènes impliqués dans le métabolisme du carbone sont réprimés, tandis que certains régulateurs transcriptionnels (facteurs sigmas) et certains gènes de protection contre la photoinhibition sont surexprimés (*hli* chez *Prochlorococcoccus* et *psbA*), ce qui montre une réaction au stress oxydant engendré par la carence (Tolonen *et al.*, 2006; Su *et al.*, 2006). Une bonne partie de ces gènes comprennent une séquence de régulation fixant NtcA dans leur promoteur (Su *et al.*, 2005, 2006), mais certains dépendent d'autres régulateurs, notamment la protéine P_{II} (Scanlan *et al.*, 2009).

De manière relativement similaire, *Synechococcus* et *Prochlorococcus* augmentent fortement leurs capacités à assimiler le phosphate quand sa concentration diminue. Plusieurs études montrent en effet une augmentation très forte de l'affinité pour le Pi (diminution du K_M) et de la vitesse d'assimilation (augmentation de v_{max}) en conditions de limitation en phosphate (Ikeya *et al.*, 1997; Donald *et al.*, 1997; Fu *et al.*, 2006; Krumhardt *et al.*, 2013). L'augmentation de v_{max} passe par l'augmentation du nombre de transporteurs transmembranaires, quand l'augmentation de l'affinité est probablement due à l'expression de différentes isoformes de PstS (qui lie le Pi) en fonction de la concentration en phosphate, comme cela a été observé chez *Synechocystis* PCC6803 (Pitt *et al.*, 2010). Les changements d'expression des gènes lors de la limitation en phosphate montrent en

effet une surexpression des gènes *pstS* et *pstABC* (Martiny *et al.*, 2006; Tetu *et al.*, 2009; Mazard *et al.*, 2012b). Les gènes de phosphatases permettant l'utilisation du phosphate organique sont également surexprimés, en accord avec l'augmentation de l'activité enzymatique de la phosphatase alcaline (Moore *et al.*, 2005). Enfin, dans les souches de *Synechococcus* qui les possèdent, les gènes de motilité sont également surexprimés (Tetu *et al.*, 2009; Mazard *et al.*, 2012b), ce qui suggère que ces souches présentent un chimiotactisme vis-à-vis du phosphate comme cela a été observé pour l'azote (Willey and Waterbury, 1989).

La détection de la concentration extérieure en phosphate se fait par la protéine PhoR, membre d'un système à deux composants dont le régulateur de réponse est PhoB, qui active la transcription des gènes évoqués ci-dessus (*pstS*, *pstABC*, *phoA*, etc.) comprenant dans leur promoteur une « pho box » (Su *et al.*, 2003, 2007; Tetu *et al.*, 2009). Un autre régulateur, PtrA, a également été identifié et semble être nécessaire pour l'activation des gènes de phosphatase chez les souches le possédant, suggérant une réponse en deux temps : dans un premier temps le système PhoR/PhoB active le transport de forte affinité du phosphate, et dans un deuxième temps PtrA active les gènes de phosphatases pour récupérer le phosphate organique (Ostrowski *et al.*, 2010). De manière intéressante, ces régulateurs ne sont pas toujours présents dans toutes les souches, ce qui suggère une modification ou une perte du système de régulation, notamment chez les *Prochlorococcus* LL et les souches côtières de *Synechococcus* (Moore *et al.*, 2005; Scanlan *et al.*, 2009). Cette observation a amené à formuler l'hypothèse selon laquelle la colonisation de milieux plus stables et où le phosphate n'est pas limitant (plus grande concentration en P en milieu côtier et lumière limitante en profondeur) aurait entrainé la perte des capacités de régulation dans ces souches.

3.2.4 Le cas des micronutriments

Les descriptions détaillées effectuées ci-dessus montrent quelques exemples l'omniprésence des micronutriments, en particulier des métaux traces, dans les enzymes du métabolisme cellulaire et leurs cofacteurs. Outre le fer, présent dans de nombreux cofacteurs de la chaîne de transport d'électrons (Ferrédoxine, protéines Fe-S, etc.) mais également dans les voies d'assimilation de l'azote (voir ci-dessus), le magnésium est présent dans la chlorophylle, le cuivre dans la plastocyanine, le zinc dans l'anhydrase carbonique, le cobalt dans la vitamine B12, le manganèse dans le complexe d'évolution de l'oxygène, etc. (Scanlan *et al.*, 2009). Ces micronutriments sont présents en très faible concentration dans l'environnement (de l'ordre du pico- ou nano-molaire), et les voies utilisées pour les piéger et les assimiler ainsi que les effets d'un manque ou d'un excès de ces éléments sont assez mal connus. L'exemple le plus documenté est de loin celui du fer, probablement du fait de son implication dans la photosynthèse et de son rôle dans la limitation de la croissance du phytoplancton *in situ*, y compris celle des picocyanobactéries (Mann and Chisholm, 2000; Shilova *et al.*, 2017).

Le fer est très présent dans les différents intermédiaires de la chaîne de transport d'électrons, avec 2 ou 3 Fe par PSII, 5 Fe par cytochrome b_6f, 12 Fe par PSI et 2 Fe par ferrédoxine (Raven *et al.*, 1999). En conditions normales (c.-à-d. quand le fer n'est pas limitant), le nombre de PSI est supérieur au nombre de PSII de manière à limiter la production de ROS (voir partie 3.1) en assurant

la réoxydation du PSII, ce qui augmente d'autant plus le besoin en Fe. On comprend donc aisément que la phase claire de la photosynthèse soit particulièrement touchée en cas de carence en fer : outre une diminution de taille cellulaire fréquemment observée, les cellules voient leur pigmentation (et la fluorescence associée) chuter suite à la diminution du nombre de phycobilisomes (les protéines de leurs voies de biosynthèse contenant du fer) et de photosystèmes, en particulier de PSI très riche en fer, en lien avec la répression des gènes associés (Bibby *et al.*, 2003; Thompson *et al.*, 2011; Morrissey and Bowler, 2012; Mackey *et al.*, 2015). La réduction du *pool* de protéines contenant du fer permet de limiter les besoins en Fe, mais dérègle la chaîne de transport d'électrons, et peut ainsi induire un stress oxydant (Latifi *et al.*, 2005): l'augmentation du rapport PSII/PSI augmente la probabilité de bloquer la chaîne à l'état réduit et donc de produire des ROS. Pour faire face à ces dérèglements, plusieurs mécanismes sont mis en place. Tout d'abord, les cellules stressées par le manque de fer surexpriment les gènes codant pour le transporteur de fer (*idiA/futB/futC*), en particulier *idiA* dont le niveau d'expression est directement relié à la concentration en fer chez *Synechococcus* (Webb *et al.*, 2001). Ensuite, la perte de certaines protéines contenant du fer peut être compensée par l'expression de protéines à la fonction équivalente mais dépourvue de Fe : la ferrédoxine (PetF) est remplacée par la flavodoxine (IsiB) et le cytochrome c_6 (PetJ) est remplacé par la plastocyanine (PetE, qui contient du cuivre) chez les souches possédant les gènes correspondant. Enfin, les cellules mettent également en place des réponses pour limiter le stress oxydant, notamment en exprimant certaines protéines HliP. Certaines souches de *Prochlorococcus* et *Synechococcus* expriment également la protéine IsiA (chez *Prochlorococcus,* elle est homologue du gène codant pour l'antenne majeure de collection de lumière, *pcb*), qui fixe la chlorophylle et forme un cercle de 18 protéines autour du PSI, augmentant son efficacité et participant ainsi à protéger la cellule du stress oxydant (Bibby *et al.*, 2001; Boekema *et al.*, 2001; Bibby *et al.*, 2003; Thompson *et al.*, 2011). L'ensemble de ces modifications transcriptionnelles sont probablement régulées par le régulateur transcriptionnel Fur, dont tous les génomes possèdent au moins une copie (Scanlan *et al.*, 2009).

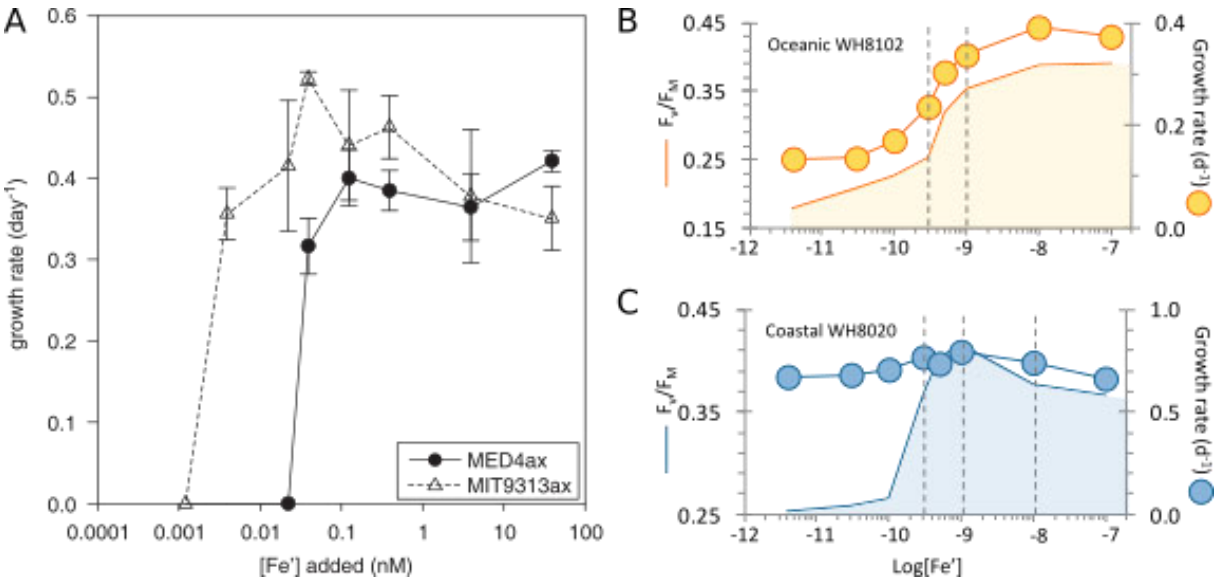


Figure 31 Croissance de différentes souches de *Prochlorococcus* et *Synechococcus* en fonction de la disponibilité en fer. A. Taux de croissance de deux souches de *Prochlorococcus,* MED4 (HLI) et MIT9313 (LLIV) en fonction de la concentration en fer dans le milieu. Figure issue de (Thompson *et al.*, 2011). B. Taux de croissance (points) et efficacité photosynthétique du photosystème II (courbe) en fonction de la concentration en fer dans le milieu (en mol/L) chez deux souches de *Synechococcus*, l'une isolée au large (WH8102, clade III) et l'autre en région côtière (WH8020, clade I). Figure issue de (Mackey *et al.*, 2015).

Si les picocyanobactéries possèdent un certain nombre de protections contre le manque de fer, les différentes souches de *Synechococcus* et *Prochlorococcus* n'ont pas les mêmes capacités de réponse. Ainsi la comparaison de souches HLI (MED4) et LLIV (MIT9313) de *Prochlorococcus* a montré d'une part que la souche LLIV survit à une concentration en fer près de 10 fois plus faible (Figure 31), et d'autre part que cette même souche mettait en place plus de mécanismes de réponse (Thompson *et al.*, 2011). De manière similaire, la comparaison de la souche côtière de *Synechococcus* sp. WH8020 (clade I) à la souche isolée au large WH8102 (clade III) a montré que la première résistait beaucoup mieux aux faibles concentrations en Fe (Figure 31) et qu'elle était capable de mettre en place une réponse hiérarchisée en plusieurs étapes en fonction de l'intensité du stress, quand WH8102 ne montrait qu'une réponse minimale (Mackey *et al.*, 2015). Dans les deux cas, les différences observées ont été reliées à la disponibilité en fer dans le lieu d'isolement des souches. De manière générale, les souches côtières de *Synechococcus* possèdent un certain nombre de gènes de réponse souvent absents chez celles isolées au large, en particulier des gènes de stockage du fer (ferritine et bacterioferritine qui peuvent être présente en nombreuses copies) et certains transporteurs supplémentaires (comme *feoB*, codant pour un transporteur de fer ferreux), ainsi que *isiA* et *isiB*. A quelques nuances près, ces différences ont également été observées *in situ* (Rivers *et al.*, 2009). Enfin chez *Prochlorococcus,* les populations HLIII et HLIV des régions HNLC (limitées en fer) semblent avoir perdu un certain nombre de gènes codant pour des protéines contenant du fer (Yooseph *et al.*, 2010). Cette variabilité de contenu en gènes est en accord avec le fait qu'un certain nombre d'entre eux sont localisés dans des îlots génomiques (voir partie 2.3) (Scanlan *et al.*, 2009). La présence et l'absence de certains gènes liés au métabolisme du fer serait donc une adaptation à sa disponibilité dans le milieu, qui est variable dans les régions côtières et

faible dans les régions HNLC et en profondeur, alors qu'elle peut être relativement élevée et stable au large en surface grâce à l'apport de fer par le dépôt de poussières.

L'effet des autres métaux traces est peu documenté, mis à part le cuivre dont la toxicité est prouvée et dépend également des souches : la souche LL de *Prochlorococcus* testée s'est avérée plus sensible au Cu que la souche HL, ce qui a été relié au fait que la concentration en Cu est généralement bien plus faible en profondeur. *Synechococcus* est de façon générale moins sensible au Cu, ce qui pourrait être dû à la production de ligands permettant de le piéger à l'extérieur de la cellule (Mann *et al.*, 2002), mais présente malgré tout des différences entre souches, les souches côtières semblant plus résistantes (Tetu *et al.*, 2009).

3.2.5 L'intégration du métabolisme des nutriments au niveau cellulaire

Les parties précédentes ont montré quelques aspects du métabolisme cellulaire des picocyanobactéries marines, et il ressort de ces exemples la forte intrication des différentes voies métaboliques, toutes interdépendantes. La photosynthèse apparaît comme étant la clef de voûte de ce métabolisme à la fois comme source d'énergie et de matière carbonée, mais également par sa sensibilité et le danger qu'elle représente quand elle est perturbée. L'exemple de l'interdépendance de la photosynthèse et de l'assimilation de l'azote est assez parlant. Les produits de la photosynthèse sont nécessaires à l'assimilation de l'azote (ATP notamment pour l'import des molécules et pouvoir réducteur pour la réduction des composés oxydés et l'incorporation dans les acide-aminés), et l'azote est un composant essentiel des protéines photosynthétiques (notamment les phycobilisomes chez *Synechococcus*) et est nécessaire pour résister à la photoinhibition par la néosynthèse de D1 (Figure 32). De plus, l'incorporation de l'azote se fait par fixation d'ammonium sur le 2-oxoglutarate, qui est un composé issu du métabolisme du carbone (Figure 32). C'est d'ailleurs par le biais de cette molécule que le régulateur NtcA (et la protéine régulatrice PII) est capable de « sentir » le besoin en azote (Muro-pastor et al., 2001; Flores et al., 2005) : dans le cas d'une augmentation du rapport C/N, la quantité de 2-oxoglutarate augmente, et sa fixation sur NtcA entraine l'activation des nombreux gènes évoqués ci-dessus. Dans ce cadre il est particulièrement intéressant de voir les différences de réaction à la carence en azote entre 2 souches de *Prochlorococcus* appartenant aux écotypes HL et LL : (Tolonen *et al.*, 2006) notent que la répression des gènes impliqués dans la fixation de carbone et son stockage est beaucoup plus forte et plus rapide chez la souche HL MED4 que chez la souche LL MIT9313, ainsi que l'activation des gènes de dégradation du glycogène (qui fournit du 2-oxoglutarate). MED4 est donc beaucoup plus efficace pour répondre au manque d'azote et diversifier les sources d'apport en azote, ce qui peut être vu comme une adaptation pour protéger les photosystèmes qui sont plus susceptibles d'être endommagés en surface qu'en profondeur. Au contraire la souche LL est, dans l'environnement, limitée par la faible quantité de lumière disponible en profondeur et donc à la fois moins soumise à la photoinhibition et moins efficace pour fixer le carbone. Dans ces circonstances, il ne serait donc pas nécessaire de diminuer la fixation de carbone, ce qui pourrait expliquer la différence de régulation (Tolonen *et al.*, 2006). De telles différences n'ont pas été observées chez

Synechococcus, mais il est fort probable que cela soit dû au faible nombre de souches caractérisées. En effet étant donnée la répartition des populations de *Synechococcus* dans l'océan, on peut s'attendre à des différences physiologiques dans leur utilisation des macronutriments.

Les micronutriments sont également un bon exemple de l'intégration au niveau cellulaire : du fait de leur présence dans de nombreux cofacteurs enzymatiques, leur absence peut perturber de nombreux processus cellulaires (Figure 32).

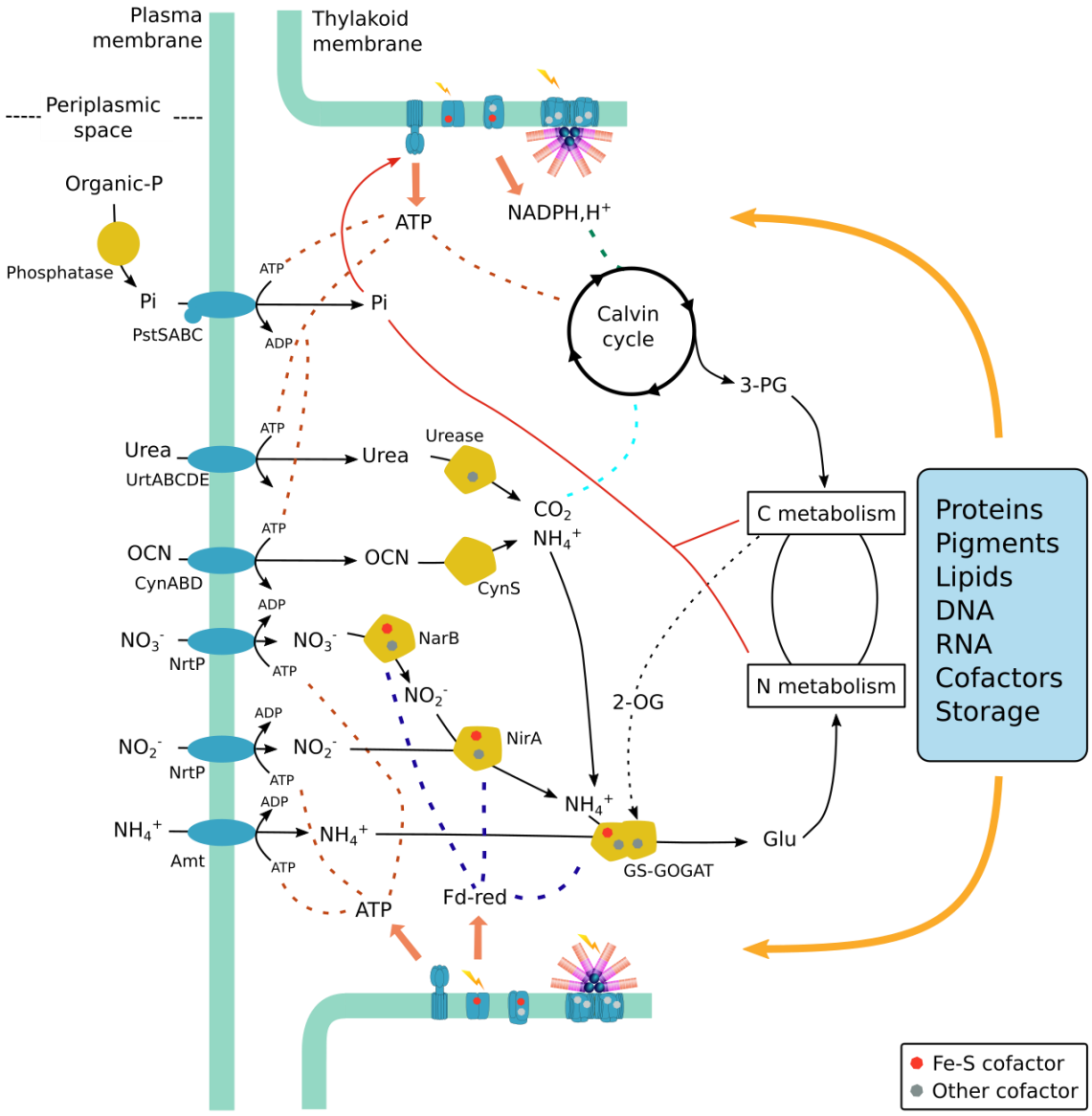


Figure 32 Intégration du métabolisme des nutriments au niveau cellulaire. Les liens avec la photosynthèse sont plus particulièrement mis en avant. Pi : Phosphate inorganique. Fd : Ferrédoxine

Le fait que la photosynthèse soit la clef de voute du métabolisme cellulaire en fait la première victime de toute dérégulation. Ainsi, la perturbation des voies métaboliques peut rapidement se traduire par la production de ROS qui réagissent avec de nombreux composants cellulaires en oxydant les lipides, les protéines et les acides nucléiques, mais dont la première cible sera le PSII.

Pour cette raison, l'efficacité photosynthétique du photosystème II (mesurée par fluorescence) sert fréquemment de proxy de l'état physiologique des cellules.

3.3 Adaptation et acclimatation différentielles à la température

L'étude de la répartition des picocyanobactéries marines (partie 2.2) a révélé l'influence majeure de la température de l'eau sur la répartition des différents clades de *Synechococcus* et *Prochlorococcus*. Malgré cette influence, très peu d'études se sont intéressées aux modifications physiologiques induites par un changement de température chez les picocyanobactéries marines.

3.3.1 Les thermotypes de *Synechococcus* et *Prochlorococcus*

Le point qui a été le plus étudié concerne les différences de capacité d'acclimatation[11] entre souches à une gamme de températures. Chez *Prochlorococcus,* des différences sont observables à deux niveaux entre les souches HLI (clade retrouvé à plus haute latitude) et les souches HLII (clade retrouvé aux basses latitudes) (Figure 33) (Johnson *et al.*, 2006; Zinser *et al.*, 2007). La première différence concerne les limites des gammes de températures dans lesquelles les souches sont capables de croître : les souches HLI sont capables de s'acclimater à des températures approchant les 10 degrés, alors que les souches HLII sont incapables de croître à 15°C. A l'inverse, les souches HLII supportent des températures supérieures ou égales à 30°C, contrairement aux souches HLI. La deuxième différence se trouve dans le taux de croissance : si les souches des deux clades ont un optimum de croissance autour de 25°C, le taux de croissance des souches HLII est supérieur à celui des souches HLI quand la température dépasse 25°C, alors que les souches HLI ont un taux de croissance supérieur quand la température passe sous les 18°C. Ces résultats s'accordent bien avec la répartition latitudinale de ces deux clades (partie 2.2). De la même façon, les gammes de température auxquelles les souches de *Synechococcus* étudiées sont capables de s'acclimater correspondent à la latitude à laquelle elles ont été isolées (Pittera *et al.*, 2014) : les souches isolées à haute latitude ont des courbes de croissance en fonction de la température décalées vers les basses températures en comparaison des souches isolées à basse latitude, et celles isolées à des latitudes moyennes montrent un profil intermédiaire (Figure 33). De façon intéressante, une certaine variété de réponse est observée pour le clade I dont 3 souches isolées à des latitudes plus ou moins élevées étaient étudiées. Ces mêmes auteurs ont également démontré que les souches de clades I et IV de *Synechococcus* ont été isolées dans des eaux plus froides (température toujours inférieure à 19°C) que les souches des clades II et III (température presque toujours supérieure à 19°C), ce qui confirme les résultats obtenus par les études de répartition globale et les a amené à qualifier les clades I et IV de « thermotypes » froids.

[11] L'acclimation désigne la platicité phénotypique d'un organisme, qui modifie son phénotype en réponse à une variation environnementale.

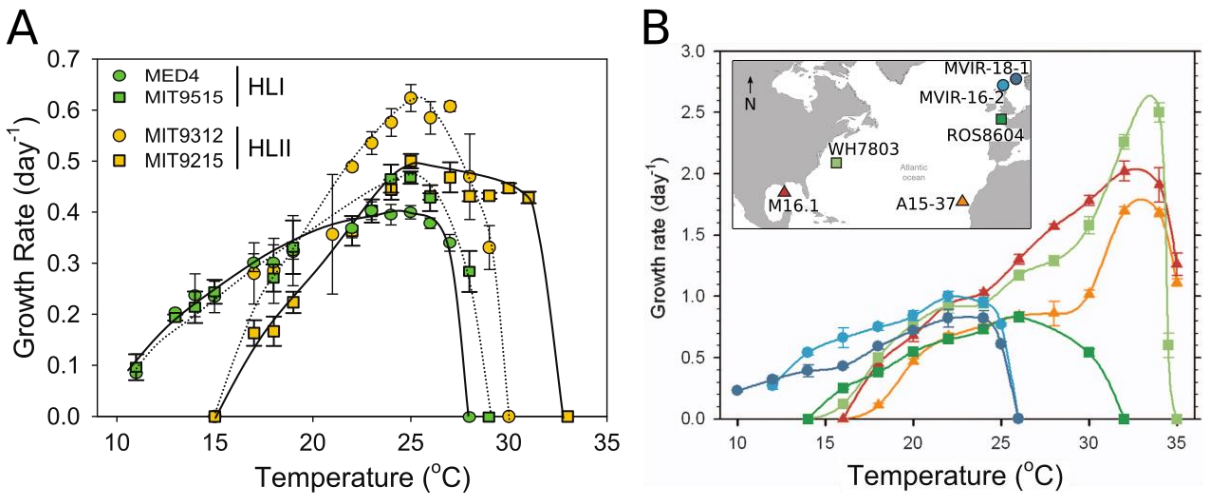


Figure 33 Preferenda thermiques de différentes souches de *Prochlorococcus* et *Synechococcus*. A. Taux de croissance en fonction de la température pour 4 souches de *Prochlorococcus*. Figure modifiée à partir de (Johnson *et al.*, 2006). B. Taux de croissance en fonction de la température pour 5 souches de *Synechococcus* dont le lieu d'isolement est indiqué sur la carte en encart : MVIR-18-1 (clade I), MVIR-16-2 (clade I), ROS8604 (clade I), M16.1 (clade II), A15-37 (clade II) et WH7803 (clade V). Figure modifiée à partir de (Pittera *et al.*, 2014).

3.3.2 Effets des changements de température sur la physiologie cellulaire

Malgré la distinction claire de thermotypes chez les picocyanobactéries marines, peu d'études se sont intéressées aux bases physiologiques des différences phénotypiques observées. Seules quelques études récentes portant sur *Synechococcus* décrivent l'effet d'un changement de température de croissance sur le métabolisme cellulaire, et commencent à expliquer comment les picocyanobactéries marines s'acclimatent à différentes températures (en particulier aux faibles températures).

De façon générale, le taux croissance augmente avec température (Figure 33), ce qui est sans doute dû au fait que les températures plus élevées augmentent la vitesse des réactions chimiques, et donc favorisent la production de biomasse. Cette augmentation s'arrête brutalement au-delà d'une température limite dépendant de la souche. On considère en général que cette limite correspond à la température à partir de laquelle certaines enzymes clé sont dénaturées, empêchant les cellules de croître et/ou à une température où les membranes deviennent trop fluides. Les basses températures ont quant à elles les effets inverses, d'une part le ralentissement des réactions métaboliques, et d'autre part la rigidification des membranes. Ce dernier aspect a un effet important sur la photophysiologie cellulaire : les cellules soumises à un stress froid (diminution brutale de la température de culture) montrent une diminution de l'efficacité photosynthétique du PSII (Pittera *et al.*, 2014), qui a été reliée à la perturbation des transitions d'état à basse température (Mackey *et al.*, 2013; Pittera *et al.*, 2014). Cette perturbation de la photosynthèse induit un stress oxydant (Pittera *et al.*, 2014), en particulier chez les souches de basse latitude, visible par la diminution du *pool* de D1 (Pittera *et al.*, 2014; Varkey *et al.*, 2016). Comme lors d'un stress de forte intensité

lumineuse, les phycobilisomes sont également touchés, avec un démantèlement des parties distales de leur bras (Pittera *et al.*, 2014). En explorant cette question de façon plus précise, il a ensuite été montré que les phycobilisomes (en particulier la phycocyanine) de souches de haute latitude étaient plus flexibles (moins thermostables) que ceux de souches de basse latitude, limitant les pertes d'énergie au sein du PBS à faible température et constituant donc une adaptation spécifique des thermotypes froids (Pittera *et al.*, 2016).

Pour pallier à la rigidification, les cellules modifient la composition lipidique de leurs membranes, notamment en augmentant le taux d'insaturation des lipides, ce qui a pour effet d'augmenter la fluidité membranaire (Pittera *et al., in revision*; Varkey *et al.*, 2016). Cette réponse semble également varier selon les souches, ce qui apporte une deuxième explication aux différences de capacités d'acclimatation observées (Varkey *et al.*, 2016).

3.3.3 Intégration au niveau cellulaire

La température a des effets sur de nombreuses voies métaboliques. Une faible température ralentit la transcription et la traduction, non seulement parce qu'elle diminue la vitesse des réactions mais aussi par exemple parce qu'elle favorise la stabilisation des structures secondaires des ARN. L'assimilation des nutriments est elle aussi perturbée, à la fois par le ralentissement général et par la moindre efficacité de la photosynthèse. Ces modifications métaboliques ont été invoquées comme expliquant la surexpression à 18°C par rapport à 22°C d'hélicases (qui rétablissent les structures des acides nucléiques) et de protéines impliquées dans l'assimilation de l'azote chez la souche de thermotype froid BL107 (clade IV) (Varkey *et al.*, 2016). La photosynthèse reste cependant le processus le plus susceptible d'être impacté : comme nous l'avons vu, la modification de la fluidité de la membrane peut perturber l'équilibre du flux d'électrons membranaires, ce qui favorise la dégradation du PSII et la formation de ROS. Cet effet est fortement accentué par le fait que la faible température limite la néo-synthèse de la protéine D1 et la réparation des PSII endommagés (Murata *et al.*, 2007). Les symptômes d'un stress de basse température sont donc relativement équivalents à ceux d'un stress de forte intensité lumineuse.

Du fait de la relation entre la température et la photophysiologie, il y a une forte interaction entre les effets de la température et ceux de la lumière : les cellules phytoplanctoniques de manière générale supportent mieux les basses températures quand elles sont cultivées à une faible intensité lumineuse ou, de manière équivalente, supportent mieux les intensités lumineuses élevées si la température n'est pas trop faible. Ainsi en moyenne, les espèces phytoplanctoniques voient leur température optimale de croissance diminuer de 5°C en conditions de limitation par la lumière (Edwards *et al.*, 2016).

Problématique et objectifs de la thèse

Cette introduction nous a permis de dresser un tableau non exhaustif des connaissances actuelles sur les picocyanobactéries marines, en mettant l'accent sur leur diversité. Cette diversité est relativement bien établie au niveau génétique, mais l'est moins en termes d'écologie (c'est-à-dire de l'intégration de la diversité génétique dans son environnement) et de physiologie. La question qui anime cette thèse se place dans la continuité des travaux effectués ces dernières années : quel est le lien entre la diversité génétique et l'adaptation à la niche écologique chez les picocyanobactéries marines ? Autrement dit, en quoi la diversité génétique de ce groupe permet-elle d'expliquer son succès écologique, qui se traduit par la colonisation de niches variées ? Cet objectif général se décline en plusieurs questions qui restent à explorer.

La première est la question de l'échelle taxonomique permettant de décrire l'adaptation à la niche écologique chez les picocyanobactéries marines. Nous avons souligné au cours de l'introduction l'importance du choix du niveau taxonomique utilisé pour comprendre la répartition de *Prochlorococcus* et *Synechococcus* et cette notion sera au cœur de la première partie, dont l'objectif est d'améliorer la délimitation des niches écologiques des picocyanobactéries marines, et en particulier de *Synechococcus,* genre pour lequel ces limites sont plus floues.

La deuxième question concerne les liens entre la mise en place de la diversité génétique et la mise en place d'entités adaptées à des niches écologiques distinctes. Quels sont les mécanismes de diversification génétique en œuvre, et en quoi cela nous informe-t-il sur les bases génétiques de l'adaptation à la niche écologique ? Nous nous appuierons sur les résultats de la première partie pour explorer les données de génomique disponibles et apporter des premiers éléments de réponse à cette question dans la deuxième partie de la thèse.

Enfin, il y a un décalage flagrant entre notre connaissance de la diversité génétique et génomique des picocyanobactéries marines et la disponibilité de données expérimentales permettant de les interpréter en termes d'adaptations physiologiques. Plus particulièrement, nos connaissances sur leur physiologie reposent sur l'étude de quelques souches modèles, voire sur d'autres organismes proches (notamment les cyanobactéries d'eau douce), qui ne sont probablement que peu représentatives de la diversité du groupe. Une question majeure qui reste à explorer est donc l'étendue de leur diversité physiologique au regard de leur diversité génétique. Nous exposerons dans la troisième partie les expérimentations mises en œuvre pour améliorer notre connaissance de la diversité des réponses physiologiques face aux stress environnementaux au sein du genre *Synechococcus*.

A travers l'utilisation de divers outils méthodologiques, ce travail de thèse explore donc la question de l'adaptation à la niche environnementale par une approche intégrative, allant de l'expression des gènes en condition contrôlée à la détermination de niches écologiques au niveau de l'océan mondial.

CHAPITRE I

GLOBAL DISTRIBUTION OF MARINE PICOCYANOBACTERIA

Context of the work and personal contribution

In order to better describe the relationship between organisms and their surrounding environment, a fundamental concept of ecology was invented in the early 20th century: the concept of ecological niche. Its most popular definition was given by Hutchinson in 1957: when considering an n-dimensional space in which each dimension is an environmental parameter (biotic or abiotic), the *fundamental* niche of a species corresponds to the part of the space (i.e. the n-dimensions hypervolume) in which the species can survive (Hutchinson, 1957). In his definition, Hutchinson considered that all species are dimensions of the space except the one under study. For practical reasons, this n-dimensional space is usually reduced to a smaller number of dimensions and biotic components are not always considered: a niche is often defined with regard to a few parameters of interest that can be controlled in the lab.

The concept of *realized* niche, derived from the fundamental niche, allows to take into account the combined effect of all environmental parameters, by measuring the niche directly in the environment. For each parameter (i.e. each dimension of the niche space), the range of values in which the species is observed *in situ* is used to delimit the niche. The realized niche is indeed likely to be different from the fundamental niche measured in the lab, because of the effects of both biotic interactions and dispersion that can either enlarge or reduce niche breadth.

In this chapter, we will explore different aspects of the realized niches of marine picocyanobacteria. As stated in the INTRODUCTION, previous studies have shown that the distribution of *Synechococcus* and *Prochlorococcus* populations are at least in part determined by the abiotic environmental conditions. This is particularly clear for *Prochlorococcus*, in which phylogenetic clades have specific distributions with regards to water temperature, light intensity and iron availability (Biller *et al.*, 2014a). The distribution of *Synechococcus* is in contrast less understood, first because it was less studied, but also because the diversity of this genus is much higher, making niches harder to delineate within this genus. We also noted in the INTRODUCTION that the taxonomic level used to study the distribution of picocyanobacterial taxa is fundamental. In this chapter, we will explore their distribution at different taxonomic levels, in particular using the gene encoding the cytochrome b_6, *petB*, which displays a high phylogenetic resolution for marine picocyanobacteria (Mazard *et al.*, 2012a), with the aim to detect more precise associations between picocyanobacterial populations and environmental (abiotic) conditions, by considering that each clade could potentially comprise several ecological units.

Biogeographical studies of microorganisms are most often based on the amplification of a marker gene, a method known as metabarcoding. This necessitates the use of PCR primers designed from sequences of known isolates, which can lead to several biases: some environmental populations might not be amplified and are thus overlooked, and the amplification might have different efficiencies depending on populations, leading to misestimated relative abundances. As more and more metagenomics data are being available form oceanic samples, some of these issues can be circumvented by looking for sequences of interest directly within metagenomes (Logares *et al.*, 2013). This method is particularly well suited for marine picocyanobacteria because they are sufficiently abundant to constitute a large part of metagenomic sequences, and we used it in parts

I and III to analyze the distribution of *Prochlorococcus* and *Synechococcus* using metagenomic data from two global surveys of the Ocean, *Tara* Oceans (Karsenti *et al.*, 2011; Pesant *et al.*, 2015; Sunagawa *et al.*, 2015) and the Ocean Sampling Day (Kopf *et al.*, 2015).

Personal Contribution

In the first part, which presents the analysis of *Tara* Oceans metagenomic dataset with an approach of recruitment of a marker gene from metagenomes, most of the recruitment work as well as the assembly of new sequences from environmental reads was performed by Gregory Farrant, a previous PhD student of the MaPP team. I was personally involved in the downstream analysis of the distribution of picocyanobacteria, the definition of ecologically significant taxonomic units and the delineation of their niches, as well as in the writing of the manuscript. In the second part, presenting an analysis of the *Synechococcus* community in the Svalbard area, I was involved in the part concerning the identification of OTUs and the analysis of their differential distribution as well as in the writing of the paper. Finally, I performed most of the work exposed in part 3 except experimental work on cultures acclimated to different temperatures that was mostly done by Justine Pittera and Solène Breton.

1. Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria

Résumé de l'article en français

Prochlorococcus et *Synechococcus* sont les organismes phytoplanctoniques les plus abondants et les plus répandus dans l'océan mondial. Afin de mieux comprendre les facteurs controlant leur biogéographie, une base de données de référence de séquences du gène marqueur de forte résoltion taxonomique *petB*, codant pour le cytochrome b_6, a été utilisée pour recruter les lectueres de 109 métagénomes issus de l'expédition *Tara* Oceans. Cela a permis de mettre en évidence une diversité insoupçonnée au sein de chaque genre, y compris pour les clades les plus abondants and les mieux caractérisés, et 136 séquences divergentes de *petB* ont été assemblées à partir des lectures métagénomiques, enrichissant significativement la base de données. Nous avons ensuite défini des Unités Taxonomiques Ecologiquement Significatives (UTES), c'est-à-dire des organismes appartenant au même clade et occupant une même niche océanique. Trois assemblages majeurs d'UTES ont été identifiés le long du trajet de la campagne *Tara* Oceans chez *Prochlorococcus,* et huit chez *Synechococcus*. Bien que les UTES HLIIIA et HLIVA de *Prochlorococcus* co-dominent les régions pauvres en fer de l'Océan Pacifique, le clade CRD1 et le clade non-cultivé EnvB de *Synechococcus* se sont avérés largement répandus dans cette région, avec trois UTES différentes de CRD1 et d'EnvB occupant des niches écologiques distinctes différant par la disponibilité en fer et la température. Des changements de communautés nets ont également été identifiés sur de courte distance géographique, par exemple au niveau des îles Marquises ou entre les océans Indien et Atlantique, révélant une corrélation étroite entre les assemblages d'UTES et des paramètres physico-chimiques donnés. Globalement, cette étude démontre l'existence d'une microdiversité au sein des écotypes de picocyanobactéries définis jusqu'alors, apportant de nouvelles connaissances sur l'écologie, la diversité et la biologie des deux organismes phototrophes les plus abondants sur Terre.

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]

[a]Sorbonne Universités, Université Paris 06, CNRS, UMR 7144, Station Biologique, CS 90074 Roscoff, France; [b]Department of Marine Biology and Oceanography, Institute of Marine Sciences (ICM), Consejo Superior de Investigaciones Científicas (CSIC), Barcelona ES-08003, Spain; [c]Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, NSW 2109, Australia; [d]School of Life Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom; [e]Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), Institut de Génomique, Genoscope, 91057 Evry, France; and [f]Stazione 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** b_6**, 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.I., 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).

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

[2]Present address: Food Safety, Environment, and Genetics, Matís Ltd., 113 Reykjavík, Iceland.

[3]To 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.

clades I–IX, which are congruent between different genetic markers (13, 18–21)—as one ecotype—that is, a group of phylogenetically related organisms sharing the same ecological niche (4, 22). However, the use of a high taxonomic resolution marker, the core, single-copy *petB* gene encoding cytochrome b_6, has revealed different spatially structured populations (subclades) within the major *Synechococcus* clades that were adapted to distinct niches (12), suggesting that the "clade" level might not be the most ecologically relevant taxonomic unit. Moreover, the systematic use of probes and/or PCR amplification might have led some to overlook some important genetic diversity, a drawback potentially resulting in a poor assessment of the relative proportion of cooccurring populations at any given station. In this context, the occurrence of a huge microdiversity within wild *Prochlorococcus* populations was recently demonstrated by estimating the genomic diversity within coexisting members of the HLII clade using a large-scale, single-cell genomics approach (23). Still, the congruency of phylogenies based on whole genome and internally transcribed spacer (ITS) suggests that ITS ribotype clusters coincide, in most cases, with distinct genomic backbones that would have diverged at least a few million years ago and the relative abundance of which varies through temporal and local adjustments (23). Thus, approaches using a single marker gene remain valid, but fine spatial, temporal, and taxonomic resolution is required to better understand how divergent picocyanobacterial lineages have adapted to different niches in the global ocean.

Here, we analyzed 109 metagenomic samples collected during the 2.5-y *Tara* Oceans circumnavigation (24, 25), a project surveying the diversity of marine plankton that produced nearly 11 times more nonredundant sequences than the previous Global Ocean Sampling (GOS) expedition (14). To retrieve taxonomically relevant information for picocyanobacteria and to avoid PCR-amplification biases, reads targeting the high-resolution *petB* gene (12) were recruited using a $_{mi}$Tag approach (26). Even though this approach did not give us access to the rare biodiversity, these analyses unveiled a previously unsuspected genetic diversity within both *Prochlorococcus* and *Synechococcus* genera. Clustering based on the distribution patterns of picocyanobacterial communities allowed us to define Ecologically Significant Taxonomic Units (ESTUs)—that is, genetically related subgroups within clades that cooccur in the field. Analyses of the biogeography of ESTU assemblages showed that they were strongly correlated with specific environmental cues, allowing us to define distinct realized environmental niches for the major ESTUs.

Results

Revealing Novel Picocyanobacterial Diversity Using *petB*-$_{mi}$Tags and Newly Assembled Sequences. To evaluate the taxonomic resolution potential of *petB* $_{mi}$Tags for assessing picocyanobacterial genetic diversity, simulated 100-bp reads (i.e., the minimum size of the *Tara* Oceans merged metagenomic reads) were generated by fragmenting sequences from our reference database (Datasets S1 and S2). This analysis showed that *petB* reads can be assigned reliably at the finest taxonomic level—that is, subclade (12)—over most of the gene length (Fig. S1). The *petB*-$_{mi}$Tags approach was therefore applied to the whole *Tara* Oceans transect (66 stations, 109 metagenomes, 20.2 ± 9.9 Gb of metagenomic data per sample). With the exception of the Southern Ocean and its vicinity (TARA_082 to TARA_085), for which no *petB* reads were recruited, picocyanobacteria were present at all sampled *Tara* Oceans stations. From 119 to 14,139 picocyanobacterial *petB* reads (average: 3,309; median: 2,545) (Dataset S3) were recruited per sample using a nonredundant reference database of 585 high-quality *petB* sequences, representing most of the genetic diversity identified so far among *Prochlorococcus* and *Synechococcus* isolates and environmental clone libraries (Fig. 1). Interestingly, most *petB* sequences in our database recruited at least one read from the *Tara* Oceans metagenome as best hit, with the notable exception of some sequences of the cold water-adapted *Synechococcus* clade I, likely due to the limited sampling performed at high latitudes during the *Tara* Oceans expedition (27). This suggests that most genotypes known so far are sufficiently well represented in the marine environment to be detected by this approach. Still, we cannot exclude that this preliminary analysis provides a somewhat biased picture of the diversity toward the "already known," as most current reference sequence databases are potentially skewed by culture isolation and/or amplification biases.

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

To have a more realistic and exhaustive view of this diversity, we assembled 136 distinct nearly complete *petB* sequences from environmental reads (121 *Prochlorococcus* and 15 *Synechococcus*), corresponding to the most divergent genotypes present in the whole *Tara* Oceans dataset. By adding these novel sequences to the reference database (see Dataset S1 and sequences in white or grey in Fig. 1), we significantly improved taxonomic assignments of *petB*-$_{mi}$Tags, as 80.3% of the *Prochlorococcus* and 90.2% of the *Synechococcus* environmental *petB* reads were found to display >94% identity with their best hits in the enriched reference database, an increase of about 11% and 7% compared with our initial assessment, respectively (Fig. 2*B* and Fig. S2). Interestingly, quite a few highly divergent sequences from *Prochlorococcus* HLIII, HLIV, and LLI as well as *Synechococcus* CRD1 were assembled from TARA_052, located East of Madagascar, a station exhibiting a picocyanobacterial community atypical for this oceanic area. Although most of these additional sequences fell into known phylogenetic clades, they allowed us to better assess the extent of genetic diversity within both *Prochlorococcus* and *Synechococcus* (Fig. 1). Although only a few *petB* sequences, all coming from cultured strains, were available for the *Prochlorococcus* HLI and LLI clades before this study, we added 43 HLI sequences (within-clade nucleotide identity range: 87–99.6%), 29 LLI sequences (within-clade identity range: 85.5–99.6%), as well as 11 sequences of the uncultured HLIII and IV clades, some of which form distinct monophyletic branches comprised entirely of novel sequences (Fig. 1 and Dataset S1). Although many HLII sequences were recently obtained by high-throughput single-cell genomics focused on this clade (23), assembly of *Tara* Oceans reads allowed us to retrieve several divergent HLII sequences (within-clade identity range: 86.2–99.8%) including a previously unidentified well-supported group (corresponding to ESTU HLIIC), located at the base of the HLII radiation. Similarly for *Synechococcus*, newly assembled sequences allowed us to refine the taxonomy of several taxa, notably for CRD1 and EnvB clades as well as subcluster 5.3, three ecologically important but previously overlooked phylogenetic lineages.

Using Global Picocyanobacterial Distribution Patterns to Define ESTUs. As expected from previous literature (1, 2, 5, 28), *Prochlorococcus* was the most abundant picocyanobacterium at the global scale, representing ~91% of all *petB* reads from the bacterial size fraction, compared with 9% for *Synechococcus* (Fig. S3*A*). These percentages compare fairly well with the global contribution of *Prochlorococcus* and *Synechococcus* estimated from flow cytometry data as 80.6% ($2.9 \pm 0.1 \times 10^{27}$ cells) and 19.4% ($7.0 \pm 0.3 \times 10^{26}$ cells), respectively (1). The apparent lower contribution of *Synechococcus* in our dataset might be due to the fact that the *Tara* Oceans sampling was not made at random in the ocean, as

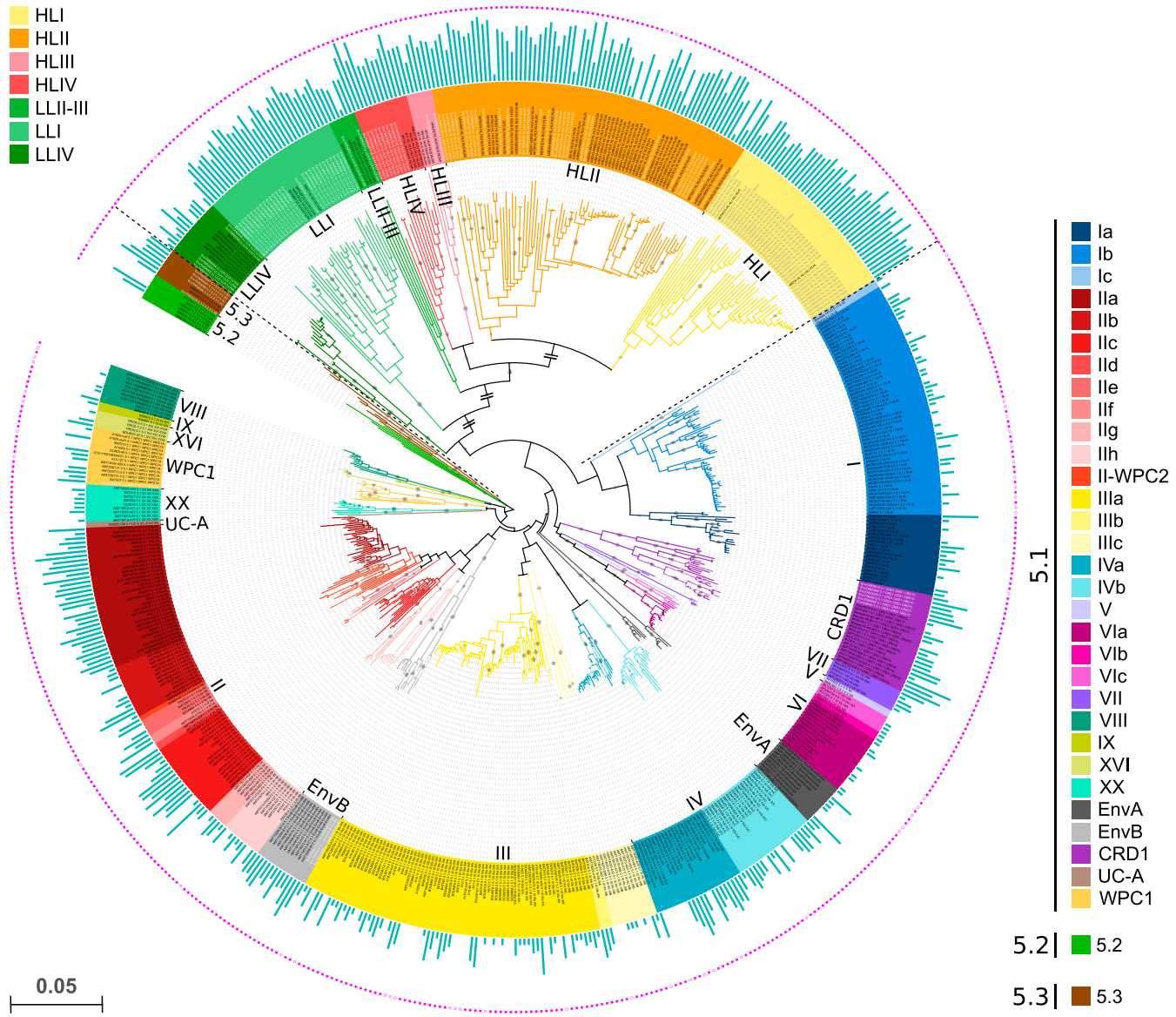


Fig. 1. Neighbor joining tree of *Synechococcus* and *Prochlorococcus* lineages based on *petB* gene sequences from both isolates and environmental sequences. Diamonds at nodes indicate bootstrap support over 70%. Taxonomic assignments are given by the color codes at clade level for *Prochlorococcus*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

most stations were located in the intertropical zone and/or selected for displaying specific traits of interest (e.g., upwelling, fronts, island proximity, etc.), whereas Flombaum et al.'s (1) dataset included many data from temperate stations, where *Synechococcus* is often abundant.

To study the global distribution of these organisms at a finer taxonomic resolution, we then examined whether *Prochlorococcus* and *Synechococcus* clades and/or subclades were ecologically meaningful. To do this, we analyzed the distribution patterns along the *Tara* Oceans transect of within-clade Operational Taxonomic Units (OTUs), as defined using a cutoff at 94% nucleotide identity (Fig. 2*C*, Fig. S4, and Dataset S4). Although for some clades OTUs displayed a homogeneous pattern over their geographical distribution area (e.g., *Prochlorococcus* HLIII and IV; Fig. S4) or were too scarce to reliably distinguish ESTUs (*Synechococcus* subcluster 5.2 and clades I, V–VIII, WPC1, EnvA, IX, XVI, XX, UC-A, and *Prochlorococcus* clades LLII–IV), most of the prevalent clades encompassed several coherent OTU clusters displaying distinct distribution patterns (and thus likely occupying different ecological niches) that were gathered into independent ESTUs (Fig. 2*C* and Fig. S4). For instance, OTUs within *Synechococcus* clade CRD1 could be split into three ESTUs (CRD1A–C) based on clustering of their abundance per station. Some of these ESTUs corresponded to previously described clades (e.g., *Prochlorococcus* HLIIIA and HLIVA) or subclades (e.g., *Synechococcus* IVC), whereas others gathered subclades having similar distribution patterns. For instance, *Synechococcus* ESTU IIA encompasses subclades IIa–d and IIf, and ESTU IIB gathers subclades IIe and

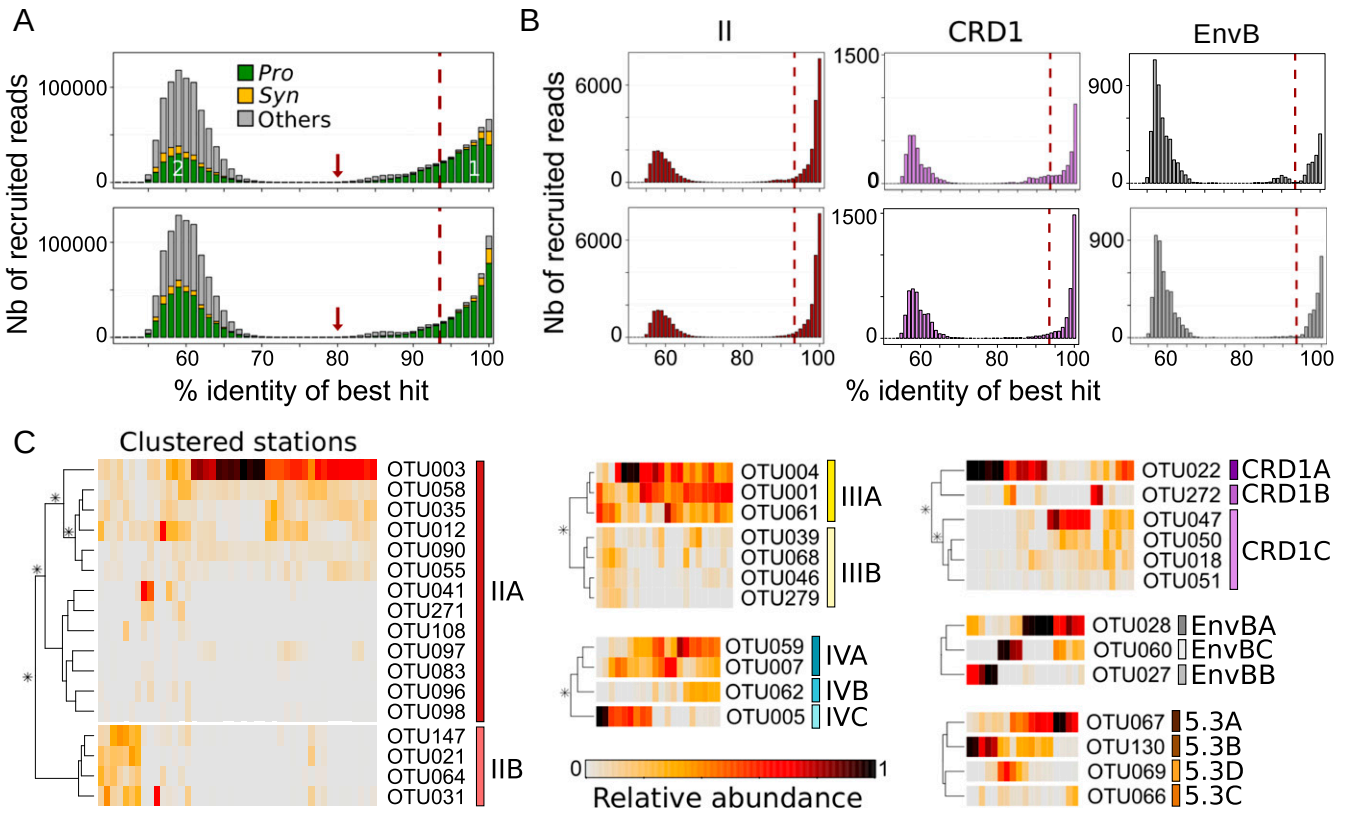


Fig. 2. Percent identity of *Tara* Oceans *petB*-$_{mi}$tags versus sequences of the reference database and abundance at different stations along the transect of OTUs clustered into ESTUs. (*A*) Distribution of the percent identity of best hits of all *petB* candidate reads recruited from the *Tara* Oceans bacterial-size fraction metagenomes against the *petB* reference database. Populations 1 and 2 correspond, respectively, to genuine *petB* reads and to nonspecific signal, due either to *petB* reads from organisms not included in the reference database or to *petB*-related genes. The gray part in population 1 corresponds to *petB* reads attributable to photosynthetic organisms of the reference database other than *Prochlorococcus* and *Synechococcus*. The red arrow shows the 80% cutoff used to separate the *petB* signal from noise. The *Top* and *Bottom* panels correspond to recruitments made before and after addition of the 136 newly assembled environmental *petB* sequences, respectively. (*B*) Same as *A* but for some selected *Synechococcus* taxa (see Fig. S2 for all other picocyanobacterial taxa). (*C*) Determination of ESTUs based on the distribution patterns of within-clade 94% OTUs. At each station, the number of reads assigned to a given OTU is normalized by the total number of reads assigned to the clade in this station. Stations and OTUs are filtered based on the number of reads recruited and hierarchically clustered (Bray–Curtis distance) according to distribution pattern. Only *Synechococcus* clades split into different ESTUs are shown (see Fig. S4 for *Prochlorococcus*). Stars indicate nodes supported by a P value < 0.05 as determined using similarity profile analysis (SIMPROF; test not applicable to pair comparisons).

IIh, as previously defined by Mazard et al. (12). Thus, although most previous field diversity studies on picocyanobacteria focused on clades (5, 13, 17, 20, 21), which were generally considered as distinct "ecotypes" (*sensu*, ref. 19), our data indicate that ESTUs provide a finer estimate of *Prochlorococcus* and *Synechococcus* ecotypes than do clades. These ESTUs were then used to study the biogeography of marine picocyanobacteria by clustering together stations exhibiting similar ESTU assemblages (Figs. 3*A* and 4*A*).

Biogeography of *Prochlorococcus* Reveals the Occurrence of Minor ESTUs with Unexpected Distribution Patterns. Most major *Prochlorococcus* clades (HLI, HLII, and LLI) could be split into several ESTUs, although for the former two, one ESTU was clearly predominant (Fig. 3*A* and Figs. S5 and S6). Only three major ESTU assemblages were identified in surface samples: (*i*) dominance of HLIA ESTU in temperate waters (above 35°N and 32°S); (*ii*) dominance of HLIIA in warm and iron-replete waters between 30°S and 30°N, with mixed HLIA–HLIIA profiles at intermediate latitudes; and (*iii*) cooccurrence of HLIIIA and IVA at a ratio of ca. 1:2.6 (±0.7) in warm, HNLC areas. The low abundance of LLII–IV clades in the whole *Tara* Oceans dataset (Fig. S3 *A*–*C*) is likely due to the fact that they usually thrive below the deep chlorophyll maximum (DCM) (5, 29)—that is, at depths not sampled during the expedition. In contrast, most LLI ESTUs were very abundant in subsurface waters (Figs. S3 and S6) and sometimes even reached the surface (e.g., at TARA_066-070; Fig. 3*A*), as expected from the ability of members of the LLI clade to tolerate a strong mixing rate and short-term exposure to high light (5, 8, 29, 30).

HLIIIA and HLIVA ESTUs altogether contributed to 15.5% of the *Prochlorococcus* community in *Tara* Oceans samples—that is, about as much as HLI (17%) or LLI (15.2%) (Fig. S3*A*). This value is slightly higher than the 9% previously estimated for HLIII–IV clades from the analysis of GOS samples (11). Consistent with previous studies (11, 15, 31, 32), we show here that their distribution covers most of the warm (>25 °C), low-Fe equatorial Pacific zone from 13°S (TARA_100) to 14°N (TARA_137), where they constitute the vast majority of the *Prochlorococcus* community in surface waters. In the Indian Ocean, we only observed them at two stations near the northern coast of Madagascar (TARA_052 and TARA_056), in agreement with a previous report that found them at two sites located further east (31), all these sites likely being influenced by the Indonesian throughflow originating from the tropical Pacific Ocean (33). Thus, HLIII/IV seemingly occurs over a much thinner latitudinal band (centered around 15°S) in the Indian compared with the Pacific Ocean, and they are apparently very scarce in the part of the Atlantic Ocean explored by the *Tara* schooner, even though the area around stations TARA_072 and TARA_070 is known to be iron-depleted (see figure S1 in ref. 17). Altogether, the distribution patterns of the dominant *Prochlorococcus* HL ESTUs seem to be mainly driven by temperature and iron availability, as confirmed by nonmetric multidimensional scaling

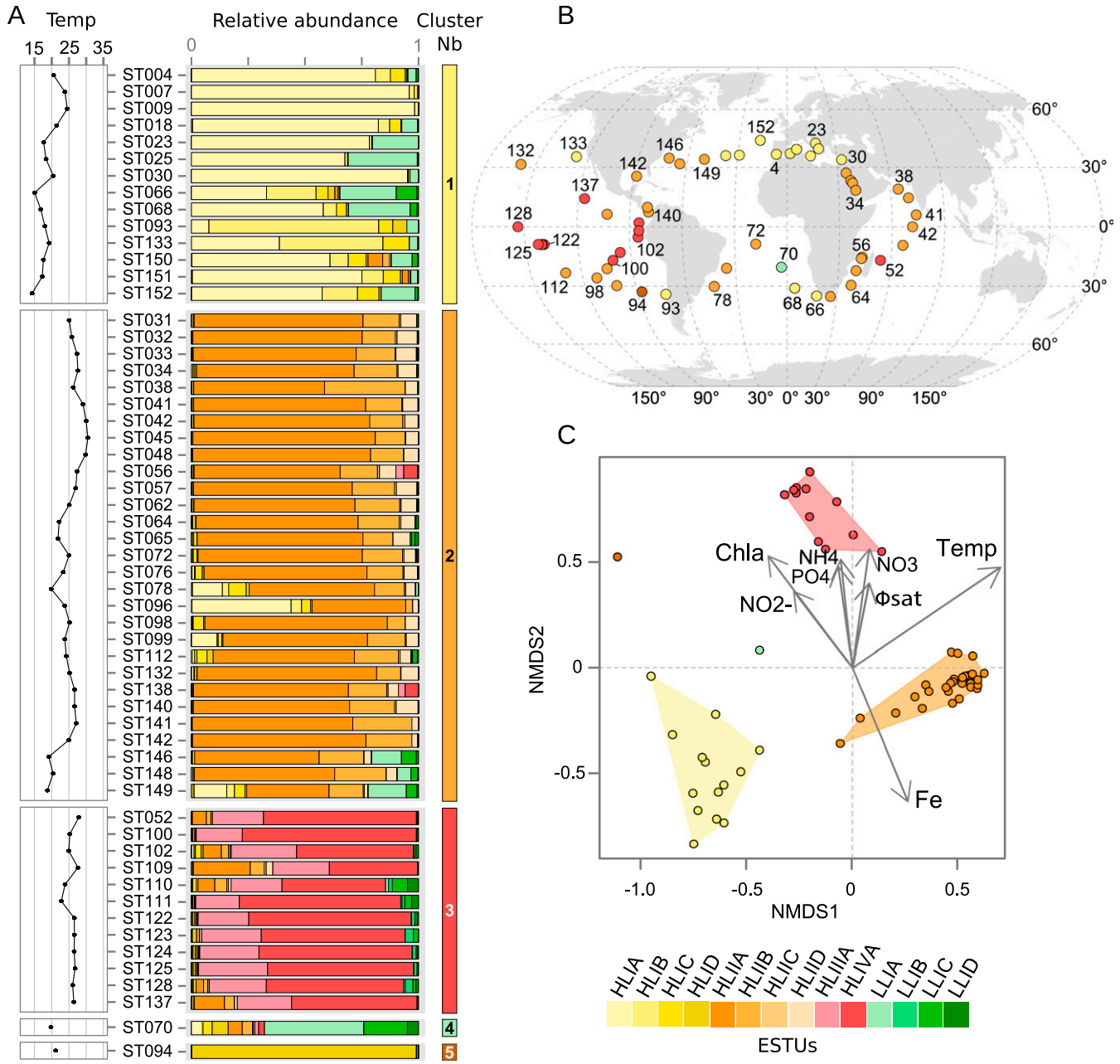


Fig. 3. Biogeography of *Prochlorococcus* ESTUs in surface *Tara* Oceans metagenomes and relation to physico-chemical parameters. (*A*) Histograms of the relative abundance of *Prochlorococcus* ESTUs at each station sorted by similarity, as determined by hierarchical clustering (Bray–Curtis distance).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

(NMDS) analyses (Fig. 3*C*). These results are globally consistent with previous reports that analyzed *Prochlorococcus* clades (5, 8, 15, 29, 31), indicating that the latter studies actually targeted the dominant ESTUs.

In contrast, a number of minor ESTUs were found to display distribution patterns very different from the major ESTUs of the same clade. For instance, the relative contribution of the previously mentioned novel HLIIC ESTU was highest at the DCM in the equatorial Indian Ocean (TARA_041-042; Fig. S6), suggesting that members of this ESTU are adapted to middepth waters, much like members of the LLI clade (5, 29). Similarly, ESTUs HLIB and -D can sometimes take over the prevalent HLIA populations and become abundant in surface waters at specific locations (e.g., at TARA_093 and TARA_094, respectively). In contrast, HLIC, which comprises a complex microdiversity (10 OTUs; Fig. S4), was found to exhibit a particularly large niche, cooccurring with HLIA at high latitude but also being present as the major HLI population in warm oligotrophic waters, where HLIIA dominated the *Prochlorococcus* community (e.g., in the Indian Ocean; Fig. S7*A*). This suggests that members of the HLIC ESTU might have a larger tolerance to temperature than the globally dominant HLIA. It is also worth noting that among the four ESTUs defined within the LLI clade, LLIB, which is entirely comprised of newly assembled *petB* sequences, dominates the LLI

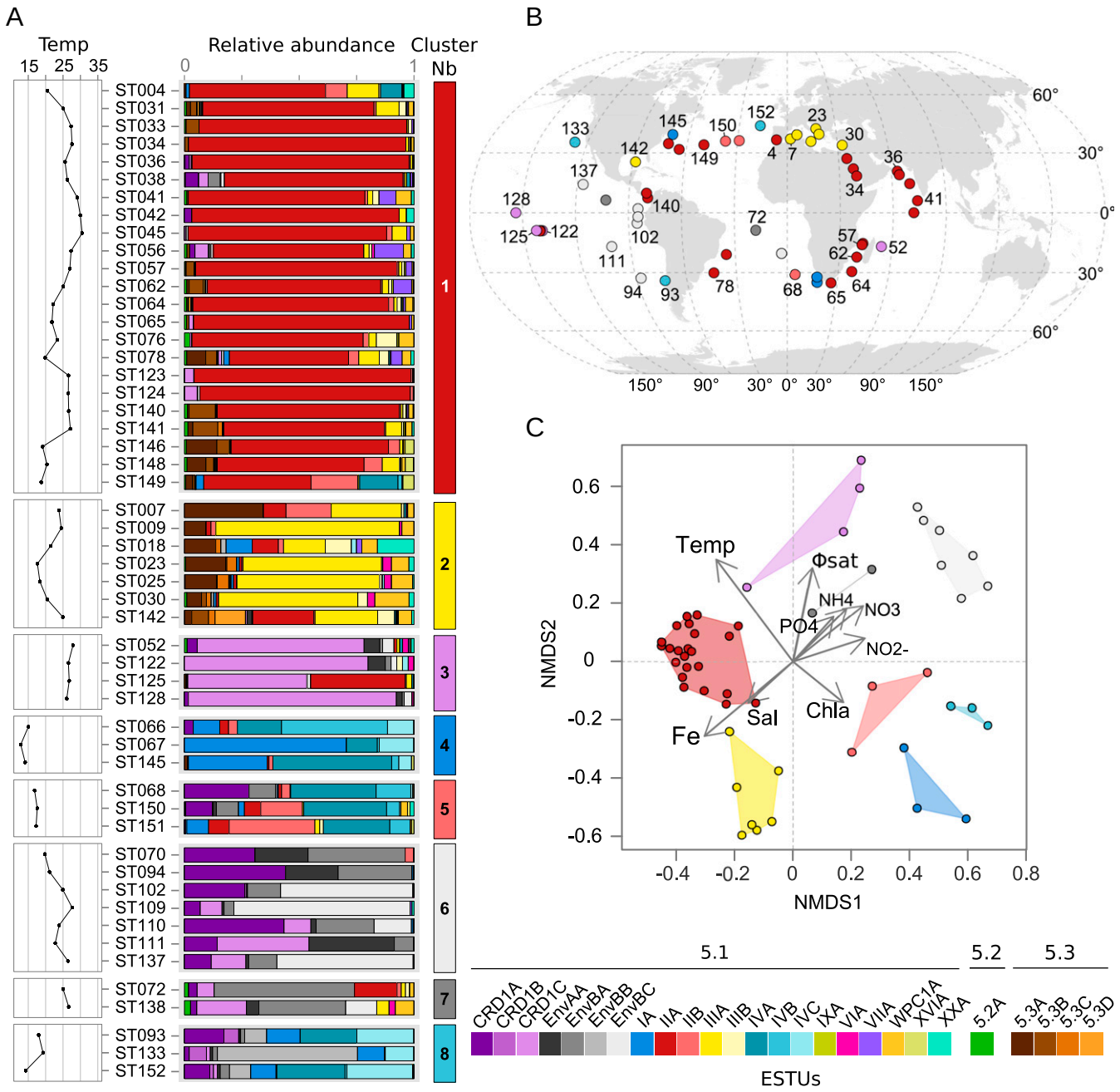


Fig. 4. Same as Fig. 3 but for *Synechococcus*. NMDS stress value: 0.1369.

population in surface iron-limited HNLC areas in both the equatorial/tropical Pacific Ocean (TARA_110 to 128) and Indian Ocean (TARA_052) (Fig. S7*B*). Thus, adaptation to low iron conditions in *Prochlorococcus* might not be an exclusive trait of HLIIIA and HLIVA.

CRD1 and EnvB ESTUs Are the Dominant *Synechococcus* Lineages in the Pacific Ocean. *Synechococcus* assemblages were much more diverse than *Prochlorococcus*, with eight distinct ESTU clusters observed along the *Tara* Oceans transect (Fig. 4 *A* and *B*). None of these assemblages were specific to a given oceanic region, although cluster 2 was mainly found in the Mediterranean Sea. ESTUs IA and IVA, IVB, and/or IVC dominated at most stations within clusters 4, 5, and 8 that were typical of cold, coastal, or mixed open ocean waters at high latitude, in agreement with previous reports on the distribution of clades I and IV (11–13, 17). In contrast, ESTU IIA, dominated by a single OTU (OTU003; Fig. 2*C*), was by far the major component of cluster 1, an assemblage characteristic of most warm, mesotrophic, and oligotrophic iron-replete waters that encompass the vast majority of the Atlantic and Indian Oceans (Fig. 4*B*). Consistently, NMDS analysis showed that the occurrence of clusters 4, 5, and 8, on the one hand, and cluster 1, on the other hand, was associated with both temperature and Chl *a*, but in opposite ways (Fig. 4*C* and Fig. S8). Interestingly, although ESTU IIA was typical of warm waters, the minor ESTU IIB was found to be restricted to fairly cold (14.1–17.5 °C), mixed waters and to cooccur with IVA and -B (Fig. 4).

Several other salient features arose from analyses of the *Tara* Oceans metagenomes. First, ESTU IIIA, the major contributor of cluster 2, was found only in the Mediterranean Sea (TARA_007 to 030) and the Gulf of Mexico (TARA_142) (Fig. 4 *A* and *B*). Both areas are known to be P-depleted (34, 35), suggesting that the

dominance of this ESTU could be linked to a specific adaptation to P limitation, as confirmed by the inverse correlation of cluster 2 with P concentrations (Fig. 4*C*) and correlation analyses between IIIA and individual physico-chemical parameters (Fig. S8). The differential availability of this nutrient on both sides of the Suez Canal is therefore probably responsible for the strong community shift from a IIIA- to a IIA-dominated assemblage between the Mediterranean and Red Sea (Fig. S5), although one cannot exclude that other specific characteristics of the Mediterranean Sea, such as the presence in the eastern basin of copper, a trace metal toxic to a number of phytoplankton species (36), might also be involved. Although the dominance of clade III in the Mediterranean Sea is consistent with previous studies (13, 37), it was also reported in fair abundance along a N–S transect in the northern Atlantic Ocean in fall 2004 (AMT15) as well as in subtropical waters of the Pacific and Atlantic oceans (12, 13), whereas we found it only as a minor component of the *Synechococcus* community in these areas. It is possible that the relative contribution of clade III might have been overestimated using PCR-based or dot-blot hybridization approaches. A more likely explanation is that this clade is subject to seasonality, as suggested by a year-round survey in the Red Sea showing that clade III abundance peaks occur during summer and stratified conditions and remains at low concentrations over the rest of the year (19, 38). In this context, it is important to note that during *Tara* Oceans, the north and south Atlantic as well as the southern Indian Ocean were all sampled during winter or early spring, whereas the Mediterranean Sea was sampled in fall (Dataset S3). Hence, this warrants future global metagenomic studies at various seasons or studies at finer geographical scale looking at seasonal variations in community structure.

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

Another striking result of this study was the strong global contribution of the cooccurring clades CRD1 and EnvB (8.4% and 5.4% of total *Synechococcus* reads, respectively; Fig. S3 *D* and *E*). Recently, low-Fe regions of the western equatorial Pacific (5°S–10°N) and southeastern Atlantic Oceans (15–20°S) were shown to be dominated by CRD1 (16, 17), a clade that was previously thought to be specific to the Costa Rica dome, where *Synechococcus* cell densities are known to be the highest worldwide (40, 41). Here, we show that CRD1 and EnvB ESTUs actually codominate the *Synechococcus* community over most of the Pacific Ocean from 33°S to 35°N and can also be prevalent in both the South (TARA_068-072) and North Atlantic (TARA_150-152) as well as in the Indian Ocean (TARA_052) but are seemingly absent from the Mediterranean Sea (Fig. 4 *A* and *B*). So it seems that, in contrast to *Prochlorococcus* HLIII/IV, the distribution of CRD1 in the Pacific Ocean extends way beyond HNLC areas. Furthermore, we show here that both the CRD1 and EnvB clades actually encompassed three distinct ESTUs, displaying partially overlapping niches and falling into five clusters (3 and 5–8; Fig. 4*A*) that were also split far apart by NMDS analyses (Fig. 4*C*). CRD1B and EnvBB were restricted to high latitude, cold, mixed waters (cluster 8), where they systematically codominated with ESTU IA, IVA, and IVC. This includes TARA_093 located in the Chilean upwelling, TARA_152 in North Atlantic, as well as TARA_068 in South Atlantic corresponding to a young Agulhas ring (42). In contrast, CRD1C and EnvBC preferentially thrived in warm HNLC regions (cluster 3 and the warmest stations of cluster 6), with CRD1C largely dominating the *Synechococcus* population in the Pacific intertropical area as well as at the Indian Ocean station TARA_052. Comparatively, CRD1A and EnvBA, which were found in both kinds of environments, appear to be much more ubiquitous and to tolerate a much wider temperature range, not only than other CRD1 and EnvB ESTUs but also more generally than all other *Synechococcus* strains characterized so far in culture (9, 10). Several previous studies also reported the presence of CRD2, cooccurring with CRD1 mainly in the Costa Rica dome area and in equatorial waters and generally constituting around 10–15% of the total *Synechococcus* surface population (16, 17). It is tempting to speculate that the *petB*-defined EnvB clade, which had so far only been reported at one station in the middle of the North Atlantic basin (12), corresponds to the ITS-defined CRD2 clade. However, the different proportions of EnvB and CRD2 relative to CRD1 strongly suggests that the quantitative PCR primers used in these studies targeted only a fraction of the CRD2/EnvB population, possibly corresponding to EnvBC, which like CRD2, is positively correlated with temperature (17) (Fig. S8). Alternatively, seasonal variations might also explain the differences observed between these two datasets.

Discussion

The comprehensive nature of the *Tara* Oceans dataset, analyzed here at high taxonomic resolution, has markedly improved our current knowledge of the global phylogeography of marine picocyanobacteria and highlighted the key role of environmental parameters in shaping their distribution patterns. Indeed, by assigning *petB*-$_{mi}$Tags recruited for each clade to narrow OTUs, then clustering those sharing a similar ecological distribution into the same ESTU, we showed that despite a wide genetic diversity, *Prochlorococcus* and *Synechococcus* communities can be split into a fairly limited number of characteristic ESTU assemblages, often dominated by one or two major ESTUs. This includes the codominating *Prochlorococcus* HLIIIA–HLIVA, which occurred at a fairly constant ratio (1:2.6) throughout low-Fe regions (Fig. 3*A*); *Synechococcus* IIIA, which was abundant all over the Mediterranean Sea; and CRD1 and EnvB ESTUs, codominating the *Synechococcus* community in vast expanses of the Pacific Ocean (Fig. 4*A*). Interestingly, we also showed that most picocyanobacterial clades encompass minor ESTUs that occupy niches distinct from dominant ones. This indicates that there is ecologically meaningful fine-scale diversity within currently defined *Synechococcus* or *Prochlorococcus* clades, even though the latter have often be referred to as ecotypes (5, 29). In this context, it is important to note that the *Prochlorococcus* genus is thought to have occurred concomitantly to the major diversification event that also led to the splitting of *Synechococcus* subcluster 5.1 into about 15 distinct clades (20, 43, 44), suggesting that, from a phylogenetic point of view, the whole *Prochlorococcus* genus is actually equivalent to a single *Synechococcus* clade, explaining why linking clades to a given ecological niche is trickier for the latter genus. In *Prochlorococcus*, several physico-chemical parameters have seemingly played a decisive role in the genetic diversification of this genus, at distinct periods of its evolutionary history, starting with light (split between LL and HL lineages), then iron availability (HLIII/IV vs. other HL), and temperature (HLI vs. HLII) (18, 21, 45). In contrast, nitrogen and phosphorus availability influenced genetic diversification only in the "leaves" of the *Prochlorococcus* radiation, through lateral transfers of gene cassettes conferring on populations the ability to adapt to local N- or P-depleted niches (46, 47). Despite this apparent solid relationship between *Prochlorococcus* phylogeny and community structure, a recent study looking at the genomic diversity of individual *Prochlorococcus* cells in a single water sample highlighted a huge microdiversity within the HLII clade (23). This microdiversity seemingly allows cells to adapt to slightly different selective pressures, such as biotic factors

(phages, grazing, etc.). Here, we also observed large microdiversity within the HLII lineage, with 25 OTUs comprising four ESTUs, but in agreement with a recent study (48), there were only subtle differences between the distribution patterns of these intraclade groups (except for ESTU HLIIC, represented by a single OTU; Fig. S4), confirming that abiotic factors have only marginally affected the genetic diversification within this clade. In contrast, the microdiversity that we identified within HLI and LLI has seemingly allowed members of these clades to colonize ecological niches clearly different from that of the dominant ESTUs, extending the global niche occupied by these lineages. This includes LLIB, which seems to be adapted to Fe-limited surface waters, much like HLIIIA–IVA, as well as HLIC, which thrives not only in cold temperate waters, as do the more typical HLIA, but also in warm subtropical waters, where it cooccurs with the dominant HLIIA (Fig. S7*A*). This is consistent with the recent finding that HLI subclades are driven by distinct environmental traits (48) and that even in HLII-dominated waters, HLI is never competed to extinction (7).

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

Focusing on shifts in community composition associated with changes in local environmental conditions or to physical barriers (Figs. S5 and S6) provided additional insights into this global picture and revealed that some ESTUs behave as opportunists. For instance, this is the case off the Marquesas Islands, where the proximity of the coast induced an iron enrichment at TARA_123 and 124 compared with a typical HNLC situation at TARA_122 and TARA_128. Although CRD1C dominated at the latter stations, ESTU IIA took over this local population in these iron-replete patches (with an intermediate situation at TARA_125; Fig. S5). By comparison, the *Prochlorococcus* abundance drastically dropped at TARA_123 but without any significant change in the community structure, suggesting that the minor HLIIA component of this assemblage was not responsive enough to local Fe enrichment to outcompete the dominant HLIIIA/IVA population. Another abrupt shift in community composition occurred at the Agulhas choke point off the southern tip of Africa, where huge anticyclonic rings (i.e., Agulhas rings) are formed in the Indian Ocean and then drift across the South Atlantic (42, 52). The strong drop in temperature, occurring within the youngest ring (TARA_068), was likely responsible for a large part in the shift from a typical subtropical ESTU assemblage in the Indian Ocean, dominated by *Prochlorococcus* HLIIA–B and *Synechococcus* IIA (TARA_064-065), to a cold-water ESTU assemblage (HLIA, LLIA, CRD1A, EnvBA, and IVA–B) at TARA_068 (Fig. S5), suggesting that the latter ESTUs might also have an opportunistic behavior with regard to their warm-waters counterparts. Although these two examples correspond to biogeochemical processes likely occurring at different time scales, the observed ESTU assemblage changes likely result from differences in the intrinsic dynamics of ESTUs

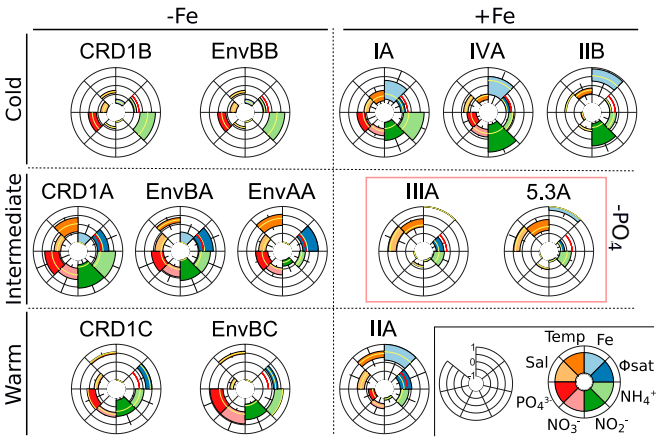


Fig. 5. Realized environmental niche of the major *Synechococcus* ESTUs in surface waters. For each ESTU, stations were sorted by order of normalized abundance, and only stations cumulating 80% of the total abundance were used to draw the graph. Boxplots represent the range of each parameter (in relative units) tolerated by any given ESTU, and the median is indicated by a yellow line. ESTUs are organized according to their relative temperature range (cold, intermediate, or warm), tolerance to iron limitation (–Fe, +Fe), and tolerance to phosphate limitation (–PO_4). Please note that the two proxies used to estimate Fe limitation ([Fe] derived from the ECCO2-Darwin model and the Φsat index; the red line indicates the 1.4% value above which iron is considered limiting) (56) are sometimes contradictory—for example, for CRD1B and EnvBB.

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

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

In conclusion, although very few studies have so far combined information from high-resolution phylogenetic markers and geographical distribution to detect ecologically coherent taxonomic groups (e.g., refs. 48, 53), we show here that this approach can bring invaluable insights for deciphering the links between genetic diversity and niche occupancy. Indeed, the definition of

within-clade ESTUs using a reference *petB* database enriched with ecologically relevant and distantly related sequences assembled from *Tara* Oceans reads has allowed us to obtain clear-cut spatial distribution patterns for taxa within both *Prochlorococcus* and *Synechococcus* genera, indicating that we explored the diversity of the picocyanobacterial community at the right taxonomic resolution. Additionally, in contrast to other phytoplankton groups, such as diatoms (54), these biogeographical patterns were found to be tightly controlled by environmental factors. Besides helping to refine models of picocyanobacterial distributions and predicting their behavior in response to ongoing climate change, knowledge of the oceanic areas where poorly characterized ESTUs predominate will also guide future strain isolation (e.g., for the yet uncultured EnvA and EnvB) and sequencing efforts. Characterizing and comparing such ecologically representative strains will help further unveil the basis of niche partitioning.

Materials and Methods

Genomic Material. This study focused on 109 *Tara* Oceans metagenomes corresponding to 66 stations along the *Tara* Oceans transect for which a "bacterial size fraction" was available (i.e., 0.2–1.6 μm for TARA_004 to TARA_052 and 0.2–3 μm for TARA_056 to TARA_152). Water samples were collected at two depths, surface and DCM, the latter sample sometimes being merely collected in the upper mixed layer, when the DCM was not clearly delineated (Dataset S3). Metagenomes were sequenced using the Illumina technology as overlapping paired reads of ∼100/108 bp with various sequencing depths, ranging from 16×10^6 to 258×10^6 reads after quality control, corresponding to an average 20.2 ± 9.9 Gb of sequence data per sample. Reads were merged using FLASH v1.2.7 with default parameters (55) and cleaned based on quality using CLC QualityTrim v4.10.86742 (CLC Bio), resulting in 100–215-bp fragments. Dataset S3 describes all metagenomic samples with location and sequencing effort. All metagenomes and corresponding environmental parameters measured during the *Tara* Oceans expedition are available at www.pangaea.de/, except for the iron and ammonium data, which were simulated with the Estimating the Circulation and Climate of the Ocean (ECCO2)-Darwin model and the iron limitation index Φsat (56) and are available in Dataset S3.

Building of the PetB Database. To recruit and taxonomically assign metagenomics reads targeting the high-resolution *petB* gene marker, we analyzed 1,091 sequences of the *petB* gene from cultured isolates and environmental samples and built a reference database including all nonredundant high-quality sequences of this marker available for the marine picocyanobacteria *Prochlorococcus* (69 sequences covering 7 clades) and *Synechococcus* (399 sequences covering 3 subclusters, 22 clades, and 30 subclades). The dataset also includes outgroup sequences from publicly available cyanobacteria, including marine (13 sequences) and freshwater isolates (40 sequences), as well as representatives of the main marine eukaryotic phytoplankton taxa and eukaryotic cyanobionts (64 plastid *petB* sequences), raising the number of *petB* sequences to 585 (Datasets S1 and S2). To avoid differential alignment effects at the edge of the reference sequences, all sequences were aligned and trimmed to 557 bp. This database was secondarily complemented by 136 *petB* sequences assembled from selected *Tara* Oceans reads displaying less than 94% identity with previously known *petB* sequences (yet some of these sequences could exhibit more than 94% identity with one another).

Read Recruitments. Targeted *petB* fragment recruitments were performed using a two-step protocol. To maximize the diversity while reducing the weight of the resulting tabulated files, translated sequences of the nonredundant *petB* database were used to recruit candidate *petB* gene fragments by BLASTX (v2.2.28+) using default parameters but by limiting the results to one target sequence. These *petB* candidates were then compared with the full reference *petB* database using BLASTN (v2.2.28+) with sensitive configuration (–task blastn –gapopen 8 –gapextend 6 –reward 5 –penalty -4 –word_size 8) and cutoffs to reduce the weight of resulting tabulated files (–perc_identity 50 –evalue 0.0001).

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

Phylogenetic and Statistical Analyses. Phylogenetic reconstructions were based on multiple alignments of *petB* nucleotide sequences generated using MAFFT v7.164b with default parameters (57). A maximum likelihood tree was inferred using PHYML v3.0–20120412 (58), with the HKY + G substitution model, as determined using jModeltest v2.1.4 (59), and the estimation of the gamma distribution parameter of the substitution rates among sites and of the proportion of invariables sites. Confidence of branch points was determined by performing bootstrap analyses including 1,000 replicate datasets. Phylogenetic trees were edited using the Archaeopteryx v0.9901 beta program (60) and drawn using iTOL (itol.embl.de) (61). OTUs for the *petB* reference dataset at 94% were defined by nucleotide identity using Mothur v1.34.4 (62).

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

Hierarchical clustering and NMDS analyses of stations were performed using R packages *cluster* v1.14.4 (64) and *MASS* v7.3–29 (65), respectively. *petB*-$_{mi}$Tag contingency tables aggregated at the ESTU level were filtered as above and normalized using Hellinger transformation that gives lower rates to rare ESTUs. The Bray–Curtis distance was then used for both clustering (*agnes* function, default parameters) and ordination (*isoMDS* function; maxit, 100; k, 2). All displayed clusters were significant ($P < 0.01$, permutation tests). Fitting of environmental parameters on NMDS ordination was performed with function *envfit* in the vegan v2.2–1 package, and *P* value based on 999 permutations was used to assess the significance of the fit, and only environmental parameters showing an adjusted *P* value below 0.05 were used.

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

ACKNOWLEDGMENTS. We thank M. Follows and O. Jahn for providing us with ECCO2-Darwin simulation values for iron and S. Speich for fruitful discussions on oceanographic context. We also thank the support and commitment of the *Tara* Oceans coordinators and consortium, agnès b. and E. Bourgois, the Veolia Environment Foundation, Region Bretagne, Lorient Agglomeration, World Courier, Illumina, the EDF Foundation, FRB, the Prince Albert II de Monaco Foundation, and the *Tara* schooner and its captains and crew. *Tara* Oceans would not exist without continuous support from 23 institutes (oceans.taraexpeditions.org). This work was supported by the French "Agence Nationale de la Recherche" Programs SAMOSA (ANR-13-ADAP-0010) and France Génomique (ANR-10-INBS-09), the French Government "Investissements d'Avenir" Program OCEANOMICS (ANR-11-BTBR-0008), UK Natural Environment Research Council Grants NE/I00985X/1 and NE/J02273X/1, and the European Union's Seventh Framework Programs FP7 MicroB3 (Grant 287589) and MaCuMBA (Grant 311975). This article is contribution number 41 of *Tara* Oceans.

1. Flombaum P, et al. (2013) Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc Natl Acad Sci USA* 110(24): 9824–9829.
2. Partensky F, Hess WR, Vaulot D (1999) *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* 63(1):106–127.
3. Dutkiewicz S, et al. (2015) Impact of ocean acidification on the structure of future phytoplankton communities. *Nat Clim Chang* 5:10002–11009.
4. Coleman ML, Chisholm SW (2007) Code and context: *Prochlorococcus* as a model for cross-scale biology. *Trends Microbiol* 15(9):398–407.
5. Johnson ZI, et al. (2006) Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* 311(5768):1737–1740.
6. Moore LR, Rocap G, Chisholm SW (1998) Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* 393(6684):464–467.
7. Chandler JW, et al. (2016) Variable but persistent coexistence of *Prochlorococcus* ecotypes along temperature gradients in the ocean's surface mixed layer. *Environ Microbiol Rep* 8(2):272–284.
8. Zinser ER, et al. (2007) Influence of light and temperature on *Prochlorococcus* ecotype distributions in the Atlantic Ocean. *Limnol Oceanogr* 52(5):2205–2220.

9. Mackey KR, et al. (2013) Effect of temperature on photosynthesis and growth in marine *Synechococcus* spp. *Plant Physiol* 163(2):815–829.
10. Pittera J, et al. (2014) Connecting thermal physiology and latitudinal niche partitioning in marine *Synechococcus*. *ISME J* 8(6):1221–1236.
11. Huang S, et al. (2012) Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *ISME J* 6(2):285–297.
12. Mazard S, Ostrowski M, Partensky F, Scanlan DJ (2012) Multi-locus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*. *Environ Microbiol* 14(2):372–386.
13. Zwirglmaier K, et al. (2008) Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ Microbiol* 10(1):147–161.
14. Rusch DB, et al. (2007) The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biol* 5(3):e77.
15. West NJ, Lebaron P, Strutton PG, Suzuki MT (2011) A novel clade of *Prochlorococcus* found in high nutrient low chlorophyll waters in the South and Equatorial Pacific Ocean. *ISME J* 5(6):933–944.
16. Ahlgren NA, et al. (2014) The unique trace metal and mixed layer conditions of the Costa Rica upwelling dome support a distinct and dense community of *Synechococcus*. *Limnol Oceanogr* 59(6):2166–2218.
17. Sohm JA, et al. (2016) Co-occurring *Synechococcus* ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. *ISME J* 10(2):333–345.
18. Kettler GC, et al. (2007) Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet* 3(12):e231.
19. Post AF, et al. (2011) Long term seasonal dynamics of *synechococcus* population structure in the gulf of aqaba, northern red sea. *Front Microbiol* 2(2):131.
20. Ahlgren NA, Rocap G (2012) Diversity and distribution of marine *Synechococcus*: Multiple gene phylogenies for consensus classification and development of qPCR assays for sensitive measurement of clades in the ocean. *Front Microbiol* 3:213.
21. Biller SJ, Berube PM, Lindell D, Chisholm SW (2015) *Prochlorococcus*: The structure and function of collective diversity. *Nat Rev Microbiol* 13(1):13–27.
22. Koeppel AF, et al. (2013) Speedy speciation in a bacterial microcosm: New species can arise as frequently as adaptations within a species. *ISME J* 7(6):1080–1091.
23. Kashtan N, et al. (2014) Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* 344(6182):416–420.
24. Armbrust EV, Palumbi SR (2015) Marine biology. Uncovering hidden worlds of ocean biodiversity. *Science* 348(6237):865–867.
25. Karsenti E, et al.; Tara Oceans Consortium (2011) A holistic approach to marine ecosystems biology. *PLoS Biol* 9(10):e1001177.
26. Logares R, et al. (2014) Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environ Microbiol* 16(9):2659–2671.
27. Sunagawa S, et al.; Tara Oceans coordinators (2015) Ocean plankton. Structure and function of the global ocean microbiome. *Science* 348(6237):1261359.
28. Bouman HA, et al. (2006) Oceanographic basis of the global surface distribution of *Prochlorococcus* ecotypes. *Science* 312(5775):918–921.
29. Malmstrom RR, et al. (2010) Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific oceans. *ISME J* 4(10):1252–1264.
30. Partensky F, Garczarek L (2010) *Prochlorococcus*: Advantages and limits of minimalism. *Annu Rev Mar Sci* 2:305–331.
31. Rusch DB, Martiny AC, Dupont CL, Halpern AL, Venter JC (2010) Characterization of *Prochlorococcus* clades from iron-depleted oceanic regions. *Proc Natl Acad Sci USA* 107(37):16184–16189.
32. Malmstrom RR, et al. (2013) Ecology of uncultured *Prochlorococcus* clades revealed through single-cell genomics and biogeographic analysis. *ISME J* 7(1):184–198.
33. Song Q, Gordon AL, Visbeck M (2004) Spreading of the Indonesian throughflow in the Indian Ocean. *J Phys Oceanogr* 34(4):772–792.
34. Moutin T, et al. (2002) Does competition for nanomolar phosphate supply explain the predominance of the cyanobacterium *Synechococcus? Limnol Oceanogr* 47(5): 1562–1567.
35. Popendorf KJ, Duhamel S (2015) Variable phosphorus uptake rates and allocation across microbial groups in the oligotrophic Gulf of Mexico. *Environ Microbiol* 17(10): 3992–4006.
36. Paytan A, et al. (2009) Toxicity of atmospheric aerosols on marine phytoplankton. *Proc Natl Acad Sci USA* 106(12):4601–4605.
37. Mella-Flores D, et al. (2011) Is the distribution of *Prochlorococcus* and *Synechococcus* ecotypes in the Mediterranean Sea affected by global warming? *Biogeosciences* 8: 2785–2804.
38. Fuller NJ, et al. (2005) Dynamics of community structure and phosphate status of picocyanobacterial populations in the Gulf of Aqaba, Red Sea. *Limnol Oceanogr* 50(1): 363–375.
39. Dufresne A, et al. (2008) Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol* 9(5):R90.
40. Saito MA, Rocap G, Moffett JW (2005) Production of cobalt binding ligands in a *Synechococcus* feature at the Costa Rica upwelling dome. *Limnol Oceanogr* 50(1): 279–290.
41. Gutierrez-Rodrıguez A, et al. (2014) Fine spatial structure of genetically distinct picocyanobacterial populations across environmental gradients in the Costa Rica Dome. *Limnol Oceanogr* 59(3):705–723.
42. Villar E, et al.; Tara Oceans Coordinators (2015) Ocean plankton. Environmental characteristics of Agulhas rings affect interocean plankton transport. *Science* 348(6237):1261447.
43. Urbach E, Chisholm SW (1998) Genetic diversity in *Prochlorococcus* populations flow cytometrically sorted from the Sargasso Sea and Gulf Stream. *Limnol Oceanogr* 43(7): 1615–1630.
44. Fuller NJ, et al. (2003) Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. *Appl Environ Microbiol* 69(5):2430–2443.
45. Martiny JB, Jones SE, Lennon JT, Martiny AC (2015) Microbiomes in light of traits: A phylogenetic perspective. *Science* 350(6261):aac9323.
46. Martiny AC, Huang Y, Li W (2009) Occurrence of phosphate acquisition genes in *Prochlorococcus* cells from different ocean regions. *Environ Microbiol* 11(6): 1340–1347.
47. Martiny AC, Kathuria S, Berube PM (2009) Widespread metabolic potential for nitrite and nitrate assimilation among *Prochlorococcus* ecotypes. *Proc Natl Acad Sci USA* 106(26):10787–10792.
48. Larkin AA, et al. (2016) Niche partitioning and biogeography of high light adapted *Prochlorococcus* across taxonomic ranks in the North Pacific. *ISME J*, 10.1038/ismej.2015.244.
49. Palenik B, et al. (2003) The genome of a motile marine Synechococcus. *Nature* 424(6952):1037–1042.
50. Scanlan DJ, et al. (2009) Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol Rev* 73(2):249–299.
51. Dufresne A, Garczarek L, Partensky F (2005) Accelerated evolution associated with genome reduction in a free-living prokaryote. *Genome Biol* 6(2):R14.
52. Biastoch A, Böning CW, Lutjeharms JR (2008) Agulhas leakage dynamics affects decadal variability in Atlantic overturning circulation. *Nature* 456(7221):489–492.
53. Somerfield PJ, Clarke KR (2013) Inverse analysis in non-parametric multivariate analyses: Distinguishing of groups of associated species which covary coherently across samples. *J Exp Mar Biol Ecol* 449:261–273.
54. Malviya S, et al. (2016) Insights into global diatom distribution and diversity in the world's ocean. *Proc Natl Acad Sci USA* 113(11):E1516–E1525.
55. Magoč T, Salzberg SL (2011) FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27(21):2957–2963.
56. Behrenfeld MJ, et al. (2009) Satellite-detected fluorescence reveals global physiology of ocean phytoplankton. *Biogeosciences* 6(5):779–794.
57. Katoh K, Standley DM (2014) MAFFT: Iterative refinement and additional methods. *Methods Mol Biol* 1079:131–146.
58. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52(5):696–704.
59. Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: More models, new heuristics and parallel computing. *Nat Methods* 9(8):772.
60. Han MV, Zmasek CM (2009) phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics* 10:356.
61. Letunic I, Bork P (2007) Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics* 23(1):127–128.
62. Schloss PD, et al. (2009) Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75(23):7537–7541.
63. Zhao X, Valen E, Parker BJ, Sandelin A (2011) Systematic clustering of transcription start site landscapes. *PLoS One* 6(8):e23409.
64. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K (2015) *Cluster: Cluster Analysis Basics and Extensions*. R package Version 2.0.3. Available at https://cran.r-project.org/web/packages/cluster/. Accessed April 18, 2016.
65. Venables WN, Ripley BD (2002) Modern *Applied Statistics* with S (Springer, New York), 4th Ed. 495 p.
66. Biller SJ, et al. (2014) Genomes of diverse isolates of the marine cyanobacterium *Prochlorococcus*. *Sci Data* 1:140034.
67. Choi DH, Noh JH (2009) Phylogenetic diversity of *Synechococcus* strains isolated from the East China Sea and the East Sea. *FEMS Microbiol Ecol* 69(3):439–448.

Supporting Information

Farrant et al. 10.1073/pnas.1524865113

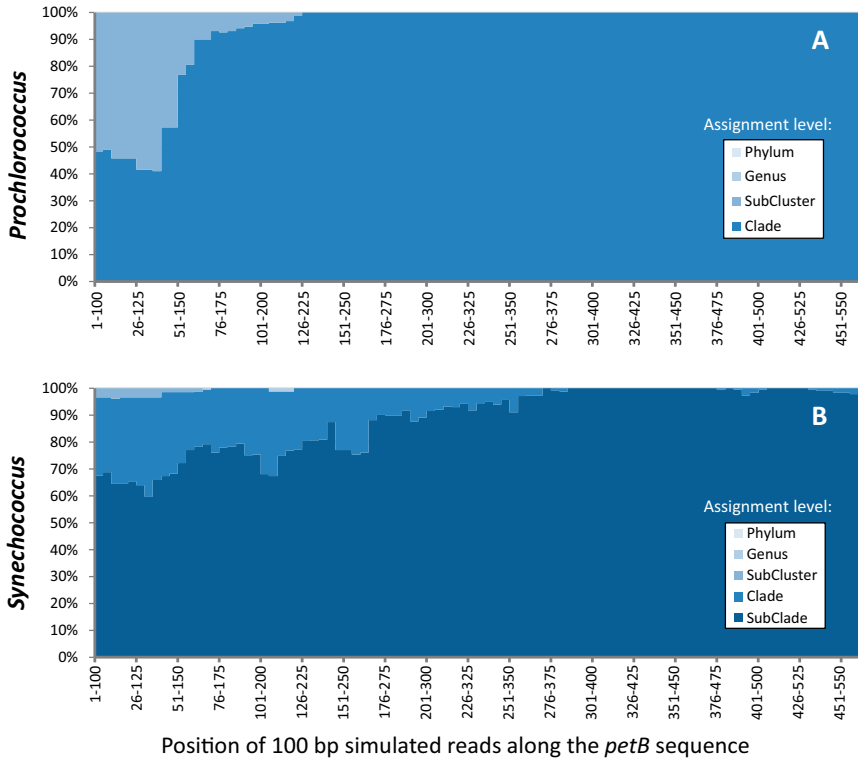


Fig. S1. Variation of the assignment ability of each individual 100-bp gene fragment along the sequence of *petB* gene using reference databases for *Prochlorococcus* (*A*) or *Synechococcus* (*B*). Simulated reads were generated by 100-bp sliding windows along the marker sequences, and the lowest taxonomic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

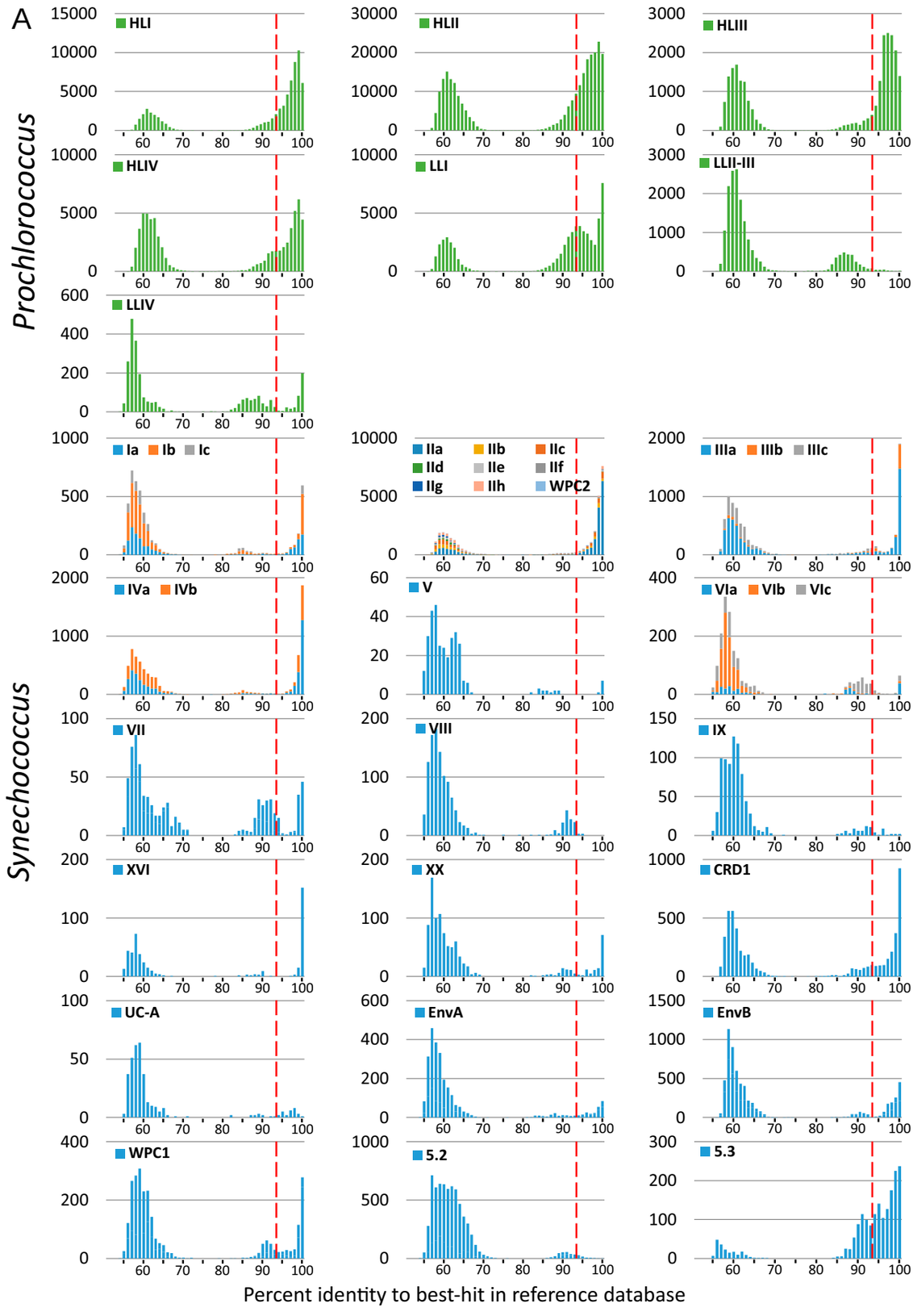


Fig. S2. (Continued)

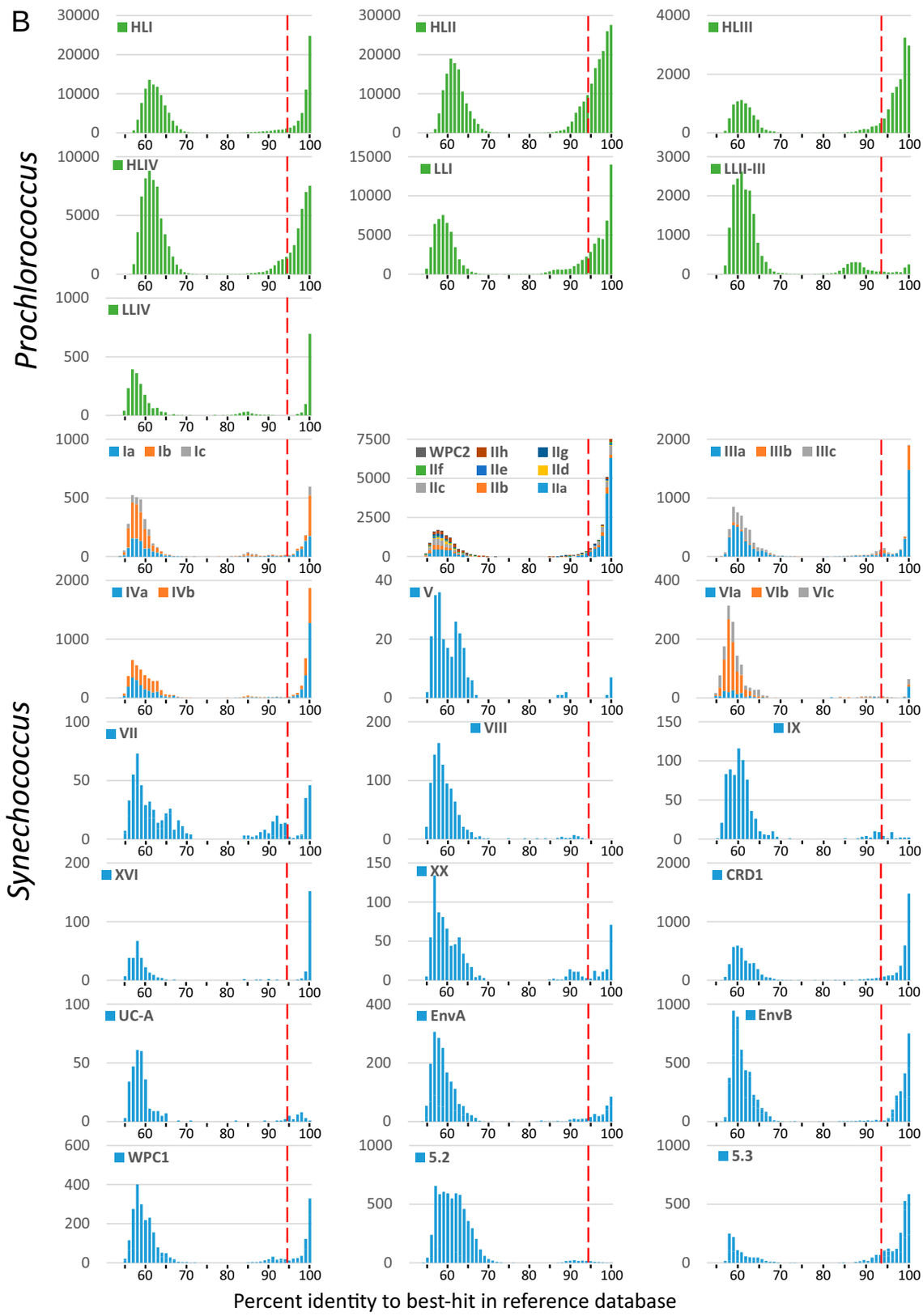


Fig. S2. (*A*) Distribution of the percent identity of *petB*-$_{mi}$tags recruited from the bacterial-size fraction of the *Tara* Oceans metagenomes with regard to their best hits in the reference database for each *Prochlorococcus* clade (top 7 graphs) and *Synechococcus* subclade (bottom 18 graphs) before addition of the 136 newly assembled environmental *petB* sequences. (*B*) Same as *A* but after addition of the 136 newly assembled environmental *petB* sequences. Note that clade XX was formerly called EnvC (12), but the name was changed here because there is at least one representative isolate (i.e., strain CC9616).

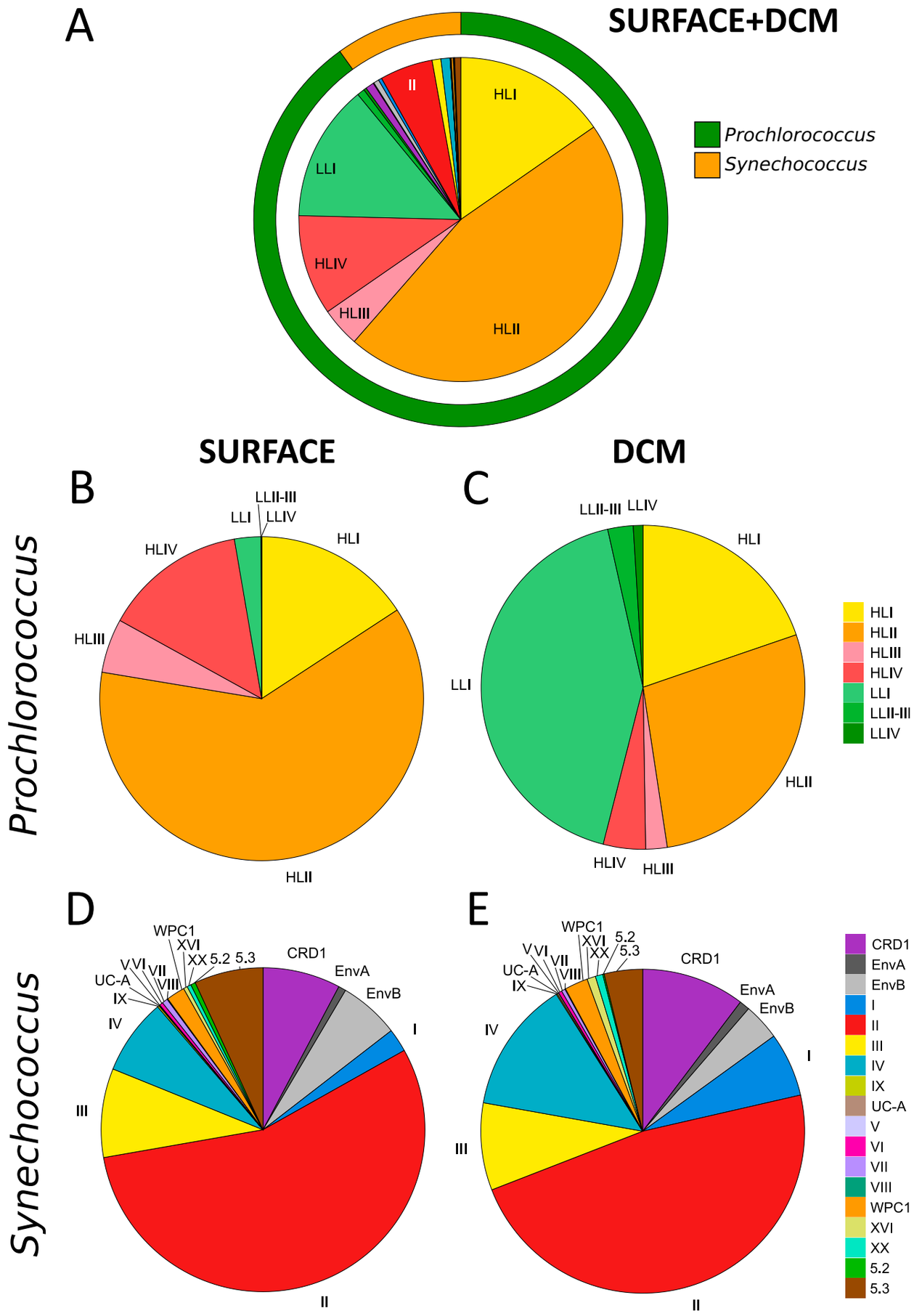


Fig. S3. Global recruitments of marine picocyanobacteria *petB*-$_{mi}$tags in the bacterial size fraction of the *Tara* Oceans metagenomes. (*A*) All picocyanobacterial clades at both sampled depths. (*B* and *C*) Percentage of each *Prochlorococcus* clade in surface (*B*) and at the DCM (*C*). (*D*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

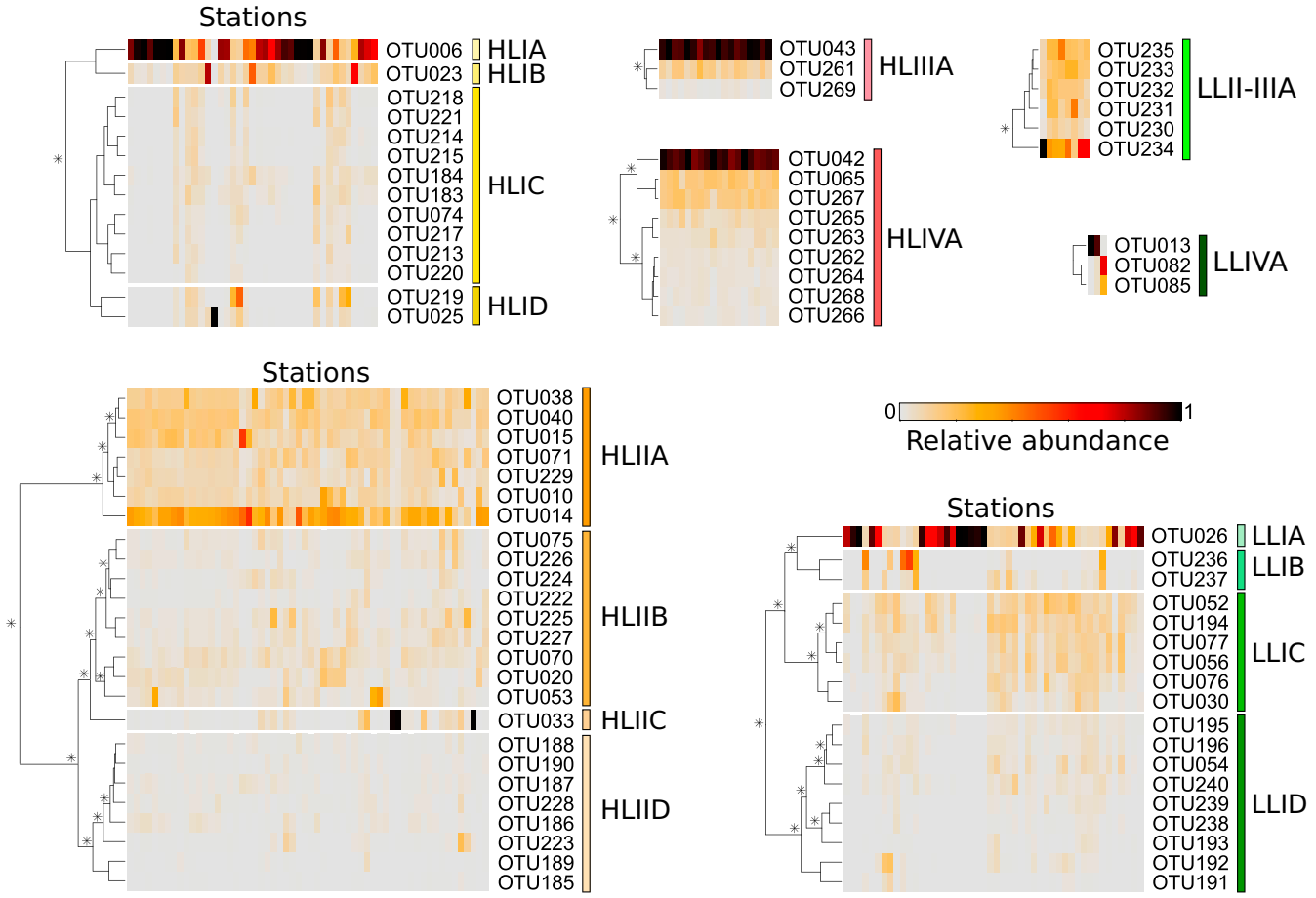


Fig. S4. *Prochlorococcus* ESTUs based on the distribution patterns of within-clade 94% OTUs. At each station, the number of reads assigned to a given OTU is normalized by the total number of reads assigned to the clade in this station. Stations and OTUs are filtered based on the number of reads recruited.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

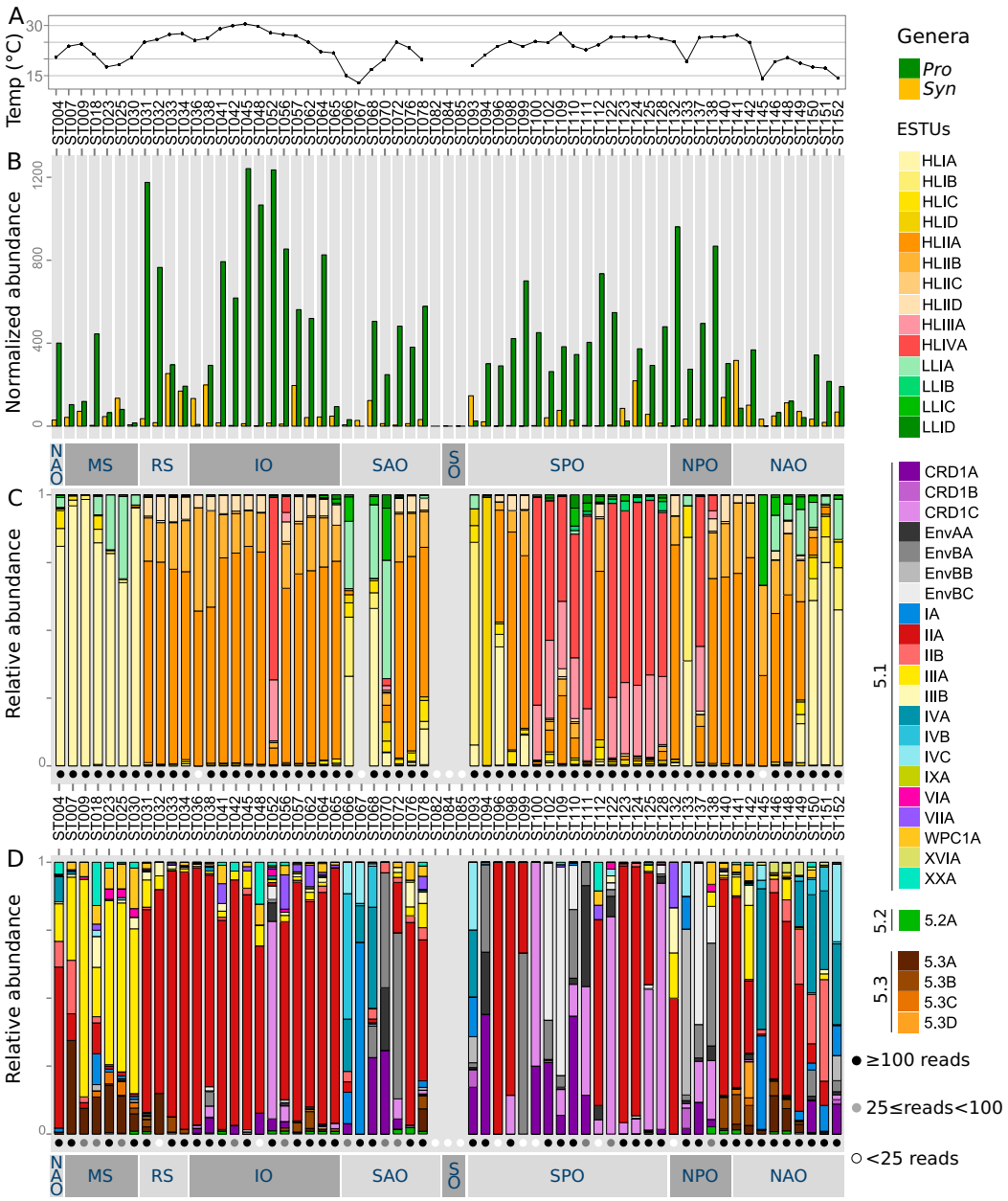


Fig. S5. Marine picocyanobacteria community structure in *Tara* Oceans surface metagenomes based on *petB*-$_{mi}$Tags recruitments. (*A*) Surface water temperature along the *Tara* Oceans transect. (*B*) Relative abundances of *Prochlorococcus* and *Synechococcus* normalized to the total number of reads at each station. (*C* and *D*) Relative abundances of *Prochlorococcus* and *Synechococcus* ESTUs, respectively. White, gray, and black dots indicate the number of reads used to build the profile, as detailed in the *Inset*. For readability, temperature for stations TARA_082 (7.3 °C), TARA_084 (1.8 °C), and TARA_085 (0.7 °C) are not shown on *A*. Abbreviations: IO, Indian Ocean; MS, Mediterranean Sea; NAO, North Atlantic Ocean; NPO, North Pacific Ocean; RS, Red Sea; SAO, South Atlantic Ocean; SO, Southern Ocean.

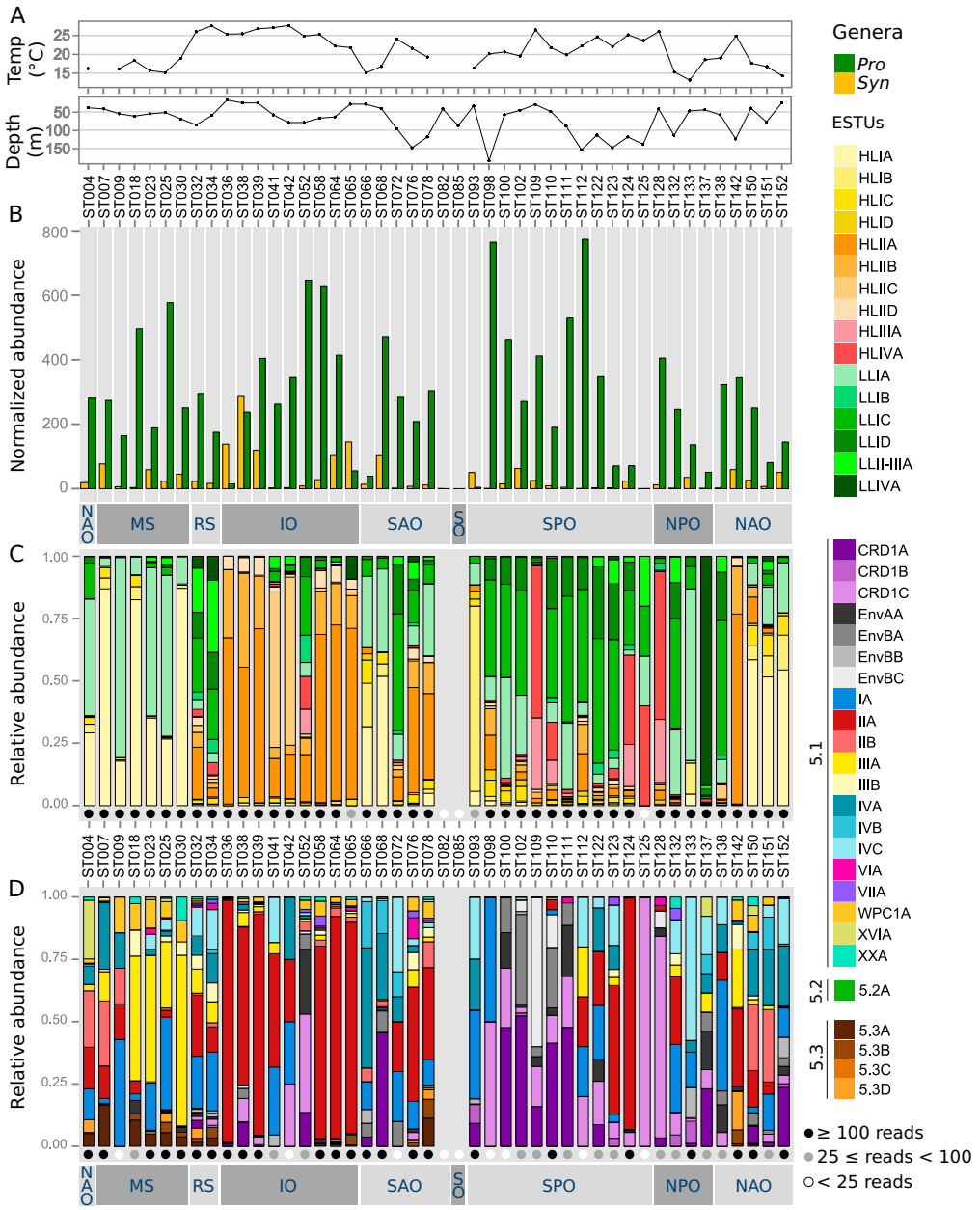


Fig. S6. Same as Fig. S5 but at the DCM. A depth profile along the *Tara* Oceans transect was added. For readability, temperature for stations TARA_082 (7.0 °C) and TARA_085 (–0.8 °C) is not shown on *A*, and the temperature is missing for station TARA_007.

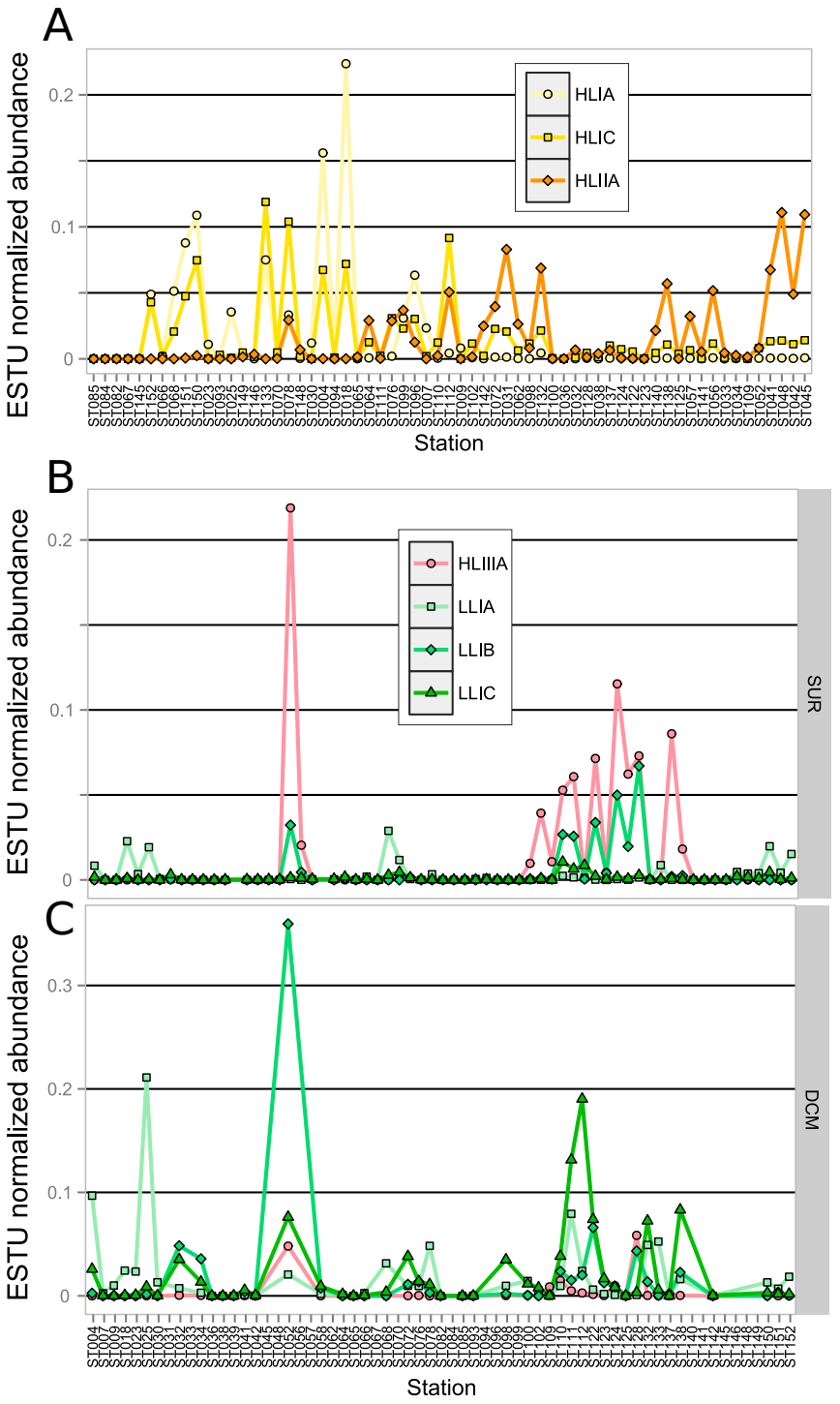


Fig. S7. Distribution of minor *Prochlorococcus* ESTUs with regard to major ESTUs in the *Tara* Oceans metagenomes. Relative abundance normalized to the total number of reads per ESTU of (*A*) ESTUs HLIA and HLIC with regard to HLIIA in surface waters and (*B* and *C*) ESTUs LLIA–C with regard to HLIIIA in surface waters and the DCM, respectively. For *A*, stations were sorted from the lowest to highest temperatures and for *B* by sampling date.

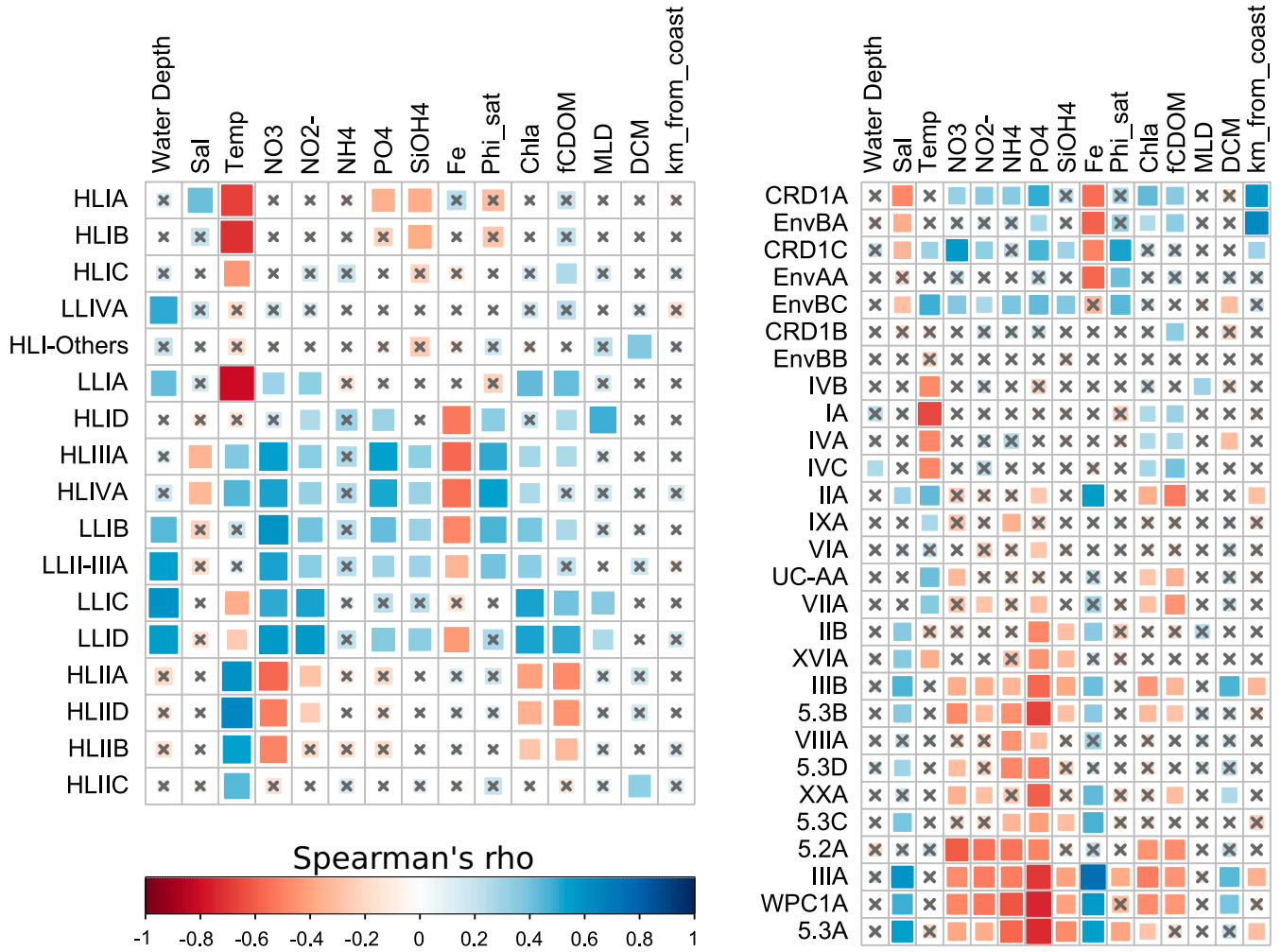


Fig. S8. Correlation analysis between marine picocyanobacterial ESTUs and environmental parameters measured along the *Tara* Oceans transect for all sampled depths. (*A*) *Prochlorococcus* ESTUs. (*B*) *Synechococcus* ESTUs. The scale shows the degree of correlation (blue) or anticorrelation (red) between the two sets of data. Correlations with an adjusted *P* value > 0.05 are indicated by gray crosses. Abbreviations: DCM, deep chlorophyll maximum; fCDOM, fluorescence, colored dissolved organic matter; MLD, mixed layer depth; Sal, salinity; Temp, temperature; Φsat, satellite-based NPQ-corrected quantum yield of fluorescence.

Dataset S1. Summary data for picocyanobacterial *petB* reference sequences used in this study, including newly assembled sequences. The table includes subclade designation based on ref. 12

Dataset S1

Dataset S2. Summary data for *petB* reference sequences for photosynthetic organisms other than marine picocyanobacteria used in this study

Dataset S2

Dataset S3. *Tara* Oceans sample description including the number of recruited *petB* reads per station. Iron and ammonium concentrations were simulated using the ECCO2-Darwin model, and an independent parameter to assess iron limitation (Φsat) was obtained using Behrenfeld et al.'s (56) equation applied to monthly averaged satellite data (AMODIS chl_ocx, nflh and ipar) retrieved from the NASA website (oceandata.sci.gsfc.nasa.gov/) for each station and corresponding sampling date. Other environmental parameters measured during the *Tara* Oceans expedition and the methods used to acquire them are available at www.pangaea.de/

Dataset S3

Dataset S4. Sequence names of the members of each OTU defined for *petB* at 94% nucleotide sequence identity

Dataset S4

2. *Synechococcus* in the Atlantic Gateway to the Arctic Ocean

Résumé de l'article en français

L'augmentation de la temperature au niveau mondial, qui a des effets prononcés aux hautes latitudes, a soulevé les questions de potentiels changements de composition en espèces et d'une augmentation possible de l'abondance des organismes phytoplanctoniques unicellulaires dans les écosystèmes marins. Dans cette étude, nous avons étudié la répartition d'un des producteurs primaires les plus petits et le plus répandu dans l'océan mondial, la picocyanobactérie *Synechococcus*, à l'interface entre les eaux issues de l'Océan Atlantique et celles de l'Océan Arctique. Contrairement à l'idée répandue selon laquelle *Synechococcus* est quasi-absent dans les régions polaires du fait des faibles températures, nous avons mesuré une abondance forte (jusqu'à 21000 cellules/mL) à 79°N et observé leur présence jusqu'à 82.5 °N. Au cours d'un cycle annuel en 2014, nous avons mis en évidence qu'en automne et en hiver, *Synechococcus* était généralement plus abondant que les picoeucaryotes qui dominent habituellement les communautés picophytoplanctoniques dans les régions arctiques. La composition des communautés de *Synechococcus* s'est trouvée altérée au cours de ce cycle, passant d'une diversité relativement élevée au cours de l'efflorescence printanière à une dominance marquée de deux unités taxonomiques opérationnelles (UTO) en automne et en hiver. Nous avons observé des abondances supérieures à 1000 cellules/mL dans des eaux à moins de 2°C dans sept stations distinctes, et des expériences de fractionnement par taille ont démontré la croissance nette de *Synechococcus* à 2°C en l'absence de brouteurs nannoplanctoniques à certaines périodes de l'année. L'analyse phylogénétique des séqences du gène *petB* a permis de montrer que ces populations de haute latitude de *Synechococcus* appartiennent aux clades I et IV, décrit précédemment comme étant adaptés aux basses tempértures, mais a également permis d'étendre notre connaissance de la diversité génétique au sein du clade I.

frontiers in Marine Science

ORIGINAL RESEARCH
published: 05 October 2016
doi: 10.3389/fmars.2016.00191

Synechococcus in the Atlantic Gateway to the Arctic Ocean

Maria L. Paulsen[1], Hugo Doré[2], Laurence Garczarek[2], Lena Seuthe[3], Oliver Müller[1], Ruth-Anne Sandaa[1], Gunnar Bratbak[1] and Aud Larsen[4]*

[1] Department of Biology, University of Bergen, Bergen, Norway, [2] Centre National de la Recherche Scientifique, UMR 7144, Station Biologique, Sorbonne Universités, UPMC Université Paris 06, Roscoff, France, [3] Department of Arctic and Marine Biology, Faculty of Biosciences, Fisheries and Economics, UiT-The Arctic University of Norway, Tromsø, Norway, [4] Hjort Centre for Marine Ecosystem Dynamics, Uni Research Environment, Bergen, Norway

Increasing temperatures, with pronounced effects at high latitudes, have raised questions about potential changes in species composition, as well as possible increased importance of small-celled phytoplankton in marine systems. In this study, we mapped out one of the smallest and globally most widespread primary producers, the picocyanobacterium *Synechococcus,* within the Atlantic inflow to the Arctic Ocean. In contrast to the general understanding that *Synechococcus* is almost absent in polar oceans due to low temperatures, we encountered high abundances (up to 21,000 cells mL^{-1}) at 79°N, and documented their presence as far north as 82.5°N. Covering an annual cycle in 2014, we found that during autumn and winter, *Synechococcus* was often more abundant than picoeukaryotes, which usually dominate the picophytoplankton communities in the Arctic. *Synechococcus* community composition shifted from a quite high genetic diversity during the spring bloom to a clear dominance of two specific operational taxonomic units (OTUs) in autumn and winter. We observed abundances higher than 1000 cells mL^{-1} in water colder than 2°C at seven distinct stations and size-fractionation experiments demonstrated a net growth of *Synechococcus* at 2°C in the absence of nano-sized grazers at certain periods of the year. Phylogenetic analysis of *petB* sequences demonstrated that these high latitude *Synechococcus* group within the previously described cold-adapted clades I and IV, but also contributed to unveil novel genetic diversity, especially within clade I.

Keywords: picocyanobacteria, picoeukaryotes, temperature adaptation, *petB* sequences, flow cytometry, high latitude ecosystems, Svalbard, West Spitsbergen Current

OPEN ACCESS

Edited by:
Connie Lovejoy,
Laval University, Canada

Reviewed by:
David J. Scanlan,
University of Warwick, UK
William Li,
Fisheries and Oceans Canada, Canada

****Correspondence:***
Maria L. Paulsen
maria.l.paulsen@uib.no

Specialty section:
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Marine Science

Received: *28 June 2016*
Accepted: *21 September 2016*
Published: *05 October 2016*

Citation:
Paulsen ML, Doré H, Garczarek L, Seuthe L, Müller O, Sandaa R-A, Bratbak G and Larsen A (2016) Synechococcus in the Atlantic Gateway to the Arctic Ocean. Front. Mar. Sci. 3:191. doi: 10.3389/fmars.2016.00191

INTRODUCTION

The widely abundant picocyanobacterium *Synechococcus* is estimated to be responsible for about 17% of ocean net primary productivity and thus to have a high impact on ocean ecosystems and biogeochemical cycles (Flombaum et al., 2013). *Synechococcus* is normally not considered to be bloom-forming even though they can appear in abundances as high as 1.2–3.7 × 10^6 cells mL^{-1} in the Costa Rica dome (Saito et al., 2005). Using 37,699 discrete global *Synechococcus* observations between 69°S and 81°N and quantitative niche models, Flombaum et al. (2013) demonstrated temperature to be the main environmental parameter explaining the global distribution of *Synechococcus*. Accordingly, the regional range of temperature was found to be a relatively good predictor for the seasonal change in *Synechococcus* abundance (Tsai et al., 2013). Although the

marine *Synechococcus* that have been studied in culture so far have a temperature optimum ranging from 20 to 33°C depending on the clade (Pittera et al., 2014), the highest annual average *in situ* cell abundances were found at temperatures around 10°C in the Indian and western Pacific Oceans, with averages of 34,000 and 40,000 cells mL^{-1}, respectively (Flombaum et al., 2013).

There has been observations of *Synechococcus* at low temperatures e.g., <4°C in low numbers (<100 cells mL^{-1}; Waterbury et al., 1986) and <2°C (Shapiro and Haugen, 1988; Gradinger and Lenz, 1995), but they are still often considered to be nearly absent from the polar ocean (Pedrós-Alió et al., 2015) in contrast to cold adapted eukaryotic picophytoplankton that occur in high abundances both in Arctic (Sherr et al., 2003; Lovejoy et al., 2007; Tremblay et al., 2009; Zhang et al., 2015) and in Antarctic waters (Doolittle et al., 2008). Only a few studies have actually documented *Synechococcus* north of 70° and none have so far described the genetic diversity of these northern populations or tested their temperature optimum.

During four expeditions Gradinger and Lenz (1995) observed maximal abundances of 5500 *Synechococcus* cells mL^{-1} in the Atlantic inflow to the Arctic Ocean west of Svalbard at 78°N, while they did not find any *Synechococcus* cells in surface samples of polar water (defined as water having Temp < 0°C; Salinity < 34). Further south, following a transect from 70.5 to 74°N, Not et al. (2005) recorded a maximum abundance of 25,000 cells mL^{-1} in the Norwegian and Barents Seas in August 2002. In the western Canadian Arctic, Cottrell and Kirchman (2012) found abundances of 40–80 cells mL^{-1} in coastal waters of the Chukchi Sea and the Beaufort Sea at 71.5°N, both during summer and winter cruises. Nelson et al. (2014) concluded in their overview that the *Synechococcus* distribution in this region is controlled mainly by inflow of the relatively warm Pacific water, but argue that water temperature alone cannot be used to define environments in which *Synechococcus* may reside as they do persist at water temperatures near the freezing point (−1.8°C) (Nelson et al., 2014).

Synechococcus is often found in Arctic lakes and rivers, and freshwater runoff may thus also represent a source of *Synechococcus* cells to the Arctic Ocean (Vincent et al., 2000). Using 16S rRNA analysis, Waleron et al. (2007) revealed that picocyanobacteria present in the Canadian Beaufort Sea originate from the Mackenzie River and other nearby inflows. High abundances of *Synechococcus* (30,000 cells mL^{-1}) were also found in the Laptev Sea, but were restricted to brackish waters near the Lena River delta, while further away from the delta, abundances decreased with increasing salinity to a total absence at salinities >20 (Moreira-Turcq and Martin, 1998). All these studies support Waterbury et al. (1986) claiming that only few brackish species tolerate wide salinity ranges and that many strains are obligate marine. Assuming that Atlantic *Synechococcus* have a low tolerance to salinity changes, the question remains whether the low salinity in the Arctic surface waters constrains their distribution in the polar ocean.

The Atlantic inflow is the main conveyor, not only of water and heat, but also of more southern species into the Arctic Ocean. *Synechococcus* has accordingly been suggested as a bio-indicator for the advection of Atlantic waters into the Arctic Ocean

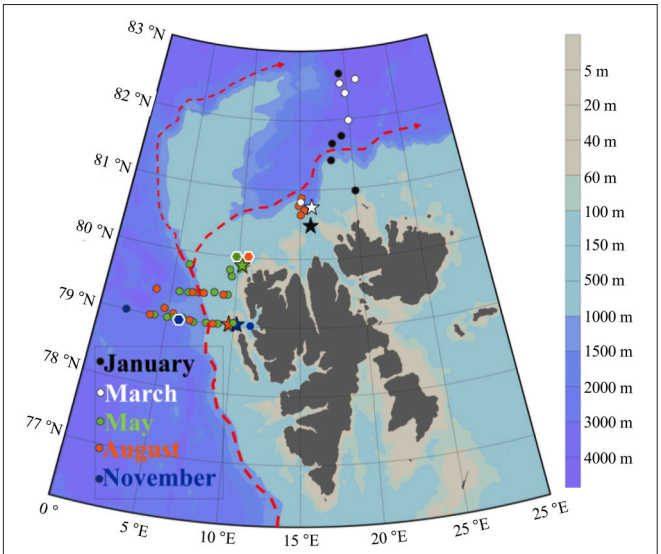


FIGURE 1 | Study area Northwest of Svalbard. Sampled stations are colored according to the month of sampling: January (black), March (white), May (green), August (orange), and November (blue). Star-symbols indicate sampling of water for fractionated growth experiments. Hexagon symbols with white border indicate sampling for molecular analysis. Red arrows indicate the main flow of the West Spitsbergen Current drawn after Cokelet et al. (2008), Randelhoff et al. (submitted).

(Murphy and Haugen, 1985; Gradinger and Lenz, 1995). The main transport follows the West Spitsbergen Current (WSC) which is an extension of the Norwegian Atlantic Current splitting up into two branches around 79–80°N (**Figure 1**). The WSC is about 100 km wide and is confined over the continental slope along the Norwegian coast. It has an average speed of 10 cm s^{-1} (Cokelet et al., 2008) but can reach a speed of up to 24–35 cm s^{-1} (Boyd and D'Asaro, 1994; Fahrbach et al., 2001). The inflow follows a strong annual cycle with maximum volume transport during winter (20 Sv in February) and minimum during summer (5 Sv in August, Fahrbach et al., 2001) (N.B. the unit sverdrup (Sv) is equal to 1 million m^3 s^{-1}). Strong variations in the strength of the Atlantic inflow combined with varying sea ice extension make it challenging to assess the spread of Atlantic organisms in this area. Little is known about how *Synechococcus* populations, originating from the Norwegian coast or further south, are affected as they are transported into the Arctic Ocean or whether some *Synechococcus* lineages are favored under the transition to more Arctic conditions.

Temperature is one of the main drivers of *Synechococcus* biogeography. Among the five globally dominating *Synechococcus* lineages (clades I, II, III, IV, and CRD1), clades I and IV dominate at high latitudes in cold and coastal waters, while clades II and III are mostly found in warm, (sub)tropical areas (Zwirglmaier et al., 2008; Farrant et al., 2016; Sohm et al., 2016). Populations adapted to distinct thermal niches were also identified within the CRD1 clade, including one co-occurring with clade I and IV in cold, mixed waters of the Pacific Ocean (Farrant et al., 2016).

Increasing ocean temperature in high latitude systems has drawn attention towards the growth of invasive organisms with higher temperature optima and subsequent ecosystem changes. In marine systems, small phytoplankton are expected to become relatively more abundant with warming (Morán et al., 2010; Tremblay et al., 2012) and it has been speculated that the warming of the Arctic Ocean could lead to a shift from picoeukaryotes to picocyanobacteria, with implications for food quality (Vincent, 2010). Flombaum et al. (2013) projected up to a 50% increase in *Synechococcus* at 60°N by the end of the twenty first century. Their models were however not able to make projections for higher latitude systems because observations in these areas are scarce. The aim of the present study is therefore to examine the distribution of *Synechococcus* in relation to environmental parameters and other microbial plankton groups within the Atlantic gateway to the Arctic. The genetic diversity of *Synechococcus* populations was also unveiled using a high resolution genetic marker, the *petB* gene (encoding the cytochrome b_6 subunit), in order to trace the geographical origin and seasonal changes of these populations.

MATERIALS AND METHODS

Locality and Sampling

This study covers the eastern part of the Fram Strait, where Atlantic water (AW) is transported northward by the West Spitsbergen Current (WSC). Data were collected during five cruises in 2014: January (06.01–15.01), March (05.03–10.03), May (15.05–02.06), August (07.08–18.08), and November (03.11–10.11). Transects were made across the core of AW inflow at 79 and 79.4°N during May, August and November. Further north (80.5 to 82.6°N) we investigated the WSC southern branch into the Arctic Ocean in January, March and August (**Figure 1**). The choice of sampling area and stations was largely determined by the extension of the sea ice (**Figure 3**). Vertical profiles of temperature, salinity and fluorescence were recorded on each sampling occasion using a SBE 911plus system. Water masses were defined based on the criteria presented in **Table 1**. Discrete water samples for analyses of nutrients ($NO_2^- + NO_3^-$, NH_4^+, PO_4^{3-}, H_4SiO_4) and enumeration of phytoplankton, viruses, bacteria, and heterotrophic nanoflagellates (HNF) were collected from 11 depths (1, 5, 10, 20, 30, 50, 100, 200, 500, 750, and 1000 m) using 10 L Niskin bottles. During the summer cruises we collected additional samples from the Deep Chlorophyll Maximum (DCM) (when different from any of the standard depths). The shallow shelf stations were sampled to near bottom and with higher sampling resolution in the surface.

Flow Cytometry

Abundances of pico- and nano-sized phytoplankton, viruses, bacteria and HNF were determined on an Attune® Acoustic Focusing Flow Cytometer (Applied Biosystems by Life technologies) with a syringe-based fluidic system and a 20 mW 488 nm (blue) laser. Samples were fixed with glutaraldehyde (0.5% final conc.) at 4°C for a minimum of 2 h, flash frozen in liquid nitrogen and stored at −80°C until analysis, except in November, when phytoplankton was enumerated using fresh samples. For analysis of HNF the samples were stained with SYBR Green I (Molecular Probes, Eugene, Oregon, USA) for 2 h in the dark and a minimum of 1 mL was measured at a flow rate of 500 μl min^{-1} following the protocol of Zubkov et al. (2007). The HNF population was discriminated from nano-sized phytoplankton based on green vs. red fluorescence and from large bacteria on a plot of side scatter vs. green fluorescence following the recommendations of Christaki et al. (2011). Pico- and nano-sized phytoplankton were counted directly after thawing and the various groups discriminated based on their side scatter signals (SSC) vs. orange fluorescence (**Figure 2A**; Marie et al., 1997; Larsen et al., 2001) as well as their red vs. orange fluorescence (**Figure 2B**). *Synechococcus* was identified in plots of orange fluorescence vs. side scatter signals (**Figure 2A**). For samples with low abundance of phytoplankton (March and January) a volume of 1.5 mL was counted, while 0.5 mL was sufficient for May, August and November-samples. Regular blank measurements using Milli-Q® water were made to ensure that there was no carry over of cells between samples and that electronic noise did not disturb the counts. Due to the inherent uncertainty connected to enumeration of cells when concentrations are low, we only included samples with > 20 cells mL^{-1} when relating counts to other environmental parameters (**Figure 5**). Samples for which 0–20 cells were detected (i.e., mainly those deeper than 500 m) are included in our total data set (**Table S1**, **Figure S2**).

TABLE 1 | Criteria determining the water masses.

Water masses	Temperature (°C)/or density (kg m^{-3})	Salinity
Atlantic water (AW)	>2°C	>34.9
cold Atlantic water (cAW)	0<T<2°C	>34.9
Intermediate water (IW)	<0°C	>34.9
Arctic water (ArW)	>27.7 kg m^{-3}	<34.9
Surface water (SW)	<27.7 kg m^{-3}	<34.9
Polar water (PW)	<0°C	<34.7

For further explanation see Cokelet et al. (2008). PW overlaps with SW and ArW.

Microscopy

The presence of *Synechococcus* was also confirmed by fluorescence microscopy (**Figure 2C**). Samples were fixed and stored as for flow cytometry. The samples were thawed, filtered onto Anodisc filters (Whatman, pore size 0.2 μm) and stained with SYBR Green I (Molecular Probes Inc., Eugene, Oregon) according to Patel et al. (2007). The samples were viewed and photographed at 400X using a Zeiss Axio Imager Z1 microscope with AxioCam MRm BW-camera, extended focus, epifluorescence illumination (HXP Illuminator) and Zeiss filter sets 09 and 43 for SYBR Green and chlorophyll fluorescence, respectively.

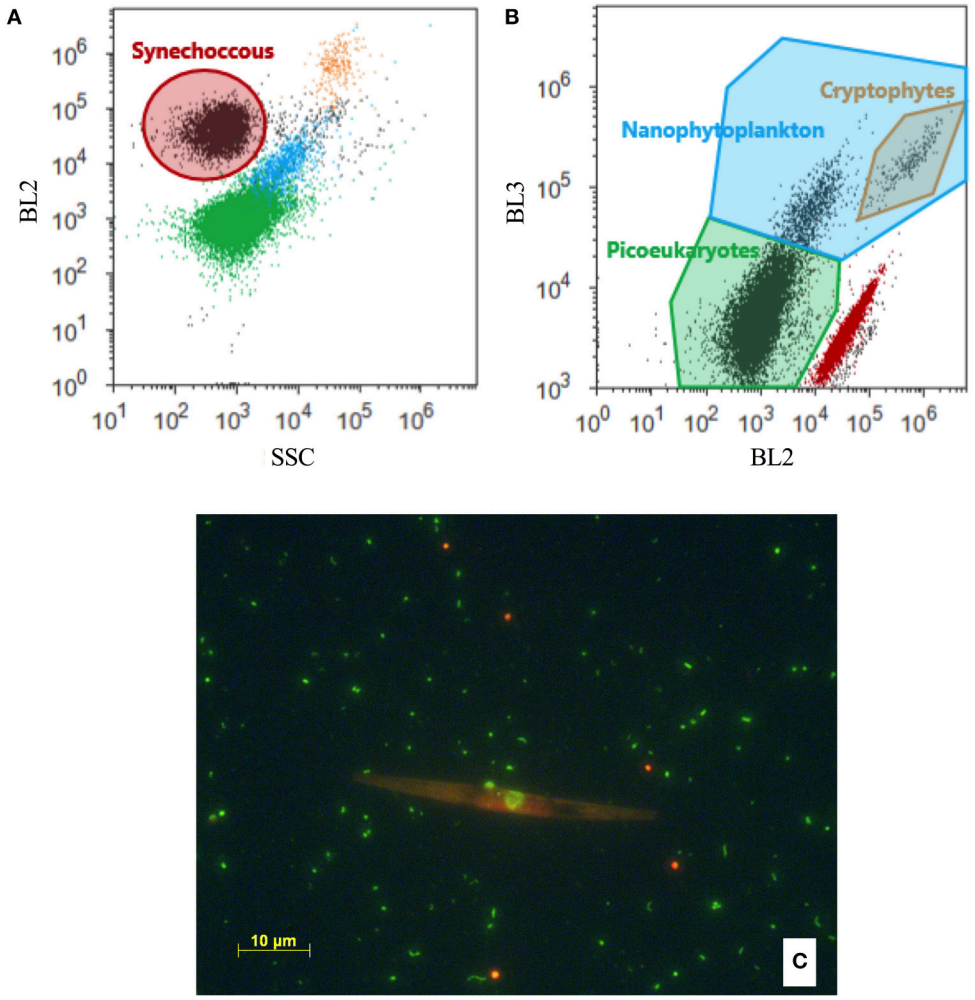


FIGURE 2 | (A,B) Biparametric flow cytometry plots with the applied grouping of the different phytoplankton groups indicated by the solid line. Groups are colored according to their grouping in order to appear on both **(A)** BL2 (orange fluorescence) vs. SSC (side scatter) and **(B)** BL3 (red fluorescence) vs. BL2. **(C)** Microscope image showing the difference in the autofluorescence spectrum of *Synechococcus* (appear orange) and picoeukaryotes (appear green). The long orange cell is a diatom.

Size-Fractionated Growth Experiments

Water fractionation experiments were used to examine interaction between different size groups of microorganisms and to estimate growth rates of the different microbial components (Simek and Chrzanowski, 1992; Jürgens et al., 2000; Christaki et al., 2001; Sato et al., 2007). Experiments were performed once every cruise using water collected from 20 m (in August and May this depth was near DCM) at stations on the shelf (marked on **Figure 1**). The water was gently screened through 3, 5, 10, and 90 µm mesh size filters by reverse filtration in order to successively exclude grazers of different sizes and thus create communities with increasing "top-predators" sizes. Water from each filtration treatment was gently transferred into triplicate 3.9 L transparent polycarbonate bottles (Nalgene®) by staggered filling using silicone tubing. The incubation experiments ran for 5 to 10 days but we show data only from the initial 5 days of incubation in order to better represent dynamics of the initial communities. Incubation water was sampled daily for enumeration of microorganisms and every second day for nutrients. Prior to setup, all bottles, carboys and silicon tubs were acid washed and then rinsed with Milli-Q® water. During the summer cruises (May and August) the experimental bottles were incubated on deck in plexiglass tanks with seawater flow-through (continuously pumped from 7 m depth), keeping the temperature close to *in situ* (May: 1.7 ± 1.6°C and August: 1 ± 0.8°C). A nylon net was wrapped around each bottle to reduce the PAR to about 30% of the surface irradiation. In the winter months (January, March, November), incubations were kept in a cooling room at a constant temperature of 2°C and in darkness, except in March were an *in situ* light cycle was set (16 h darkness and 8 h at 5 µmol photons m^{-2} s^{-1}). The fractionation experiments

provided net growth rates of *Synechococcus* and HNF by fitting exponential functions to the change in the abundance of cells every 24 h during the first 5 days of the experiments (**Figure S1**).

DNA Extraction, PCR Cloning and Phylogenetic Analysis

Environmental samples for molecular analysis were collected by filtering water onto 0.22 μm pore size Millipore® Sterivex filters. The filters were immediately flash frozen in liquid nitrogen and stored at −80°C until extraction. DNA and RNA were extracted simultaneously using the AllPrep DNA/RNA Mini Kit (Qiagen, Hilden, Germany) according to manufacturer's instructions with some optimisation for extraction from Sterivex filters as follows. The filters were thawed on ice and 1 mL extraction buffer (990 μl RLT buffer; containing guanidine isothiocyanate + 10 μL ß-mercaptoethanol) was added before incubating for 4 min on a Vortex adapter at medium speed. The resulting lysates were recovered using a 10 mL syringe and used for nucleic acids extraction. DNA samples harvested from Arctic surface water collected in May (80°N, 10.7°E at 1 m depth, 10 L water filtered), August (80°N, 10.8°E at 1 m, 7.5 L filtered) and November (79°N, 6°E, 20 m, 20 L filtered) were selected to amplify the *Synechococcus petB* marker gene (stations marked on **Figure 1** and profiles of picophytoplankton are included in **Figure S2**). Polymerase chain reaction (PCR) were performed using the *petB* primers and set-up recommended by Mazard et al. (2012) using 30–40 amplification cycles (iCycler, Bio-Rad, CA, USA). Positive PCR products were purified using the Zymo DNA Clean and Concentrator™-5 kit (Zymo research, CA, USA) and subsequently cloned with the StrataClone™ PCR Cloning Kit (Agilent Technologies, CA, USA) following the manufacturer's instructions. A total of 96 clones from each of the three samples were picked (total 288 clones) and sequenced by LCG Genomic GmbH (Berlin, Germany) using Sanger sequencing. A total of 229 *petB* sequences were obtained and deposited in the GenBank database (accession no. KX345947–KX346174). These sequences are in the following referred to as "MicroPolar sequences."

The 229 MicroPolar sequences include 174 unique full-length sequences. Together with 721 *petB* sequences from a non-redundant reference database (representing most of the genetic diversity so far identified within *Prochlorococcus* and *Synechococcus* genera; Farrant et al., 2016), the MicroPolar sequences were used to define operational taxonomical units (OTUs) at 97% identity using Mothur v1.34.4 (**Table S2**). Since all MicroPolar sequences clustered with clades I and IV reference sequences, a subset of the *petB* database, comprising only the 117 reference sequences of these clades, as well as the 174 unique MicroPolar *petB* sequences was used for a subsequent analysis. Phylogenetic reconstructions were based on multiple alignments of *petB* nucleotide sequences generated using MAFFT v7.164b with default parameters (Katoh and Standley, 2014). A maximum likelihood tree was inferred using PHYML v3.0, (Guindon and Gascuel, 2003) with the HKY + G substitution model, as determined using jModeltest v2.1.4 (Darriba et al., 2012) and estimation of the gamma distribution parameter of the substitution rates among sites and of the proportion of invariable sites. The tree was drawn using iTOL (Letunic and Bork, 2007). The 229 sequences retrieved from MicroPolar were recruited using BLASTN (v2.2.28+) against the full *petB* database: reads with more than 90% of their sequence aligned and with more than 80% sequence identity to their BLASTN best-hit were taxonomically assigned to their best-hit and subsequently used to build per-sequence read counts tables. Counts were then aggregated by OTUs and relative abundance was computed for each MicroPolar station.

Nutrients

Unfiltered seawater was filled directly from the Niskin bottles into 30 mL acid washed HDPE bottles and stored at −20°C. Nitrite and nitrate ($NO_2^- + NO_3^-$), phosphate (PO_4^{3-}) and silicic acid (H_4SiO_4) were measured on a Smartchem200 (by AMS Alliance) autoanalyser following procedures as outlined in Wood et al. (1967) for $NO_3^- + NO_2^-$, Murphy and Riley (1962) for PO_4^{3-} and Koroleff (1983) for the determination of H_4SiO_4. The determination of NO_3^- was done by reduction to NO_2^- on a built-in cadmium column, which was loaded prior to every sample run. Seven-point standard curves were made prior to every run. Two internal standards and one blank were inserted for every 8 samples and these were used to correct for any drift in the measurements. Concentration of NH_4^+ was determined directly in fresh samples using ortho-phthaladehyde according to Holmes et al. (1999).

RESULTS

Synechococcus cells were detected by flow cytometry in all samples within the upper 100 m of the water column during all seasons (**Figure 3** and **Figure S2**). The identity of *Synechococcus* was confirmed by epifluorescence microscopy (**Figure 2C**) and by sequencing of the *petB* gene (**Figure 7**). The closely related genus *Prochlorococcus* was never detected.

Synechococcus Distribution

The highest sampling frequency was obtained in May and August, when the sampling sites were restricted to latitudes below 80°N (May) and 81°N (August) by sea ice (**Figure 3**), while the most northern samples were acquired in January and March at around 82.5°N. *Synechococcus* was present in abundances higher than 50 cells mL^{-1} in 337 samples both within the Atlantic water, Arctic water, Surface water and Polar water (water mass definitions are shown in **Table 1**). Within the cold surface water (<2°C, upper 50 m), 60% of the samples contained *Synechococcus* with abundances ranging from 50 to 4300 cells mL^{-1}. *Synechococcus* was not detected in the cold Atlantic water or intermediate water masses, which comprise water collected deeper than 500 m (see temperature-salinity plots in **Figure 3**).

In January, the average abundance in the upper 100 m was 51 cells mL^{-1}, with highest abundance found at 100 m depth (maximum 106 cells mL^{-1}) and generally low numbers in the surface (**Figures 3F**, **4A**). The lowest average abundance of

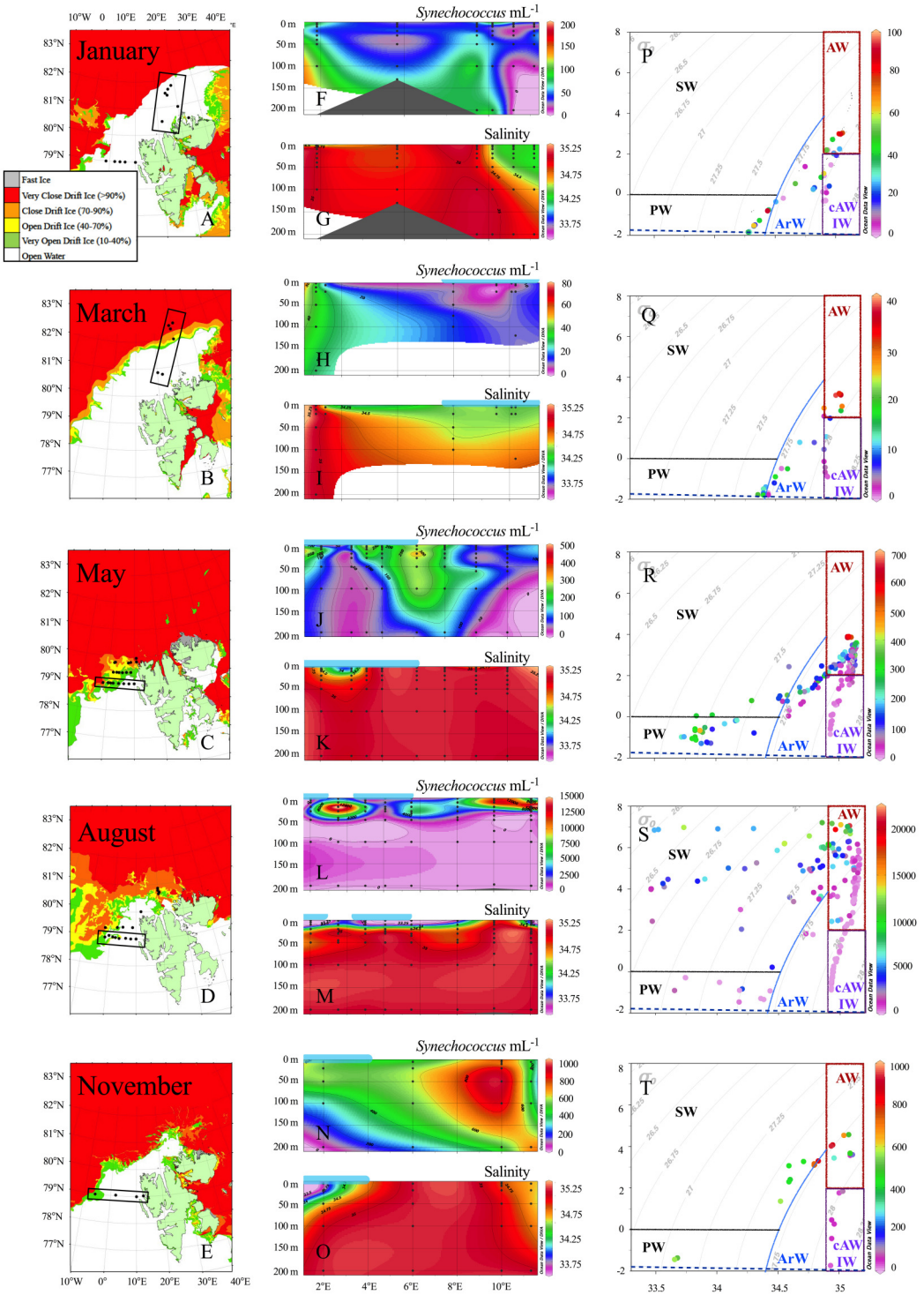


FIGURE 3 | (Left panel) Ice-maps provided from the Norwegian Meteorological Institute (istjenesten@met.no) from following dates; January 10, March 7, May 23, August 12, and November 7, 2014. Transects shown in the middle panel as contour plots are marked with black boxes. **(Middle panel)** Contour plots showing the abundance (cells mL^{-1}) of *Synechococcus* and salinity of the upper 200 m from 4 cruises. January transect stretches from North-South between 15 and 20°E, while the remaining transects expands West-East (2–11°E) following the 79°N latitude isoline (see transect marked in boxes left of plots). The horizontal light blue lines above the plots roughly mark the cover of open drift ice. Note different scales for *Synechococcus* abundance. **(Right panel)** Potential temperature and salinity (TS) diagram for each month. Data included for all depths 1–1000 m. *Synechococcus* abundance is given on the z-axis by color gradient (N.B. different scales). Potential density (σ_0, kg/m^3) isolines overlaid with gray and the surface freezing line is show in dashed blue. Following water masses (**Table 1**) are marked: Atlantic water (AW), cold Atlantic water or Intermediate water (cAW or IW), which consist mainly of deep water samples (>500 m), Arctic water (ArW), surface water (SW) and Polar water (PW).

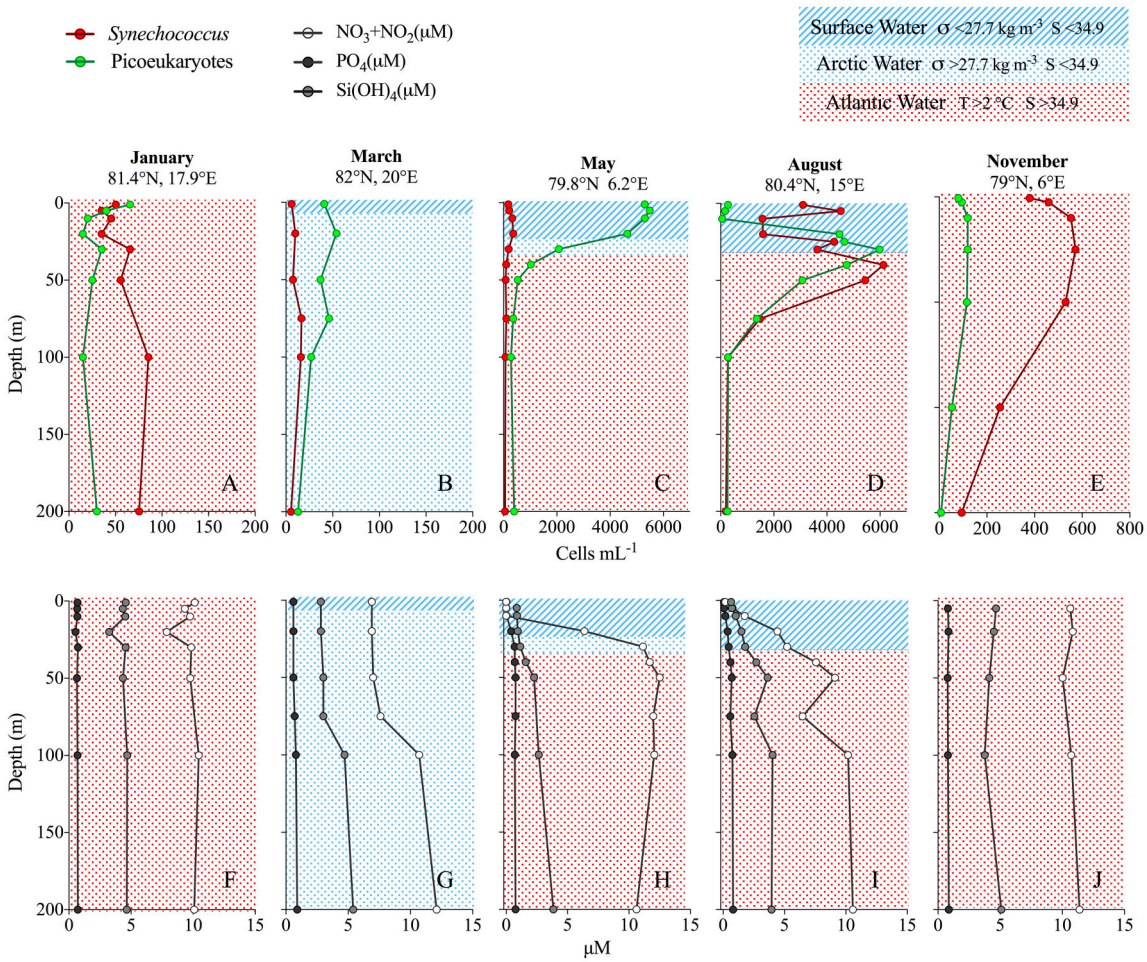


FIGURE 4 | (A–E) profiles of *Synechococcus* (red) and picoeukaryotes (green) abundance (cells mL^{-1}) from each of the cruises at given coordinates. Notice different scales on the x-axis. **(F–J)** nutrient concentrations (µM) from the same profiles. Background colors and pattern indicate water masses.

Synechococcus was found in March with 13 cells mL^{-1} and a maximum abundance that did not exceed 40 cells mL^{-1} (**Figures 3H**, **4B**). In May the maximum abundance was around 1300 cells mL^{-1} and the average ± SD was 181 ± 147 cells mL^{-1} ($n = 150$; **Figures 3J**, **4C**). The highest *Synechococcus* abundances were detected in August with a maximum of 21,300 cells mL^{-1}. When averaged for the upper 50 m at most southern stations (79–79.4°N), abundances were 5700 ± 4200 cells mL^{-1} ($n = 61$) (**Figure 3L**) over the whole transect, while abundances at the stations north of 80°N averaged to about 3000 ± 2000 cells mL^{-1} ($n = 27$). In November *Synechococcus* cells were evenly distributed down to 200 m (**Figure 4E**), with a maximum of 1000 cells mL^{-1} and an average abundance of 600 ± 250 cells mL^{-1} ($n = 18$) within the upper 200 m. The vertical distribution of *Synechococcus* varied from mainly surface peaks in May (upper 20 m) to maximum abundance at depths greater than 50 m in August and November, to a more vertically uniform distribution in January and March with maxima in abundance at around 100 m depth (**Figures 4A–E** and **Figure S2**).

Biotic and Abiotic Environment

The association between phytoplankton abundances and environmental parameters showed that the abundance of both *Synechococcus* and picoeukaryotes decreased with increasing latitude, but that picoeukaryotes were relatively more abundant at the northernmost stations (**Figure 5A**). No clear relationship was found for salinity (ranging from 31 to 35 in this study), although the highest *Synechococcus* abundances were found at salinities >34.5, while picoeukaryotes had their peak abundance at lower salinities of 33.5–34 (**Figure 5B**). Further, we found picoeukaryotes to be strongly dominant over *Synechococcus* in 14 out of 17 samples with lowest salinity (31–33), all sampled in August. The abundance of *Synechococcus* ranged from 250 to 4000 cells mL^{-1} in these low salinity samples (**Figure 3S**). The presence of sea ice had no clear effect on the vertical distribution of picophytoplankton but at the ice-covered stations, a subsurface maximum of *Synechococcus* was most prominent. On the other hand, picoeukaryotes tended to peak near surface in ice-covered stations in March and May, while in August the highest surface

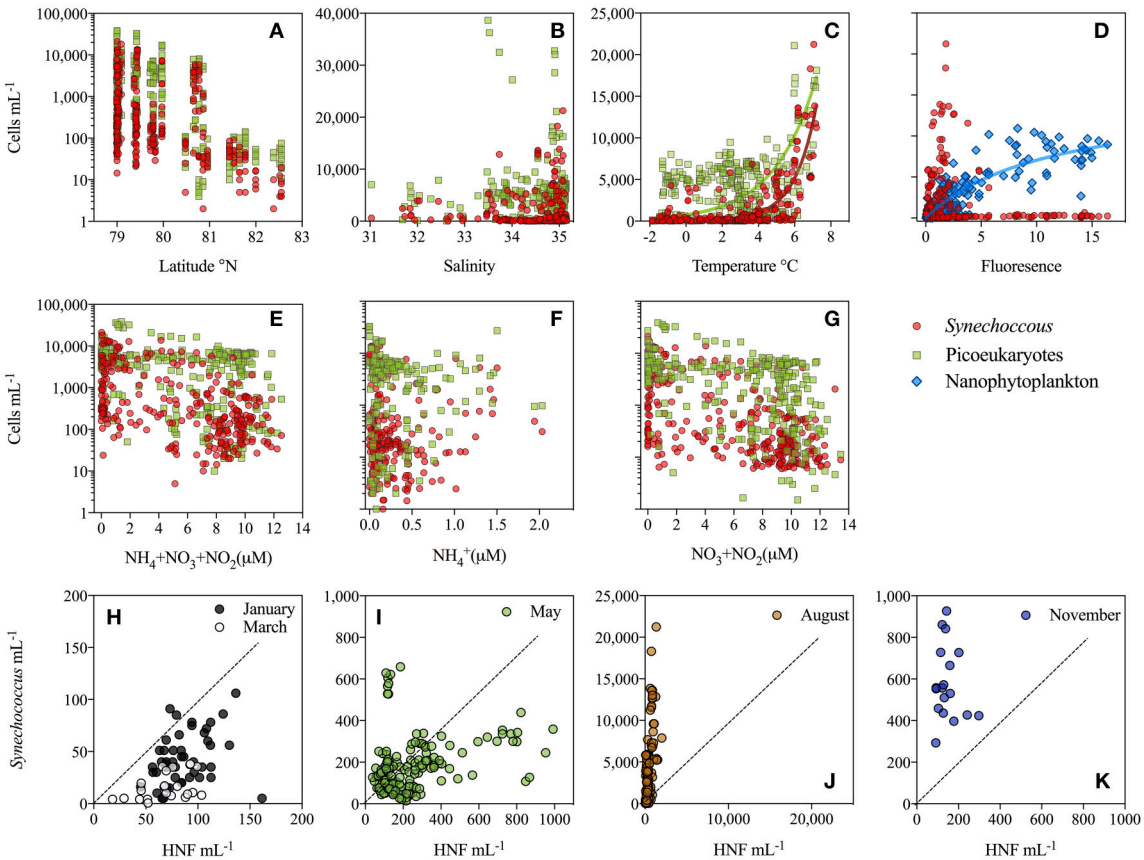


FIGURE 5 | (Two upper panels) Data collected within upper 100 m from all cruises plotted to see how abundance of *Synechococcus* (red) relates to environmental parameters, such as **(A)** latitude **(B)** salinity **(C)** temperature **(D)** chlorophyll fluorescence and **(E–G)** concentration of different sources of mineral nitrogen. Additionally **(A,B,E–G)** contain abundance of picoeukaryotes (green) and **(D)** nanophytoplankton (blue). Notice log-scale for **(A,B,E–G)**. **(Lower panel) (H–K)** *Synechococcus* abundance (mL^{-1}) plotted against HNF abundance during the different months (indicated by color) within the upper 200 m. The broken line illustrates the one-to-one line. Notice different y-axis between months. Only samples with *Synechococcus* abundance >20 cells mL^{-1} were included.

maximum of picoeukaryotes was found within the freshwater lens at stations without ice-cover (**Figure S2**).

The highest water temperatures were measured in August followed by those measured in January. The lowest surface temperatures were recorded in the ice-influenced surface waters in March and May. Temperature was the only parameter that displayed a strong relationship with *Synechococcus* abundance resulting in an exponential fit ($r^2 = 0.66$, $p < 0.005$, $n = 346$; **Figure 5C**), while picoeukaryotes did not show a similar strong relationship ($r^2 = 0.31$, $p < 0.005$, $n = 372$; **Figure 5C**). *Synechococcus* was more dominant at stations with low chlorophyll *a* (i.e., chl *a* fluorescence) compared to larger nanophytoplankton, which correlated positively to chl *a* (**Figure 5D**).

Nutrients were evenly distributed over the upper 200 m in the winter months (January to November), although a slightly lower concentration was observed in March within the upper 100 m (**Figures 4F,G,J**). In May and August all nutrients were depleted in the upper 10–20 m, with NO_3^- reaching the lowest values (**Figures 4H,I**). NH_4^+ reached the highest values around 2 µM in August at depths below 20 m. At high N concentrations (>2 µM; **Figure 5E**) *Synechococcus* were generally less abundant than picoeukaryotes, while under low N conditions they were equally numerous. When looking at the N sources separately it appears that at $NH_4^+ > 0.5$ µM, *Synechococcus* increased at higher NH_4^+ levels, whereas they decreased with increasing $NO_3^- + NO_2^-$ (**Figures 5F,G**).

The abundance of HNF increased during the summer months, from less than 200 cells mL^{-1} in the winter months up to 1000 and 1500 HNF mL^{-1} in May and August, respectively (**Figures 5H–K**). *Synechococcus* and HNF abundances generally showed a positive relationship within the upper 100 m. In January and March, *Synechococcus* and HNF cell numbers were within the same order of magnitude, but with slightly more HNF than *Synechococcus* (i.e., below the dotted line; **Figure 5H**). In May highest *Synechococcus* abundances were found at the lowest HNF abundance and vice versa (**Figure 5I**). In August, *Synechococcus* was generally 10 times more abundant than HNF, a trend also observed in November, although less pronounced (**Figures 5J,K**).

Growth and Microbial Interactions

Net-growth rates of *Synechococcus* and HNF were estimated from four different size fractions (<3, <5, <10, and <90 μm) from each of the five cruises and summarized in **Figure 6** (for abundances during incubation see **Figure S1**). *Synechococcus* showed positive net growth in 9 out of 20 experiments mainly in January and March. Positive growth rates ranged from 0.01 to 0.13 d^{-1}. HNF showed positive growth in 14 out of 20 experiments, displaying a maximum growth rate of 0.45 d^{-1} when water was filtered through a 5 μm mesh (Treat < 5 μm) in January, otherwise the highest HNF growth rates were measured in May ranging from 0.13 to 0.3 d^{-1}. In January, November, and August HNF growth was reduced to close to zero after filtering in the Treat < 90 μm and in March HNF showed negative growth in all treatments. *Synechococcus* showed positive growth in January, March in all size-fractions and in the Treat < 90 μm in May, which became strongly dominated by *Phaeocystis* sp. and where both HNF, picoeukaryotes and heterotrophic bacteria increased in abundance simultaneously. *Synechococcus* had the strongest negative growth in August and in the Treat < 3 μm in May (**Figure 6**). In summary, we measured a positive growth of *Synechococcus* and negative growth of HNF in March, but in general negative *Synechococcus* and positive HNF growth in May, August, and November. Only in January and in the May <90 μm treatment, both *Synechococcus* and HNF displayed positive net growth. Corresponding to the seasonal changes in abundance (**Figures 5H–K**) the prey:HNF ratio (prey being the sum of all picoplankton; *Synechococcus* + picoeukaryotes + heterotrophic bacteria) of the initial community was highest in May i.e., most prey per HNF grazer and lowest in August, when HNF were more abundant. Generally, the maximum growth rates of *Synechococcus* were found when prey:HNF was at its highest (**Figure S1**).

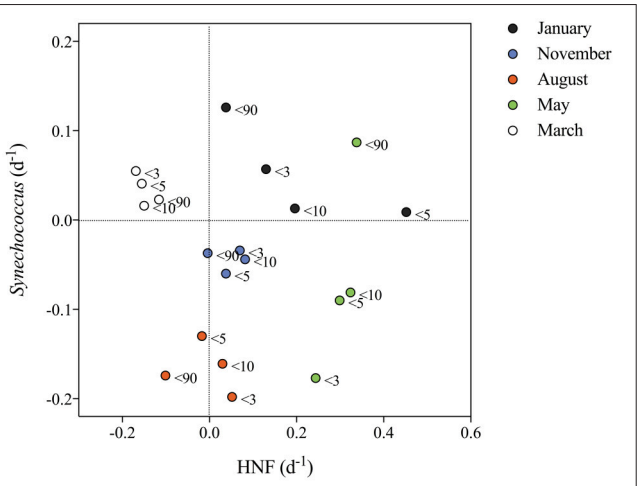


FIGURE 6 | Net growth rates (d^{-1}) of *Synechococcus* plotted against HNF net-growth. Net growth rates are obtained from the fractionation experiments from each cruise where an exponential growth curve was fitted to the change in abundance of the respective cells during a 5-day period (**Figure S1**). The color indicates the month and the legend at each point indicates the size-fraction treatment from which the values were obtained.

Synechococcus Diversity

The gene *petB*, which has proved to display a high taxonomic resolution for picocyanobacteria (Mazard et al., 2012), was used as phylogenetic marker for *Synechococcus* genetic diversity. Only *petB* sequences related to clade I and IV were retrieved from our dataset (MicroPolar). Based on a *petB* reference database (including 117 sequences from clade I and IV, described in Farrant et al., 2016), enriched with the 174 unique *petB* sequences retrieved from MicroPolar samples, 41 OTUs were defined at 97% ID within clade I and IV (**Figure 7**). The *petB* sequences obtained in the present study correspond more specifically to sub-clades Ib and IVb, with a clear dominance of subclade Ib. Although none of these 41 OTUs form a new subclade, 17 OTUs were composed of only MicroPolar sequences and were not represented in the previous reference database (colored branches, **Figure 7**). In May sub-clade Ib was the only one present, whereas subclade IVb appeared in August and increased in relative abundance in November, indicating seasonal changes in the community composition. Seasonality was also found within subclade Ib. The majority of sequences obtained in August and November belonged to two specific OTUs (Arctic732-2b_Ib_IA and Arctic732-35b_Ib_IA), which mostly gather reference sequences from the Barents Sea ("Arctic," 72.5°N, 19.57°W) and the North Atlantic Ocean (The Extended Ellett Line; "EEL," 57–63°N). In contrast, sequences retrieved from samples harvested in May were more evenly distributed over all other OTUs defined within subclade Ib that mainly gather sequences from The Atlantic Meridional Transect ("AMT," http://www.amt-uk.org/) and the North Sea ("MICROVIR" cruise 50–60°N).

In order to assess whether the genetic populations sampled in MicroPolar cruises could be related to other cold-water populations, we also recruited Illumina reads from 62 surface water metagenomes collected during the *Tara* Oceans cross-ocean ecosystem study using the same *petB* database (Karsenti et al., 2011; Farrant et al., 2016) (https://doi.pangaea.de/10.1594/PANGAEA.840718, note that the *Tara* Oceans samples analyzed here do not include recent Arctic samples from the latest *Tara* Ocean cruise as they are not yet published). These data showed that the two most abundant MicroPolar OTUs in subclade Ib (Arctic732-2b_1b_1A and Arctic732-35b_1b_1A) had a low relative abundance in *Tara* Oceans stations. Other OTUs identified in MicroPolar samples were also poorly represented in the *Tara* Oceans dataset, with the notable exception of two subclade Ib OTUs, "MP_may_P1_1m_E08_Ib_I" and "MP_may_P1_1m_D10_Ib_IA" (formed only of MicroPolar sequences), that were dominant in *Tara* Oceans coldest stations (<14°C), and of the subclade IVb OTU "Ellet21_IVb_IVC" present in the *Tara* Oceans dataset at all temperatures and especially at cold (<14°C) and intermediate (18–22°C) temperatures. Other MicroPolar OTUs were detected at a similarly low level in all temperature ranges of *Tara* Oceans stations.

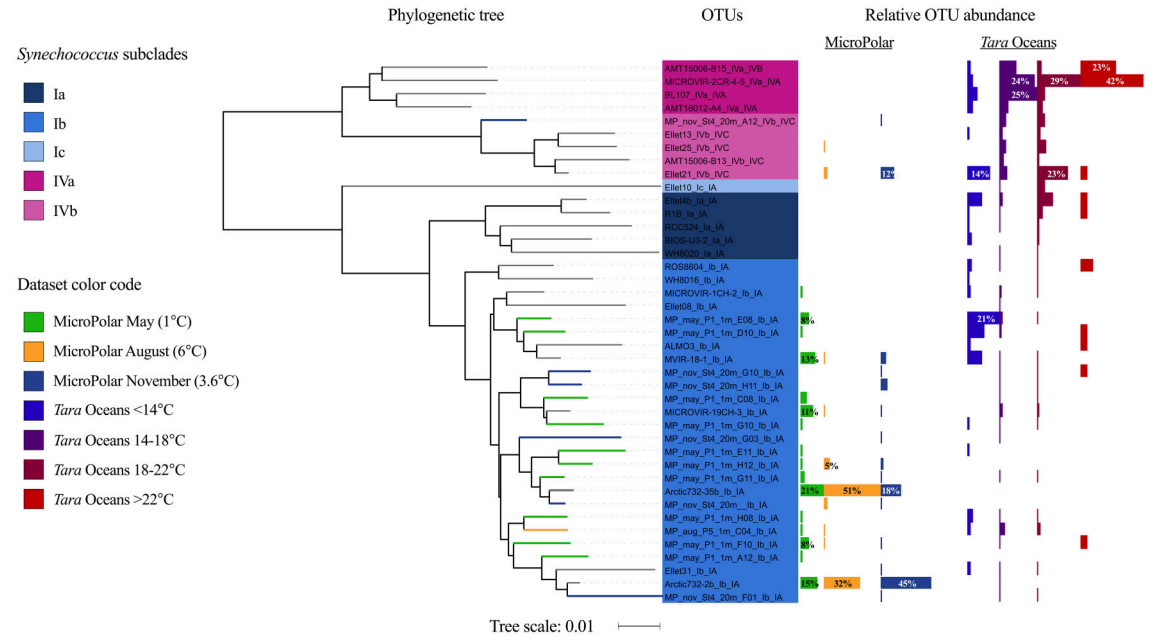


FIGURE 7 | (Left panel) Phylogenetic tree of the 41 *petB* OTUs at 97% ID, representative of the genetic diversity of *Synechococcus* within clades I and IV in the reference database (117 sequences) and in three MicroPolar samples (174 sequences). The colored thick branches indicate OTUs that are unique to the MicroPolar study (named MP_month_station_depth_) and colored according to the sampling month: May (green), August (orange) and November (blue). **(Right panel)** Relative abundance of the OTUs in each MicroPolar months (following the same color code as the branches) and in *Tara* Oceans stations grouped by surface water temperature ranges of the sampling site. The relative abundance is additionally given in percentage on selected bars.

DISCUSSION

Arctic Adaptation; *Synechococcus vs. Micromonas*

For the first time we here documented a high abundance of *Synechococcus* in the Atlantic gateway to the Arctic Ocean north of 79°N. *Synechococcus* is generally not thought to be part of the picophytoplankton community in Arctic water masses (e.g., Pedrós-Alió et al., 2015), which has repeatedly been found to be dominated by picoeukaryotes, such as *Micromonas* spp. (Not et al., 2005). Li et al. (2013) do document their existence in the Canadian Basin of the Arctic proper although as a very small fraction (2%) of the total picophytoplankton community at the only one station higher north than 70°. Arctic *Micromonas* spp. differ from *Micromonas* genotypes identified elsewhere in the World Ocean (Lovejoy et al., 2007), with these Arctic types being adapted to low temperatures. Similarly, by combining our observations with data from the *Tara* Ocean we confirmed the latitudinal shift previously described within the *Synechococcus* genus between the warm-adapted clades II and III and the cold-adapted clades I and IV (Zwirglmaier et al., 2008; Mazard et al., 2012; Farrant et al., 2016). Interestingly, clade IV was clearly dominating in Atlantic waters from the *Tara* Oceans dataset and its relative contribution seemed to increase with temperature in August and November in MicroPolar samples, while clade I appeared to dominate in colder Arctic waters. Thus, although this would need to be confirmed by physiological characterization of representative strains, it suggests that clade I could be adapted to colder waters than clade IV. Overall, it seems that temperature is the main driver of *Synechococcus* abundance and diversity in this area.

In laboratory experiments using isolates from tropical sites, *Synechococcus* has been found not to grow at temperatures below 10°C (Mackey et al., 2013), even though they have been observed in nature at temperatures as cold as 2°C (Shapiro and Haugen, 1988), and 0°C (Gradinger and Lenz, 1995). Our deck incubation experiments showed that northern *Synechococcus* populations can actually grow at 2°C, although with a quite low growth rate (maximum of 0.13 d^{-1}), suggesting a physiological adaptation of Arctic populations to low temperatures that further supports the existence of *Synechococcus* thermotypes (Pittera et al., 2014). This hypothesis is strengthened by our findings that many MicroPolar sequences formed new OTUs, unveiling an important novel genetic diversity (especially within clade I), which seems to be specific to this geographic area (17 OTUs out of the 41 OTU identified within clades I and IV). Furthermore, sequences obtained from August and November are mainly found in two OTUs within subclade Ib, gathering reference sequences retrieved only at high latitude from the Barents Sea (72°N) and the North Atlantic Ocean (57°N), but hardly detected in the *Tara* Oceans dataset. Altogether, these results point toward the existence of *Synechococcus* populations endemic to these Arctic or subarctic areas.

The peak-values of *Synechococcus* were clearly associated with the Atlantic inflow (salinity > 34.9) and abundances decreased exponentially with decreasing temperature and were

most often low in ice-associated water. This, along with the tendency of decreasing concentrations with decreasing salinity, is in accordance with the suggestion of *Synechococcus* being an indicator of saline Atlantic water transported into the Arctic (Murphy and Haugen, 1985; Gradinger and Lenz, 1995) as well as the low tolerance to wide salinity ranges of obligate marine *Synechococcus* (Waterbury et al., 1986). It should also be noted that although *Synechococcus* peak abundances were found in the relatively warm, saline Atlantic water, equally high abundances were observed in discrete samples from non-Atlantic water masses throughout the year (**Figure 3**), indicating the potential of *Synechococcus* to adapt to cold, low saline water, as also suggested by Nelson et al. (2014) for Canadian Arctic *Synechococcus*. The observed maximum abundance of picoeukaryotes, on the other hand, was found at a salinity of 33.5 and they were in general less affected by low salinities than *Synechococcus*. The dominance of picoeukaryotes over *Synechococcus* in the Arctic region may thus be connected to their capacity to stand a wide range of salinities in addition to an adaptation to low temperature. As only a few of our samples had a low salinity (17 surface samples in August have salinity <33), more efforts are needed to confirm this trend. In the Canada Basin of the Arctic Ocean proper *Synechococcus* abundance of 60 cell mL^{-1} was found at salinities substantially lower than 33 (Li et al., 2013).

The extreme changes in light conditions in polar environments may also have been a driver for the diversification of the *Synechococcus* populations. However, in contrast to *Prochlorococcus*, obvious light partitioning is usually not observed for *Synechococcus* (Scanlan et al., 2009) since only one study reported a vertical partitioning of some *Synechococcus* genotypes so far (Gutiérrez-Rodríguez et al., 2014). In our incubations *Synechococcus* surprisingly showed a net growth in January and March when light was absent or low, respectively, while picoeukaryotes did not grow (data not shown) (**Figure 6**). The ability of *Synechococcus* to grow under very low light conditions is presumably related to their capacity to consume dissolved organic matter (Palenik et al., 2003; Cottrell and Kirchman, 2009). Yelton et al. (2016) indeed found that the genetic potential for mixotrophy in picocyanobacteria (through osmotrophy) is globally distributed. Although this still needs to be confirmed by laboratory experiments, it is possible that *Synechococcus* OTUs detected in November, when there is no light, belong to mixotrophic populations that are adapted to slow growth in the dark. Picoeukaryotes may use another mixotrophic strategy, i.e., bacterial grazing, to sustain growth during dark months (Sanders and Gast, 2012). Our observations that *Synechococcus* can be more abundant than picoeukaryotes in the Arctic in autumn and winter (**Figure 4**) are consistent with previous results [Gradinger and Lenz, 1995; unpublished results from Adventfjorden, Svalbard (I. Kessel Nordgård, personal communication)] and may suggest that cyanobacterial osmotrophy is a more efficient strategy than picoeukaryotic phagotrophy to survive in the dark.

Grazing on *Synechococcus*

The highest *Synechococcus* abundances were observed when NO_3^- concentrations were low. Hence, there is no reason to believe that they were resource controlled. The tendency of increased growth when potential grazers were removed, rather points at a top-down control. The all-year-round presence of heterotrophic flagellates (HNF), considered to be their main predators (Sanders et al., 1992; Christaki et al., 2001; Kuipers et al., 2003; Zwirglmaier et al., 2009) indeed allows for grazer control of the *Synechococcus* populations. Still, grazing losses of *Synechococcus* are challenging to estimate as potential grazers can include various nano—but also microzooplankton and the specific loss also depends on the presence of other prey types (i.e., bacteria and picoeukaryotes; Pernthaler, 2005). This is illustrated by the different outcomes of successively removing various grazer fractions, which in March, August and November did not result in different growth patterns, but in January and May led to higher growth rates of *Synechococcus* when organisms larger than 90 μm were removed (**Figure S1**). Thus, this may reflect a trophic cascade where the microzooplankton graze on HNF and thereby release picoplankton from grazing pressure in the <90 μm fraction. In March, August and November, however, there was little effect of size fractionation, which indicates that small HNF (<3 μm) were the main grazers of picoplankton and that these were not grazer-controlled themselves. Exactly "who" were the most important *Synechococcus* grazers is not possible to deduce from the presented data, and probably varies over the season. In addition, infection by viruses probably also functions as a top down regulator of these *Synechococcus* populations (Sandaa and Larsen, 2006), however virus counts remained relatively constant in all five experiments (data not shown). Still, we did find the highest net growth rates for *Synechococcus* when the HNF abundance was lowest (January and March) as well as the highest *Synechococcus in situ* abundance in water with low HNF concentration (and vice versa), which is in accordance with the view that HNF control their abundance and distribution at large. The picoeukaryote abundance did not follow the same patterns (data not shown), suggesting that they may have different predators. The fact that autotrophs, such as *Synechococcus* and picoeukaryotes, persist during winter in very low abundances further suggests that low encounter rates between predator and prey in the highly diluted wintry environment release the picophytoplankton from grazing pressure and allows survival despite adverse growth conditions (Kiørboe, 2008). The experiments also illustrate that *Synechococcus* in both January, March and August have the highest growth rates in the fractions where the total prey:HNF ratio is highest, indicating that *Synechococcus* might escape the grazers when other potential prey organisms are relatively abundant.

Synechococcus As an Active Player in the Arctic and Future Implications

It may be questioned whether the observed occurrence of *Synechococcus* was simply a result of advection and passive transport via the Atlantic water inflow. Since the highest measured abundances were found within the core of the Atlantic water, this probably represents the major source. The seasonal maximum *Synechococcus* abundance, which was observed in August, does however coincide in time with the seasonal *Synechococcus* bloom further south along the Norwegian coast. Given the average transportation time is at its minimum in summer (Fahrbach et al., 2001), it seems unlikely that the

encountered seasonal change in *Synechococcus* community we observed was a mere product of advection of Atlantic water. Moreover, the spatial and temporal distribution of clades and OTUs as well as the observed growth at low temperatures when released from grazing pressure, rather suggests that at least some of the observed *Synechococcus* populations are adapted to Arctic conditions and are indigenous to these waters.

Due to their small size (1.1 ± 0.4 μm diameter in the subarctic Atlantic; Paulsen et al., 2015), *Synechococcus* cells are largely grazed by HNF and microzooplankton (Christaki et al., 1999, 2005). This implies that their biomass production will be largely recycled in the microbial food web and thus be of minor contribution to higher trophic levels in the grazing food web. Even at the highest abundances observed in this study, *Synechococcus* only constitutes a minor part of the Arctic epipelagic carbon and energy pool (e.g., 21,000 cells mL^{-1} is equal to 2.3 μg C L^{-1}, assuming a diameter of 1.1 μm and 250 fg C μm^{-3}; Kana and Glibert, 1987) relative to the total phytoplankton biomass of 42 μg C L^{-1} (assuming a carbon to chl *a* conversion of 30). A warmer Arctic ocean that may favor *Synechococcus* at the expense of larger phytoplankton species (Flombaum et al., 2013) implies that more energy and carbon could be retained within the microbial food web, further reducing the contribution of Arctic primary production to the top of the food chain.

AUTHOR CONTRIBUTIONS

MP led the collection and analysis of data, and the writing of the paper. All other authors contributed to writing the paper and in addition AL, OM, RS, and LS helped collecting data and performing experiments. LG, HD, OM, and GB helped analyse the data and prepare figures.

FUNDING

The work was conducted by the projects MicroPolar (RCN 225956) and CarbonBridge (RCN 226415), both funded by the Norwegian Research Council. HD and LG were supported by the French "Agence Nationale de la Recherche" Program SAMOSA (ANR-13-ADAP-0010).

ACKNOWLEDGMENTS

We thank the helpful crew of RV Helmer Hanssen and RV Lance and to the MicroPolar and CarbonBridge teams for always being helpful during sample collection. A special thank you to Jean-Éric Tremblay for providing ammonium measurements and to Sophie Radeke for steadfast assistance at the flow cytometer and to Hilde M. K. Stabell for making the clone library. We also thank Daniel Vaulot for fruitful discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmars.2016.00191

Figure S1 | The abundance (cells mL^{-1}) of *Synechococcus* (red) and HNF (blue) plotted on the left y-axis during the first 5 days of fractionation experiments performed during the 5 cruises. The fractions <90 μm **(A-E)**, <10 μm **(F-J)**, <5 μm **(K-O)** and <3 μm **(Q-U)** are represented on each row. Exponential functions were fitted (lines) to the abundance providing the net growth rates (μ) given in the upper left corner for *Synechococcus* (red) and HNF (blue). The total prey (sum of *Synechococcus*, picoeukaryotes and heterotrophic bacteria) to HNF ratio is plotted for each triplicate on he right y-axis (open black circles), the black line connects the daily average prey:HNF ratio.

Figure S2 | The abundance (cells mL^{-1}) of *Synechococcus* (red) and picoeukaryotes (green) for all months within the upper 500 m, except for March where profiles are shown down to 1000 and 3000 m. Horizontal light blue lines mark the stations that were influenced by sea ice. Note the different x-axis for different months. Coordinates are given for each station above each graph.

Table S1 | Environmental from the cruises containing: dates (mm/dd/yy), latitude and longitude of stations (decimal degrees), depth (m), flow cytometer counts of *Synechococcus*, picoeukaryotes, nanophytoplankton, heterotrophic bacteria, and nanoflagellates (cells mL^{-1}), the growth rates *Synechococcus* and HNF (d^{-1}) from the <90 μm incubation, salinity, temperature and potential temperature (°C), CTD-fluorescence (RUF), total chl a and the chl a fraction >10 μm (μg L^{-1}), and nutrients (NH_4^+, NO_3^-, NO_2^-, PO_4^+, $Si(OH)_4$ (μM). N.B. nutrients from January, May and August are not included here but will be available in Randelhoff et al. submitted.

Table S2 | Sequence ID of the members of each Operational Taxonomical Unit (OTU) defined for *petB* at 97% nucleotide sequence identity.

REFERENCES

Boyd, T. J., and D'Asaro, E. A. (1994). Cooling of the West Spitsbergen Current: wintertime observations west of Svalbard. *J. Geophys. Res.* 99, 22597–22618. doi: 10.1029/94JC01824

Christaki, U., Courties, C., Massana, R., Catala, P., Lebaron, P., Gasol, J. M., et al. (2011). Optimized routine flow cytometric enumeration of heterotrophic flagellates using SYBR Green I. *Limnol. Oceanogr. Methods* 9, 329–339. doi: 10.4319/lom.2011.9.329

Christaki, U., Giannakourou, A., Wambeke, F. V., and Grégori, G. (2001). Nanoflagellate predation on auto- and heterotrophic picoplankton in the oligotrophic Mediterranean Sea. *J. Plankton Res.* 23, 1297–1310. doi: 10.1093/plankt/23.11.1297

Christaki, U., Jacquet, S., Dolan, J. R., Vaulot, D., and Rassoulzadegan, F. (1999). Growth and grazing on *Prochlorococcus* and *Synechococcus* by two marine ciliates. *Limnol. Oceanogr.* 44, 52–61. doi: 10.4319/lo.1999.44.1.0052

Christaki, U., Vázquez-Domínguez, E., Courties, C., and Lebaron, P. (2005). Grazing impact of different heterotrophic nanoflagellates on eukaryotic (*Ostreococcus tauri*) and prokaryotic picoautotrophs (*Prochlorococcus* and *Synechococcus*). *Environ. Microbiol.* 7, 1200–1210. doi: 10.1111/j.1462-2920.2005.00800.x

Cokelet, E. D., Tervalon, N., and Bellingham, J. G. (2008). Hydrography of the West Spitsbergen Current, Svalbard Branch: autumn 2001. *J. Geophys. Res. Ocean.* 113, 1–17. doi: 10.1029/2007JC004150

Cottrell, M., and Kirchman, D. (2012). Virus genes in Arctic marine bacteria identified by metagenomic analysis. *Aquat. Microb. Ecol.* 66, 107–116. doi: 10.3354/ame01569

Cottrell, M. T., and Kirchman, D. L. (2009). Photoheterotrophic microbes in the arctic ocean in summer and winter. *Appl. Environ. Microbiol.* 75, 4958–4966. doi: 10.1128/AEM.00117-09

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772–772. doi: 10.1038/nmeth.2109

Doolittle, D. F., Li, W. K. W., and Wood, A. M. (2008). Wintertime abundance of picoplankton in the Atlantic sector of the Southern Ocean by. *Nov. Hedwigia* 133, 147–160.

Fahrbach, E., Meincke, J., Østerhus, S., Rohardt, G., Schauer, U., Tverberg, V., et al. (2001). Direct measurements of volume transports through Fram Strait. *Polar Res.* 20, 217–224. doi: 10.1111/j.1751-8369.2001.tb00059.x

Farrant, G. K., Doré, H., Cornejo-Castillo, F. M., Partensky, F., Ratin, M., Ostrowski, M., et al. (2016). Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proc. Natl. Acad. Sci. U.S.A.* 113, E3365–E3374. doi: 10.1073/pnas.1524865113

Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincón, J., Zabala, L. L., Jiao, N., et al. (2013). Present and future global distributions of the marine Cyanobacteria *Prochlrococcus* and *Synechococcus*. *Proc. Natl. Acad. Sci. U.S.A.* 110, 9824–9829. doi: 10.1073/pnas.1307701110

Gradinger, R., and Lenz, J. (1995). Seasonal occurrence of picocyanobacteria in the Greenland Sea and central Arctic Ocean. *Polar Biol.* 15, 447–452. doi: 10.1007/BF00239722

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704. doi: 10.1080/10635150390235520

Gutiérrez-Rodríguez, A., Slack, G., Daniels, E. F., Selph, K. E., Palenik, B., and Landry, M. R. (2014). Fine spatial structure of genetically distinct picocyanobacterial populations across environmental gradients in the Costa Rica Dome. *Limnol. Oceanogr.* 59, 705–723. doi: 10.4319/lo.2014.59.3.0705

Holmes, R. M., Aminot, A., Kérouel, R., Hooker, B. A., and Peterson, B. J. (1999). A simple and precise method for measuring ammonium in marine and freshwater ecosystems. *Can. J. Fish. Aquat. Sci.* 56, 1801–1808. doi: 10.1139/f99-128

Jürgens, K., Gasol, J. M., and Vaqué, D. (2000). Bacteria-flagellate coupling in microcosm experiments in the Central Atlantic Ocean. *J. Exp. Mar. Biol. Ecol.* 245, 127–147. doi: 10.1016/S0022-0981(99)00156-2

Kana, T. M., and Glibert, P. M. (1987). Effect of irradiances up to 2000 μE $m^{-2}s^{-1}$ on marine *Synechococcus* WH7803-I. Growth, pigmentation, and cell composition. *Deep Sea Res.* 34, 479–495. doi: 10.1016/0198-0149(87)90001-X

Karsenti, E., Acinas, S. G., Bork, P., Bowler, C., de Vargas, C., Raes, J., et al. (2011). A holistic approach to marine eco-systems biology. *PLoS Biol.* 9:e1001177. doi: 10.1371/journal.pbio.1001177

Katoh, K., and Standley, D. M. (2014). MAFFT: iterative refinement and additional methods. *Methods Mol. Biol.* 1079, 131–146. doi: 10.1007/978-1-62703-646-7_8

Kiørboe, T. (2008). "Zooplankton feeding rates and bioenergetics," in *A Mechanistic Approach to Plankton Ecology* (Oxfordshire: Princeton University Press), 224.

Koroleff, F. (1983). "Determination of nutrients," in *Methods of Seawater Analysis*, eds K. Grasshoff, M. Erhardt, and K. Kremling (Weinheim: Verlag Chemie), 125–187.

Kuipers, B., Witte, H., van Noort, G., and Gonzalez, S. (2003). Grazing loss-rates in pico- and nanoplankton in the Faroe-Shetland Channel and their different relations with prey density. *J. Sea Res.* 50, 1–9. doi: 10.1016/S1385-1101(03)00043-1

Larsen, A., Castberg, T., Sandaa, R. A., Brussaard, C. P. D., Egge, J., Heldal, M., et al. (2001). Population dynamics and diversity of phytoplankton, bacteria and viruses in a seawater enclosure. *Mar. Ecol. Prog. Ser.* 221, 47–57. doi: 10.3354/meps221047

Letunic, I., and Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23, 127–128. doi: 10.1093/bioinformatics/btl529

Li, W. K. W., Carmack, E. C., Mclaughlin, F. A., Nelson, R. J., and Williams, W. J. (2013). Space-for-time substitution in predicting the state of picoplankton and nanoplankton in a changing Arctic Ocean. *J. Geophys. Res. Ocean* 118, 5750–5759. doi: 10.1002/jgrc.20417

Lovejoy, C., Vincent, W. F., Bonilla, S., Roy, S., Martineau, M.-J., Terrado, R., et al. (2007). Distribution, phylogeny, and growth of cold-adapted picoprasinophytes in Arctic seas. *J. Phycol.* 43, 78–89. doi: 10.1111/j.1529-8817.2006.00310.x

Mackey, K. R. M., Paytan, A., Caldeira, K., Grossman, A. R., Moran, D., McIlvin, M., et al. (2013). Effect of temperature on photosynthesis and growth in marine *Synechococcus* spp. *Plant Physiol.* 163, 815–829. doi: 10.1104/pp.113.221937

Marie, D., Partensky, F., Jacquet, S., and Vaulot, D. (1997). Enumeration and cell cycle analysis of natural populations of marine picoplankton by flow cytometry using the nucleic acid stain SYBR Green I. *Appl. Environ. Microbiol.* 63, 186–93.

Mazard, S., Ostrowski, M., Partensky, F., and Scanlan, D. J. (2012). Multi-locus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*. *Environ. Microbiol.* 14, 372–386. doi: 10.1111/j.1462-2920.2011.02514.x

Morán, X. A. G., López-Urrutia, À., Calvo-Díaz, A., and Li, W. K. W. (2010). Increasing importance of small phytoplankton in a warmer ocean. *Glob. Change Biol.* 16, 1137–1144. doi: 10.1111/j.1365-2486.2009.01960.x

Moreira-Turcq, P. F., and Martin, J. M. (1998). Characterisation of fine particles by flow cytometry in estuarine and coastal Arctic waters. *J. Sea Res.* 39, 217–226. doi: 10.1016/S1385-1101(97)00053-1

Murphy, J., and Riley, J. P. (1962). A modified single solution method for the determination of phosphate in natural waters. *Anal. Chim. Acta* 26, 31–36. doi: 10.1016/S0003-2670(00)88444-5

Murphy, L. S., and Haugen, E. M. (1985). The distribution and abundance of phototrophic ultraplankton in the North Atlantic. *Limnol. Oceanogr.* 30, 47–58. doi: 10.4319/lo.1985.30.1.0047

Nelson, R. J., Ashjian, C. J., Bluhm, B. A., Conlan, K. E., Gradinger, R. R., Grebmeier, J. M., et al. (2014). "Biodiversity and biogeography of the lower trophic taxa of the Pacific Arctic Region: sensitivities to climate change," in *The Pacific Arctic Region: Ecosystem Status and Trends in a Rapidly Changing Environment*, eds J. M. Grebmeier and W. Maslowski (Dordrecht, NL: Springer Science+Business Media Dordrecht 2014), 269–336.

Not, F., Teissier, P. G., Massana, R., Latasa, M., Marie, D., Colson, C., et al. (2005). Late summer community composition and abundance of photosynthetic picoeukaryotes in Norwegian and Barents Seas. *Limnol. Oceanogr.* 50, 1677–1686. doi: 10.4319/lo.2005.50.5.1677

Palenik, B., Brahamsha, B., Larimer, F. W., Land, M., Hauser, L., Chain, P., et al. (2003). The genome of a motile marine *Synechococcus*. *Nature* 424, 1035–1037. doi: 10.1038/nature01883.1

Patel, A., Noble, R. T., Steele, J. A., Schwalbach, M. S., Hewson, I., and Fuhrman, J. A. (2007). Virus and prokaryote enumeration from planktonic aquatic environments by epifluorescence microscopy with SYBR Green I. *Nat. Protoc.* 2, 269–276. doi: 10.1038/nprot.2007.6

Paulsen, M. L., Riisgaard, K., Thingstad, T. F., and St. John, M. (2015). Winter−spring transition in the subarctic Atlantic: microbial response to deep mixing and pre-bloom production. *Aquat. Microb. Ecol.* 76, 49–69. doi: 10.3354/ame01767

Pedrós-Alió, C., Potvin, M., and Lovejoy, C. (2015). Diversity of planktonic microorganisms in the Arctic Ocean. *Prog. Oceanogr.* 139, 233–243. doi: 10.1016/j.pocean.2015.07.009

Pernthaler, J. (2005). Predation on prokaryotes in the water column and its ecological implications. *Nat. Rev. Microbiol.* 3, 537–546. doi: 10.1038/nrmicro1180

Pittera, J., Humily, F., Thorel, M., Grulois, D., Garczarek, L., and Six, C. (2014). Connecting thermal physiology and latitudinal niche partitioning in marine *Synechococcus*. *ISME J.* 8, 1221–1236. doi: 10.1038/ismej.2013.228

Saito, M. A., Rocap, G., and Moffett, J. W. (2005). Production of cobalt binding ligands in a *Synechococcus* feature at the Costa Rica upwelling dome. *Limnol. Oceanogr.* 50, 279–290. doi: 10.4319/lo.2005.50.1.0279

Sandaa, R.-A., and Larsen, A. (2006). Seasonal variations in virus-host populations in Norwegian coastal waters: focusing on the cyanophage community infecting marine *Synechococcus* spp. *Appl. Environ. Microbiol.* 72, 4610–4618. doi: 10.1128/AEM.00168-06

Sanders, R. W., Caron, D. A., and Berninger, U.-G. (1992). Relationships between bacteria and heterotrophic nanoplankton in marine and fresh waters: an inter-ecosystem comparison. *Mar. Ecol. Prog. Ser.* 86, 1–14.

Sanders, R. W., and Gast, R. J. (2012). Bacterivory by phototrophic picoplankton and nanoplankton in Arctic waters. *FEMS Microbiol. Ecol.* 82, 242–253. doi: 10.1111/j.1574-6941.2011.01253.x

Sato, M., Yoshikawa, T., Takeda, S., and Furuya, K. (2007). Application of the size-fractionation method to simultaneous estimation of clearance rates by heterotrophic flagellates and ciliates of pico- and nanophytoplankton. *J. Exp. Mar. Biol. Ecol.* 349, 334–343. doi: 10.1016/j.jembe.2007.05.027

Scanlan, D. J., Ostrowski, M., Mazard, S., Dufresne, A., Garczarek, L., Hess, W. R., et al. (2009). Ecological genomics of marine picocyanobacteria. *Microbiol. Mol. Biol. Rev.* 73, 249–299. doi: 10.1128/MMBR.00035-08

Shapiro, L. P., and Haugen, E. M. (1988). Seasonal distribution and temperature tolerance of *Synechococcus* in Boothbay Harbor, Maine. *Estuar. Coast. Shelf Sci.* 26, 517–525.

Sherr, E. B., Sherr, B. F., Wheeler, P. A., and Thompson, K. (2003). Temporal and spatial variation in stocks of autotrophic and heterotrophic microbes in the upper water column of the central Arctic Ocean. *Deep Sea Res. I* 50, 557–571. doi: 10.1016/S0967-0637(03)00031-1

Simek, K., and Chrzanowski, T. (1992). Direct and indirect evidence of size-selective grazing on pelagic bacteria by fresh-water nanoflagellates. *Appl. Environ. Microbiol.* 58, 3715–3720.

Sohm, J. A., Ahlgren, N. A., Thomson, Z. J., Williams, C., Moffett, J. W., Saito, M. A., et al. (2016). Co-occurring *Synechococcus* ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. *ISME J.* 10, 333–345. doi: 10.1038/ismej.2015.115

Tremblay, G., Belzile, C., Gosselin, M., Poulin, M., Roy, S., and Tremblay, J. (2009). Late summer phytoplankton distribution along a 3500 km transect in Canadian Arctic waters: strong numerical dominance by picoeukaryotes. *Aquat. Microb. Ecol.* 54, 55–70. doi: 10.3354/ame01257

Tremblay, J.-É., Robert, D., Varela, D. E., Lovejoy, C., Darnis, G., Nelson, R. J., et al. (2012). Current state and trends in Canadian Arctic marine ecosystems: I. Primary production. *Clim. Change* 115, 161–178. doi: 10.1007/s10584-012-0496-3

Tsai, A. Y., Gong, G. C., Sanders, R. W., and Chiang, K. P. (2013). Relationship of *Synechococcus* abundance to seasonal ocean temperature ranges. *Terr. Atmos. Ocean. Sci.* 24, 925–932. doi: 10.3319/TAO.2013.06.17.01(Oc)

Vincent, W. F. (2010). Microbial ecosystem responses to rapid climate change in the Arctic. *ISME J.* 4, 1087–1090. doi: 10.1038/ismej.2010.108

Vincent, W. F., Bowman, J. P., Rankin, L. M., and Mcmeekin, T. A. (2000). "Microbial biosystems," in *New Frontiers Proceedings of the 8th International Symposium on Microbial Ecology*, eds C. R. Bell, M. Brylinsky, and P. Johnson-Green (Halifax, NS: Atlantic Canada Society for Microbial Ecology).

Waleron, M., Waleron, K., Vincent, W. F., and Wilmotte, A. (2007). Allochthonous inputs of riverine picocyanobacteria to coastal waters in the Arctic Ocean. *FEMS Microbiol. Ecol.* 59, 356–365. doi: 10.1111/j.1574-6941.2006.00236.x

Waterbury, J. B., Watson, S. W., Valois, F. W., and Franks, D. G. (1986). "Biological and ecological characterisation of the marine unicellular cyanobacterium *Synechococcus*," in *Photosynthetic Picoplankton*, eds T. Platt and W. K. W. Li (Ottawa, ON: Canadian Bulletin of Fisheries and Aquatic Sciences), 71–120.

Wood, E. D., Armstrong, F. A. J., and Rich, F. A. (1967). Determination of nitrate in seawater by cadmium-copper reduction to nitrate. *J. Biol. Assoc. UK.* 47, 23–31.

Yelton, A. P., Acinas, S. G., Sunagawa, S., Bork, P., Pedrós-Alió, C., and Chisholm, S. W. (2016). Global genetic capacity for mixotrophy in marine picocyanobacteria. *ISME J.* doi: 10.1038/ismej.2016.64. [Epub ahead of print].

Zhang, F., He, J., Lin, L., and Jin, H. (2015). Dominance of picophytoplankton in the newly open surface water of the central Arctic Ocean. *Polar Biol.* 38, 1081–1089. doi: 10.1007/s00300-015-1662-7

Zubkov, M. V., Burkill, P. H., and Topping, J. N. (2007). Flow cytometric enumeration of DNA-stained oceanic planktonic protists. *J. Plankton Res.* 29, 79–86. doi: 10.1093/plankt/fbl059

Zwirglmaier, K., Jardillier, L., Ostrowski, M., Mazard, S., Garczarek, L., Vaulot, D., et al. (2008). Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ. Microbiol.* 10, 147–161. doi: 10.1111/j.14622920.2007.01440.x

Zwirglmaier, K., Spence, E., Zubkov, M. V., Scanlan, D. J., and Mann, N. H. (2009). Differential grazing of two heterotrophic nanoflagellates on marine *Synechococcus* strains. *Environ. Microbiol.* 11, 1767–1776. doi: 10.1111/j.1462-2920.2009.01902.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplemental Information

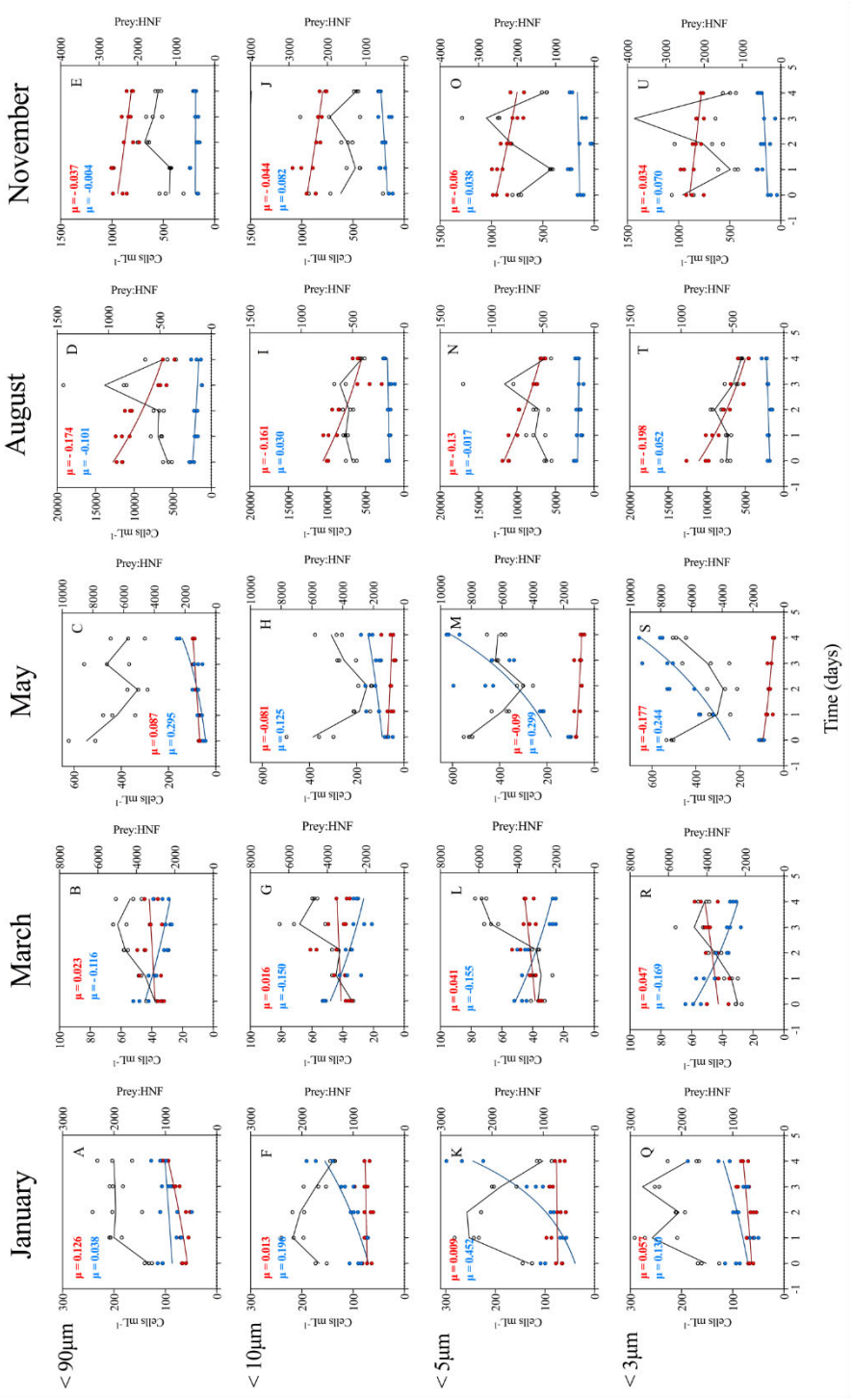


Figure S1. The abundance (cells mL^{-1}) of *Synechococcus* (red) and HNF (blue) plotted on the left y-axis during the first 5 days of fractionation experiments performed during the 5 cruises. The fractions < 90 µm (A-E), < 10 µm (F-J), < 5 µm (K-O) and < 3 µm (Q-U) are represented on each row. Exponential functions were fitted (lines) to the abundance providing the net growth rates (µ) given in the upper left corner for *Synechococcus* (red) and HNF (blue). The total prey (sum of *Synechococcus*, picoeukaryotes and heterotrophic bacteria) to HNF ratio is plotted for each triplicate on the right y-axis (open black circles), the black line connects the daily average prey:HNF ratio.

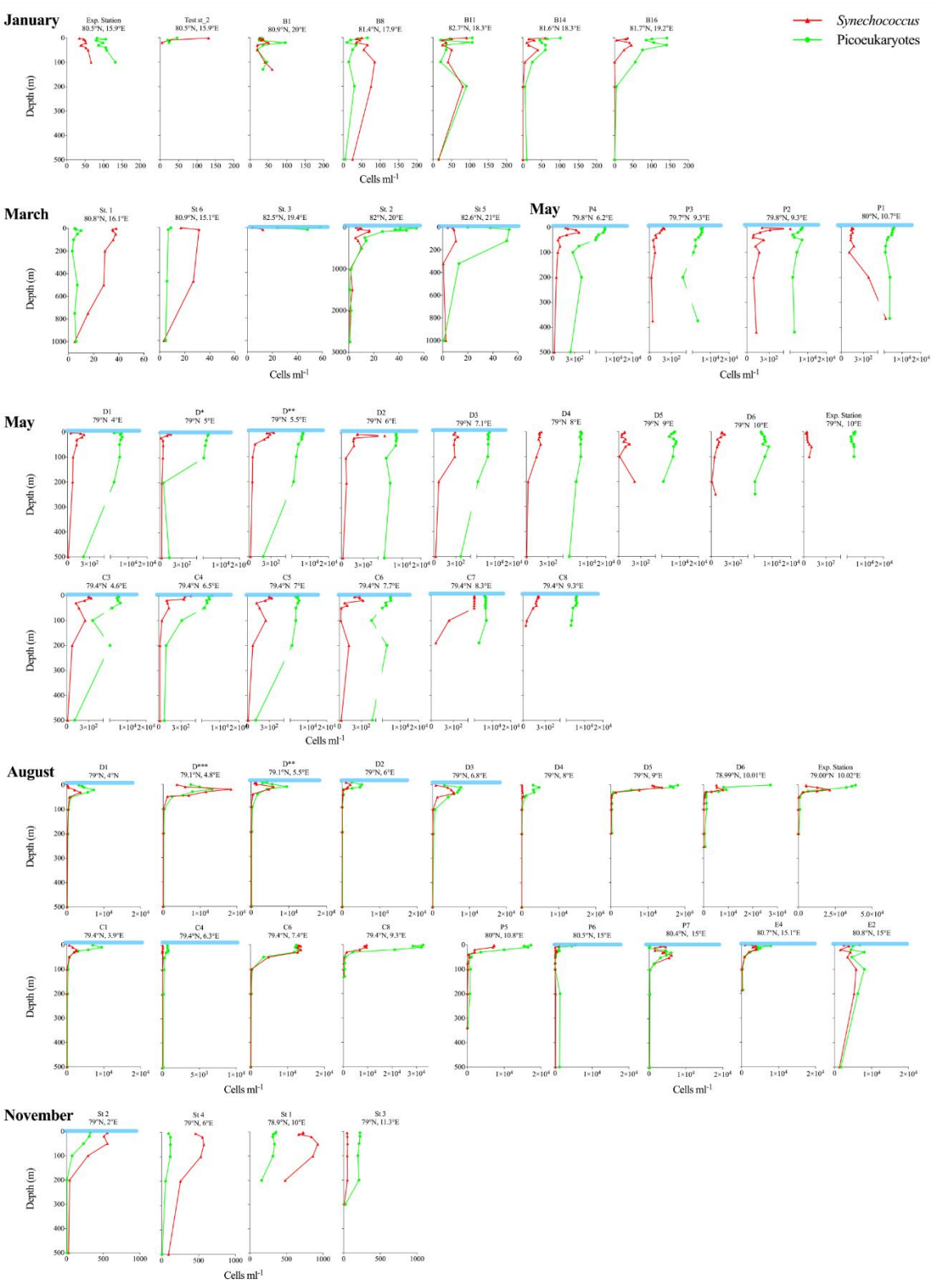


Figure S2. The abundance (cells mL^{-1}) of *Synechococcus* (red) and picoeukaryotes (green) for all months within the upper 500 m, except for March where profiles are shown down to 1000 and 3000 m. Horizontal light blue lines mark the stations that were influenced by sea ice. Note the different x-axis for different months. Coordinates are given for each station above each graph.

3. Global distribution of marine *Synechococcus* populations in coastal areas from Ocean Sampling Day metagenomes

Global distribution of marine *Synechococcus* populations in coastal areas from Ocean Sampling Day metagenomes

Running Title: Global *Synechococcus* distribution in coastal areas

Hugo Doré[1], Solène Breton[1], Justine Pittera[1], Morgane Ratin[1], David J. Scanlan[2], Frédéric Partensky[1], Christophe Six[1] and Laurence Garczarek[1]

[1] Sorbonne Universités, UPMC Université Paris 06, CNRS, UMR 7144, Station Biologique, CS 90074, Roscoff, France

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

IN PREPARATION

Significance

Marine *Synechococcus* constitute a numerically and ecologically prominent group, playing a major role in both carbon cycling and trophic networks in all oceanic regions except the Southern Ocean. Despite their high abundance in coastal areas, our knowledge of *Synechococcus* communities in these environments is only based on few local studies. Here, we use a publicly available global metagenome dataset of coastal stations sampled worldwide on a single day (Ocean Sampling Day 2014) to get a snapshot of the taxonomic composition of coastal *Synechococcus* communities at the global scale, by recruitment on a reference database of 54 *Synechococcus* genomes, representative of the whole genus diversity. This allowed us to tackle the main differences between open-ocean and coastal communities worldwide, but also to unravel drastic community shifts over small to medium scale gradients of environmental factors, in particular along European coasts. Combining analysis of the clade distribution of natural populations with the physiological characterization of thermal *preferenda* of nine strains, representative of cold or warm *Synechococcus* thermotypes, brought some additional clues about niche partitioning in this ecologically significant lineage.

Summary:

The marine cyanobacteria *Synechococcus* are among the most abundant photosynthetic organisms on earth and are widely distributed in the oceans, a success that is believed to be due to their high genetic diversity, with about 15 phylogenetic clades defined so far. The worldwide biogeography of these clades is only beginning to be understood, mostly through open-ocean cruises. Using the Ocean Sampling Day metagenomics dataset, we mapped the *Synechococcus* community at the clade level in a large range of coastal areas over the word ocean, and identified significant shifts in community composition, notably along the European coasts. Correlation of *Synechococcus* clade distribution with temperature and salinity measurements, combined with experimental characterization the thermal *preferenda* of nine strains representative of most abundant clades *in situ*, brought new insights into potential abiotic factors driving the observed community changes. Altogether, this work broadens our knowledge on the factors driving the *Synechococcus* distribution and community composition in coastal areas, where it is often the dominant cyanobacterial lineage.

Introduction

In a time of global changes that particularly affect the oceans, it is necessary to improve our knowledge of marine ecosystems in order to better anticipate their reactions. Marine picocyanobacteria, and in particular their two most abundant representatives *Prochlorococcus* and *Synechococcus*, are key players of these ecosystems: they are together responsible for around 25% of the ocean net primary productivity, and are present worldwide except in the Antarctic Ocean (Flombaum et al., 2013). *Prochlorococcus* distribution is restricted to a latitudinal band between 45°N and 45°C, and is specialized to oligotrophic open ocean areas (Partensky *et al.*, 1999b). In this genus, numerous environmental and physiological studies have allowed the identification of physiologically and genetically distinct ecotypes showing a differential niche partitioning (e.g (Moore et al., 1998; Johnson et al., 2006; Rusch et al., 2010; West et al., 2011)). In contrast much less is known about *Synechococcus*. This clade is more ubiquitous, and its ability to colonize a wide range of ecological niches has been related to its huge genetic diversity (Dufresne *et al.*, 2008; Mazard *et al.*, 2012a): between 10 and 15 phylogenetic clades have been determined in this genus, depending on the phylogenetic marker used (Fuller *et al.*, 2003; Zwirglmaier *et al.*, 2008; Mazard *et al.*, 2012a). Some global distribution patterns of *Synechococcus* populations have been determined, showing that clades I, II, III, IV and CRD1 (*sensu* (Mazard *et al.*, 2012a)) usually dominate *Synechococcus* communities in the oceans (Zwirglmaier et al., 2007, 2008; Huang et al., 2012; Sohm et al., 2015; Farrant et al., 2016). However, most of the global studies were carried out in the open ocean and much less is known about the distribution of *Synechococcus* populations in coastal areas at a global scale, despite the fact that *Synechococcus* is often the dominant cyanobacterial group in these environments (Partensky *et al.*, 1999a).

Among the environmental factors that have been shown to drive the distribution of *Synechococcus* populations *in situ*, water temperature is one of the strongest (Zwirglmaier et al., 2008; Sohm et al., 2015; Farrant et al., 2016). More specifically, clades I and IV have been shown to dominate *Synechococcus* communities in cold and nutrient-rich areas, while clades II, III and CRD1 are found in warmer waters. These differences in adaptation to temperature have been confirmed by the experimental study of the physiology of 6 strains, which showed that distinct thermotypes could be distinguished based on their thermal physiology, in accordance to their latitude of isolation (Pittera et al., 2014).

Here, we use the metagenomics data from the Ocean Sampling Day (OSD) 2014 campaign (Kopf et al., 2015) to extend our knowledge of *Synechococcus* distribution in coastal areas. All samples in this dataset were taken on June 21st, 2014 using the same protocol, allowing a fair comparison of sampling sites. Most of the samples were taken near coast, and the participation of many laboratories around Europe allowed to get a good spatial resolution in this region of the globe. Using a Whole Genome Recruitment approach against a database of publicly available marine prokaryotic genomes, we assess the distribution of *Synechococcus* clades in OSD samples, and analyze the results in the light of new experimental data on the ability of different clade representatives to acclimate to a range of temperatures. The good spatial resolution achieved in Europe allowed us to reveal several spatial community shifts in its coastal waters, with a potential role of temperature and salinity as drivers of community composition.

Results and Discussion

Global geographic distribution of coastal communities of marine picocyanobacteria

The 150 samples that were collected in the framework of the OSD 2014 campaign (Kopf et al., 2015) for which metagenomes were available, corresponding to a total of 41 Gbp (168.7 million reads), were used in this study. This campaign is complementary to other recent global expeditions, such as *Tara* Oceans (Sunagawa et al., 2015; Farrant et al., 2016) and Malaspina (Pernice et al., 2015), in the sense that most of the sampled sites correspond to coastal areas (average distance to the nearest coast: 6.3 nautical miles, median: 0.29 nautical miles, only 17 stations located over 11 nautical miles from nearest coast). While few of the sampled sites were located in the southern hemisphere (7 out of 150), this dataset displays a very good spatial resolution in some regions of the world ocean and in particular along European and Eastern United States coasts (Fig. S1). Here, we applied a Whole Genome Recruitment (WGR) approach against a reference genome database encompassing 97 genomes of marine picocyanobacteria as well as 722 cyanobacterial or other aquatic microbial genomes used as outgroups to assess the relative abundance of *Synechococcus* and *Prochlorococcus* clades (Fig. 1). *Prochlorococcus* was only abundant at a few stations, likely due to the coastal localization of the sampling sites (Partensky *et al.*, 1999a). On the contrary, *Synechococcus* that is known to outnumber *Prochlorococcus* in coastal areas (Partensky *et al.*, 1999a), was detected with sufficient coverage to perform reliable taxonomic assignment at the clade level (see Experimental Procedures) in 102 out of 150 OSD samples. At most stations, *Synechococcus* community was dominated by one or two taxa among clades I, II, III, IV and subclusters 5.2 and 5.3 (Fig. 1). Consistent with previous studies of the picocyanobacterial distribution in the open ocean (Zwirglmaier et al., 2008; Mella-Flores et al., 2011; Sohm et al., 2015; Farrant et al., 2016), clades I and IV dominated at high latitudes (above 35°N, except in the Mediterranean Sea) and clade II at low latitudes (below 35°N), while clade III was present almost exclusively in the Mediterranean Sea. It is also worth noting that the co-occurrence of clades I and IV at the few stations above 35° in the Southern hemisphere mirrored the profiles obtained at the same latitude in the Northern hemisphere, in agreement with their characterization as “cold thermotypes” (Pittera et al., 2014). Coastal *Synechococcus* communities also exhibited some specificities as compared to open ocean populations, notably the very low amount of clade CRD1, which was shown to be prevalent in large regions of the open ocean that are limited in iron (Sohm et al., 2015; Farrant et al., 2016), as well as the dominance of subcluster 5.2 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 an 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 (Fuller et al., 2003; Chen et al., 2006; Dufresne et al., 2008). This hypothesis was further confirmed by clustering stations according to the relative abundance profiles of *Synechococcus* clades, which clearly separated stations dominated by subcluster 5.2 and showed that they had a lower salinity than most other stations (Fig. 2). Finally clades V and VI, which were most often not distinguished from clade VII (and CRD1) in previous surveys of *Synechococcus* distribution, the V/VI/VII group being considered to be widely distributed in oceanic waters (Zwirglmaier et al., 2008; Mella-Flores et al., 2011), were found to be locally abundant in our dataset, revealing their potential preference for coastal areas.

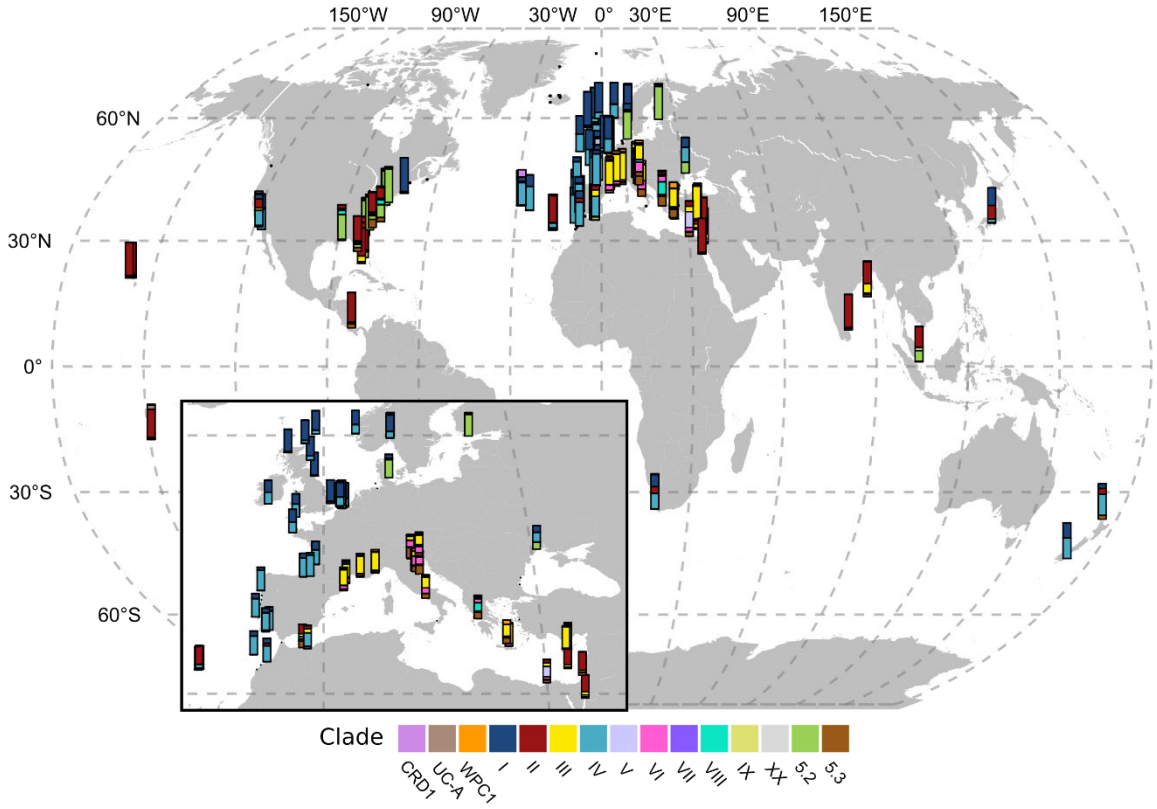


Figure 1 Relative abundance of marine *Synechococcus* clades in OSD stations. Stations with insufficient coverage are represented by a dot. Stations are located at the bottom of histograms of relative abundance. The insert shows a close-up version of Europe. Station numbers are shown in Figure S1.

A progressive latitudinal shift in communities along the coast of Europe

Besides the above mentioned specificities of coastal regions in terms of *Synechococcus* 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 southwestern coast, *Synechococcus* communities were dominated by clade IV (e.g. at OSD92, close to the Moroccan coast), 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. 2). 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 16.5°C, median 17°C) than all other clusters except cluster 3 dominated by clade IV (average temperature 19.1°C, median 18.7°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 water temperature seem to be sufficient to fully explain the observed changes.

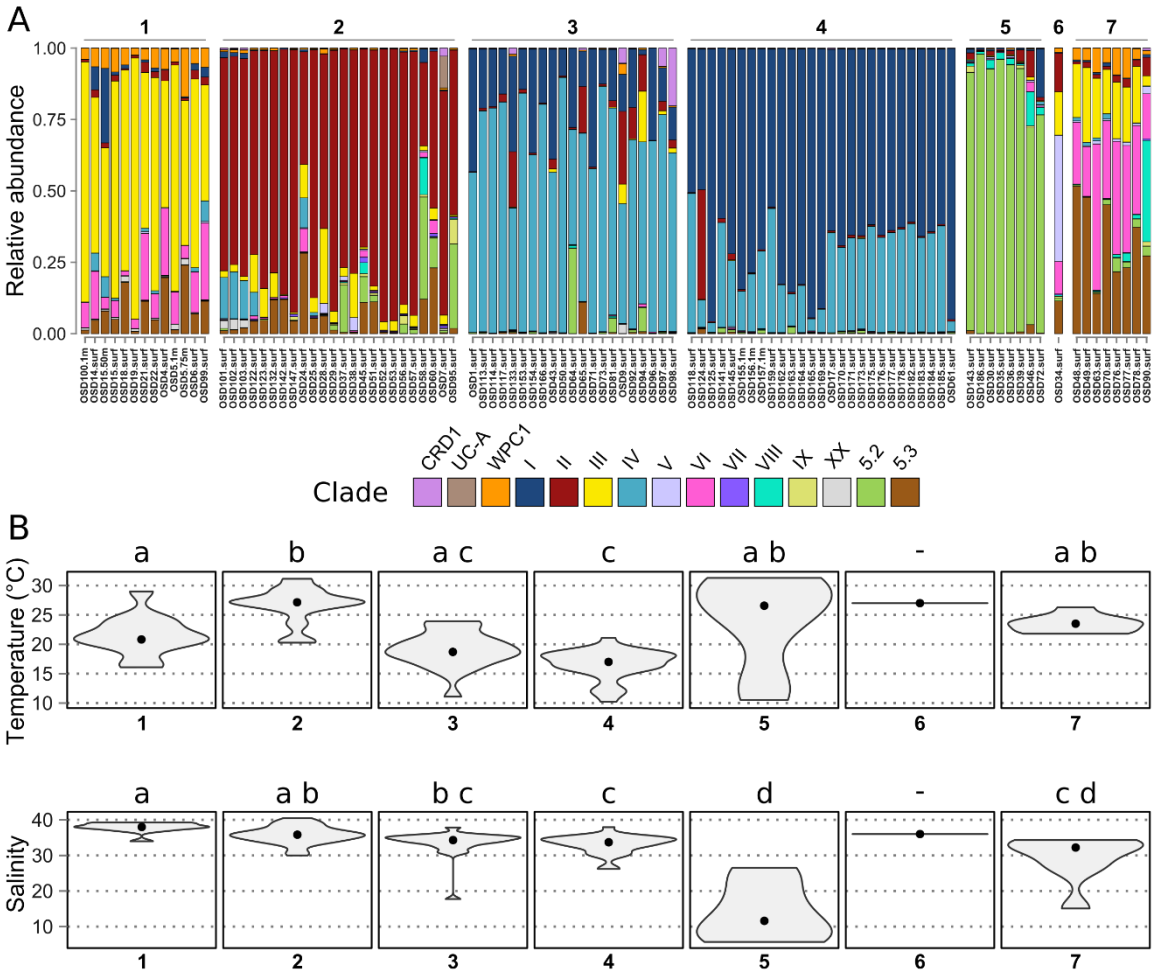


Figure 2 Clusters of OSD stations based on relative abundance profiles of *Synechococcus* clades and corresponding environmental conditions. A- OSD stations were clustered based on the relative abundance profiles of marine *Synechococcus* clades using Bray-Curtis distance. B. Violin plots showing distribution of temperature (°C) and salinity for each cluster. 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.

Several other potential reasons had been previously advanced to explain co-occurrence of clades I and IV and variations in their relative abundance, including difference in metal concentration requirements (Tai and Palenik, 2009; Sohm et al., 2015), transport of populations by advection as suggested for the Korean sea, where clades I, II and IV were found to co-occur (Choi et al., 2013) or for waters located 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 (Paulsen et al., 2016). In this context, it is interesting to note that the only Japanese station in the OSD dataset (OSD124, Fig. S1) was described by samplers 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). Alternatively, some studies also suggested that clade I could be a more coastal and opportunistic clade (Tai and Palenik, 2009; Dufresne *et al.*, 2008), 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.

In order to further explore whether temperature could have a role in the observed northward latitudinal shift between clades IV and I, we characterized the thermal *preferenda* of nine strains,

including two clade I and two clade IV strains (Fig 3). While several strains belonging to clade I were previously shown to stand colder temperatures than their tropical clade II counterparts (Pittera et al., 2014), no studies on growth optima and boundary limits for temperature was available for clade IV strains. We show here that clades I and IV strains showed quite similar thermal preferences, with an optimal temperature for growth of 24°C according to our model fit (Fig. 3 and Table 1), and they were all able to grow at the lowest tested temperature, i.e. 10°C, which is also the lowest temperature measured in the OSD stations, where *Synechococcus* community was analyzed. In comparison, strains of clades II and III were not able to grow at temperatures of 13°C and below. While this brings an experimental evidence that clades I and IV are both cold thermotypes, and partially explains their global dominance at latitudes above 35°, it also tends to confirm that temperature alone do not explain their differential distribution, with the caveat that only two strains per clade were analyzed. Thus, it seems other parameters or combinations of parameters varying with latitude should be taken into account to explain this shift. Notably, an interaction between light and temperature on phytoplankton physiology has been described in a number of species (Edwards et al., 2016), and these two parameters vary greatly with latitude. Alternatively, as previously suggested for clade I in a number of environmental studies (Tai *et al.*, 2011; Paulsen *et al.*, 2016; Choi *et al.*, 2016; Xia *et al.*, 2017), it is possible that clade I and potentially clade IV are constituted of distinct genotypes exhibiting distinct thermal preferenda and colonizing different temperature niches. Yet no isolate representative of the very cold niches found at high latitudes is currently available in culture, which would allow to confirm this hypothesis. Finally, we cannot rule out the hypothesis that the co-occurrence of clades I and IV might be due to distinct yet co-occurring environmental microniches.

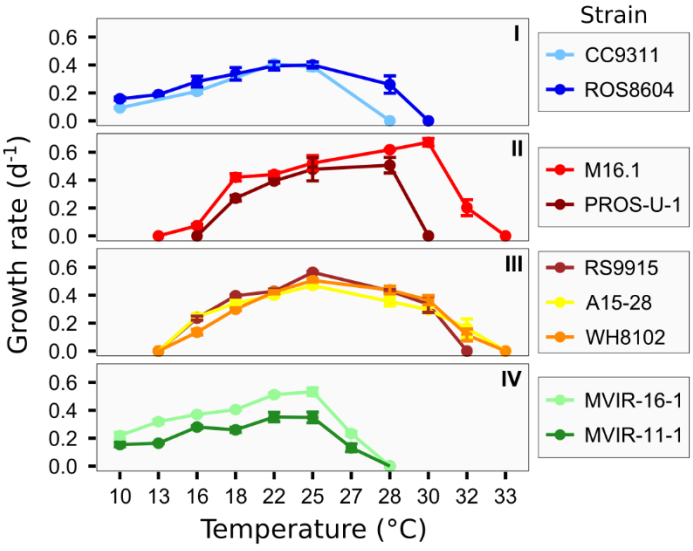


Figure 3 Temperature preferenda of 9 marine *Synechococcus* strains. Growth rate as a function of temperature of acclimated growth. At least 2 strains were chosen within each of the four major clades I, II, III and IV (top to bottom). All cultures were grown at a light intensity of 20 µmol photons m^{-2} s^{-1}.

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* clades. Most stations belonged to cluster 1, dominated by clade III with a low relative abundance of clades VI, WPC1 and subcluster 5.3. This clade composition is quite similar to that recently described by (Farrant et al., 2016) for open waters of the Mediterranean Sea, which was

suggested to be related to specific features of this sea and notably its low phosphate concentration (Zwirglmaier et al., 2008; Mella-Flores et al., 2011; Farrant et al., 2016), a parameter that was not available in our dataset. Most of the stations of the Adriatic Sea formed a distinct cluster (7), where the same clades were present but in different proportions, clade VI and subcluster 5.3 taking over clade III. Finally, station OSD34, located on the Egyptian coast, the only station of the OSD dataset comprising a high proportion of clade V, formed a cluster on its own (cluster 6). While these three clusters (cluster 1, 6 and 7) are specific to the Mediterranean Sea, it is worth noting that two stations at the easternmost end of the Mediterranean Sea (OSD123 and OSD 132, Fig. 2 and Fig. S1) fell into cluster 2, dominated by clade II, and showed a clade composition very similar to the sample collected in the Red Sea, suggesting that Israeli coastal waters are strongly influenced by waters entering the Mediterranean Sea via the Suez Canal, consistent with previous finding by (Feingersch et al., 2010) for *Synechococcus* as well as for larger organisms (see e.g. (Katsanevakis et al., 2014).

Interestingly, the three specific clusters identified in the Mediterranean Sea displayed different temperature and salinity characteristics (Fig. 2B). The salinity range of stations in cluster 1 (dominated by clade III) was narrow (average salinity 37.9 psu, median 37.98 psu) and significantly higher than that of cluster 7 (dominated by clade VI and subcluster 5.3, average salinity 29.38 psu, median 32.19 psu), suggesting that clade VI and subcluster 5.3 are able to cope with quite low salinities. Consistently, subcluster 5.3 was recently found to encompass members colonizing freshwater lakes (Cabello-Yeves et al., 2017) although in the marine environment, this subcluster was so far only reported in strictly marine waters (Huang et al., 2012; Farrant et al., 2016). Our study also brings new insight onto the ecological niche occupied by clade VI, whose distribution was so far poorly known, and that appears to be restricted to coastal regions of intermediate salinity. All stations of the Adriatic Sea constituting cluster 7 were sampled in the northwestern part of the Sea, where the influence of the Po River plume is important, particularly in summer (Lipizer et al., 2014). Thus, besides a lower salinity, other local specificities linked to this riverine input might also explain the predominance of subcluster 5.3 and clade VI in these coastal areas.

Table 1 Parameters of growth *versus* temperature 8 *Synechococcus* strains representative of the four main clades. 2 strains were chosen within each of the four major clades I, II, III, and IV. All cultures were grown at a light intensity of 20 µmol photons m^{-2} s^{-1}. T_{max}: maximal temperature for growth; T_{opt}: optimal temperature for growth; T_{min} (minimal temperature for growth) is not shown because it could not be constrained by the model. "measured": measured values, see Fig. 3. "model": values estimated by a model of growth *versus* temperature fitted on the data shown in Fig. 3. "range": confidence intervals of model predictions.

	Strain	T_{opt} measured	T_{opt} model	T_{opt} range	T_{max} measured	T_{max} model	T_{max} range
Clade I	CC9311	22	24.6	[22.98-25.91]	25	27.7	[26.23-28.97]
	ROS8604	25	23.7	[22.23-24.80]	28	29.7	[28.19-31.36]
Clade IV	MVIR-16-1	25	24.3	[23.37-25.09]	27	27.3	[27.06-27.47]
	MVIR-11-1	22	23.8	[22.42-25.05]	27	27.3	[26.93-27.63]
Clade III	RS9915	25	24.8	[23.11-28.44]	32	32.2	[25.37-34.88]
	A15-28	25	23.4	[22.97-23.72]	32	33.9	[32.79-34.82]
Clade II	M16.1	30	29.0	[27.99-30.19]	32	32.3	[31.93-32.51]
	PROS-U-1	28	26.8	[25.58-27.96]	28	30.0	[3.04-39.90]

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 26.6°C, median 27.1°C). This suggests that the shift observed at the Easternmost part of the Mediterranean Sea from dominance of clade III to a local dominance of clade II (stations OSD123 and OSD 132, Fig. 1 and Fig. S1) might be related to a difference in water temperature, which is potentially exceeding the tolerance range of clade III at the warmest stations of the Mediterranean Sea. Interestingly, in contrast to clades I and IV that often co-occur, clades II and III seem to be mutually exclusive, at least in the Mediterranean Sea, and the temperature limit above which clade II dominates seems to lie around 25°C (Fig. 2). 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, consistent with the higher optimal temperature of clade II compared to clade III strains (Table 1). Altogether, temperature and salinity appear as major factors driving the composition of *Synechococcus* communities in coastal waters of the Mediterranean Sea, although we cannot rule out the potential role of other co-varying parameters, including biotic factors.

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 open oceans (Sunagawa *et al.*, 2015; Farrant *et al.*, 2016; Pernice *et al.*, 2015). 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 few physico-chemical parameters were collected, this dataset allowed us to considerably improve our knowledge of the distribution of *Synechococcus* clades in coastal areas, to gain insights into the environmental *preferenda* of some clades that were previously poorly known such as clade VI, as well as to reinforce hypotheses about thermal niche differentiation that were supported by physiological measurements on a set of representative strains. Maintenance of such a global instantaneous survey of microbial diversity in coastal areas on the long term would be invaluable to monitor the evolution of microbial communities in relation to global climate changes.

Acknowledgements

We thank the OSD Consortium for sampling, sequencing and making freely available the data analyzed in this paper, David Demory for his help to fit the model on temperature data as well as the Roscoff Culture Collection (http://roscoff-culture-collection.org/) for providing *Synechococcus* strains used in this study. Analyses were performed thanks to the bioinformatics capacities of the ABiMS platform in Roscoff (http://abims.sb-roscoff.fr/). Financial support for 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" program SAMOSA (ANR-13-ADAP-0010).

Experimental Procedures

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. Details about sampling methods can be found in (Ten Hoopen et al., 2014) and at https://www.microb3.eu/sites/default/files/osd/OSD_Handbook_June_2015.pdf. 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 EBI Metagenomics portal under the project accession number ERP009703 (processed data). Data were downloaded from the EBI for the 150 stations for which a "Processed reads without annotation" file was available. Reads were processed 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, then trimmed for low quality ends and sequences with more than 10% undetermined nucleotides were removed using Trimmomatic (Bolger et al., 2014), then reads shorter than 100 nucleotides were discarded. Contextual data at all OSD stations were retrieved from PANGAEA (https://doi.pangaea.de/10.1594/PANGAEA.854419;) (Ocean Sampling Day Consortium, 2015) and those used in this study are listed in Dataset 1. A map of numbered OSD samples used in this study is available as Fig. S1.

Taxonomic assignment of metagenomic reads

BLASTN (v2.2.28+) (Altschul et al., 1990; Camacho et al., 2009) was used to align metagenomic reads against a database of 819 complete genomes of aquatic bacteria (Dataset 2), including 97 genomes of marine picocyanobacteria and 722 outgroup genomes, among which 185 cyanobacterial genomes other than *Prochlorococcus* and marine *Synechococcus* were downloaded from NCBI ftp on May 4, 2017 and 537 genomes of aquatic microbes were downloaded from proGenomes on June 1, 2017 ("aquatic representatives", 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 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). BLASTN results were then parsed using the Lowest Common Ancestor method (Huson et al., 2007). For each read, blast matches with over 80% ID 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 *sensu* (Farrant et al., 2016)). Counts of reads assigned to the strain or clade level were ultimately aggregated by clade, while others were discarded.

Analysis of picocyanobacterial community composition

In order to take into account the potential variation in genome length between clades, we divided read counts by the average genome length within each clade. To minimize the noise in recruitment data, we then removed stations with less than 600 recruited reads per million bp (corresponding to a genome coverage of 15%, since reads are 242 bp in average) from the analysis. Read counts in each station were

further normalized by the total number of recruited reads in the station to assess relative abundances of taxa. We used the R packages cluster v1.14.4 (Maechler et al., 2013) and vegan v2.2-1 (Oksanen et al., 2015) to cluster stations according to the Bray-Curtis distance. Figures were drawn in R v3.03 with package ggplot2 v1.0.1 (Wickham, 2009).

Temperature preferenda of representative strains

Two strains of each of the most abundant *Synechococcus* clades in Fe replete areas (clades I to IV) were chosen and retrieved from the Roscoff culture collection (Table 1; http://roscoff-culture-collection.org/; (Vaulot et al., 2004)). Strains were grown in polystyrene flasks in PCR-S11 medium (Rippka et al., 2000) with the following modifications: 2 µM Na_2EDTA-$FeCl_3$, 1µg/L vitamin B12, twice the concentration of the Gaffron+Se trace-metal solution and addition of 1mM sodium nitrate. The seawater was reconstituted using Red Sea Salt (Houston, TX, USA) and distilled water. Cultures of the eight strains were long-term 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 (Marie et al., 1999). 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. Growth rates were computed as the slope of a ln(Nt) *vs* time plot (with Nt the cell concentration at time t) on the linear part of the curve (i.e. in exponential growth phase only).

Supplementary information

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

Dataset 1: OSD samples used in this study. Accession numbers, location and physico-chemical parameters of all stations used in this study.

Dataset 2: Reference genomes used for whole-genome recruitment. Strain name and accession number for all the genomes used in this study.

References

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

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

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

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

Chen F, Wang K, Kan J, Suzuki MT, Wommack KE. (2006). Diverse and Unique Picocyanobacteria in Chesapeake Bay, Revealed by 16S-23S rRNA Internal Transcribed Spacer Sequences. *Appl Environ Microbiol* **72**: 2239–2243.

Choi DH, Noh JH, An SM, Choi YR, Lee H, Ra K, et al. (2016). Spatial distribution of cold-adapted *Synechococcus* during spring in seas adjacent to Korea. *Algae* **31**: 231–241.

Choi DH, Noh JH, Shim J. (2013). Seasonal changes in picocyanobacterial diversity as revealed by pyrosequencing in temperate waters of the East China Sea and the East Sea. *Aquat Microb Ecol* **71**: 75–90.

Dufresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik BP, et al. (2008). Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol* **9**: R90.

Edwards KF, Thomas MK, Klausmeier CA, Litchman E. (2016). Phytoplankton growth and the interaction of light and temperature : A synthesis at the species and community level. *Limnol Oceanogr* **61**: 1232–1244.

Farrant GK, Doré H, Cornejo-castillo FM, Partensky F, Ratin M, Garczarek L. (2016). Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proc Natl Acad Sci* **113**: E3365-74.

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

Flombaum P, Gallegos JL, Gordillo R a, Rincón J, Zabala LL, Jiao N, et al. (2013). Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc Natl Acad Sci U S A* **110**: 9824–9.

Fuller NJ, Marie D, Partensky F, Vaulot D, Post AF, Scanlan DJ. (2003). Clade-Specific 16S Ribosomal DNA Oligonucleotides Reveal the Predominance of a Single Marine *Synechococcus* Clade throughout a Stratified Water Column in the Red Sea. *Appl Environ Microbiol* **69**: 2430–2443.

Ten Hoopen P, Cochrane G, Micro B3Consortium. (2014). Ocean Sampling Day Handbook - Version of June 2014. http://www.microb3.eu/ sites/default/files/osd/OSD_Handbook_v2.0.pdf. 2014.

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

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

Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward EMS, Chisholm SW. (2006). Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* **311**: 1737–1740.

Katsanevakis S, Coll M, Piroddi C, Steenbeek J, Ben Rais Lasram F, Zenetos A, et al. (2014). Invading the Mediterranean Sea: biodiversity patterns shaped by human activities. *Front Mar Sci* **1**: 1–11.

Kopf A, Bicak M, Kottmann R, Schnetzer J, Kostadinov I, Lehmann K, et al. (2015). The ocean sampling day consortium. *Gigascience* **4**: 27.

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

Maechler M, Peter R, Struyf A, Hubert M, Hornik K. (2013). cluster: Cluster Analysis Basics and Extensions.

Marie D, Brussaard CPD, Partensky F, Vaulot D. (1999). Flow cytometric analysis of phytoplankton, bacteria and viruses. In: Robinson JP (ed) Vol. 11.11. Current protocols in cytometry. John Wiley & Sons: New York, pp 1–15.

Mazard S, Ostrowski M, Partensky F, Scanlan DJ. (2012). Multi-locus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*. *Environ Microbiol* **14**: 372–86.

Mella-Flores D, Mazard S, Humily F, Partensky F, Mahé F, Bariat L, et al. (2011). Is the distribution of *Prochlorococcus* and *Synechococcus* ecotypes in the Mediterranean Sea affected by global warming? *Biogeosciences* **8**: 2785–2804.

Moore LR, Rocap G, Chisholm SW. (1998). Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393**: 464–467.

Ocean Sampling Day Consortium P. (2015). Registry of samples and environmental context from the Ocean Sampling Day 2014. e-pub ahead of print, doi: 10.1594/PANGAEA.854419.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, et al. (2015). vegan: Community Ecology Package. http://cran.r-project.org/package=vegan.

Partensky F, Blanchot J, Vaulot D. (1999a). Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters: a review. *Bull l'institut océanographique* **19**: 457–475.

Partensky F, Hess WR, Vaulot D. (1999b). *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* **63**: 106–127.

Paulsen ML, Doré H, Garczarek L, Seuthe L, Müller O, Sandaa R-A, et al. (2016). *Synechococcus* in the Atlantic Gateway to the Arctic Ocean. *Front Mar Sci* **3**. e-pub ahead of print, doi: 10.3389/fmars.2016.00191.

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

Pittera J, Humily F, Thorel M, Grulois D, Garczarek L, Six C. (2014). Connecting thermal physiology and latitudinal niche partitioning in marine *Synechococcus*. *ISME J* **8**: 1221–1236.

Rippka R, Coursin T, Hess W, Lichtlé C, Scanlan DJ, Palinska KA, et al. (2000). *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* **50**: 1833–1847.

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

Sohm J a, Ahlgren N a, Thomson ZJ, Williams C, Moffett JW, Saito M a, et al. (2015). Co-occurring *Synechococcus* ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. *ISME J* **10**: 1–13.

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

Tai V, Burton R, Palenik B. (2011). Temporal and spatial distributions of marine *Synechococcus* in the Southern California Bight assessed by hybridization to bead-arrays. *Mar Ecol Prog Ser* **426**: 133–147.

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

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

West NJ, Lebaron P, Strutton PG, Suzuki MT. (2011). A novel clade of *Prochlorococcus* found in high nutrient low chlorophyll waters in the South and Equatorial Pacific Ocean. *ISME J* **5**: 933–944.

Wickham H. (2009). ggplot2: elegant graphics for data analysis. *Springer New York*.

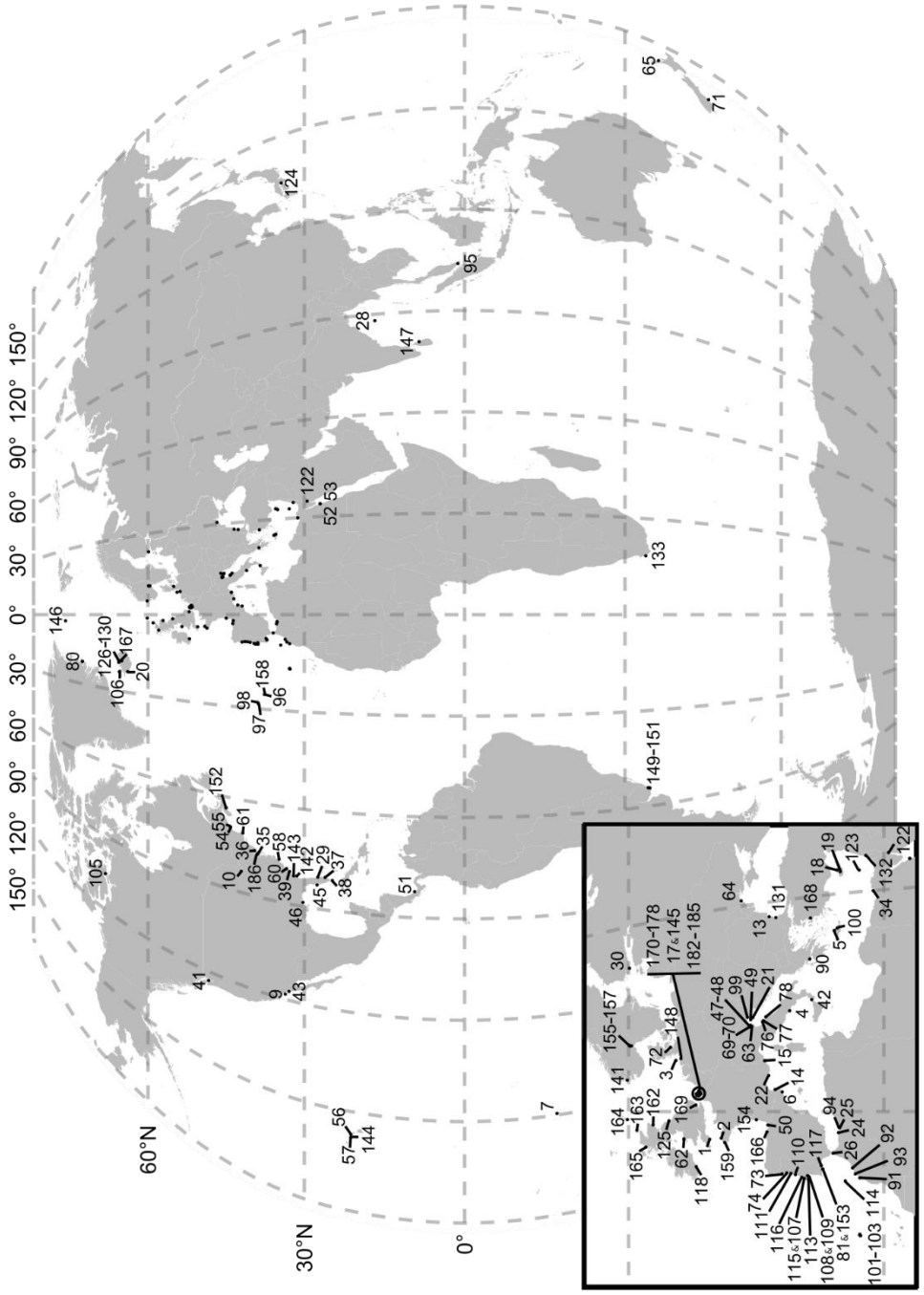
Xia X, Partensky F, Garczarek L, Suzuki K, Guo C, Cheung SY, et al. (2017). Phylogeography and pigment type diversity of *Synechococcus* cyanobacteria in surface waters of the northwestern Pacific Ocean. *Environ Microbiol* **19**: 142–158.

Zwirglmaier K, Heywood JL, Chamberlain K, Woodward EMS, Zubkov M V, Scanlan DJ. (2007). Basin-scale distribution patterns of picocyanobacterial lineages in the Atlantic Ocean. *Environ Microbiol* **9**: 1278–1290.

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

Supplementary Information

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



CHAPITRE II

GENETIC BASES OF ADAPTATION IN MARINE PICOCYANOBACTERIA

Context of the work and personal contribution

The picture of global distribution of marine picocyanobacteria that was drawn in the previous chapter clearly showed that the two genera *Synechococcus* and *Prochlorococcus* each encompasses genetically distinct populations that are adapted to distinct niches. More precisely, we delineated the global abiotic niche of each clade that were globally in agreement with previous studies on the genetic diversity and distribution of marine picocyanobacteria (Johnson *et al.*, 2006; Zwirglmaier *et al.*, 2008; Sohm *et al.*, 2015; Kent *et al.*, 2016), but also deciphered the niches of smaller taxonomic units defined within each clade, emphasizing the wide ecological microdiversity of these genera.

Two questions arise from this global picture that was drawn from a purely taxonomic point of view: what are the genetic bases of the adaptation to different ecological niches, and what evolutionary mechanisms allowed this ecological diversification? To answer these questions, it is necessary to get one step further by studying adaptation at the genomic scale. Indeed, as we stated in the INTRODUCTION, both genera display a large genomic diversity both in terms of gene content and sequence divergence (Coleman *et al.*, 2006; Kettler *et al.*, 2007; Dufresne *et al.*, 2008).

In this chapter, we will address these questions through the analysis of 97 genomes of marine picocyanobacteria, of which 31 are unpublished *Synechococcus* genomes that were sequenced and assembled before I started my PhD. These 97 genomes have been integrated into the information system Cyanorak v2 (http://sb-roscoff.fr/Phyto/cyanorak), a platform in continuous development in Roscoff, in collaboration with the bioinformatics platform ABiMS. Genes predicted in the 97 genomes were grouped into clusters of orthologous genes, two genes in two different strains being orthologous if they evolved from a common ancestral gene by speciation. A Cyanorak cluster thus gathers a set of genes corresponding to the different versions of a same gene in different strains. The orthologous genes are supposed to have the same biological function, and each cluster is annotated with the putative function of the genes it contains. Some genes have been experimentally characterized, but most annotations are estimated from similarity with sequences in public databases and are thus putative. Furthermore, a vast majority of genes cannot be assigned any function because of the lack of similarity to any known gene and the absence of characterized motifs or domains, and these are flagged as “hypothetical” or “uncharacterized” genes. The Cyanorak information system is particularly useful with regard to the issue of gene annotation for two reasons. First, it facilitates the comparison of orthologous gene sequences, and allows one to edit manually both the gene sequences, their functional annotation and their clustering in orthologous groups. Second, the rich manually refined annotations made at the cluster level can be transferred to all sequences of the cluster to which it belongs, allowing to annotate 97 genomes at a time. Thanks to these functionalities and to the manual curation of several experts through years, the annotation of available genomes has been widely improved, even if most hypothetical genes remain of unknown function.

Besides allowing to refine gene annotation, the Cyanorak information system automatically generates various exports containing functional genomic information about each strain and each cluster. We used these exports to perform comparative genomic analyses in order to explore

mechanisms involved in niche-related genome diversification (part 1), but also as references to analyze meta-omic data and start unveiling the genetic bases of niche adaptation (part 2 and 3). As part of this functional (meta-) genomics approach, clusters of genes involved in the phycobilisome biosynthesis were also used to study the global distribution of *Synechococcus* pigment types (see the article (Grébert *et al., in revision*) included as ANNEXES1). In brief, by using three distinct genetic markers (*cpcBA, mpeBA and mpeW*), we estimated the relative abundance of all *Synechococcus* pigment types known to date, which allowed us to draw a first map of their distribution in the global ocean. This work notably highlighted the ubiquity and the abundance of chromatic acclimaters (pigment type 3d) in the oceans, revealing the adaptive potential of this phenotype.

Personal contribution

I was not personally involved in genome preparation, sequencing, annotation and incorporation into Cyanorak v2, but participated actively to the improvement of this database as part of the collaboration between the MaPP team and the ABiMS platform. In the work presented here, I performed most of the analyses of comparative genomics and all those based on meta-omics data. Concerning the study on the global distribution of pigment types (Grébert *et al., in revision*), I did all preliminary analyses showing the relative abundance and differential distribution of CA4-A and CA4-B chromatic acclimaters.

1. Genome evolution in the highly diverse marine picocyanobacteria: diversification is not only a matter of gene content.

Genome evolution in the highly diverse marine picocyanobacteria: diversification is not only a matter of gene content.

Hugo Doré[1,2], Gregory K. Farrant[1,2], Frédéric Partensky[1,2], Ulysse Guyet[1,2], Florian Humily[1,2], Morgane Ratin[1,2], Sophie Mazard[3,4], Martin Ostrowski[3,4], Mark Hoebeke[5], Gildas Le Corguillé[5], Erwan Corre[5], Christophe Caron[5], Karine Labady[6], Jean-Marc Aury[6], Patrick Wincker[6], David J. Scanlan[3] and Laurence Garczarek[1,2]

[1] Sorbonne Universités, UPMC Université Paris 06, CNRS, UMR 7144, Station Biologique, CS 90074, Roscoff, France

[2] CNRS, UMR 7144 Adaptation and Diversity in the Marine Environment, Oceanic Plankton Group, Marine Phototrophic Prokaryotes team, Place Georges Teissier, CS 90074, 29688 Roscoff cedex, France

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

[4] Current address: Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, NSW 2109, Australia

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

[6] Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), Institut de Génomique, Genoscope, 91057 Evry, France

IN PREPARATION

Abstract

Background:

Marine picocyanobacteria *Synechococcus* and *Prochlorococcus* are estimated to be the most abundant photosynthetic organisms on earth. They are present in most oceanic regions, a success that is believed to be linked to their tremendous genetic diversity. These two genera indeed encompass a number of phylogenetic clades, some of them showing a clear partitioning in the ocean. While previous genomic studies have identified the genetic bases of some phenotypic traits, the processes that governed the diversification of this cyanobacterial group are still poorly known.

Results:

In this work, we use a dataset of 81 non-redundant picocyanobacterial genomes, of which 31 are newly sequenced *Synechococcus* genomes, to further assess the diversity of these genera and how it relates to niche adaptation. With several genomes in most phylogenetic clades, we found a limited number of genes potentially related to the niche occupied by each clade, questioning the role of gene content variation in the diversification of picocyanobacteria. By adding a time calibration to the phylogenetic tree, we show that the rate of long-term gene fixation (0.3 and 0.16 gene per million year (My) in *Prochlorococcus* and *Synechococcus*, respectively) is much lower than the rate of fixation of amino-acid substitutions in core genes (64.5 and 15 substitutions per My, respectively).

Conclusions:

These results refine our understanding of the relative role of gene content variation and sequence substitutions in the diversification of this major group of marine phytoplankton, and have important consequences on our interpretation of the adaptive role of the flexible genome.

Keywords: Marine cyanobacteria, *Prochlorococcus*, *Synechococcus*, genome, adaptation, diversity, evolution

Background

Understanding how micro-organisms adapt to the marine environment, a dynamic system through time and space, is a major challenge, notably in a context of rapid global change. One way to better understand these processes consists in looking at the current genomic diversity and niche occupancy of these organisms. Still, such an approach requires complete genomes with representatives of distinct ecological niches, a resource which remains limited even with the advent of next-generation sequencing. Due to their abundance *in situ* and the quite good knowledge on the influence of environmental parameters on their biogeography, marine picocyanobacteria constitute very good models to tackle this question.

The two marine picocyanobacterial genera, *Synechococcus* and *Prochlorococcus,* are estimated to be the two most abundant photosynthetic organisms in the oceans (Flombaum et al., 2013). As such, they are key players in climate regulation, notably through their pivotal role in carbon uptake (Guidi et al., 2016), but also in marine trophic networks as major primary producers (Flombaum et al., 2013). These organisms are ubiquitous and often co-occur in tropical and temperate waters. While *Synechococcus* is present from the equator to the poles, *Prochlorococcus* distribution is restricted to the latitudinal band between 45°N and 40°S (Flombaum *et al.*, 2013; Johnson *et al.*, 2006; Paulsen *et al.*, 2016). This large distribution implies that these two genera are able to survive in a large range of environmental niches along *in situ* gradients of temperature, light intensity as well as micro- and macro-nutrients (Scanlan, 2012).

The ability of marine picocyanobacteria to occupy various niches is likely related to the high genetic diversity of these taxa, with distinct lineages occupying specific ecological niches and representative strains exhibiting distinct preferenda. The *Synechococcus* radiation has been split into 3 main groups, subclusters (hereafter SC) 5.1 to 5.3. While SC 5.2 are restricted to near coastal and estuarine areas, SC 5.1 and 5.3 are essentially marine, with SC 5.1 dominating in most oceanic waters and showing the highest genetic diversity with about 15 distinct clades and 40 sub-clades (Mazard *et al.*, 2012a). The *Prochlorococcus* genus forms a branch within the *Synechococcus* SC 5.1 radiation and from a phylogenetic viewpoint, it is actually equivalent to a single *Synechococcus* clade, with 7 major *Prochlorococcus* lineages, usually referred to as "clades" (Huang *et al.*, 2012; Biller *et al.*, 2014a). Lineages thriving in the upper mixed layer, so-called high light (HL) clades, are genetically distinct from those occupying the bottom of the euphotic zone, so-called low light (LL) clades. Furthermore, while members of HLI were shown to colonize temperate waters, HLII to IV are adapted to higher temperatures (Johnson *et al.*, 2006; Zinser *et al.*, 2007; Martiny *et al.*, 2009c), with HLII colonizing N-poor areas and HLIII-HLIV being restricted to iron(Fe)-limited environments (Rusch et al., 2010; Malmstrom et al., 2012). For *Synechococcus*, the distribution and environmental preferenda have only been well characterized for the 5 dominant clades in the field (clades I to IV and CRD1). Members of clades I and IV have been shown to be cold thermotypes that dominate in coastal, mixed and/or high latitude, nutrient-rich waters, while clades II and III are warm thermotypes (Zwirglmaier et al., 2007, 2008; Pittera et al., 2014; Sohm et al., 2015), thriving in oligotrophic intertropical areas , with clades II predominating in N-poor and and clade III in P-poor regions (Zwirglmaier et al., 2008; Scanlan et al., 2009; Farrant et al., 2016).

Finally, members of clade CRD1 were more recently shown to be dominant in large Fe-depleted areas of the world Ocean (Sohm et al., 2015; Farrant et al., 2016). If clades globally occupy distinct niches, it was also shown that populations within *Prochlorococcus* and *Synechococcus* clades can display distinct distribution patterns (Mazard *et al.*, 2012a; Kashtan *et al.*, 2014; Larkin *et al.*, 2016), sub-populations within clades II and CRD1 colonizing for instance different thermal niches (Farrant *et al.*, 2016).

Despite a good knowledge on both their genetic diversity and environmental preferenda, little is known about how marine picocyanobacteria diversified and how evolution has influenced their community structure. The recent rise of the genomic era, thanks to the development of high throughput sequencing, now allows one to start addressing such questions. Comparative genomics approaches applied to bacteria revealed the high variability of microbial gene content, even for closely related strains sometimes displaying identical 16S sequence and allowed to define i) the core genome, the part of the genome that contains conserved genes for a given group and that is considered to contain genes coding for essential functions, and ii) the flexible genome, the content of which being much more variable and dependent on the local environment (Lan and Reeves, 2000; Cordero and Polz, 2014). In cyanobacteria, the first studies based on multiple genomes comparison have revealed that these organisms have a so-called 'open pangenome', which means that each newly sequenced genome brings novel genes without saturation (Tettelin et al., 2005; Simm et al., 2015), and this holds true for *Prochlorococcus* and *Synechococcus*, for which 17 (Kettler et al., 2007) and 14 genomes (Dufresne et al., 2008) have been compared so far, respectively. These studies clearly highlighted that the genomic diversity of natural populations is still largely under-sampled, which strongly limits the interpretation of comparative genomic analyses. Here, we use a dataset of 81 non-redundant genomes of mostly marine picocyanobacteria, of which 31 are newly sequenced *Synechococcus* genomes, to further assess the genomic diversity within these genera and how it relates to their current niche occupancy. This unprecedented genome dataset allowed us to show that although both gene gain/loss and genetic divergence co-occur in the adaptation of marine picocyanobacterial clades to environmental niches, variation in nucleotide sequence seems to better explain their diversification on the long term.

Material and Methods

Genome sequencing and assembly

32 *Synechococcus* were chosen for sequencing based on their phylogenetic position and pigmentation (Fig. 1, Table S1). Strains were retrieved from Roscoff Culture Collection (RCC) and cloned by three successive transfers on agarose (0.6%) to reduce contamination by heterotrophic bacteria, since none were axenic. Shotgun sequencing reads with various lengths were generated for 25 novel *Synechococcus* genomes at the Genoscope (CEA, Paris-Saclay, France) based on two libraries: a short-insert forward-reverse pair-end (PE) library and a long-insert reverse-forward mate-pair library, both sequenced by Illumina™ technology. Additionally, shotgun sequencing reads (250 bp) were generated for 7 other genomes at the Center for Genomic Research (University of Liverpool, UK). Single or PE reads were assembled into contigs using the

CLC AssemblyCell© 4.10 (CLC Bio, Aarhus, Denmark). In order to select only contigs corresponding to *Synechococcus* (and not to co-occurring heterotrophic bacterial contaminants), PE reads were re-mapped onto contigs using CLC mapper and average coverage values for each contig were obtained using CLC assembly info. Based on the average coverage of *Synechococcus* contigs, threshold values were applied to select a subset of contigs. Contigs were then scaffolded and most genomes closed using WiseScaffolder (Farrant et al., 2015). Base numbering of the circularized genomes was set at 174 bp before dnaN start, corresponding approximately to the origin of replication. Automatic functional annotation of the genomes was then realized using the IGS Manatee pipeline (University of Maryland School of Medicine, Baltimore, USA) (Galens et al., 2011).

Clustering of orthologous genes

New protein sequences were assigned to previously defined clusters of orthologs (CLOGS) available in Cyanorak information system (Dufresne et al., 2008) using HMMER and an e-value threshold of 10^{-5} (Eddy, 2011). Genes that could not be assigned this way were then clustered together using all-against-all BLASTp+ or BLASTn+ (for protein-coding and tRNAs/rRNAs, respectively) and the OrthoMCL algorithm using an alignment e-value threshold of 10^{-5} and a 50% sequence length aligned cut-off (Li et al., 2003). The custom-designed information system CYANORAK v2 (http://www.sb-roscoff.fr/cyanorak/) was used for the manual curation of orthologous gene families.

This CLOGs clustering allowed to build a phyletic pattern (i.e. the number of copies of each CLOG in each genome), which was used to extract lists of genes shared at different taxonomic levels. Core genomes were defined at the genus, sub-cluster and clade levels when more than 3 genomes were available. When a given taxon contained more than 5 strains, we considered a gene as core if it was present in at least 90% of the strains of this taxon, in order to take into account potential mis-annotations or issues in the clustering of orthologs. As only few genes were found to be strictly specific to previously defined clades, we also used relaxed rules to find genes shared between a set of strain: in Dataset 1, genes are found in at least 80% of the strains within a clade, and no other strain of the genus. In Dataset 2, sets of clades of *Synechococcus* were chosen according to co-occurrence patterns described in (Farrant et al., 2016), and genes present in more than 90% of strains within a set and less than 10% outside the set were extracted. When this yielded less than 10 CLOGs, these cut-offs were lowered to more than 80% of strains within a set and less than 20% outside (as indicated in the column "cut-off used", Dataset 2). These relaxed rules allow to overcome possible issues resulting from the clustering of orthologous genes and to take into account clades for which the ecological niche is poorly known.

The phyletic pattern was also used to estimate the size of the pangenome and core genome of each genus. Sampling of strain combinations to draw pangenome curves was performed using the software PanGP (Zhao et al., 2014) with parameters 'Totally Random', SR=100 and SS=1000 and curves were drawn with R 3.3.1. The results of PanGP exponential fits were used as estimates of the asymptotic number of core genes.

ANI/AAI calculation

Whole-genome Average Nucleotide Identity (ANI) and percentage of conserved DNA between pairs of strains were calculated following the method described in (Goris et al., 2007), and ANI values were condisered only if the percentage of conserved DNA was over 10%, a value below which ANI values cannot be trusted (Goris et al., 2007). AAI was calculated following the same method as (Konstantinidis and Tiedje, 2005b).

Phylogeny

The phylogeny shown on Fig. 1 is based on a set of 230 public sequences of the gene *petB*, which are representative of the *Synechococcus* diversity found in the oceans (Mazard *et al.*, 2012a). The tree was built with PhyML (Guindon and Gascuel, 2003) with the HKY model and estimation of gamma parameters and proportion of invariant sites from the dataset and 100 bootstraps. Additionally, a set of 821 core proteins with no paralog was concatenated, resulting in a total of 226778 amino acids. A phylogenetic tree was built with PhyML with the WAG model and estimation of parameters of the gamma distribution and of the proportion of invariant sites. For this second tree, bootstrap values were computed by randomly sampling 1% of sites before building the tree, and repeating this operation 100 times. All alignments were performed with MAFFT v7.164b (Katoh and Standley, 2014).

The phylogeny based on gene content was built as follows: a Jaccard distance matrix was computed from the phyletic pattern, and then used by a Neighbor-Joining algorithm to generate the tree with 100 bootstraps. All calculations were made using R packages *vegan* (Oksanen et al., 2015) and *ape* (Paradis et al., 2004).

Tree comparisons

The phylogenetic tree based on the alignement of the protein sequences of 821 core genes was compared to the tree based on phyletic pattern using R package *dendextend* v.1.3.0 (Galili, 2015). We compared branch lengths using custom python scripts based on python ete toolkit (ete2, (Huerta-Cepas et al., 2010)). Briefly, for each node at the base of a clade (highlighted in Fig. 5), the average distance from the node to the descending leaves (‘external’ length) and the distance to the parent node (‘internal’ length) were calculated. Boxplots of the ratio of clade external to internal branch lengths were drawn in R for both trees, and the difference between the mean ratios was assessed by a paired Mann-Whitney-Wilcoxon test.

Estimation of gene gain and loss and identification of genomic islands

Estimation of the number of gene gains and losses were estimated from the phyletic pattern using the software Count (Csurös, 2010) that implements a Maximum Likelihood method for estimation of the ancestral states (presence, absence or multiple copies) of every CLOG in the dataset, using a phylogenetic tree based on the concatenated alignement of 821 protein sequences of core genes (see Phylogeny) as reference, and allowing four categories for the gamma distribution of duplications and branch lengths. We used a cut-off on posterior probability of 95%, which allowed to obtain a number of CLOGs at the root of the tree similar to the average number of CLOGs in current *Synechococcus* strains. Using this cut-off, we extracted the state of presence-absence of

each family at every nodes of the tree, and used it to calculate the number of gene gains and losses on every branch. These estimations were used to identify genomic islands, defined as regions of the genome with a high turnover of genes (Kettler et al., 2007). The number of genes gained over the evolutionary history of each strain was calculated on a sliding window of size 10kb with a step of 100 bp, and regions with more than 5 genes gained per 10 kb were considered as genomic islands.

Time calibration of the tree

The phylogeny based on the sequence of 821 core proteins was used as input for *reltime* algorithm (Tamura et al., 2012) and the JTT matrix-based model (Jones et al., 1992), as implemented in MEGA7 (Kumar et al., 2016), with default parameters and SC 5.3 designated as outgroup. Two calibration points were used, based on (Sánchez-Baracaldo, 2015) and www.timetree.org: the first calibration was set on node n2 (see Fig. S1), i.e. the common ancestor of strains WH5701 and WH8102 between 582 and 878 million years (My), and the second on node n4 (i.e. the common ancestor of strains CC9311 and WH8102) was set between 252 and 486 My. This method allowed us to relate these gain/loss events with the time elapsed on each branch of the tree, taking into account the higher evolution rate of protein coding genes in *Prochlorococcus* than in *Synechococcus* (Dufresne et al., 2005). We also calculated the number of substitutions for each branch of the tree by multiplying branch length by the total number of residues in the alignment, and divided it by the time elapsed on the branch to obtain a substitution rate per My.

Estimation of the number of fixed genes and fixed substitutions specific to a taxon

At a given node of the phylogenetic tree based on 821 core proteins, genes that were found in all descending leaves and no other strain in the dataset were considered as fixed genes specific to this node. Similarly, considering the concatenated alignment of 821 core proteins, every positions that showed the same amino-acid in all leaves below a node and another amino-acid in every other strains were considered as fixed substitutions specific to this node. Terminal branches were not taken into account in these calculations since by definition strain-specific amino acids or genes cannot be considered as fixed.

Results

Picocyanobacteria exhibit a very wide genomic diversity even at intra-clade level

To compensate for the under-coverage of *Synechococcus* in the available marine picocyanobacterial genomes, 31 new strains were sequenced from cultured isolates, resulting in more than doubling the number of genomes for this genus. Strains were selected in order to cover at best the phylogenetic and pigment diversity described in *Synechococcus,* as well as their geographic origin and trophic regime of their isolation site (Fig. 1, Table S1), though no culture isolates are available yet for EnvA and EnvB clades (Mazard *et al.*, 2012a; Farrant *et al.*, 2016). The use of Wisescaffolder (Farrant et al., 2015) allowed us to close 28 out of the 31 genomes, with only 1 to 3 gaps remaining in the others, corresponding to high repeat regions within single genes. Thus, this high-quality genome dataset constitutes a key asset for comparative genomics analyses.

Consistent with the genome streamlining that occurred in most *Prochlorococcus* lineages (Dufresne et al., 2005, 2008; Kettler et al., 2007), genome size and GC% are expectedly lower in *Prochlorococcus* (average genome size 1.815 Mb, average GC% 34.77) than in *Synechococcus* and *Cyanobium* (average genome size 2.537 Mb, average GC% 59.29), with genome sizes ranging from 1,625 Mb for *Prochlorococcus* strain GP2 to 3,342 Mb for *Cyanobium gracile* PCC 6307 and GC% from 30,8% (EQPAC1, MED4 and MIT9515) to 68,7% (PCC 7001 and PCC 6307). While the large genome size of the sole freshwater strains of the dataset (PCC 6307) is expected, the second largest one is a CRD1 strain isolated from the border of the south Pacific gyre (BIOS-E4-1, 3.314 Mb), a very oligotrophic, iron depleted area.

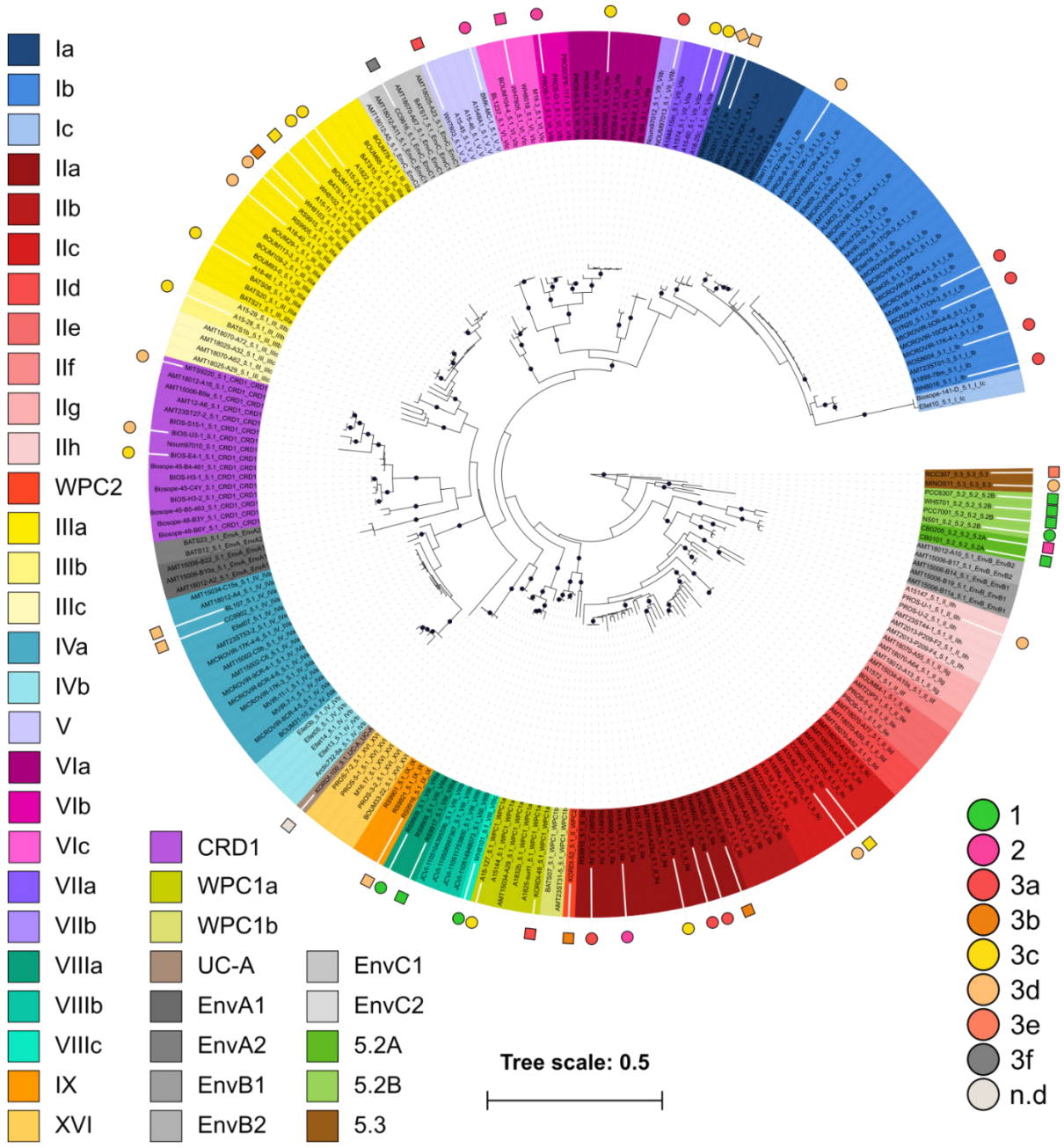


Figure 1 Phylogenetic tree of marine picocyanobacteria showing the genomes analyzed in this study. Maximum-Likelihood tree of 231 *petB* sequences from marine *Synechococcus*. Diamonds indicate bootstrap support over 70%. Colored circles indicate newly sequenced genomes, while squares indicate previously available ones. The diamond used for strain WH8020 indicates that this genome was not used in our study due to its poor quality. Symbols are colored according to their pigment type as defined in (Six *et al.*, 2007c).

Although they all belong to a monophyletic, long diverged branch within Cyanobacteria (Shih et al., 2013; Sánchez-Baracaldo, 2015), picocyanobacterial genomes show a tremendous diversity in terms of both nucleotide sequences and gene content. Average nucleotide identity (ANI) between pairs of genomes, for pairs for which more than 10% of the genome could be aligned, indeed ranged from 83.8 to 99.9% and within-clade ANI was in average 91.1% (Fig. 2A), showing

that each clade and even most sub-clades are at least as diverse genetically as microbial species, as defined by the 95% ANI criterion (Goris et al., 2007). In order to manually refine the annotation of these genomes and ease comparative genomic analyses in terms of gene content, all genomes were included in the Cyanorak v2 information system (www.sb-roscoff.fr/cyanorak), in which predicted genes were grouped into Clusters of Orthologs (CLOGs) by sequence similarity. This allowed us to determine the core genome (CLOGs present in all strains) and the pangenome (all CLOGs present in at least one strain) at various phylogenetic depths (Tettelin et al., 2005). When considering the whole dataset, the number of core CLOGs as a function of the number of genomes showed an asymptotic decline, tending toward a core set of 911 genes (Fig. 2B). In contrast, the pangenome of marine picocyanobacteria, containing 27376 CLOGs, is still far from saturation, revealing that even with 81 genomes, every newly sequenced genome still brings a significant number of new genes. This holds true when considering *Prochlorococcus* (7537 CLOGs) and *Synechococcus* (20986 CLOGs) independently, indicating that we still miss an important part of the diversity within both genera that is yet to be sequenced from the field.

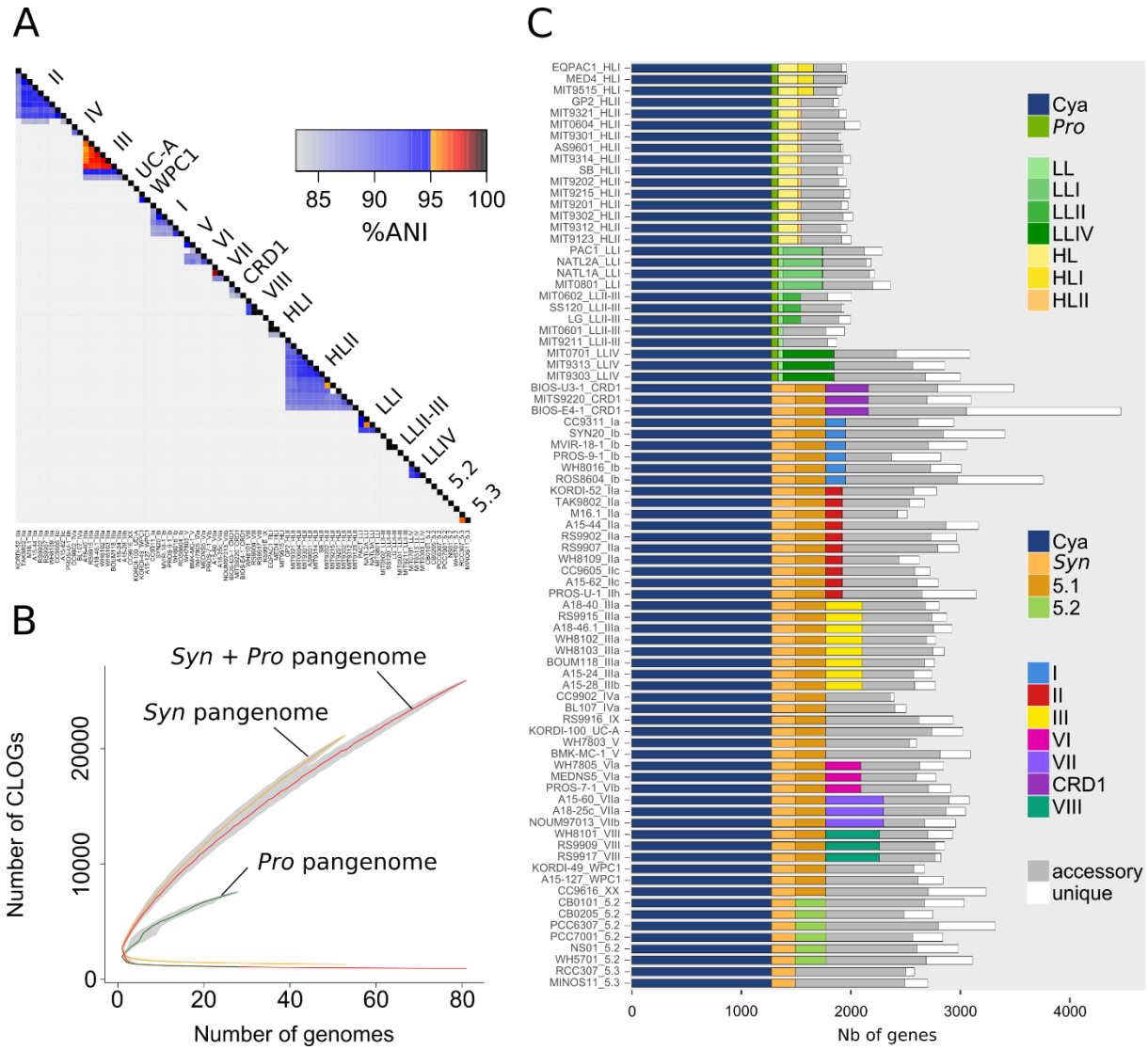


Figure 2 Genomic diversity of marine picocyanobacteria. A. Heatmap of Average Nucleotide Identity (ANI) between pairs of strains. Each column corresponds to a strain, and strains are ordered according to their phylogenetic relatedness. B. Pan- and core genomes for an increasing genome number of picocyanobacteria (red, 81 genomes), *Synechococcus* (orange, 53 genomes) and *Prochlorococcus* (green, 28 genomes). The grey zone around each curve represents the first and third quartiles around the median of 1000 samplings by randomly modifying the order of integration of genomes. C. Distribution of Clusters of Orthologous Genes (CLOGs) in picocyanobacterial genomes. A CLOG is considered as core of a taxonomical group if it is present in >= 90% of the strains within this group. Sets of core CLOGS are inferred only for taxonomical groups with more than 3 strains. In A and C strains are as labeled as strain_subclade.

Role of gene content in the adaptation to specific niches

A major asset brought by the 31 newly sequenced *Synechococcus* genomes is the availability of several genomes per clade, which allowed us to estimate the relative sizes of the core set of CLOGs at different taxonomic levels (genus, SC and clades), the accessory genome (i.e. CLOGs shared by at least two strains but not core) and unique genes (CLOGs present in a single strain; Fig. 2C). While the proportion of accessory genes was pretty constant between genomes, constituting in average 18.5 % (standard deviation 3%) and 24.2 % (6%) of the *Prochlorococcus* and

Synechococcus genomes respectively, unique genes constitute the most variable part of the genomes and clearly determine genome sizes (Fig. 2C). The newly sequenced strain BIOS-E4-1 contains by far the largest gene number of the genome dataset (4452 genes), with a large proportion (31.7 %) being unique. Noteworthy, a significant part of CLOGs is shared within a given clade (for instance 335 genes for *Synechococcus* clade III, or 143 genes for *Prochlorococcus* clade HLI) and could be involved in the adaptation of these taxa to specific environmental conditions. However, it must be noted that most of these CLOGs are also present in other picocyanobacterial taxa and only few CLOGs (32 and 12 for clades III and HLI, respectively) are quasi-specific to each clade within a genus (i.e. present in at least 80% of the strains of a clade, and absent from other strains of the genus, Dataset 1).

In order to unveil whether the presence or absence of genes might be related to the adaptation to specific niches, we defined sets of clades co-occurring in the field and occupying similar niches, based on the assemblages of Ecologically Significant Taxonomic Units (ESTUs) defined in (Farrant et al., 2016). We then searched for genes occurring in at least 90% (or 80%) of strains within a set and in less than 10% (20%) of all other picocyanobacterial strains (Dataset 2, see Material and Methods for the choice between the two cut-offs). The latter analyses were focused on *Synechococcus* since *Prochlorococcus* genomic diversity has already been extensively studied (Kettler *et al.*, 2007; Partensky and Garczarek, 2010; Biller *et al.*, 2014a). Using this relaxed, niche-related definition of specificity, 18 genes were found to be specific of members of cold thermotypes, clades I and IV, among which only 6 had a putative function, though with seemingly no direct relationship with adaptation to low temperature. However, the set of 21 CLOGs specific to clade I includes a particular isoform of DnaK (DnaK4, CK_00056929, Dataset 1), in addition to the 3 gene copies present in most *Synechococcus* SC 5.1 strains, which might be involved in protein folding in cold conditions (Genevaux et al., 2007).

Members of clades III, WPC1 and SC 5.3, co-occurring in warm, P-depleted oligotrophic waters, were found to share a much higher number of genes (85), among which 2 were previously reported to be related to phosphate availability: a yet uncharacterized gene (CK_00002088) found to be downregulated in early phosphate stress (Tetu et al., 2009) and a chromate transporter (ChrA), which was recently suggested to be involved in *Prochlorococcus* phosphate acquisition, based on its enrichment in P-poor oligotrophic areas (Kent et al., 2016). Clades III and WPC1 also share a cluster of 12 consecutive genes potentially involved in capsular polysaccharide synthesis and export (including genes similar to *kps* genes in *E. coli* K1*,* responsible of the formation of a polysialic acid extracellular capsule; CK_00043447, CK_00002152, CK_00002153, CK_00057342, CK_00002158, CK_00036768, CK_00002156, CK_00002160, CK_00009077, CK_00002161, CK_00002162, CK_00002307, Dataset 2) and another cluster of 7 genes that might be involved in the use of organic N sources since it encompasses a putative nitrilase (CK_00002256). Additionally, 32 genes were found to be specific to the 8 clade III strains, including the cyanate transporter genes (*cynABD; (Kamennaya and Post, 2011)*) as well as the phosphate starvation-induced protein (PsiP1) and a specific alkaline phosphatase (CK_00052500) that allows to use extracellular organic phosphate. Similarly, the two closely related members of SC 5.3 also share a large number of strictly specific genes (215), including a regulator of phosphate

uptake (PhoU) as well as two putative phosphatases (CK_00005504, CK_00005619) and a putative pyrophosphatase (CK_00005811), in addition to the 4 phosphatases present in most picocyanobacterial genomes. Thus, occurrence of these genes might contribute to the success of clade III, WPC1 and SC 5.3 in very oligotrophic, P depleted environments such as the Mediterranean Sea in summer (Farrant et al., 2016), and suggests that the three clades have adopted partially different strategies to cope with P depletion.

Genes potentially involved in niche adaptation were also found in all 3 strains of the CRD1 clade, known to dominate in oceanic regions depleted in iron, which share a quite high number of specific CLOGs (81, Dataset 2), but most of them have no known function. Among the characterized ones, there is a second copy of flavodoxin (IsiB), a Cu-containing protein known to replace ferredoxin in iron depleted conditions (Erdner and Anderson, 1999), the ferrous iron transport protein FeoA, an iron-sulfur cluster biosynthesis family protein (CK_00008433) as well as 3 specific Hlips that might provide protection from oxidative stress to photosystems (He et al., 2001).

Finally clades VIII and SC 5.2, both known to thrive in brackish waters, share 28 genes including a few strictly specific genes such as a fatty acid hydroxylase (CK_00002851), involved in lipid biosynthesis, and one or two copies of a P-type ATPase (CK_00045881), a protein family known to transport a variety of different ions and phospholipids across membranes (Axelsen and Palmgren, 1998). It is also worth noting that SC 5.2 and clade VIII share a fair number of genes potentially involved in the adaptation to low salinity with members of clades V, VI and sometimes VII, whose ecological niches are still poorly known (Zwirglmaier et al., 2008; Farrant et al., 2016) and possibly encompass environments with variable salinity. This includes a small-conductance mechanosensitive ion channel (MscS family) that might be involved in the response to osmotic stress (CK_00056919; (Haswell et al., 2011)) and a bacterial regulatory protein of the ArsR family that besides regulating the efflux of arsenic and arsenite was suggested to participate to salt tolerance in *Staphylococcus aureus* through a Na^+ efflux activity (Scybert et al., 2003)]. In addition, members of clade VIII share 29 specific genes, including a *mscS* gene copy (CK_00056915), while members of SC 5.2 share 33 specific genes, including another *mscS* gene copy (CK_00003081), a putative chloride channel (CK_00042275) and a NAD-dependent malic enzyme, a protein that is known to be enhanced under salt stress in plants (Liu et al., 2007).

Despite these few examples, it seems that the amount of genes potentially related to the broad niche occupied by each clade or assemblage of clades is quite limited and depends on the characteristics of the niche taken into consideration. Most of the diversity in gene content is therefore not clade-dependent, but rather relies on differences between individual strains.

Dynamics of the evolution of gene content in marine picocyanobacteria

In order to better understand the evolutionary processes that led to the current distribution of gene content in marine picocyanobacterial genomes, we estimated by Maximum Likelihood the number of gene gain and loss events on each branch of a reference phylogenetic tree built from a

concatenation of 821 core proteins (Fig 3). The results obtained were consistent with the scenario of a major genome streamlining during *Prochlorococcus* evolution (Dufresne et al., 2005; Kettler et al., 2007), an excess of gene loss being observed at the base of this radiation (Fig. 3). Globally, the number of genes gained and lost on each branch of the picocyanobacterial tree was quite variable. While on internal branches the number of gains and losses remained limited and balanced for both *Prochlorococcus* and *Synechococcus* SC 5.1 (gains ≤ 359, losses ≤ 402), there was generally a higher number of events on terminal branches as well as an excess of gain as compared to loss events, with up to 1,678 genes gained on the branch leading to *Synechococcus* BIOS-E4-1 (SC 5.1) and 446 on the one leading to *Prochlorococcus* MIT9303, for 109 and 56 lost genes, respectively.

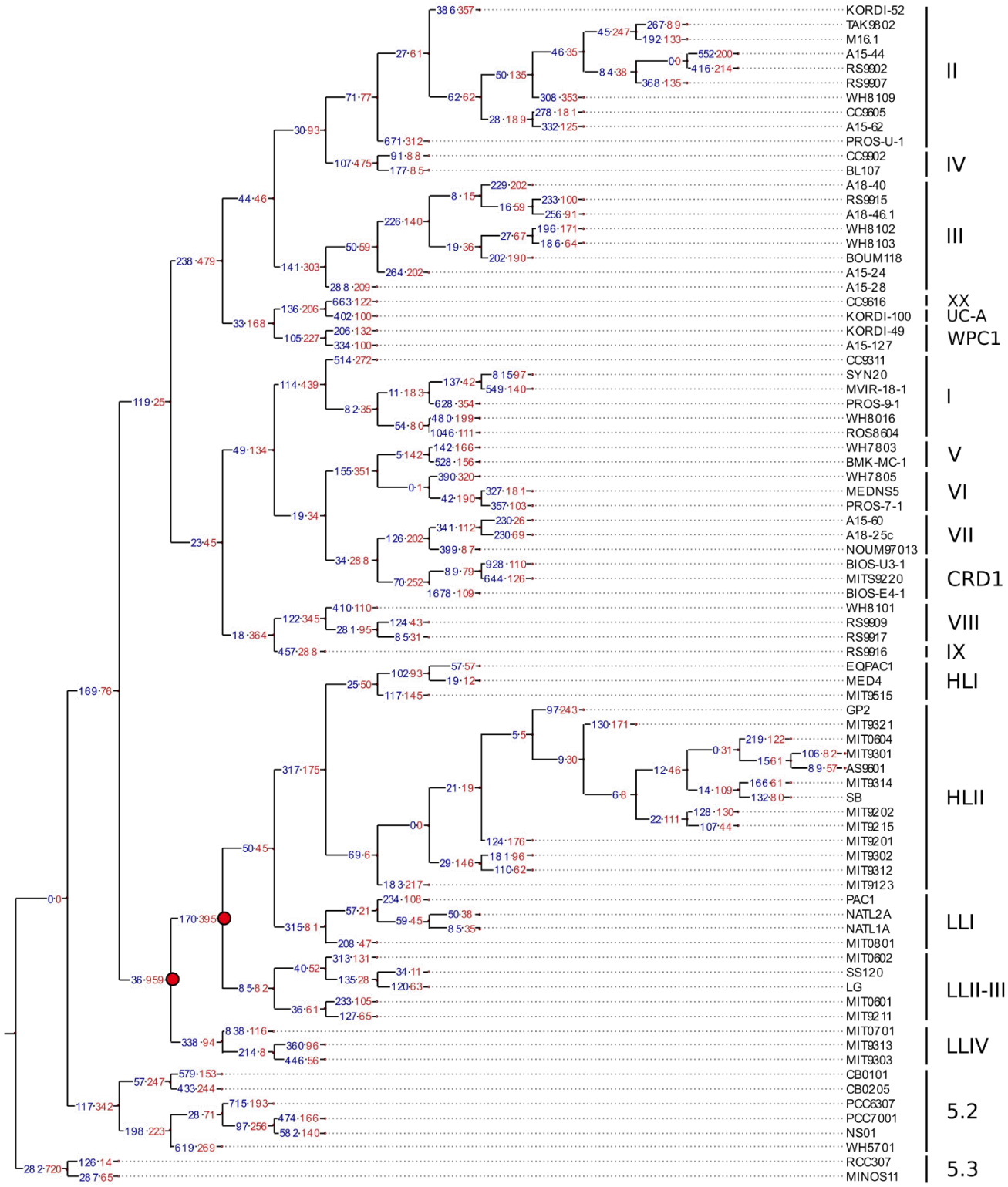


Figure 3 Gene gain and loss along the phylogenetic tree of marine picocyanobacteria. The ancestral state of presence/absence of every Cluster of Orthologous Genes (CLOG) was inferred using Count (Csurös, 2010) and used to infer the number of gains and losses of gene families on each branch of the tree. Gain and loss numbers are labelled in blue and red, respectively. Nodes highligted in red correspond to *Prochlorococcus* genome streamlining.

By using calibration time points from a previous study (Sánchez-Baracaldo, 2015), we estimated that per million year (My), this corresponds to about 0.68 and 3.95 genes gained (1.33 and 1.03 genes lost) on internal and terminal branches for *Synechococcus* SC 5.1, respectively, while internal and terminal branches of *Prochlorococcus* HL gained 1.45 and 4.69 genes (0.81 and 3.94 lost; Table 1). The higher values observed for the terminal branches are related to the high number of strain-specific genes and shows that most of the variability in gene content occurs at the 'leaves' of the tree. If we suppose the rate of gene gain to be constant over time, this suggests that most of the genes gained on internal branches have been secondarily lost and are not represented in our genomic dataset.

Table 1 Estimation of rates of substitution, gene gain, gene loss, gene fixation and substitution fixation per million year. Rates are estimated by linear regression. *p-value*: F-test p-value. "internal": rates were estimated using internal branches only. "leaves": rates were estimated on terminal branches only (i.e. branches leading to leaves).

		HL internal	HL leaves	5.1 internal	5.1 leaves
Estimated substitutions number	value	239.71	256.44	71.61	68.80
	standard error	0.95	4.56	1.07	1.50
	adj R^2	1.00	1.00	0.99	0.98
	p-value	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$
Specific positions fixation	value	64.52	-	15.00	-
	standard error	4.46	8.20	0.69	0.88
	adj R^2	0.93	0.87	0.91	0.95
	p-value	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$
Estimated Gain number	value	1.452	4.690	0.681	3.950
	standard error	0.086	0.552	0.134	0.607
	adj R^2	0.950	0.826	0.362	0.463
	p-value	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$
Estimated Loss number	value	0.811	3.937	1.332	1.030
	standard error	0.226	0.403	0.123	0.148
	adj R^2	0.441	0.863	0.725	0.496
	p-value	0.003	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$
Specific CLOGs fixation	value	0.322	-	0.156	-
	standard error	0.034	0.269	0.063	0.372
	adj R^2	0.853	0.759	0.103	0.385
	p-value	$< 10^{-5}$	$< 10^{-5}$	0.018	$< 10^{-5}$

As genomic islands have been shown to play a key role as repositories of laterally transferred genes in marine picocyanobacteria (Coleman et al., 2006; Kettler et al., 2007; Dufresne et al., 2008), we explored the similarity of gene contents of genomic islands between strains, islands being defined as regions of the genome enriched in gained genes (Kettler et al., 2007). The pattern obtained showed that more closely related strains share more island genes, while a low proportion of island genes are also shared with more distant strains. This could be either interpreted as a higher rate of exchange within closely-related groups, or as a vertical inheritance of genes found in islands, but shows the relatively low exchange of genes between distantly related clades. This is particularly visible for *Prochlorococcus* HL streamlined genomes that share only a low proportion of island genes with *Synechococcus*. A notable exception is the *Synechococcus* clade VIII, which shares more island genes with strains of SC 5.2 than with most SC 5.1 strains, an expected pattern since, as mentioned above, these groups co-occur in waters of variable salinity and share a number of genes related to their adaptation to this specific niche.

Relative contribution of variability at sequence and gene content levels in the evolution of picocyanobacteria

The relatively low rate of gene acquisition and the low number of genes specific of phylogenetic clades (or niches) raises the question of the relative weights of gene content variation *vs.* substitutions in nucleotide sequence in the long term diversification of these organisms. Fig. 5 compares a phylogenetic tree based on a concatenation of 821 picocyanobacterial core genes to a dendrogram based on the phyletic pattern (i.e. the pattern of presence/absence of each CLOG in each strain). Topologies of the two trees were globally similar, which reveals that the gene content reflects quite well the evolutionary history of the group, i.e. that fixation of genes and fixation of mutations occurred concomitantly in the evolutionary history of this group. Yet, a notable difference between these trees concerns the position of *Synechococcus* SC 5.2 and 5.3 that both fall within the SC 5.1 radiation, thus separating *Synechococcus* and *Prochlorococcus* in two monophyletic genera. Indeed, as previously reported in a study using 11 *Synechococcus* genomes (Dufresne et al., 2008), SC 5.1 clade VIII and SC 5.2 are share a number of genes, in relation with the similarity of the brackish environmental niche occupied by these two groups (see above). Interestingly, clades V and VI, whose niche is poorly known, also cluster with these two groups, indicating that they may occupy a niche of low salinity.

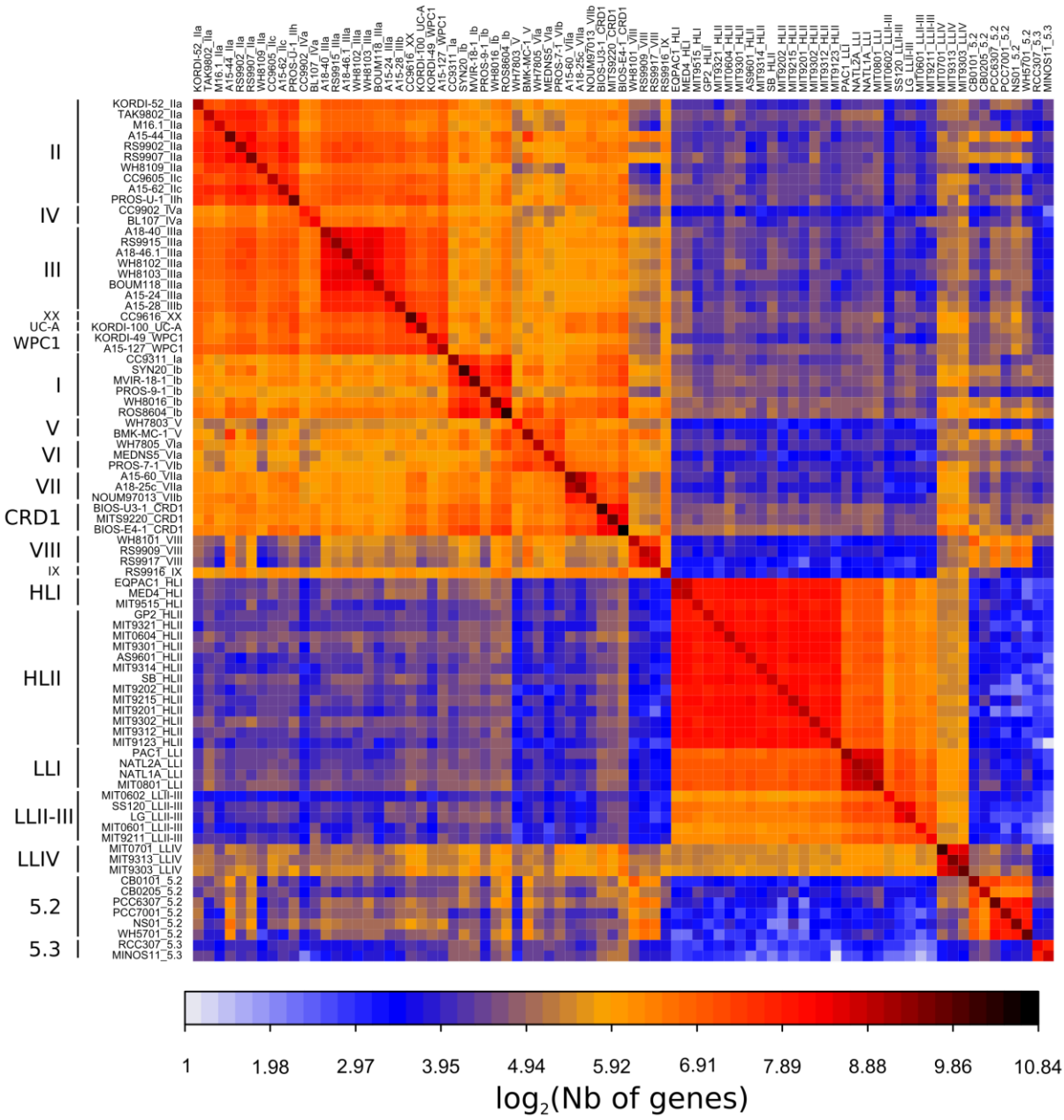


Figure 4 Number of gained genes included in genomic islands for all 81 sequenced picocyanobacterial strains. The color scale indicates the total number of gained genes (log_2) predicted to be in a genomic island for both genomes in each pair of strains. The diagonal color is thus representative of the number of such genes in each genome.

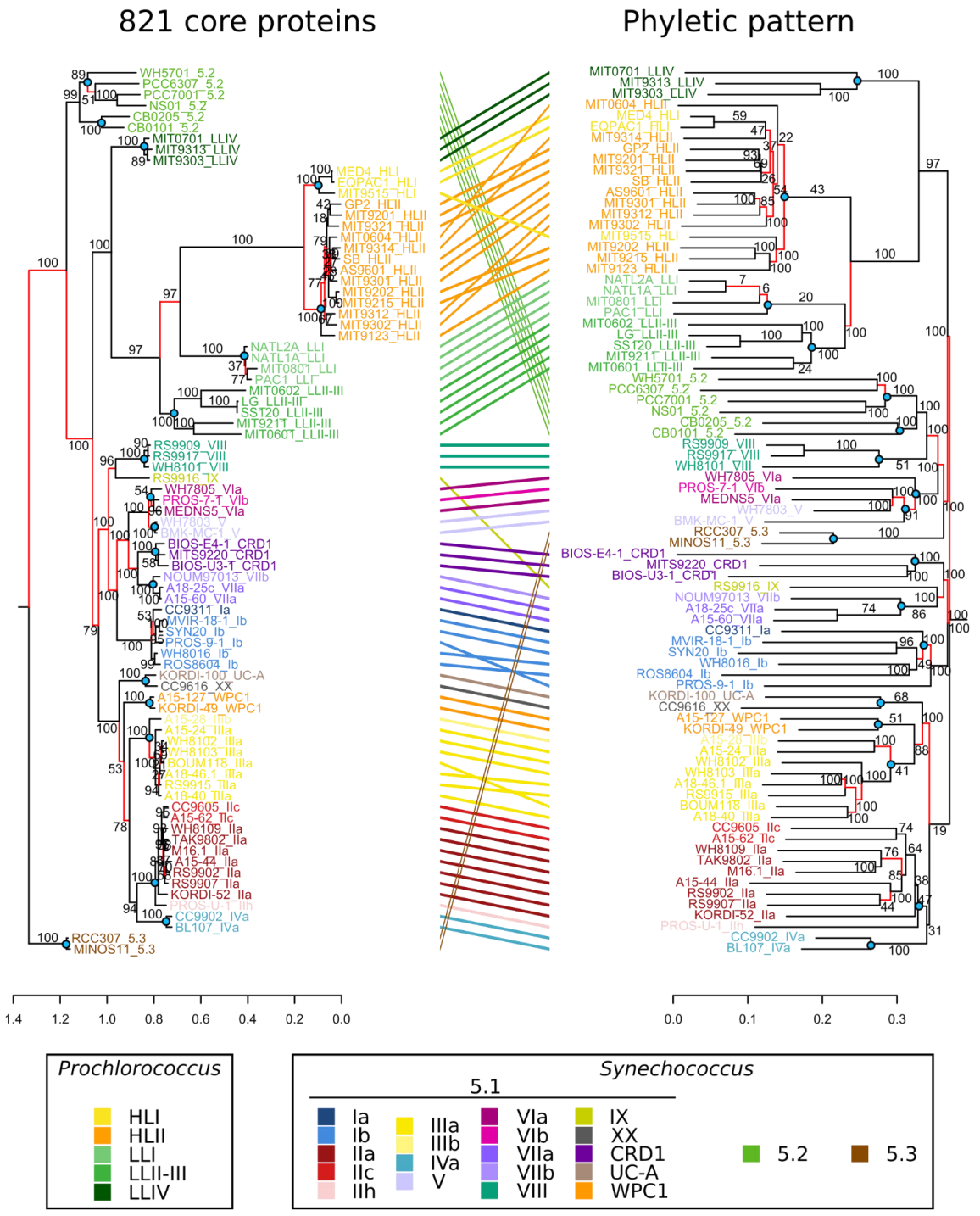


Figure 5 Comparison of a phylogenetic tree based on the alignement of core proteins and a tree based on phyletic pattern. Left tree is a Maximum Likelihood tree based on the alignement of 821 concatenated core proteins. Right tree is a Neighbor-Joining tree based on the Jaccard distance between the phyletic patterns of 27376 gene families found in the 81 picocyanobacterial genomes. Labels are colored according to the strain subclade. Red branches indicate discrepancies between the topology of the two trees. Nodes highlighted in blue were used for branch length comparisons in Fig. 6.

Among the *Synechococcus* SC 5. 1 and *Prochlorococcus* radiations, besides a few incongruences within *Synechococcus* clades I and III and *Prochlorococcus* HLII that are likely due to the relatively low number of genes shared by all strains within these clades, it is worth noting that some clades have lost their monophyly in the tree based on phyletic pattern such as the group of clades WPC1-EnvC-UC-A in *Synechococcus*, with WPC1 being found closer to clade III than in the tree based on core genes, and *Prochlorococcus* clade HLI being partially mixed with clade HLII. This last example is particularly interesting, since despite their clearly distinct phylogenetic clustering based on protein sequences and proved ecological and physiological differences (Johnson *et al.*, 2006; Martiny *et al.*, 2009c), these two clades have a quite similar gene content, with only few genes present in all HLII strains but not in all HL strains (29, Fig. 2C). Similarly, *Prochlorococcus* clade LLII-III clusters together with LLI on the base of phyletic pattern, which confirms previous literature showing that gene content at least partially explains the differentiation between HL clades (colonizing surface waters) and LL clades (thriving deeper in the water column; (Kettler et al., 2007; Scanlan et al., 2009; Partensky and Garczarek, 2010)).

Another major difference between these trees concerns branch lengths. By computing for each node at the base of a clade (colored in Fig. 5) the average length from the node to its descending leaves (terminal length), and the length from the node to its parent node (internal length), we showed that the ratio of terminal to internal branch lengths was significantly higher (Mann-Whitney paired test, p-value < 0.001) in the phyletic pattern tree than in the core tree (Fig. 6). This suggests that there is more sequence evolution between clades than between strains within each clade, whereas there is more gene content variation between strains than between clades. In other words, this comparison reveals that most of the changes that are fixed on the long term by evolution are substitutions and not changes in gene content. In order to quantify more precisely this difference, we compared the number of amino acid changes in core proteins occurring per million years (My), estimated from branch lengths in the core phylogenetic tree to the estimated number of gained genes per My (Table 1). It is important to note that the rates of gene gain/loss and gene substitutions calculated this way should only be considered as lower bound estimates for several reasons. Indeed, since we only have access to the present-day genomes and not to ancestral ones, our measure of gain rate rather corresponds to the rate of genes gained and successfully retained over time in at least one strain. Second, the substitution rate was measured on core genes, which likely undergo a strong purifying selection but some sites could still be saturated by substitutions. Thus, both rates were probably underestimated. This, together with the longer generation time of picocyanobacteria compared to other model bacteria (e. g *Escherichia coli* or *Bacillus subtilis*), and their very large population size (Partensky *et al.*, 1999a; Dufresne *et al.*, 2005; Flombaum *et al.*, 2013), could explain why estimated rates were lower than for other lineages (e.g. (Lawrence and Ochman, 1998; McDonald and Currie, 2017)).

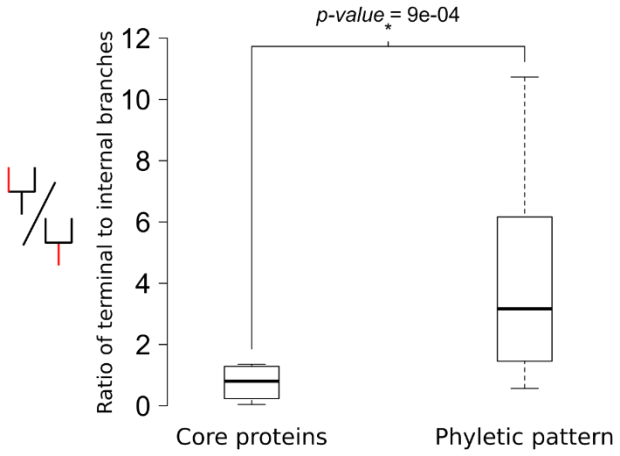


Figure 6 Comparison of within and between clades evolution. For each node highlighted in Fig. 5, the average distance from the node to the descending leaves and the distance to the parent node were calculated. The boxplots show the ratio of clade external to internal branch lengths calculated from the tree based on core proteins "Core proteins" and from the tree based on phyletic pattern ("phyletic pattern"). Difference between the mean ratios was assessed by a paired Mann-Whitney-Wilcoxon test ($p \leq 0.0009$).

With this caveat in mind, in *Prochlorococcus* HL, 165 times more amino acid substitutions than gene gains were estimated for internal branches, and 54.5 times for terminal branches, essentially due to a higher rate of gene gain in terminal branches. In *Synechococcus* SC 5.1, a ratio of 105 was obtained for internal branches and 17 for terminal branches, the difference between the two genera being due to the higher rate of protein sequence evolution observed in *Prochlorococcus* (Dufresne et al., 2005).

We also compared at each node the number of fixed amino acids in core proteins (i.e. amino acids in the alignment that are identical in all the descending strains and different in all other strains) to the number of fixed genes (i.e present in all the descending strains and in no other strain). 200 times more amino acids than genes are fixed per million year in *Prochlorococcus* HL (96-fold for *Synechococcus* SC 5.1). This corresponds to a fixation rate of 64.5 and 15 amino acid changes in core genes per My for *Prochlorococcus* HL and *Synechococcus* SC 5.1, respectively, while a gene is fixed once every 3 My for *Prochlorococcus* HL and once every 6.4 My for *Synechococcus* SC 5.1.

Discussion

The availability of 81 complete and closed picocyanobacterial genomes, including 31 novel *Synechococcus*, with manually refined annotations, constitutes a key asset for comparative genomics analyses. With regard to previous studies (see e.g. (Kettler et al., 2007; Dufresne et al., 2008; Scanlan et al., 2009)), sequencing of several strains for most major *Synechococcus* clades revealed that the within-clade diversity remains tremendous since ANI and AAI were well below the cut-off of 94 % (Fig. 2A and Fig. S2), usually considered to be the limit between bacterial species (Konstantinidis and Tiedje, 2005b, 2005a). Thus, most clades would rather correspond to

one or several species, and several genera might be defined within *Synechococcus* and *Prochlorococcus* depending on the cut-off used (Fig. S2). It is important to keep in mind this level of divergence if one wants to compare results from the present study to those obtained in highly sequenced bacteria such as pathogens and commensals (e.g. (Harris et al., 2010; Kennemann et al., 2011; Mather et al., 2013)). While high divergence and associated low level of synteny somehow limit the application of classical population genetics approaches, such as calculation of recombination rates (McDonald and Currie, 2017), this dataset is in contrast well suited to study the long term evolutionary processes that have shaped the genomes of these abundant and ubiquitous organisms in relation with their ecological niche occupancy.

Comparative genomic analyses on marine picocyanobacteria have so far mainly focused on comparing gene repertoires from strains isolated from distinct niches, with the idea that niche adaptation mainly relies on differential gene content (Rocap et al., 2003; Palenik et al., 2006; Kettler et al., 2007; Dufresne et al., 2008). Here, the comparison of several strains per clade led in most cases to the identification of relatively few specific genes of known function that may confer a trait important for niche adaptation, even using relaxed stringency criteria (e.g. by selecting genes present in > 90 % of strains within a clade assemblage and in < 10 % of others; Dataset 2). This may be due to the existence of an extended within-taxa microdiversity (Martiny *et al.*, 2009c; Kashtan *et al.*, 2014; Larkin *et al.*, 2016; Farrant *et al.*, 2016): the more genomes are added into a taxon, the lower the number of genes found in all strains of this taxon. This fairly low number of niche-specific genes might also suggest that gene gain/loss, and fixation of these events during evolution, is a less prominent process to explain niche adaptation of marine picocyanobacteria than previously thought. Indeed, although lateral gene transfer is often considered to "commonly" occur between cells, and was notably shown to be involved in adaptation to nitrogen- or phosphorus-poor conditions in *Prochlorococcus,* no previous study explicitly stated the evolutionary time scale at which these adaptations take place (Martiny *et al.*, 2006, 2009b; Kettler *et al.*, 2007; Dufresne *et al.*, 2008; Scanlan *et al.*, 2009; Berube *et al.*, 2014). While the higher estimated rate of gene gain on the terminal branches of the phylogenetic tree (Fig. 3, Table 1) indicates that most detectable events occurred relatively recently with regard to the long evolutionary history of both genera, adding time calibration to the tree questions the dynamics of adaptation by changes in gene content. With an estimation of 4.7 and 4 genes gained per My on terminal branches in *Prochlorococcus* HL and *Synechococcus* SC 5.1 strains, respectively, gene gains actually appear to be rather rare events. Even though these numbers are approximate due to uncertainties in time calibration and probably underestimated, they are quite in line with those estimated for *Prochlorococcus* HLII populations, thought to have diverged a few million years ago but only possessing a dozen unique genes (Kashtan et al., 2014). Furthermore, in accordance with previous studies on other bacterial groups (Lerat et al., 2005; Ochman et al., 2005; Nowell et al., 2014; McDonald and Currie, 2017), the fact that rates of gene gain/loss are estimated to be higher on terminal branches of the tree (Fig. 3), together with the high number of unique genes in every sequenced strains (Fig. 2C), clearly shows that most recently acquired genes will not be kept on the long term in both genera. Our calculation indeed gives an approximate value of 1.45 and 0.68 genes gained and subsequently fixed per My in *Prochlorococcus* HL and *Synechococcus* SC 5.1, respectively. This low fixation rate suggests that most of the recently gained genes have no or little beneficial effect on the fitness on the long

term and that we observe them in genomes because purging selection has not deleted them yet (Hao and Golding, 2006; Abby and Daubin, 2007; Rocha, 2008).

This result also has important consequences on our interpretation of the role of the flexible genome in terms of adaptation to distinct niches. Indeed, genes conferring adaptation to a specific niche are mixed in the genomes with genes of no or little beneficial effect and are thus difficult to identify – in particular when they have only a putative function. The dynamics of gene content we exposed also implies that flexible genes that are fixed within a clade (i.e. clade-specific genes) were gained tens of millions years ago, and thus might be more reflective of past selective forces than of recent adaptation to newly colonized niches. In this context, genes specifically found in *Synechococcus* clade VIII and SC 5.2 suggest that adaptation to low salinity environments was a key factor in their differentiation from other taxa. Similarly, adaptation to phosphorus-depleted oligotrophic areas probably drove the differentiation of *Synechococcus* clade III, as revealed by the fairly large number of P- and other nutrient-uptake genes specific of this clade. Interestingly, co-occurring ESTUs IIIA, WPC1A and SC 5.3A, do not share many genes that could explain a common adaptation to this environment. Instead these ESTUs seem to have independently acquired different sets of genes to perform similar functions (see Results and Dataset 2). It is also noteworthy that in *Prochlorococcus,* P-uptake and assimilation is not clade-related but dependent on within-clade variability in genomic islands (Martiny *et al.*, 2006, 2009b), thus highlighting the variety of evolutionary paths that led to adaptation to low-P in these different lineages.

As proposed recently for other bacterial model organisms (Thrash et al., 2014; McDonald and Currie, 2017), sequence alterations could also have a key role in the genome diversification of marine picocyanobacteria and drive their adaptation to specific environments. Indeed, in the time period necessary for one gene to be gained, we found that 20 to 60 substitutions accumulate in any picocyanobacterial genome, and a much higher proportion of substitutions than gained genes appear to be fixed on the long term. This finding brings new evidence in support of the "Maestro Microbe" model of bacterial genome evolution recently proposed by (Larkin and Martiny, 2017), which posits that some phenotypic trait, such as thermal preferences, evolve by progressive fitness optimization of protein sequences rather than gene gain and loss. This theory is mainly based on the lack of specific genes that may explain trait differences between closely related organisms inhabiting distinct niches, and one of the best examples concerns *Prochlorococcus* clades colonizing cold (HLI) and warm (HLII) environments (Coleman et al., 2006; Martiny et al., 2006; Kettler et al., 2007; Larkin and Martiny, 2017), which were undifferentiated on our tree based on gene content despite a clear phylogenetic clustering with core marker genes (Fig. 5). The sequencing of new *Synechococcus* genomes also allows us to extend this hypothesis to *Synechococcus* thermotypes (Zwirglmaier et al., 2008; Pittera et al., 2014), since very few genes were found to be specific to cold-adapted clades I and IV (Dataset 2). In comparison, 5,352 and 3,381 amino-acid substitutions in core genes were found to be specific to the cold thermotypes *Prochlorococcus* HLI and *Synechococcus* clade I, respectively. In line with this, a recent study identified two specific substitutions (also retrieved in our analysis) in the genes encoding the two subunits of phycocyanin in *Synechococcus,* which could be involved in the adaptation to low temperature (Pittera et al., 2016). Several examples have also shown that mutations can rapidly rise

in a *Prochlorococcus* population as an adaptation to selective conditions such as UV radiations (Osburne et al., 2010), antibiotics (Osburne et al., 2011) or phage pressure (Avrani et al., 2011), emphasizing the role of substitutions in adaptation. Interestingly, clade-specific substitutions were found in our dataset to preferentially target the same genes in all genomes, more particularly genes involved in DNA repair, cell cycle as well as cobalamin biosynthesis (not shown). Although the adaptive potential of these fixed, specific substitutions remains to be confirmed, our results are in support of a combined role of gene acquisition by horizontal gene transfer and gene sequence modification in the adaptation of bacteria to environmental changes at long evolutionary time scales. It is important to note that both processes leave a quite congruent phylogenetic signal (as exemplified by the correspondence between the two trees in Fig. 5), confirming that gene content evolution is not decoupled from sequence evolution (Snel et al., 1999; Ochman et al., 2005; Kettler et al., 2007; Dufresne et al., 2008).

Thus altogether, we might consider clades as the survivors of a former set of "backbone" populations (as defined by (Kashtan et al., 2014)) that appeared hundreds of millions years ago, and then optimized their sequence, while eventually losing part of the genes that initially allowed niche colonization (Lawrence, 2002; Cohan and Koeppel, 2008; Polz et al., 2013; Kashtan et al., 2014). More recently, each of these clades further diversified into a number of new backbone populations, which correspond to the within-clade microdiversity recently described in *Prochlorococcus* and *Synechococcus* (see e.g. (Martiny *et al.*, 2009c; Kashtan *et al.*, 2014; Larkin *et al.*, 2016; Farrant *et al.*, 2016)). One explanation for the topology of the phylogenetic tree based on core proteins (short branches at the leaves of the tree and long branches at the base of clades, Fig. 5) would be the occurrence of periods of rapid diversification, such as the one that gave rise to the different *Synechococcus* clades within SC 5.1 and to *Prochlorococcus* radiation (Urbach et al., 1998; Dufresne et al., 2008) and longer periods during which each population stays relatively genetically homogeneous (e.g. by homologous recombination or by frequent genomic sweeps). Alternatively, and perhaps more likely, picocyanobacterial populations might undergo a constant diversification, and the diversity is purged during rare but severe events, leaving traces only of the surviving ones. Such an event would have occurred for example just before the within-clade diversification (explaining why clades have long branches at their base). While it is tempting to relate these events (diversification or purge) to past geological and climatic shifts, this would need a more thorough examination with an improved time calibration.

One of the next challenges will be to more precisely relate variants (genes or SNPs) to a particular niche. This can be achieved by comparative genomics but usually necessitates hundreds to thousands of closely related genomes (for review see (Read and Massey, 2014; Chen and Shapiro, 2015)), as well as a refined phenotypic characterization of the sequenced strains. Alternatively, one can search directly *in situ* data for genes or SNPs related to a particular niche or environmental parameter (see e.g. (Kent et al., 2016)). Given the wealth of metagenomics data from the oceans that are becoming available, such an approach will probably bring important novel insights on this topic in the forthcoming years.

Acknowledgements

This work was supported by the French "Agence Nationale de la Recherche" BioAdapt Program (SAMOSA, ANR-13-ADAP-0010), the Genoscope project METASYN (http://www.genoscope.cns.fr/spip/Variety-of-photosynthetic-pigments.html). We warmly thank the Roscoff Culture Collection for maintaining the *Synechococcus* strain used in this study.

Supplemental information

Figure S1 Phylogenetic tree of the 81 picocyanobacterial strains based on 821 concatenated core proteins, with internal nodes named. Maximum-likelihood tree, only the topology is given. Node names used in the text are indicated.

Figure S2 Average Amino-acid Identity (AAI) between pairs of picocyanobacterial strains. A. Heatmap showing the AAI between pairs of strains. B. Relationship between AAI and ANI (shown on Fig 2). No clear discontinuity in AAI or ANI allows to delineate distinct genera at a fixed cut-off.

Table S1 Accession numbers and characteristics of the genomes used is this study.

References

Abby S, Daubin V. (2007). Comparative genomics and the evolution of prokaryotes. *Trends Microbiol* **15**: 135–141.

Avrani S, Wurtzel O, Sharon I, Sorek R, Lindell D. (2011). Genomic island variability facilitates *Prochlorococcus*-virus coexistence. *Nature* **474**: 604–608.

Axelsen KB, Palmgren MG. (1998). Evolution of substrate specificities in the P-type ATPase superfamily. *J Mol Evol* **46**: 84–101.

Berube PM, Biller SJ, Kent AG, Berta-Thompson JW, Kelly L, Roggensack SE, et al. (2014). Physiology and evolution of nitrogen acquisition in *Prochlorococcus*. *ISME J* **9**: 1195–1207.

Biller SJ, Berube PM, Lindell D, Chisholm SW. (2014). *Prochlorococcus*: the structure and function of collective diversity. *Nat Rev Microbiol* **13**: 13–27.

Chen PE, Shapiro BJ. (2015). The advent of genome-wide association studies for bacteria. *Curr Opin Microbiol* **25**: 17–24.

Cohan FM, Koeppel AF. (2008). The Origins of Ecological Diversity in Prokaryotes. *Curr Biol* **18**: 1024–1034.

Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, Delong EF, et al. (2006). Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311**: 1768–1770.

Cordero OX, Polz MF. (2014). Explaining microbial genomic diversity in light of evolutionary ecology. *Nat Rev Microbiol* **12**: 263–273.

Csurös M. (2010). Count: Evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* **26**: 1910–1912.

Dufresne A, Garczarek L, Partensky F. (2005). Accelerated evolution associated with genome reduction in a free-living prokaryote. *Genome Biol* **6**: R14.1-10.

Dufresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik BP, et al. (2008). Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol* **9**: R90.

Eddy SR. (2011). Accelerated profile HMM searches. *PLoS Comput Biol* 7. e-pub ahead of print, doi: 10.1371/journal.pcbi.1002195.

Erdner DDL, Anderson DMD. (1999). Ferredoxin and flavodoxin as biochemical indicators of iron limitation during open-ocean iron enrichment. *Limnol Oceanogr* **44**: 1609–1615.

Farrant GK, Doré H, Cornejo-castillo FM, Partensky F, Ratin M, Garczarek L. (2016). Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proc Natl Acad Sci* **113**: E3365-74.

Farrant GK, Hoebeke M, Partensky F, Andres G, Corre E, Garczarek L. (2015). WiseScaffolder: an algorithm for the semi-automatic scaffolding of Next Generation Sequencing data. *BMC Bioinformatics* **16**: 281.

Flombaum P, Gallegos JL, Gordillo R a, Rincón J, Zabala LL, Jiao N, et al. (2013). Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc Natl Acad Sci U S A* **110**: 9824–9.

Galens K, Orvis J, Daugherty S, Creasy HH, Angiuoli S, White O, et al. (2011). The IGS Standard Operating Procedure for Automated Prokaryotic Annotation. *Stand Genomic Sci* **4**: 244–251.

Galili T. (2015). dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **31**: 3718–3720.

Genevaux P, Georgopoulos C, Kelley WL. (2007). The Hsp70 chaperone machines of *Escherichia coli*: A paradigm for the repartition of chaperone functions. *Mol Microbiol* **66**: 840–857.

Goris J, Konstantinidis KT, Klappenbach J a., Coenye T, Vandamme P, Tiedje JM. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57**: 81–91.

Guidi L, Chaffron S, Bittner L, Eveillard D. (2016). Plankton networks driving carbon export in the oligotrophic ocean. *Nature* **in press**.

Guindon S, Gascuel O. (2003). A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Syst Biol* 52: 696–704.

Hao W, Golding G. (2006). The fate of laterally transferred genes: life in the fast lane to adaptation or death. *Genome Res* 16: 636–643.

Harris SR, Feil EJ, Holden MTG, Quail MA, Nickerson EK, Chantratita N, et al. (2010). Evolution of MRSA During Hospital Transmission and Intercontinental Spread. *Science (80-)* **327**: 469 LP-474.

Haswell ES, Phillips R, Rees DC. (2011). Mechanosensitive channels: What can they do and how do they do it? *Structure* **19**: 1356–1369.

He Q, Dolganov N, Bjo O, Grossman AR, Natl P, Sci A. (2001). The High Light-inducible Polypeptides in *Synechocystis* PCC6803. *J Biol Chem* **276**: 306–314.

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

Huerta-Cepas J, Dopazo J, Gabaldón T. (2010). ETE: a python Environment for Tree Exploration. *BMC* Bioinformatics **11**: 24.

Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward EMS, Chisholm SW. (2006). Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* **311**: 1737–1740.

Jones DT, Taylor WR, Thornton JM. (1992). The rapid generation of mutation data matricies from protein sequences. *Comput Appl Biosci* **8**: 275–282.

Kamennaya NA, Post AF. (2011). Characterization of cyanate metabolism in marine *Synechococcus* and *Prochlorococcus* spp. *Appl Environ Microbiol* **77**: 291–301.

Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A, et al. (2014). Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* **344**: 416–20.

Katoh K, Standley DM. (2014). MAFFT: Iterative refinement and additional methods. In: D. R (ed) Vol. 1079. Methods in Molecular Biology (Methods and Protocols). Humana Press: Totowa, NJ, pp 131–146.

Kennemann L, Didelot X, Aebischer T, Kuhn S, Drescher B, Droege M, et al. (2011). *Helicobacter pylori* genome evolution during human infection. *Proc Natl Acad Sci* **108**: 5033–5038.

Kent AG, Dupont CL, Yooseph S, Martiny AC. (2016). Global biogeography of *Prochlorococcus* genome diversity in the surface ocean. *ISME J* **10**: 1856–1865.

Kettler GC, Martiny AC, Huang K, Zucker J, Coleman ML, Rodrigue S, et al. (2007). Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet* **3**: e231.

Konstantinidis KT, Tiedje JM. (2005a). Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A* **102**: 2567–2572.

Konstantinidis KT, Tiedje JM. (2005b). Towards a genome-based taxonomy for prokaryotes. *J Bacteriol* **187**: 6258–6264.

Kumar S, Stecher G, Tamura K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* **33**: 1870–1874.

Lan R, Reeves PR. (2000). Intraspecies variation in bacterial genomes: The need for a species genome concept. *Trends Microbiol* **8**: 396–401.

Larkin AA, Blinebry SK, Howes C, Lin Y, Loftus SE, Schmaus CA, et al. (2016). Niche partitioning and biogeography of high light adapted *Prochlorococcus* across taxonomic ranks in the North Pacific. *ISME J* 1–13.

Larkin AA, Martiny AC. (2017). Microdiversity shapes the traits, niche space, and biogeography of microbial taxa. *Environ Microbiol Rep* **9**: 55–70.

Lawrence JG. (2002). Gene transfer in bacteria: speciation without species? *Theor Popul Biol* **61**: 449–460.

Lawrence JG, Ochman H. (1998). Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci U S A* **95**: 9413–9417.

Lerat E, Daubin V, Ochman H, Moran NA. (2005). Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol* **3**: 0807–0814.

Li L, Stoeckert CJJ, Roos DS. (2003). OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res* **13**: 2178–2189.

Liu S, Cheng Y, Zhang X, Guan Q, Nishiuchi S, Hase K, et al. (2007). Expression of an NADP-malic enzyme gene in rice (Oryza sativa. L) is induced by environmental stresses; over-expression of the gene in Arabidopsis confers salt and osmotic stress tolerance. *Plant Mol Biol* **64**: 49–58.

Malmstrom RR, Rodrigue S, Huang KH, Kelly L, Kern SE, Thompson A, et al. (2012). Ecology of uncultured *Prochlorococcus* clades revealed through single-cell genomics and biogeographic analysis. *ISME J* **7**: 184–98.

Martiny AC, Coleman ML, Chisholm SW. (2006). Phosphate acquisition genes in *Prochlorococcus* ecotypes: Evidence for genome-wide adaptation. *Proc Natl Acad Sci U S A* **103**: 12552–12557.

Martiny AC, Huang Y, Li W. (2009a). Occurrence of phosphate acquisition genes in *Prochlorococcus* cells from different ocean regions. *Environ Microbiol* **11**: 1340–7.

Martiny AC, Tai APK, Veneziano D, Primeau F, Chisholm SW. (2009b). Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. *Environ Microbiol* **11**: 823–32.

Mather AE, Reid SWJ, Maskell DJ, Parkhill J, Fookes MC, Harris SR, et al. (2013). Distinguishable Epidemics of Multidrug-Resistant *Salmonella Typhimurium* DT104 in Different Hosts. *Science (80-)* **341**: 1514–1517.

Mazard S, Ostrowski M, Partensky F, Scanlan DJ. (2012). Multi-locus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*. *Environ Microbiol* **14**: 372–86.

McDonald BR, Currie CR. (2017). Lateral gene transfer dynamics in the ancient bacterial genus *Streptomyces*. *MBio* **8**: e00644-17.

Nowell RW, Green S, Laue BE, Sharp PM. (2014). The extent of genome flux and its role in the differentiation of bacterial lineages. *Genome Biol Evol* **6**: 1514–1529.

Ochman H, Lerat E, Daubin V. (2005). Examining bacterial species under the specter of gene transfer and exchange. *Proc Natl Acad Sci* **102**: 6595–6599.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, et al. (2015). vegan: Community Ecology Package. http://cran.r-project.org/package=vegan.

Osburne MS, Holmbeck BM, Coe A, Chisholm SW. (2011). The spontaneous mutation frequencies of *Prochlorococcus* strains are commensurate with those of other bacteria. *Environ Microbiol Rep* **3**: 744–749.

Osburne MS, Holmbeck BM, Frias-Lopez J, Steen R, Huang K, Kelly L, et al. (2010). UV hyper-resistance in *Prochlorococcus* MED4 results from a single base pair deletion just upstream of an operon encoding nudix hydrolase and photolyase. *Environ Microbiol 12*: 1978–1988.

Palenik B, Ren Q, Dupont CL, Myers GS, Heidelberg JF, Badger JH, et al. (2006). Genome sequence of *Synechococcus* CC9311: Insights into adaptation to a coastal environment. *Proc Natl Acad Sci* **103**: 13555–13559.

Paradis E, Claude J, Strimmer K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**: 289–290.

Partensky F, Blanchot J, Vaulot D. (1999). Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters: a review. *Bull l'institut océanographique 19*: 457–475.

Partensky F, Garczarek L. (2010). *Prochlorococcus*: advantages and limits of minimalism. *Ann Rev Mar Sci* **2**: 305–31.

Paulsen ML, Doré H, Garczarek L, Seuthe L, Müller O, Sandaa R-A, et al. (2016). *Synechococcus* in the Atlantic Gateway to the Arctic Ocean. *Front Mar Sci* **3**. e-pub ahead of print, doi: 10.3389/fmars.2016.00191.

Pittera J, Humily F, Thorel M, Grulois D, Garczarek L, Six C. (2014). Connecting thermal physiology and latitudinal niche partitioning in marine *Synechococcus*. *ISME J* **8**: 1221–1236.

Pittera J, Partensky F, Six C. (2016). Adaptive thermostability of light-harvesting complexes in marine picocyanobacteria. *ISME J* 1–13.

Polz MF, Alm EJ, Hanage WP. (2013). Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends Genet* **29**: 170–175.

Read TD, Massey RC. (2014). Characterizing the genetic basis of bacterial phenotypes using genome-wide association studies : a new direction for bacteriology. *Genome Med* **6**: 1–11.

Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA, et al. (2003). Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**: 1042–1047.

Rocha EP. (2008). Evolutionary patterns in prokaryotic genomes. *Curr Opin Microbiol* **11**: 454–460.

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

Sánchez-Baracaldo P. (2015). Origin of marine planktonic cyanobacteria. *Sci Rep* **5**: 17418.

Scanlan DJ. (2012). Marine Picocyanobacteria. In: Whitton BA (ed). Ecology of Cyanobacteria II: Their Diversity in Space and Time. Springer Netherlands: Dordrecht, pp 503–533.

Scanlan DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR, et al. (2009). Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol Rev* **73**: 249–99.

Scybert S, Pechous R, Sitthisak S, Nadakavukaren MJ, Wilkinson BJ, Jayaswal RK. (2003). NaCl-sensitive mutant of Staphylococcus aureus has a Tn917-lacZ insertion in its ars operon. *FEMS Microbiol Lett* **222**: 171–176.

Shih PM, Wu D, Latifi A, Axen SD, Fewer DP, Talla E, et al. (2013). Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci U S A* **110**: 1053–8.

Simm S, Keller M, Selymesi M, Schleiff E. (2015). The composition of the global and feature specific cyanobacterial core-genomes. *Front Microbiol* **6**: 1–21.

Six C, Thomas J-C, Garczarek L, Ostrowski M, Dufresne A, Blot N, et al. (2007). Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.: a comparative genomics study. *Genome Biol* **8**: R259.

Snel B, Bork P, Huynen M a. (1999). Genome phylogeny based on gene content. *Nat Genet* **21**: 108–110.

Sohm J a, Ahlgren N a, Thomson ZJ, Williams C, Moffett JW, Saito M a, et al. (2015). Co-occurring *Synechococcus* ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. *ISME J* **10**: 1–13.

Tamura K, Battistuzzi FU, Billing-Ross P, Murillo O, Filipski A, Kumar S. (2012). Estimating divergence times in large molecular phylogenies. *Proc Natl Acad Sci U S A* **109**: 19333–8.

Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, et al. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial 'pan-genome'. *Proc Natl Acad Sci U S A* **102**: 13950–5.

Tetu SG, Brahamsha B, Johnson DA, Tai V, Phillippy K, Palenik B, et al. (2009). Microarray analysis of phosphate regulation in the marine cyanobacterium *Synechococcus* sp. WH8102. *ISME J* **3**: 835–849.

Thrash JC, Temperton B, Swan BK, Landry ZC, Woyke T, DeLong EF, et al. (2014). Single-cell enabled comparative genomics of a deep ocean SAR11 bathytype. *ISME J* **8**: 1440–51.

Urbach E, Scanlan DJ, Distel DL, Waterbury JB, Chisholm SW. (1998). Rapid diversification of marine picophytoplankton with dissimilar light- harvesting structures inferred from sequences of *Prochlorococcus* and *Synechococcus* (cyanobacteria). *J Mol Evol* **46**: 188–201.

Zhao Y, Jia X, Yang J, Ling Y, Zhang Z, Yu J, et al. (2014). PanGP: A tool for quickly analyzing bacterial pan-genome profile. *Bioinformatics* **30**: 1297–1299.

Zinser ER, Johnson ZI, Coe A, Karaca E, Veneziano D, Chisholm SW. (2007). Influence of light and temperature on *Prochlorococcus* ecotype distributions in the Atlantic Ocean. *Limnol Oceanogr* **52**: 2205–2220.

Zwirglmaier K, Heywood JL, Chamberlain K, Woodward EMS, Zubkov M V, Scanlan DJ. (2007). Basin-scale distribution patterns of picocyanobacterial lineages in the Atlantic Ocean. *Environ Microbiol* **9**: 1278–1290.

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

Supplemental Information

Table S1 Accession numbers and characteristics of the genomes used is this study.

Taxonomy						Accession	Genome caracteristics							
Strain Name	Genus	Sub-cluster	Clade	Sub-clade	Pigment type	GenBank ID	Sequencing center	Genome size	Region origin	Depth isolation	Date isolation	Lat (°N)	Long (°E)	References
EQPAC1	*Prochlorococcus*	HL	HLI	HLI	Low b/a	JNAG00000000	MIT BioMicro Center	1 655 089						Biller et al., 2014
MED4	*Prochlorococcus*	HL	HLI	HLI	Low b/a	BX548174	JGI	1 657 991						Rocap et al., 2003; Kettler et al., 2007
MIT9515	*Prochlorococcus*	HL	HLI	HLI	Low b/a	CP000552	JCVI	1 704 176						Kettler et al., 2007
AS9601	*Prochlorococcus*	HL	HLII	HLII	Low b/a	CP000551	JCVI	1 669 886						Kettler et al., 2007
GP2	*Prochlorococcus*	HL	HLII	HLII	Low b/a	JNAH00000000	MIT BioMicro Center	1 624 810						Biller et al., 2014
MIT0604	*Prochlorococcus*	HL	HLII	HLII	Low b/a	CP007753	Center/Yale Center for Ge	1 780 061						Biller et al., 2014
MIT9123	*Prochlorococcus*	HL	HLII	HLII	Low b/a	JNAK00000000	MIT BioMicro Center	1 698 598						Biller et al., 2014
MIT9201	*Prochlorococcus*	HL	HLII	HLII	Low b/a	JNAL00000000	MIT BioMicro Center	1 673 653						Biller et al., 2014
MIT9202	*Prochlorococcus*	HL	HLII	HLII	Low b/a	ACDW00000000	JCVI	1 691 453						Unpublished
MIT9215	*Prochlorococcus*	HL	HLII	HLII	Low b/a	CP000825	JGI	1 738 790						Kettler et al., 2007
MIT9301	*Prochlorococcus*	HL	HLII	HLII	Low b/a	CP000576	JCVI	1 641 879						Kettler et al., 2007
MIT9302	*Prochlorococcus*	HL	HLII	HLII	Low b/a	JNAM00000000	MIT BioMicro Center	1 746 143						Biller et al., 2014
MIT9312	*Prochlorococcus*	HL	HLII	HLII	Low b/a	CP000111	JGI	1 709 204						Coleman et al., 2006
MIT9314	*Prochlorococcus*	HL	HLII	HLII	Low b/a	JNAO00000000	MIT BioMicro Center	1 691 306						Biller et al., 2014
MIT9321	*Prochlorococcus*	HL	HLII	HLII	Low b/a	JNAP00000000	MIT BioMicro Center	1 659 114						Biller et al., 2014
SB	*Prochlorococcus*	HL	HLII	HLII	Low b/a	JNAS00000000	MIT BioMicro Center	1 669 973						Biller et al., 2014
MIT0801	*Prochlorococcus*	LL	LLI	LLI	High b/a	CP007754	Center/Yale Center for Ge	1 929 203						Biller et al., 2014
NATL1A	*Prochlorococcus*	LL	LLI	LLI	High b/a	CP000553	JGI	1 864 731						Kettler et al., 2007
NATL2A	*Prochlorococcus*	LL	LLI	LLI	High b/a	CP000095	JGI	1 842 899						Kettler et al., 2007
PAC1	*Prochlorococcus*	LL	LLI	LLI	High b/a	JNAX00000000	MIT BioMicro Center	1 842 113						Biller et al., 2014
LG	*Prochlorococcus*	LL	LLII	LLII	High b/a	JNAT00000000	MIT BioMicro Center	1 754 713						Biller et al., 2014
SS120	*Prochlorococcus*	LL	LLII	LLII	High b/a	AE017126	Genoscope	1 751 081						Dufresne et al., 2003
MIT0601	*Prochlorococcus*	LL	LLIII	LLIII	High b/a	JNAU00000000	MIT BioMicro Center	1 707 592						Biller et al., 2014
MIT0602	*Prochlorococcus*	LL	LLIII	LLIII	High b/a	JNAV00000000	MIT BioMicro Center	1 751 318						Biller et al., 2014
MIT9211	*Prochlorococcus*	LL	LLIII	LLIII	High b/a	CP000878	JGI	1 688 963						Kettler et al., 2007
MIT0701	*Prochlorococcus*	LL	LLIV	LLIV	High b/a	JNBA00000000	MIT BioMicro Center	2 595 171						Biller et al., 2014
MIT9303	*Prochlorococcus*	LL	LLIV	LLIV	High b/a	CP000554	JCVI	2 682 675						Kettler et al., 2007
MIT9313	*Prochlorococcus*	LL	LLIV	LLIV	High b/a	BX548175	JGI	2 410 874						Rocap et al., 2003; Kettler et al., 2007
BIOS-U3-1	*Synechococcus*	5.1	CRD1	CRD1	3dA	n.a.	Genoscope	2 710 834	Chile upwelling	5	06/12/2004	-33,98	-73,39	Unpublished
MITS9220	*Synechococcus*	5.1	CRD1	CRD1	3dA	n.a.	CGR, Liverpool, UK	2 424 175						Unpublished
BIOS-E4-1	*Synechococcus*	5.1	CRD1	CRD1	3cA	n.a.	Genoscope	3 314 220	South East Pacific	40	28/11/2004	-31,87	-91,44	Unpublished
CC9616	*Synechococcus*	5.1	XX	XX	n.a.	n.a.	JGI	2 645 910						http://genome.jgi.doe.gov/Syn9616DRAFT_10047/Syn9616DRAFT_10047.info.html
CC9311	*Synechococcus*	5.1	I	Ia	3dA	CP000435	JGI	2 606 749						Palenik et al., 2006, Dufresne et al., 2008
MVIR-18-1	*Synechococcus*	5.1	I	Ib	3aA	n.a.	Genoscope	2 451 974	-	25	23/07/2007	61,00	1,98	Unpublished
PROS-9-1	*Synechococcus*	5.1	I	Ib	3dA	n.a.	CGR, Liverpool, UK	2 273 940						Unpublished
ROS8604	*Synechococcus*	5.1	I	Ib	3a	n.a.	Genoscope	2 876 904	Britanny coast	0	24/11/1986	48,72	-3,98	Unpublished
SYN20	*Synechococcus*	5.1	I	Ib	3a	n.a.	CGR, Liverpool, UK	2 728 628						Unpublished
WH8016	*Synechococcus*	5.1	I	Ib	3aA	n.a.	JGI	2 695 643						Unpublished
A15-44	*Synechococcus*	5.1	II	IIa	2	n.a.	Genoscope	2 621 965	Upwelling	20	01/10/2004	21,68	-17,83	Unpublished
KORDI-52	*Synechococcus*	5.1	II	IIa	3bB	CP006271.1	JGI	2 572 069						Unpublished; 16S rRNA and ITS phylogenies described in Choi & Noh (2009); Choi et al. (2014)
M16.1	*Synechococcus*	5.1	II	IIa	3a	n.a.	Genoscope	2 112 236	Gulf of Mexico	275	09/02/2004	27,70	-91,30	Unpublished
RS9902	*Synechococcus*	5.1	II	IIa	3c	n.a.	Genoscope	2 481 217	Gulf of Aqaba	1	29/03/1999	29,47	34,92	Unpublished
RS9907	*Synechococcus*	5.1	II	IIa	3a	n.a.	Genoscope	2 581 635	Gulf of Aqaba	10	23/08/1999	29,47	34,92	Unpublished
TAK9802	*Synechococcus*	5.1	II	IIa	3a	n.a.	Genoscope	2 190 394	Takapoto atoll	7	06/02/1998	-14,50	-145,33	Unpublished
WH8109	*Synechococcus*	5.1	II	IIa	3bB	CP006882.1	JCVI	2 111 515						Unpublished
A15-62	*Synechococcus*	5.1	II	IIc	3dB	n.a.	Genoscope	2 294 140	-	15	04/10/2004	17,62	-20,95	Unpublished
CC9605	*Synechococcus*	5.1	II	IIc	3c	CP000110	JGI	2 510 660						Dufresne et al., 2008
PROS-U-1	*Synechococcus*	5.1	II	IIh	3dB	n.a.	Genoscope	2 576 003	Maroccan upwelling	5	12/09/1999	30,13	-10,05	Unpublished
A15-24	*Synechococcus*	5.1	III	IIIa	3c	n.a.	CGR, Liverpool, UK	2 305 373						Unpublished
A18-40	*Synechococcus*	5.1	III	IIIa	3dB	n.a.	CGR, Liverpool, UK	2 401 547						Unpublished
A18-46.1	*Synechococcus*	5.1	III	IIIa	3c	n.a.	CGR, Liverpool, UK	2 471 770						Unpublished
BOUM118	*Synechococcus*	5.1	III	IIIa	3c	n.a.	Genoscope	2 326 334	-	5	01/10/2008	33,63	32,63	Unpublished
RS9915	*Synechococcus*	5.1	III	IIIa	3dB	n.a.	Genoscope	2 417 983	Gulf of Aqaba	10	18/10/1999	29,47	34,92	Unpublished
WH8102	*Synechococcus*	5.1	III	IIIa	3c	BX548020	JGI	2 434 429						Palenik et al., 2003; Dufresne et al., 2008
WH8103	*Synechococcus*	5.1	III	IIIa	3bB	n.a.	Genoscope	2 429 688	-	?	17/03/1981	28,50	-67,40	Unpublished
A15-28	*Synechococcus*	5.1	III	IIIb	3c	n.a.	Genoscope	2 341 586	Northern gyre (edge)	15	25/09/2004	31,25	-20,72	Unpublished
BL107	*Synechococcus*	5.1	IV	IVa	3dA	AATZ00000000	JCVI	2 285 035						Dufresne et al., 2008
CC9902	*Synechococcus*	5.1	IV	IVa	3dA	CP000097	JGI	2 234 829						Dufresne et al., 2008
RS9916	*Synechococcus*	5.1	IX	IX	3dA	AAUA00000000	JCVI	2 664 874						Dufresne et al., 2008
KORDI-100	*Synechococcus*	5.1	UC-A	UC-A	n.a.	CP006269.1	JGI	2 789 000						Unpublished; ITS phylogeny described in Choi et al. (2014)
BMK-MC-1	*Synechococcus*	5.1	V	V	2	n.a.	Genoscope	2 601 150	Naples bay	23	13/10/2009	40,81	14,25	Unpublished
WH7803	*Synechococcus*	5.1	V	V	3a	CT971583	Genoscope	2 366 981						Dufresne et al., 2008
MEDNS5	*Synechococcus*	5.1	VI	VIa	3c	n.a.	Genoscope	2 284 343	-	80	01/07/1993	41,00	6,00	Unpublished
PROS-7-1	*Synechococcus*	5.1	VI	VIb	2	n.a.	Genoscope	2 565 218	-	5	26/09/1999	37,40	15,62	Unpublished
WH7805	*Synechococcus*	5.1	VI	VIc	2	AAOK00000000	JCVI	2 627 147						Dufresne et al., 2008
A15-60	*Synechococcus*	5.1	VII	VIIa	3c	n.a.	Genoscope	2 543 402	-	10	04/10/2004	17,62	-20,95	Unpublished
A18-25c	*Synechococcus*	5.1	VII	VIIa	3c	n.a.	CGR, Liverpool, UK	2 511 360						Unpublished
NOUM97013	*Synechococcus*	5.1	VII	VIIb	3a	n.a.	Genoscope	2 552 712	Equatorial Pacific	0	19/11/1996	-22,33	166,33	Unpublished
RS9909	*Synechococcus*	5.1	VIII	VIIIa	1	n.a.	Genoscope	2 603 169	Gulf of Aqaba	10	07/09/1999	29,47	34,92	Unpublished
RS9917	*Synechococcus*	5.1	VIII	VIIIa	1	AANP00000000	JCVI	2 584 919						Dufresne et al., 2008
WH8101	*Synechococcus*	5.1	VIII	VIIIc	1	n.a.	Genoscope	2 623 023	East coast USA	?	01/01/1981	41,53	-70,68	Unpublished
A15-127	*Synechococcus*	5.1	WPC1	WPC1a	3c	n.a.	Genoscope	2 543 463	-	45	22/10/2004	-31,12	-3,92	Unpublished
KORDI-49	*Synechococcus*	5.1	WPC1	WPC1a	3a	CP006270.1	JGI	2 585 813						Unpublished; 16S rRNA and ITS phylogenies described in Choi & Noh (2009) ;Choi et al. (2014)
NS01	*Cyanobium*	5.2	n.a.	n.a.	1	n.a.	Genoscope	2 748 087		?				Unpublished
PCC6307	*Cyanobium*	5.2	n.a.	n.a.	1	CP003495	JGI	3 342 364						Shih et al., 2013; Gerloff et al., 1950
PCC7001	*Cyanobium*	5.2	n.a.	n.a.	1	NZ_ABSE01000000	JCVI	2 832 412						Unpublished
CB0101	*Synechococcus*	5.2	n.a.	n.a.	1	ADXL00000000	TIGR	2 691 045						Marsan et al., 2014
CB0205	*Synechococcus*	5.2	n.a.	n.a.	2	ADXM00000000	TIGR	2 431 158						Unpublished
WH5701	*Synechococcus*	5.2	n.a.	n.a.	1	AANO00000000	JCVI	3 051 556						Dufresne et al., 2008
MINOS11	*Synechococcus*	5.3	n.a.	n.a.	3dB	n.a.	Genoscope	2 284 343	-	20	19/06/1996	34,00	18,00	Unpublished
RCC307	*Synechococcus*	5.3	n.a.	n.a.	3eA	CT978603	Genoscope	2 224 915						Dufresne et al., 2008

Figure S1 Phylogenetic tree of the 81 picocyanobacterial strains based on 821 concatenated core proteins, with internal nodes named. Maximum-likelihood tree, only the topology is given. Node names used in the text are indicated.

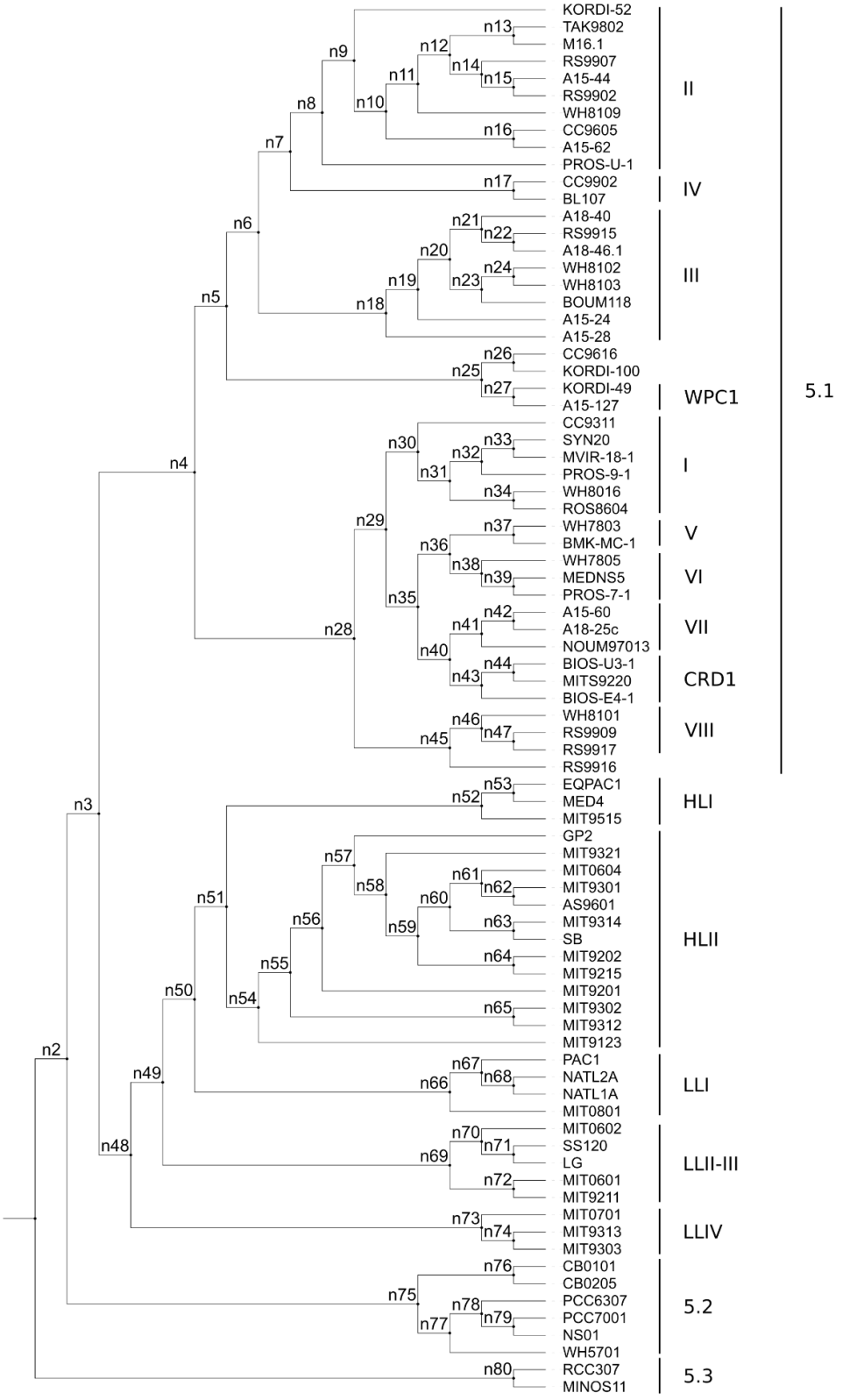
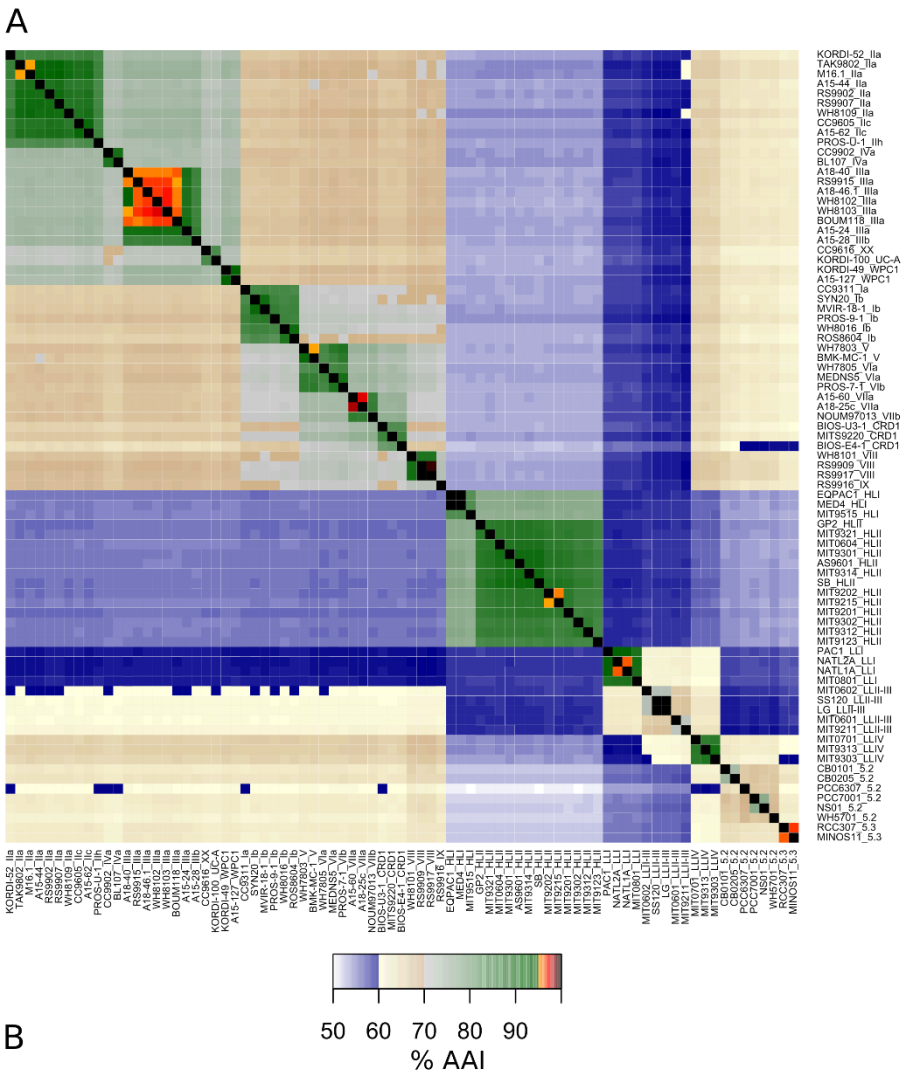


Figure S2 Average Amino-acid Identity (AAI) between pairs of picocyanobacterial strains. A. Heatmap showing the AAI between pairs of strains. B. Relationship between AAI and ANI (shown on Fig 2). No clear discontinuity in AAI or ANI allows to delineate distinct genera at a fixed cut-off.


A
KORDI-52_IIa
TAK9802_IIa
M16.1_IIa
A15-44_IIa
RS9902_IIa
RS9907_IIa
WH8109_IIa
CC9605_IIc
A15-62_IIc
PROS-U-1_IIh
CC9902_IVa
BL107_IVa
A18-40_IIIa
RS9915_IIIa
A18-46.1_IIIa
WH8102_IIIa
WH8103_IIIa
BOUM118_IIIa
A15-24_IIIa
A15-28_IIIb
CC9616_XX
KORDI-100_UC-A
KORDI-49_WPC1
A15-127_WPC1
CC9311_Ia
SYN20_Ib
MVIR-18-1_Ib
PROS-9-1_Ib
WH8016_Ib
ROS8604_Ib
WH7803_V
BMK-MC-1_V
WH7805_VIa
MEDNS5_VIa
PROS-7-1_VIb
A15-60_VIIa
A18-25c_VIIa
NOUM97013_VIIb
BIOS-U3-1_CRD1
MITS9220_CRD1
BIOS-E4-1_CRD1
WH8101_VIII
RS9909_VIII
RS9917_VIII
RS9916_IX
EQPACT_HLI
MED4_HLI
MIT9515_HLI
GP2_HLII
MIT9321_HLII
MIT0604_HLII
MIT9301_HLII
AS9601_HLII
MIT9314_HLII
SB_HLII
MIT9202_HLII
MIT9215_HLII
MIT9201_HLII
MIT9302_HLII
MIT9312_HLII
MIT9123_HLII
PAC1_LLI
NATL2A_LLI
NATL1A_LLI
MIT0801_LLI
MIT0602_LLII-III
SS120_LLII-III
LG_LLII-III
MIT0601_LLII-III
MIT9211_LLII-III
MIT0701_LLIV
MIT9313_LLIV
MIT9303_LLIV
CB0101_5.2
CB0205_5.2
PCC6307_5.2
PCC7001_5.2
NS01_5.2
WH5701_5.2
RCC307_5.3
MINOS11_5.3
50 60 70 80 90
% AAI
B
ANI
AAI
80 85 90 95 100
80 85 90 95 100

2. Global distribution of *Synechococcus* and *Prochlorococcus* gene repertoires reveals adaptive strategies of picocyanobacterial communities

Context of the work:

In the first chapter of this thesis, we revealed the quite clear-cut distribution of picocyanobacterial Ecologically Significant Taxonomic Units (ESTUs) in the global ocean. The analysis of the distribution of these ESTUs along the *Tara* Oceans transect allowed us to discriminate 8 communities of *Synechococcus* and 3 main *Prochlorococcus* communities, each dominated by one or a few ESTUs and presenting distinct *preferenda* with regard to abiotic factors. More precisely, the most discriminant environmental parameters were temperature and iron availability for *Prochlorococcus,* to which can be added phosphate concentration for *Synechococcus*. However, the extensive study of gene content of available genomes, even with several sequenced strains representative of the dominant ESTUs, revealed the difficulty of finding niche specific genes by a mere comparative genomic approach. Various reasons can explain these difficulties, including our poor knowledge of the physiological characteristics of sequenced strains, as well as the imprecise delineation of niches in particular for organisms that were scarce in the sampled environments. As we evoked in the discussion of the last paper, one way to overcome these difficulties is to analyze directly the distribution of genes in the oceans, allowing to link them to both the taxonomic composition and the physico-chemical characteristics of the sampling sites. This is what we propose in this part, in which we use the 97 picocyanobacterial genomes available in Cyanorak v2 and their functional annotation to explore the distribution of *Prochlorococcus* and *Synechococcus* gene repertoires in the *Tara* oceans metagenomic dataset.

2.1 Material and Methods:

Tara Oceans dataset

We used the 109 metagenomes from the *Tara* Oceans campaign corresponding to 66 stations for which a "prokaryotic size fraction" was available (0.2-1.6 µm for TARA_004 to TARA_052 and 0.2-3µm for TARA_056 to TARA_152). This corresponds to a subset of the data presented in (Sunagawa *et al.*, 2015). Briefly, water samples were collected at two depths, surface and Deep Chlorophyll Maximum (DCM), the latter sample sometimes being merely collected in the upper mixed layer, when the DCM was not clearly delineated. Only surface samples were analyzed in the present study. All metagenomes were sequenced as Illumina overlapping paired reads of 100-108 bp. Paired reads were merged using FLASH v1.2.7 with default parameters (Magoč and Salzberg, 2011) and trimmed based on quality using CLC QualityTrim v4.10.86742 (CLC Bio), resulting in 100-215 bp fragments. All environmental parameters were retrieved from PANGAEA

https://doi.pangaea.de/10.1594/PANGAEA.840718) except for modeled iron and ammonium concentrations that are available in (Farrant *et al.*, 2016) and the iron limitation index Φ_{sat} calculated from satellite data as in (Behrenfeld *et al.*, 2009; Farrant *et al.*, 2016).

Taxonomic assignment using *petB* marker gene.

Relative abundance of marine picocyanobacterial taxa was estimated by recruiting metagenomic reads on a database of 722 sequences corresponding to the gene *petB* (encoding the cytochrome b_6) as described in (Farrant *et al.*, 2016). Groups of *Tara* Oceans stations based on *Synechococcus* or *Prochlorococcus* communities were retrieved from (Farrant *et al.*, 2016).

Functional assignment of metagenomics reads

Metagenomics reads mapping on picocyanobacterial genomes were assigned to clusters of orthologous genes (CLOGs) defined in the information system Cyanorak V2 (http://sb-roscoff.fr/Phyto/cyanorak), which comprises 97 genomes of marine picocyanobacteria. In order to reduce the amount of reads before assignment, metagenomic reads were first recruited against a database of 97 genomes of marine picocyanobacteria using BLASTN (v2.2.28+) (Altschul *et al.*, 1990) with default parameters but limiting the results to one target sequence (--max_target_seqs 1) and keeping only results with an E-value below 0.001 (-evalue 0.001). The resulting reads were then mapped to a database comprising 282 genomes of prokaryotes among which the 97 genomes of marine picocyanobacteria included in Cyanorak along with 185 outgroup cyanobacterial genomes downloaded from NCBI ftp on May 4, 2017. Reads were recruited with the same BLASTN options as above. Reads mapping to outgroup sequences or having less than 90% of their sequence aligned were filtered out and remaining reads were assigned to *Synechococcus* or *Prochlorococcus* according to their best hit. Reads were then assigned to a CLOG according to the position of their BLAST match in the genome. More precisely, a read was assigned to a gene if at least 75% of its size was aligned inside this gene and if the percentage of identity of the blast alignment was over 80%. Finally, read counts were aggregated by CLOG and normalized by average gene length of the CLOG.

Only samples that contained at least 2 500 and 1 700 distinct CLOGs for *Synechococcus* and *Prochlorococcus,* respectively, were kept, corresponding roughly to the average number of genes in a *Synechococcus* and a *Prochlorococcus* HL genome. After this filtration step, a CLOG was kept if it showed a normalized abundance higher than 1 in at least 2 of the remaining samples. For further analyses, CLOGs abundance in each sample was divided by the total abundance of the sample to obtain relative abundance profiles. CLOGs that were found in all retained stations were removed to keep only flexible genes (i.e. genes that are differentially distributed). The resulting relative abundance profiles were used to group stations by clustering using Bray-Curtis similarity and to perform co-occurrence network analyses.

Comparison of taxonomic and functional assignment

In order to check whether the distance between stations based on *petB* picocyanobacterial communities and the distance between stations based on gene content were significantly correlated, we used a mantel test between the Bray-Curtis distance matrices (based on relative abundance of ESTUs or genes) as implemented in the R package *vegan* v2.4 with 9999 permutations (Oksanen *et al.*, 2015).

Gene co-occurrence network analysis

In a data-reduction approach, we used a weighted correlation network analysis (WGCNA, (Zhang and Horvath, 2005)) as implemented in the R package WGCNA v1.51 (Langfelder and Horvath, 2008) to build a co-occurrence network of CLOGs based on their relative abundance in *Tara* Oceans stations and to delineate modules of CLOGs (i.e. subnetworks). The adjacency matrix was calculated in "signed" mode (i.e. without considering anti-correlated CLOGs as correlated), by taking the *pearson* correlation between pairs of CLOGs (based on their relative abundance in every samples) and raising it to the power 12, which allowed to obtain a scale-free topology of the network. Modules were identified by indicating a minimal number of genes in each module of 100 and 50 for *Synechococcus* and *Prochlorococcus,* respectively, and by forcing every gene to be included in a module. The *eigengene* of each module (representative of the level of presence of its genes in each *Tara* Oceans station) was then correlated to environmental parameters and ESTUs relative abundance. Finally, genes in each module with the highest correlation to the *eigengene* (a measure called *membership*) were extracted for further analysis.

2.2 Results

Distinct picocyanobacterial communities have distinct gene repertoires

The recruitment of the *Tara* Oceans metagenomic dataset against a database of 97 picocyanobacterial genomes yielded a total of 1.07 billion reads, of which 12.3% mapped on a *Synechococcus* genome and 87.7% on a *Prochlorococcus* genome. In average, *Synechococcus* and *Prochlorococcus* reads represented 1.08 % and 6.37 % of the reads in a given sample of the bacterial size fraction (0.22-1.6 µm or 0.22-3 µm), respectively. 96.3% of the recruited reads were successfully assigned to a cluster of orthologous genes (CLOG), unassigned reads corresponding to reads mapping on intergenic regions.

In order to understand how picocyanobacterial genes were distributed in the oceans, stations were first clustered according to the relative abundance of genes they contained, resulting in 4 well-formed clusters for *Synechococcus* (left dendrogram on Figure 34A). Interestingly, this clustering matches quite well the one found in our previous study of *Tara* Oceans dataset, where stations were clustered according their taxonomic composition in picocyanobacterial lineages, as assessed using the single *petB* marker gene (Farrant *et al.*, 2016). However, in the latter study we described 8 groups of stations based on *Synechococcus* ESTU assemblages that were not all discriminated in the present analysis based on gene content. Most of these discrepancies correspond to stations at which the *Synechococcus* community displayed an ESTU content somewhat intermediate between two ESTU assemblages (e.g. Station 125, which belongs to assemblage 3 but contains a large proportion of ESTU IIA, dominating in assemblage 1, see Fig. 4 from (Farrant *et al.*, 2016)). Overall, the 4 clusters retrieved when using gene abundance profiles correspond to four broadly defined ecological niches, namely cold, coastal environments (blue and light red in Figure 1A), High-Nutrients Low Chlorophyll (HNLC) environments (purple and grey), warm, P-depleted areas (yellow) and warm, iron-replete regions (dark red), as detailed below. The same applies for *Prochlorococcus,* in which the three types of communities described in (Farrant *et al.*, 2016) were clearly discriminated when looking at gene content, with only a few discrepancies (Figure 34B). This correspondence between taxonomic and functional information was confirmed by the high concordance between distance matrices based on ESTU relative abundance and on CLOG relative abundance (p-value $< 10^{-4}$, mantel test r = 0.89 and r = 0.78 for *Synechococcus* and *Prochlorococcus,* respectively). Thus, this indicates that distinct picocyanobacterial communities, as assessed based on a single taxonomic marker, also display distinct gene repertoires.

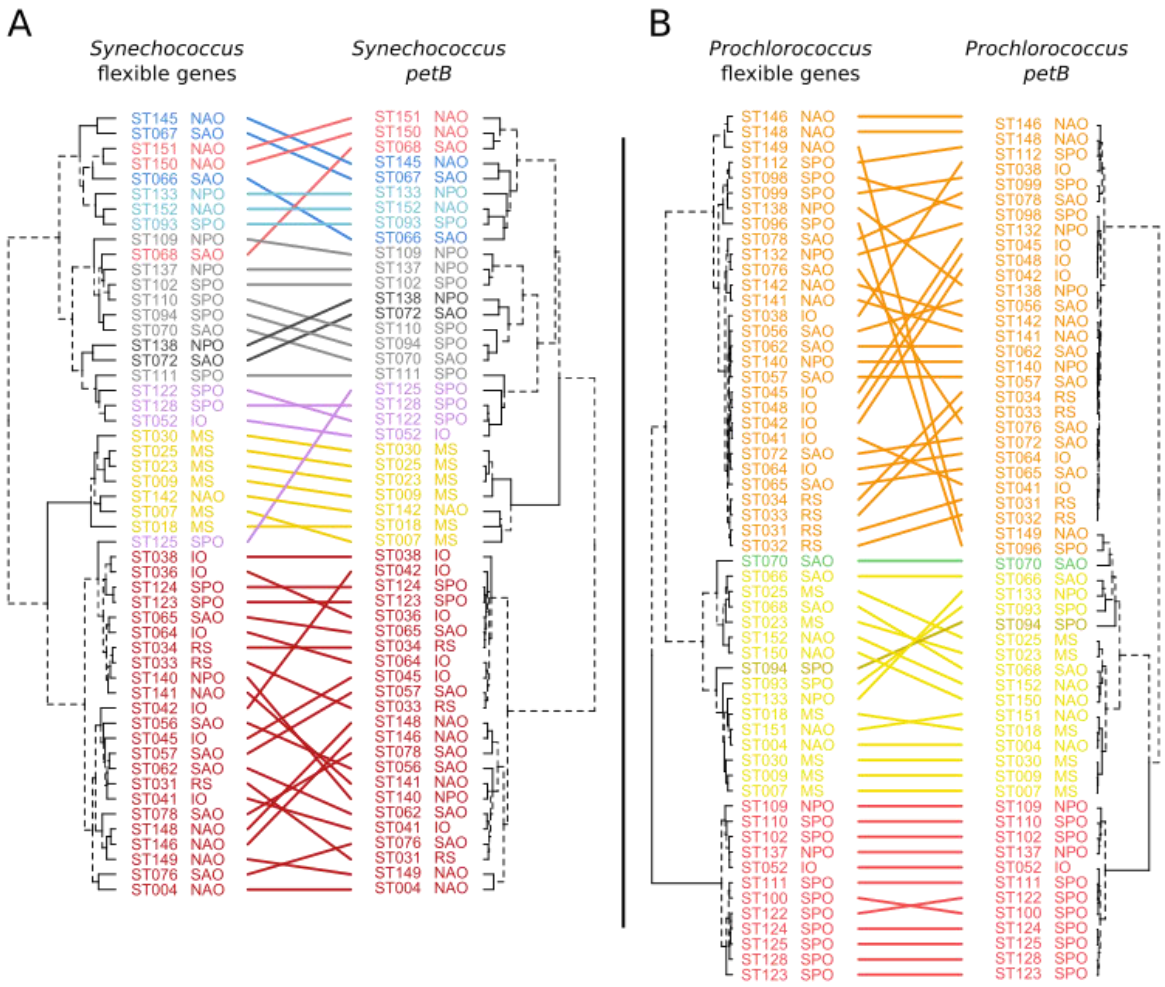


Figure 34 Comparison of clustering based on relative abundance profiles of ecologically significant taxonomic units (ESTUs) and on relative abundance profiles of flexible genes. Sampling stations are colored according the the assemblages defined in (Farrant et al., 2016). A. *Synechococcus* clustering. B. *Prochlorococcus* clustering. Taxonomic assignment (right tree in A and B) was performed using the single-copy, core, *petB* gene. Disrepancies between the clustering obtained here and the one obtained in the original publication are due to a slightly different normalization of relative abundance data. Dotted lines in dendrograms indicate discrepancies between trees. Flexible genes correspond to genes that are not detected in a least one sample.

Distribution of *Synechococcus* flexible genes is tightly linked to that of ESTUs

In order to further analyze the global distribution of gene content in light of available environmental parameters and to reduce the amount of data, a correlation network of genes was built based on relative abundance profiles of genes across *Tara* Oceans samples (see Material and Methods, Figure 35A-B and Figure 37A-B). In *Synechococcus*, this resulted into four main modules of genes, whose *eigengenes* (i.e. representative abundance profile of genes belonging to the module) show that each of them is abundant in a different set of stations, corresponding to the clusters of stations described above (Figure 35C).

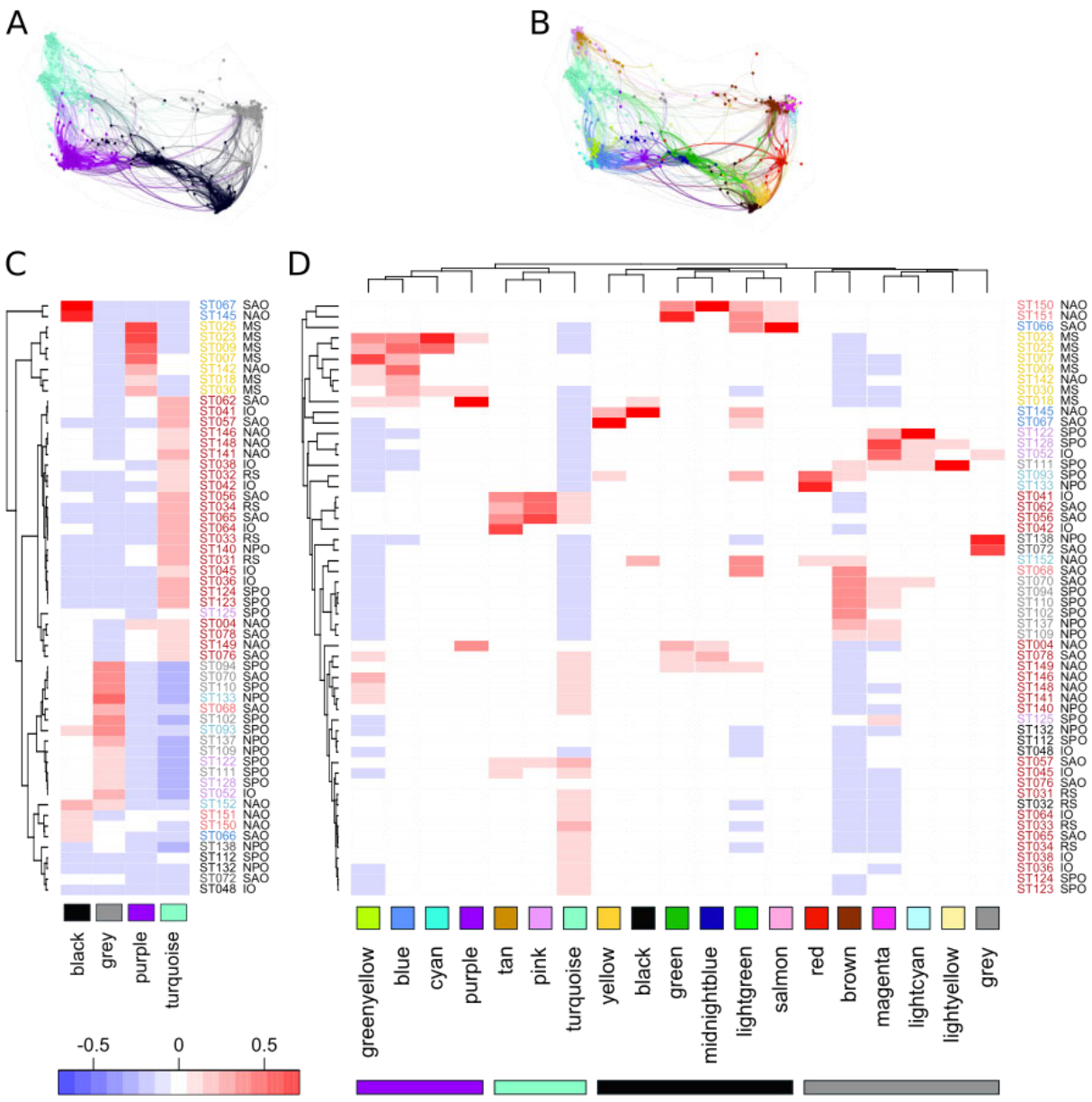


Figure 35 Correlation network based on relative abundance profiles of *Synechococcus* genes and decomposition into modules. A-B. Weighted correlation network based on flexible genes relative abundance profiles in *Tara* Oceans stations. The network contains four modules (A) that can be further decomposed into 19 submodules (B). C-D. Eigengenes representative of each module, designated by a color name. The eigengene of a module represents a consensus of the relative abundance profiles of genes within this module. Eigengenes of modules (C) and submodules (D) show the normalized abundance of genes in *Tara* Oceans stations, which are colored according to the clusters defined by (Farrant et al., 2016). Stations that were not analyzed in the latter study because of a low signal (ST032, ST048, ST112, ST132) are colored in black. The oceanic region of each station is also specified: SAO, Sout Atlantic Ocean; MS, Mediterranean Sea; NAO, North Atlantic Ocean; IO, Indian Ocean; RS, Red Sea; SPO, South Pacific Ocean; NPO, North Pacific Ocean.

Eigengenes were further correlated to available environmental parameters and relative abundance of ESTUs in every stations (Figure 36). This revealed that the *black* module corresponds to genes found in cold, mesotrophic environments with a high relative abundance of ESTUs IA and IVA-C, while the *turquoise* module is in contrast associated with warm, Fe-replete waters and mostly with the presence of ESTU IIA but also VIIA and 5.3B (Figure 36A-B). The *grey* module

comprises genes found in HNLC areas, since it is correlated to nitrate and ammonium concentrations and strongly anti-correlated to iron, and is accordingly correlated to ESTUs CRD1A, CRD1C, EnvAA and EnvBA (Figure 36A-B). Relation to environmental parameters is less marked for the *purple* module, but it is still anticorrelated to concentrations of all nutrients but iron, though this anticorrelation is only significant with ammonium concentration, and genes of this module are clearly found at stations were ESTUs IIIA, WPC1A and 5.3A were most abundant (Figure 36A).

Each module can further be divided into several sub-modules of genes showing slightly different distribution patterns (Figure 35B). Sub-modules are generally correlated to a lower number of environmental parameters and/or and more specific sets of ESTUs than the corresponding module. For instance, a sub-module can show a specific association with some weakly abundant ESTUs, such as the sub-module *cyan* that is associated with ESTU VIA, while the purple module is not. The relative abundance profiles of *Synechococcus* genes from the *cyan* sub-module at each station (Figure 35D) indeed shows its presence at stations where ESTU VIA reaches its highest abundance (ST023), while the phyletic pattern of genes of this sub-module (i.e. the number of copies of these genes in every picocyanobacterial genomes) shows that most of them are specific to genomes of clades V and VI (Dataset 1).

The same applies for the *green* sub-module, which is correlated with ESTU IIB and whose genes are often specific to the only IIB genome of our database (PROS-U-1), and for *pink* sub-module, the most associated to ESTU VIIA and whose genes are specific to one or two of the three clade VII genomes in our database.

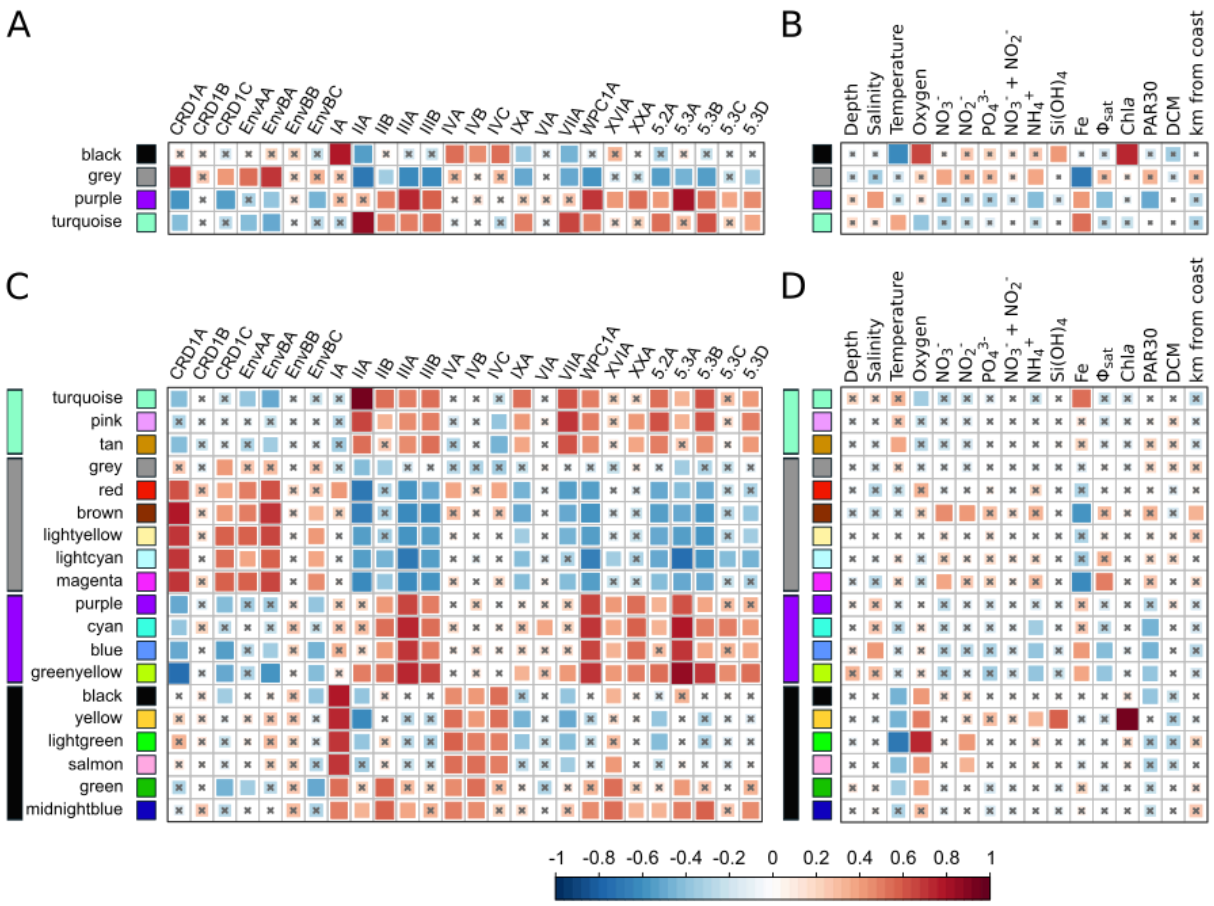


Figure 36 Correlation of *Synechococcus* module *eigengenes* to ESTU abundance and physico-chemical parameters. Correlation of module *eigenges* to relative abundance profiles of *Synechococcus* ESTUs (A) and to physico-chemical parameters (B), and correlation of sub-module *eigengenes* to ESTU relative abundance profiles (C) and to physico-chemical parameters (D). Spearman (A,C) and Pearson (B,D) correlation coefficient (R^2) is indicated by the color scale. Non-significant correlations (Student asymptotic p-value > 0.01) are marked by a cross. Φ_{sat} : index of iron limitation derived from satellite data. PAR30: satellite-derived photosynthetically available radiation at the surface, averaged on 30 days. DCM: depth of the deep chlorophyll maximum.

Thus, it seems that distinct ESTUs have different gene pools even when they co-occur, since at least part of the genes that are specific to an ESTU according to comparative genomics (including those found in a single sequenced strain) are preferentially associated with this ESTU *in situ.*

In order to identify genes related to particular environmental conditions and to specific ESTU assemblages, we correlated relative abundance profiles of each gene to the *eigengene* of its sub-module, a method allowing one to identify genes that contribute the most to the *eigengene*. Almost all genes retrieved this way were coding for hypothetical proteins (95% of 6243 genes, Dataset 1). As noted above, the phyletic pattern of these genes (i.e. their distribution in our genomic dataset) was in most cases consistent with the correlation between modules and ESTUs: for instance, members of the *black* module, which is correlated to ESTU IA, contained many genes found exclusively in some of (but usually not all) clade I genomes.

Among the genes with a functional annotation (Dataset 2), some have a function related to their realized environmental niche (*sensu* (Pearman *et al.*, 2008)). Within the *grey* module, associated to HNLC areas, one can find a gene encoding for ferritin (*fnt*), a protein involved in iron

storage, an additional copy of flavodoxin (*isiB;* a Cu-containing protein known to replace ferredoxin in response to iron depletion (Erdner and Anderson, 1999)), a putative ferrochelatase (CK_00035106) as well as a putative Zn transporter periplasmic component (*znuC*), which could be involved in the transport of other metals such as iron (Dataset 2). Thus, these genes are likely important for survival of *Synechococcus* communities thriving in these Fe-depleted areas. Also well represented in this *grey* module are genes encoding proteins of unknown function containing a nif11-like leader peptide domain (CK_00050707, CK_00047059, CK_00033157, CK_00002199), one of which being present in high copy number in CRD1 genomes (up to 53 in BIOS-E4-1, CK_00050707). Finally, a second copy of the cyanate hydratase (*cynH2*), specific to CRD1 genomes, and an accessory P-binding protein of the P transporter (*pstS2*) are also present in this module.

The *purple* module, which contains genes that are enriched in regions dominated by ESTUs IIIA, WPC1A and 5.3A, such as the P-depleted Mediterranean Sea (Figure 35C and Figure 36A), contains two genes related to phosphate metabolism, *sphX* (a cyanobacterial-specific P-binding component of the phosphate transporter) and *phoX* (an alkaline phosphatase; Dataset 2). This module also contains two genes involved in fatty acid biosynthesis, *fabF2* and the desaturase *desA3*, suggesting that these communities might have a particular lipid composition, as well as a number of genes involved in response to excess of light or oxidative stress (*ocp*, *frp*, *crtW* and one *hli*).

The *black* module, associated with cold, nutrient-rich environments contains a number of genes involved in photosynthesis, in particular those involved in type IVA chromatic acclimation, confirming that this process is mostly found in temperate to cold environments (Grébert *et al., in revision*; Xia *et al.*, 2017) (Dataset 2). This module also contains a few genes related to iron metabolism (CK_00038776 , encoding a CutA1 divalent ion tolerance family protein, a copy of the flavodoxin gene *isiB* and a copy of the ferredoxin gene CK_00008099), supporting the hypothesis drawn from experimental studies that coastal strains exhibit a higher tolerance to fluctuating metal concentrations than oceanic strains (Stuart *et al.*, 2009; Mackey *et al.*, 2015).

Finally the *turquoise* module, whose genes are associated with warm, Fe-replete environments dominated by ESTU IIA, contains genes involved in P and N uptake and assimilation such as the sensor and regulator of P stress *phoBR*, the gene encoding asparagine synthase (*asnB*) and the one encoding aspartate carbamoyltransferase (*pyrB*), the latter two being known to be downregulated in response to P stress (Tetu *et al.*, 2009), as well as two genes potentially involved in organic N scavenging, *cdd* (cytidine deaminase) and *amaB* (N-carbamoyl-L-amino acid amidohydrolase). Since this module is positively correlated to Fe concentration, its genes also show a relatively low presence in Fe-depleted regions. In this context, it is interesting to note the high number of genes (at least 12) encoding proteins using Fe or ferredoxin as a cofactor, among which the succinate dehydrogenase/fumarate reductase *sdhABC*, the glycolate oxidase genes *gclEF*, the ferredoxin-thioredoxin reductase *ftrV*, the pheophorbide *a* oxygenase *pao*, the plastoquinol terminal oxidase *ptox* and two ferredoxins *petF* (Dataset 2). Thus, this could indicate a reduction of iron quotas in *Synechococcus* communities thriving in Fe-depleted regions through the loss of genes interacting with iron, as was suggested for *Prochlorococcus* HLIII/IV (see below, (Rusch *et al.*, 2010)).

Besides these genes potentially important for adaptation to abiotic constraints, it is important to note that every modules contain a number of genes either originating from phages or coding for defense mechanisms, thus emphasizing the potential role of biotic interactions in shaping *Synechococcus* communities and distribution, as well as in the diversification of this genus.

Global distribution of *Prochlorococcus* genes sheds light on its adaptive strategies

The same analysis was performed with *Prochlorococcus* genes, which formed 3 main modules in correlation network that can be further divided in 6 sub-modules (Figure 37), showing more precise relations to ESTUs and environmental parameters. The *brown* module, which was not subdivided into sub-modules, was strongly correlated to nutrient concentrations, in particular nitrates, and strongly anti-correlated with iron concentration (Figure 38). This module thus corresponds to genes found preferentially in HNLC regions (Figure 37A,), and is most strongly associated to ESTUs thriving in these areas: the well-known HLIIIA and HLIVA ESTUs (Rusch *et al.*, 2010) but also the LLIB ESTU, previously found to dominate the LLI population in surface iron-limited HNLC areas (Farrant *et al.*, 2016) (Figure 37C-D). This module comprises a putative Mn2+/Fe2+ transporter (CK_00001683) as well as 3 genes forming an iron siderophore/vitamin B12 transporter (*febBCD*), most likely used for the uptake of complexed iron (Hopkinson and Morel, 2009) (Dataset 4). Interestingly the latter genes were not present in HNLC1 and HNLC2 strains (representative of ESTUs HLIIIA and HLIVA), which have been assembled from metagenomes and may thus be incomplete, but were recovered from single cell sequencing (Malmstrom *et al.*, 2012), showing that the metagenomic approach used here is as powerful as single cell analysis to detect genes associated to specific environmental parameters. This module also comprises genes coding for a transporter of acidic and neutral amino-acid (*natF* and *natG*) and a leucine dehydrogenase (*leuDH*, previously found to be associated with the Equatorial Pacific Ocean (Kent *et al.*, 2016)), which might indicate the use of organic nitrogen by *Prochlorococcus* communities inhabiting HNLC areas.

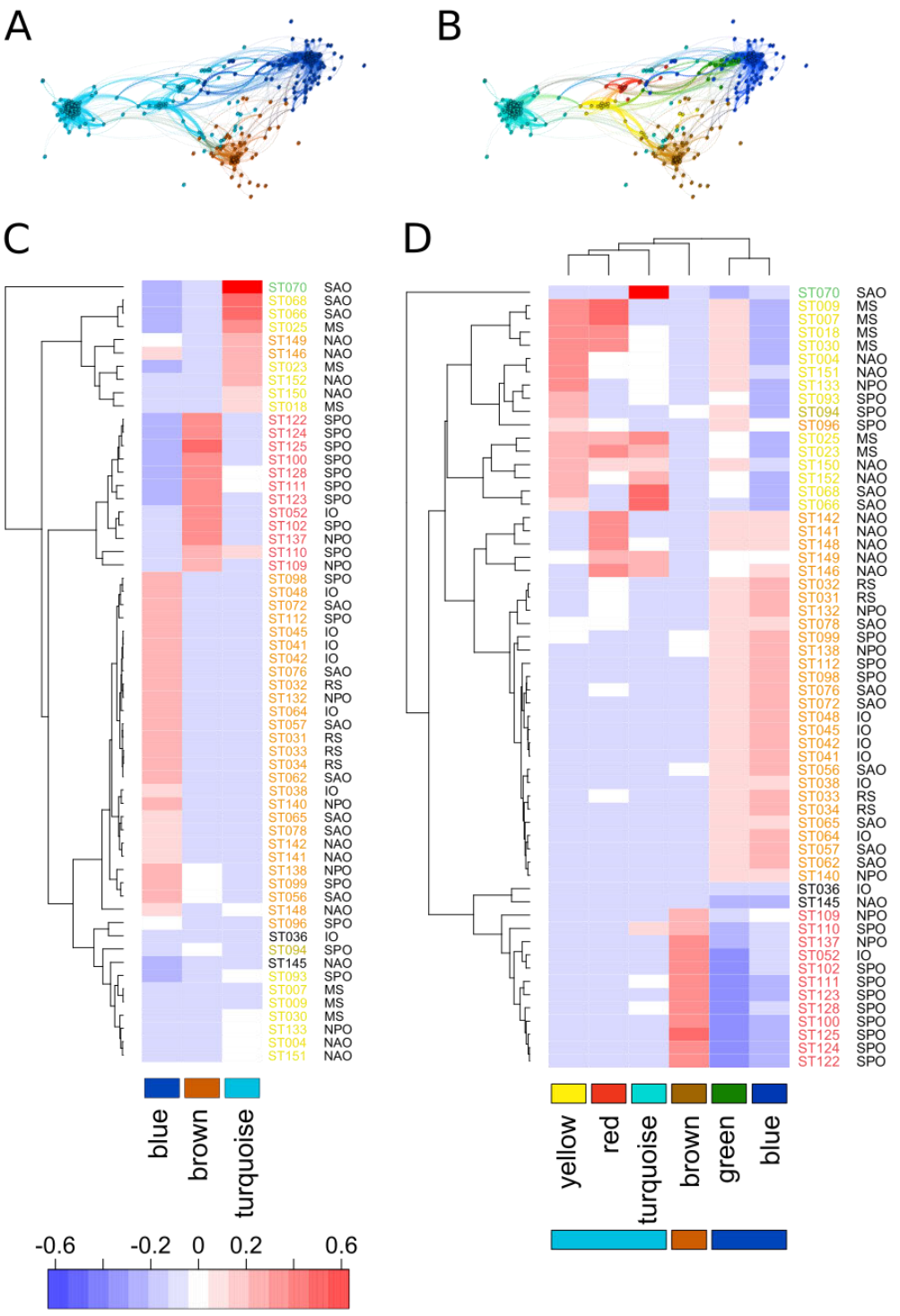


Figure 37 Correlation network based on relative abundance profiles of *Prochlorococcus* genes and decomposition into modules. A-B. Weighted correlation network based on flexible genes relative abundance profiles in *Tara* Oceans stations. The network contains three modules (A) that can be further decomposed into 6 submodules (B). C-D. Eigengenes representative of each module, designated by a color name. The eigengene of a module represents a consensus of the relative abundance profiles of genes within this module. Eigengenes of modules (C) and submodules (D) show the normalized abundance of genes in *Tara* Oceans stations, which are colored according to the clusters defined by (Farrant et al., 2016). Stations that were not analyzed in the latter study because of a low signal (ST036 and ST145) are colored in black. The oceanic region of each station is also specified: SAO, Sout Atlantic Ocean; MS, Mediterranean Sea; NAO, North Atlantic Ocean; IO, Indian Ocean; RS, Red Sea; SPO, South Pacific Ocean; NPO, North Pacific Ocean.

The *blue* sub-module is associated to warm, oligotrophic regions with low nitrate and nitrite concentrations where ESTUs HLIIA/B/D dominate *Prochlorococcus* community (Figure 38C-D). Aside from genes of unknown function, this sub-module comprises many genes involved in N uptake, including 12 genes necessary for nitrate and nitrite uptake (*nrtP*, *narB*, *narM*, *nirA*, *moaABCE*, *mobA*, *moeA* and two genes encoding a putative ABC transporter for nitrate CK_00002250 and CK_00002249) (Flores *et al.*, 2005), a nitrilase (*merR*, hydrolysis of nitrile groups) as well as an agmatinase (*speB*, degrades agmatine and products urea) (Dataset 4). Four out of these 14 genes were found to be negatively associated with the Equatorial Pacific Ocean by (Kent *et al.*, 2016), and we show here that these N-uptake genes are actually preferentially present in regions where nitrate and nitrite concentrations are among the lowest in our dataset.

The *green s*ub-module also comprises genes that, like the blue sub-module, are significantly correlated to ESTUs HLIIA/B/D (Figure 38C), but its eigengene shows that these genes are also present at stations dominated by ESTU HLIA/B (station names in yellow in Figure 37D). In comparison to the previous one, this sub-module is even more anti-correlated to nitrate and nitrite concentrations and strongly correlated to iron concentration (Figure 38D), a correlation pattern at the opposite of that observed for the eigengene of the brown module. Thus, these genes might be specifically missing from HNLC regions, as also suggested by the anti-correlation of the green sub-module to ESTUs HLIIIA and HLIVA (Figure 38C). Annotated genes in this sub-module are mostly involved in photosynthesis, including three genes coding for different ferredoxins (CK_00008097, CK_00001299 and CK_00000571), which could be a strategy to reduce iron quotas as suggested by (Rusch *et al.*, 2010). Another possible example of this strategy is that this module also comprises several small subunits of the PSII reaction center (*psbF1*, *psbY*, *psbT*, *psbK*) and components of cytochrome $b_{6/f}$ (*petL*, *petN*), which interact with Fe-containing hemes, as well as the gene encoding the fatty acid desaturase DesA3, which uses Fe as a cofactor (Dataset 4).

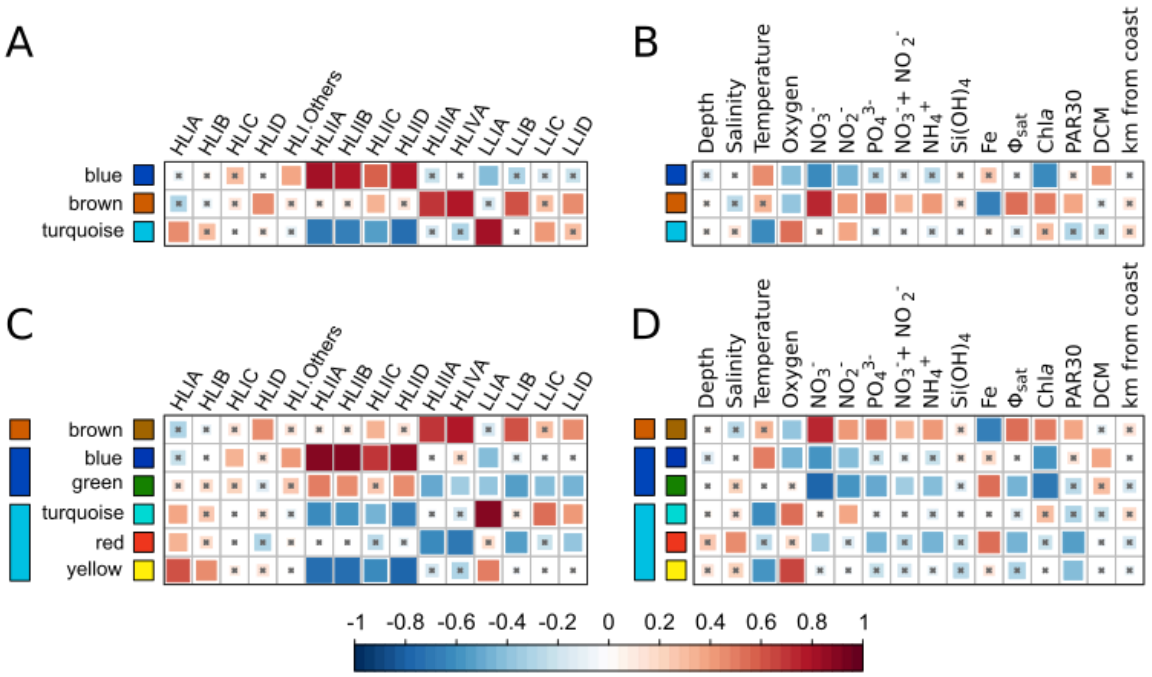


Figure 38 Correlation of *Prochlorococcus* module *eigengenes* to ESTU abundance and physico-chemical parameters. Correlation of module eigenges to relative abundance profiles of *Prochlorococcus* ESTUs (A) and to physico-chemical parameters (B), and correlation of sub-module eigengenes to ESTU relative abundance profiles (C) and to physico-chemical parameters (D). Spearman (A,C) and Pearson (B,D) correlation coefficient (R^2) is indicated by the color scale. Non-significant correlations (Student asymptotic p-value > 0.01) are marked by a cross. Φ_{sat} : index of iron limitation derived from satellite data. PAR30: satellite-derived photosynthetically available radiation at the surface, averaged on 30 days. DCM: depth of the deep chlorophyll maximum.

The *turquoise* sub-module is clearly associated to ESTU LLIA (Figure 38C), which in this dataset was found in rather cold and nitrite-rich waters (Figure 38D). This sub-module comprises by far the highest number of genes (860 out of 2044 genes), which is likely due to the fact that LLI genomes are larger than HL ones (Kettler *et al.*, 2007). Accordingly, it encompasses many genes that are present only in LLI genomes (Datasets 3 and 4). As we focused on samples from surface waters, our dataset is not well suited to study the niche-related genes of LL ecotypes, so we did not analyze this sub-module further.

The *red* sub-module is a small module of 64 genes associated to waters of high salinity with a relatively high Fe concentration and low concentrations of ammonium and phosphate (Figure 38D) corresponding in our dataset to stations of the Mediterranean Sea and of the North Atlantic Ocean (Figure 37C). Noteworthy, this sub-module is not associated to any ESTU, with only a weak correlation with ESTU HLIA (Figure 38C). Most of the annotated genes in the *red* sub-module are involved in phosphate acquisition: several genes encoding phosphonate transporters (*phnCDE*) and a phosphonate dehydrogenase (*ptxD*), the P stress sensor *phoR*, the phosphate starvation inducible protein *psip1*, an arsenate efflux pump (*arsB*), two putative phosphatases (CK_00003430, CK_00056879) and the chromate transporter *chrA* (Dataset 4). Some of these genes were previously shown to be associated with regions of low phosphate availability (Martiny *et al.*, 2009b; Kent *et al.*, 2016). Thus, these results tend to support the idea that *Prochlorococcus* P-

acquisition genes are not linked to a particular ecotype, but are rather found in different ecotypes in P-depleted areas. (Martiny *et al.*, 2006; Coleman and Chisholm, 2010; Kent *et al.*, 2016).

Finally, the last sub-module (*yellow*) is clearly associated with waters of low temperature and with ESTUs HLIA/B (Figure 38C-D). This sub-module is strikingly small (103 genes) compared to those associated to other major ESTUs, despite the fact that ESTU HLIA is quite abundant in our dataset and dominating *Prochlorococcus* community in various oceanic areas (North Atlantic Ocean, Mediterranean Sea, South Atlantic Ocean, Chili upwelling, North Pacific Ocean, see (Farrant *et al.*, 2016)). Only 16 of these genes have an annotation (Dataset 4), and most of them have an elusive function, but two were previously identified as associated to low temperature in *Prochlorococcus* (CK_00000399 encoding a protein of the 2OG-Fe(II) oxygenase superfamily and CK_00000656, a putative dihydroorotase, (Kent *et al.*, 2016)), although the adaptive potential of these genes with regard to temperature remains to be confirmed. Thus, altogether our results confirm that only few genes seem to be associated with low temperature in *Prochlorococcus* (Kettler *et al.*, 2007; Kent *et al.*, 2016).

2.3 Discussion

The global transect of the *Tara* Oceans expedition, which encompasses regions with various combinations of physico-chemical parameters, and the very deep sequencing of its metagenomes, make this dataset particularly well suited to study niche adaptation *in situ* of ubiquitous and highly abundant organisms such as the marine picocyanobacterial *Synechococcus* and *Prochlorococcus.* These assets, combined with the availability of a reference database comprising almost hundred genomes, representative of the wide diversity existing within both genera, allowed us to get a good genome coverage in all oceanic basins, a prerequisite for studying gene abundance. Despite the fact that we can only detect genes that are represented in our genome database, and thus have not access to the whole picocyanobacterial pangenome (see chapter II part 1), we showed that there is a striking correspondence between the distribution of ESTUs, as assessed by a single, high-resolution, core marker gene, and the distribution of genes constituting the flexible genome of these populations. This strong correlation between taxonomy and gene content strengthens the idea suggested by (Kent *et al.*, 2016) for *Prochlorococcus* and confirmed in the first part of this chapter from picocyanobacterial comparative genomics analysis (see chapter II part 1), that the evolution of genomes is mainly driven by vertical transmission in both genera, with a low extent of lateral gene transfer. This is particularly true for *Synechococcus*, for which every modules and sub-modules were strongly correlated to some ESTUs (Figure 36).

In *Prochlorococcus,* the flexible gene content of the different ESTUs was strongly associated with distinct combinations of physico-chemical parameters. Examination of the functions of genes specific of nutrient poor areas not only confirmed previous findings such as the presence of N- and P-uptake genes in genomes of *Prochlorococcus* cells found in areas depleted in these nutrients (Martiny *et al.*, 2006, 2009a; Berube *et al.*, 2014; Kent *et al.*, 2016) but also provided several novel insights concerning the adaptive strategies adopted by the ESTUs colonizing these niches since many more potential niche related genes were identified in the present study. The distribution

pattern with regard to nutrient was less clear for *Synechococcus*, since correlations or anticorrelations of modules to N or P concentrations were much weaker than for *Prochlorococcus,* although we found some genes involved to P metabolism to be enriched in P-deplete regions. While we cannot exclude that this low signal might be due to the module delineation (i.e. smaller modules might unveil new patterns), it more likely reflects that all *Synechococcus* ESTUs exhibit a fairly similar gene repertoire to deal with macronutrients compared to the streamlined *Prochlorococcus* genomes (Scanlan *et al.*, 2009).

In contrast to macronutrients, it has been hypothesized that survival of *Prochlorococcus* clades HLIII and HLIV in low-Fe regions was made possible through the elimination from the genomes of genes coding for proteins that contain iron as a cofactor (Rusch *et al.*, 2010; Kent *et al.*, 2016) as well as the acquisition of iron-scavenging genes (Malmstrom *et al.*, 2012). Accordingly, many proteins interacting with iron were indeed found to be anti-correlated to HNLC regions in our dataset and these stations were also found to be enriched in *Prochlorococcus* iron-scavenging genes. Thus, these results support the hypothesis that adaptation of *Prochlorococcus* HLIII/IV to Fe-deplete environments implies both a reduction of intracellular iron quotas and the use of efficient iron-scavenging systems (Malmstrom *et al.*, 2012). A similar strategy seems to have been adopted by *Synechococcus* communities found in Fe-deplete (HNLC) regions, mainly encompassing CRD1 and EnvB ESTUs, since we found several Fe transporter and storage genes positively associated to these regions, and many Fe-containing or Fe-interacting proteins negatively associated, often corresponding to genes absent in some or all available CRD1 genomes but present in most others (note that no genome representative of the EnvB clade is available yet). In both *Prochlorococcus* and *Synechococcus*, genes that have been eliminated by selection also encompass genes coding for proteins located in photosystem II reaction center or involved in electron transport, showing that adaptation to iron scarcity might also impact this highly conserved and central metabolic process for cyanobacterial life. This finding is also quite consistent with what has been observed by comparative genomic analysis in diatoms adapted to Fe scarcity (Lommer *et al.*, 2012), in which photosynthesis is remodeled to lower iron quotas.

Besides nutrient acquisition and metabolism, we also found in both *Prochlorococcus* and *Synechococcus* a module associated to low temperature. As mentioned before, the size of this module was strikingly small in *Prochlorococcus,* suggesting that few genes are involved in the adaptation to low temperature in the cold-adapted ecotype (HLI). Moreover, genes of this module, when they were annotated, gave little information on how they might play a role in adaptation to cold waters. Even though more genes were unveiled in *Synechococcus*, some of the few annotated ones were rather related to metal tolerance than to a direct adaptation to cold. Still, a number of these genes were coding for proteins involved in photosynthesis, suggesting a possible remodeling of this process in cold environmental conditions, as suggested by previous transcriptomic and photophysiological studies (Mackey *et al.*, 2013; Pittera *et al.*, 2014; Varkey *et al.*, 2016; Pittera *et al.*, 2016). Altogether, the low number of genes related to temperature adaptation tend to supports the hypothesis that adaptation to cold temperature is not mediated by evolution of gene content but rather by the evolution of protein sequences (Kettler *et al.*, 2007; see chapter II part 1).

As noted above, gene distribution is strongly correlated to the distribution of taxonomic units both in *Synechococcus* and *Prochlorococcus*. However, at least one module of *Prochlorococcus* genes, containing mostly genes involved in phosphate uptake, is not linked to a particular ESTU, suggesting a potential independent acquisition by lateral transfers in multiple populations. Thus, some genes might sweep into populations independently of the host genome and behave as entities having their own ecological niche (Polz *et al.*, 2013), though it seems to be limited to particular functional traits. Such a module of genes was not found in *Synechococcus*, in which each module and sub-module was found to be strongly correlated to the abundance of a set of ESTUs. This might reflect a difference in the evolutionary dynamics between *Prochlorococcus* and *Synechococcus*, the latter being less subjected to an evolution by acquisition of beneficiary genes. Alternatively, it might be possible to find genes dissociated from ESTUs by further splitting modules of genes and/or by adding other environmental parameters. Moreover, the signal contained in the distribution of *Synechococcus* genes is blurred by a huge noise of genes of unknown function. While some of these genes might carry an adaptive potential, many are probably of little or no beneficial use. This problem is less prominent in *Prochlorococcus,* in which genome streamlining likely frees more rapidly the genome from such genomic burdens.

An important selective force that could not be fully assessed with our dataset is the biotic pressure exerted by viruses and grazers. Our results show that each picocyanobacterial community, in particular in *Synechococcus*, has niche-specific genes originating from phages or involved in cell-surface modification or other cell defense mechanisms, including bacteriocin-type genes and toxin/antitoxin systems that are indicative of potential allelopathic interactions with other bacteria (Paz-Yepes *et al.*, 2013). These genes are often present in only one or a few sequenced genomes and are thus likely responsible for a significant part of the gene content diversity of marine picocyanobacteria. This highlights the role of biotic interactions, in particular with cyanophages, in shaping the global distribution of marine picocyanobacteria and warrants future studies of co-occurrence at the global scale.

In conclusion, our analysis of picocyanobacterial gene distribution in the global ocean reveals that each community has a specific gene repertoire, suggesting that genes of the picoyanobacterial pangenome might not be available to every picocyanobacterial cells. Despite the tremendous amount of genes of unknown function, patterns of gene distribution clearly revealed some of the adaptive strategies that allowed picocyanobacteria to colonize almost all niches in the lit layer of the ocean, while it failed to explain the adaptation to some environments, in particular to cold temperatures. In order to further understand niche adaptation, future studies should focus on the distribution of sequence variants, on the acclimation response of each community by examining corresponding metatranscriptomes and on the impact of biotic interactions, three research paths that promise important new discoveries.

Supplemental information

Dataset 1: Composition of *Synechococcus* gene modules. Only genes whose profile shows a correlation coefficient higher than 0.5 with the *eigengene* of its sub-module are shown. The first column shows the Cyanorak cluster number, and the second and third show the module and submodule to which the gene belongs. Correlation to the eigengene of each sub-module is displayed in the 19 next columns, followed by gene name (when available) and gene product. After the product, COG annotation and a custom hierarchical annotation are indicated, followed by supplemental comments. Finally, the 97 last columns indicate the phyletic pattern of each gene, i.e. the number of copies of the gene in every genomes of the database.

Dataset 2: Composition of *Synechococcus* gene modules, without hypothetical genes. Same as Dataset 1, but without hypothetical genes. Genes were kept only if some information about its function was given.

Dataset 3: Composition of *Prochlorococcus* gene modules. Same as Dataset 1, but for *Prochlorococcus,* with only 6 sub-modules.

Dataset 4: Composition of *Prochlorococcus* gene modules, without hypothetical genes. Same as Dataset 2, but for *Prochlorococcus,* with only 6 sub-modules.

3. Response of picocyanobacterial communities to a natural iron enrichment

Context of the work

The analysis of global distribution of picocyanobacterial communities in the first chapter revealed the clear partitioning of ecologically significant taxonomic units (ESTUs) in the oceans. In the previous part, we unveiled some of the adaptations underlying this niche partitioning, with a particularly clear trend concerning adaptation to Fe-replete *versus* Fe-depleted conditions. In collaboration with D. Iudicone (Stazione Zoologica Anton Dohrn, Naples) and other members of the *Tara* consortium (Caputi *et al*., *in prep*, ANNEXES 2), we explored in more details the response of picocyanobacterial communities to iron enrichment *in situ*. This work is part of a larger project, which aimed at exploring the response of the whole planktonic ecosystem to a dynamic natural perturbation related to iron bioavailability, by focusing on a few stations of the *Tara* Oceans transect that we briefly mentioned in our global analysis of picocyanobacterial distribution as showing a shift in *Synechococcus* community (Farrant et al., 2016) and located near the Marquesas archipelago (Figure 39). During the *Tara* Oceans expedition, this area was selected because of the plankton blooms that are visible from space. Although iron concentrations have never been measured around the archipelago, these blooms are believed to be the result of iron injection coupled with altered circulation due to an island mass effect (Martinez and Maamaatuaiahutapu, 2004). A first station was sampled in the vicinity of the main island (ST123), and a second in one of the small eddies formed downstream of the island (ST124, Figure 39). Because the chlorophyll-enriched patch extended over 400 km to the west of the islands (Figure 39), an additional station was sampled in the bloom area corresponding to about 2/3 weeks downstream from the islands (ST125). These sites were compared to an HNLC site upstream of the islands, ST122) where chlorophyll levels were lower than those observed inside the bloom (Figure 39).

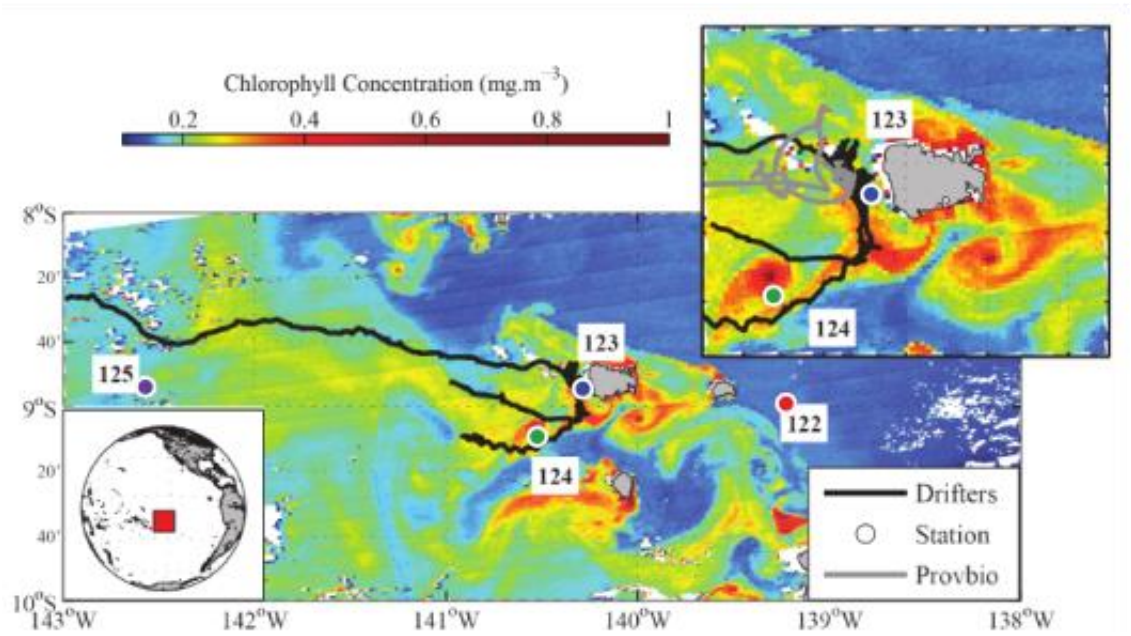


Figure 39 Context and location of sampling stations near the Marquesas Islands. Locations of the four sampling stations are indicated by colored dots. The phytoplankton bloom is clearly visible from satellite-derived measures of chlorophyll concentration. Station 122 is located at the limit of the oligotrophic gyre, with conditions typical of an HNLC area. Station 123 is located close to the main island and is strongly influenced by the island vicinity. Station 124 is located in one of the small eddies formed downstream of the islands (current is oriented from West to East). Finally station 125 is located at the western extremity of the bloom area. Figure from Caputi et al., *in prep* (Annexe 2).

As part of this project, I was involved in the analysis of the response of picocyanobacterial communities to perturbation/iron enrichment, and had access to both metagenomic and metatranscriptomic data, which allowed me to analyze changes in both community structure and gene expression.

Methods

Functional assignment of metagenomic and metatranscriptomic reads was performed using the clusters of orthologous genes (CLOGs) defined and annotated in the information system Cyanorak V2 (http://sb-roscoff.fr/Phyto/cyanorak), which comprises 97 genomes of marine picocyanobacteria. Metagenomic and metatranscriptomic reads were cleaned as described in (Farrant *et al.*, 2016) and recruited against the database of 97 genomes of marine picocyanobacteria using BLASTN (v2.2.28+) (Altschul *et al.*, 1990) with default parameters but limiting the results to one target sequence (--max_target_seqs 1) and keeping only results with an E-value below 0.001 (-evalue 0.001). Reads having less than 90% of their sequence aligned or showing less than 95% identity with their best hit were filtered out, and remaining reads were assigned to a CLOG according to the position of their BLAST match in the genome: a read was assigned to a gene if at least 90% of its size was aligned inside this gene. A second BLASTN round was then run for each CLOG with a database containing all *Prochlorococcus* and *Synechococcus* sequenced of this CLOG, and results were then parsed using the Lowest Common Ancestor method (Huson et al., 2007): for each read, blast matches with over 95%ID 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. the strain, subclade, clade, subcluster or genus *sensu* (Mazard *et al.*, 2012a)).

Finally, in each station read counts were aggregated by CLOG and by genus and normalized by average gene length of the CLOG and by the total read abundance to obtain relative abundance profiles. In order to estimate the level of gene expression, read counts obtained from metatranscriptomic data were divided by read counts obtained from metagenomic data, resulting in a proxy of the number of transcripts per number of copies of the gene.

Taxonomic and functional response of the picocyanobacterial community

Metagenomic and flow cytometry analyses indicate that the picocyanobacterial community displayed large variations of composition in the Marquesas Islands area. While *Prochlorococcus* was highly dominant over *Synechococcus* at Stations 122, 124 and 125, it was almost absent from Station 123, and all four stations were co-dominated by clades HLIII and HLIV, known to thrive in HNLC regions (Rusch *et al.*, 2010; West *et al.*, 2011) (Figure 40). In contrast, *Synechococcus* was almost absent from Station 122 and bloomed in the three stations located downstream of the islands (123-125), concomitant with a drastic shift in clade composition from a dominance of clade CRD1 (the major clade in HNLC regions (Sohm *et al.*, 2015; Farrant *et al.*, 2016)) to a dominance of clade II (a clade known to prevail in warm Fe-replete intertropical areas) (Figure 40). This difference between *Prochlorococcus* and *Synechococcus* is interesting because it suggests that *Prochlorococcus* is less responsive to environmental perturbations and that its community does not contain opportunistic populations able to rapidly take over others as is observed with *Synechococcus* ESTU IIA. This behavior of *Prochlorococcus* community echoes the one observed by (Chandler *et al.*, 2016) in a very different context: they observed an “ecotypic lag”, while following a *Prochlorococcus* community in an environment of changing temperature, i.e. a lag between the environmental change and the change in community composition. Thus, the lack of community responsiveness could be a characteristic of *Prochlorococcus* (as compared to *Synechococcus*).

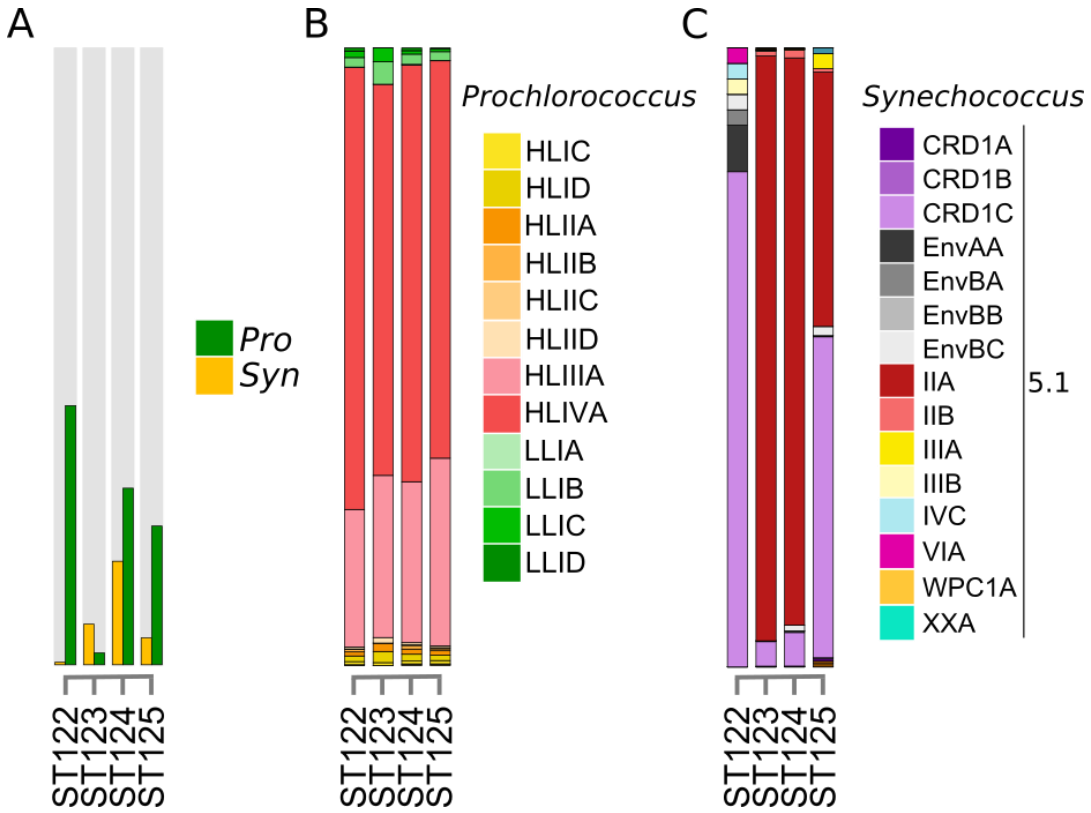


Figure 40 Taxonomic composition in the four sampling stations of the project. A. Absolute abundance (normalized by sequencing depth) of *Prochlorococcus* (green) and *Synechococcus* (orange). B. Relative abundance of *Prochlorococcus* ESTUs. C. Relative abundance of *Synechococcus* ESTUs, showing the drastic shift in community between HNLC station ST122 and iron-enriched stations ST123 and 124.

Metatranscriptomic analyses further revealed dramatic differences in the functional response of both organisms, despite their close phylogenetic relatedness (Table 1, metaT/metaG ratio). The differential expression of genes related to iron metabolism indicates that *Synechococcus* was responsive to island vicinity at Stations 123-125, in a way that supports an injection of iron into the ecosystem. This includes the replacement of the Cu-containing flavodoxin (*isiB,* downregulated 10-fold at station 123) by the Fe-containing ferredoxin (*petF;* upregulated 2-fold in stations 123-125), induction of the *petJ* gene, encoding cytochrome c_6, which was only detected at Stations 123-125 (although its Cu-containing counterpart, plastocyanin (*petE*), was also upregulated at these stations), or the replacement of the gene encoding fructose bisphosphate aldolase class I (*fba*) by the metal-dependent FBA class II (*fbaA*) (Table 1). In contrast, the iron-stress induced gene *isiA* was more expressed at the HNLC Station 122, which could help compensate for the reduction in the number of photosystem I complexes known to occur under iron limitation (Burnap *et al.*, 1993; Boekema *et al.*, 2001). Finally, *Synechococcus* also seemed to be able to efficiently scavenge and store iron in island-influenced stations, as suggested by the upregulation of genes encoding numerous iron transporters (*sufC*, *sufD*, *feoB*, *idiB* and an unnamed iron transporter of the NRAMP family) and for the iron-storage protein ferritin (*ftn*; upregulated > 10-fold) in iron-enriched stations (Table 1). In contrast, for *Prochlorococcus* several of these key genes (e.g., *petJ*, *fbaA, isiA*) could not be detected in meta-omics datasets nor in HLIII/IV genomes (assembled from metagenomes)

available to date (Rusch *et al.*, 2010) and, when present, only a slight differential expression was observed between Station 122 and the three stations under island influence (Table 1). Consistent with the results obtained from the GOS dataset (Pierella Karlusich *et al.*, 2015), no negative correlation between the presence of flavodoxin and iron bioavailability could be observed for *Prochlorococcus,* and ferredoxin encoding genes (*petF*) were only expressed at a low level at all four stations. Similarly, the expression levels of genes involved in iron storage (ferritin, *ftn*) and transport (*sufC, sufD, feoB,* etc.) remained fairly constant in all stations, suggesting that *Prochlorococcus* was not able to increase its iron-uptake at stations 123-125. Thus, in contrast to *Synechococcus*, no clear acclimation process to iron availability was detected for *Prochlorococcus.* Furthermore, analysis of the mapping intensity of metagenomic reads (Table 1, metaG) also provides insights into the differential adaptation of communities present at strictly HNLC *vs.* stations influenced by the Marquesas Islands. In particular, several *Synechococcus* genes involved in iron metabolism (*fbaA*, *isiA*, *ftn*, and the transporters *sufC*, *sufD*, *feoB*) show a higher mapping intensity of metagenomic reads at station 122 than at stations 123-125, suggesting that the community present at this HNLC station (dominated by clade CRD1) has a higher copy number of these genes in their genome compared to those found in island-influenced stations (dominated by clade II). In contrast, some genes were seemingly more abundant (or only detected) at stations 123-125 (*petJ*, *fda* and the NRAMP-like transporter). This suggests that the two communities display distinct genetic potentials, and reveals a possible adaptation of clade CRD1 to iron-limited (HNLC) areas, and/or an adaptation of clade II to iron-richer conditions, as suggested by recent studies (Sohm *et al.*, 2015; Farrant *et al.*, 2016).

In conclusion, the combination of metatranscriptomics and metagenomics allowed us to decipher both an adaptive response (illustrated by the change of *Synechococcus* community) and an acclimation response (i.e. a change in cell physiology) in reaction to a natural environmental perturbation. From our results, it seems that the iron-enrichment induced by the vicinity of the Marquesas Island provoked a dramatic shift in the *Synechococcus* community, from a typical HNLC population at station 122 to a population more adapted to iron-replete conditions at stations 123-125, associated with a strong transcriptomic response to iron enrichment, while *Prochlorococcus* community remained constant and seemingly failed to acclimate to the iron-replete conditions. Altogether, these differences in both adaption and acclimation responses might explain the opportunistic behavior of *Synechococcus* (and in particular clade II) in these iron-enriched waters

Table 1 Mapping intensity of metagenomic and metatranscriptomic reads assigned to *Prochlorococcus* and *Synechococcus* genes involved in iron metabolism.

			Prochlorococcus										*Synechococcus*									
			MetaT				MetaG				MetaT/MetaG				MetaT				MetaG [2]				MetaT/MetaG			
Category	Gene [1]	Product	122	123	124	125	122	123	124	125	122	123	124	125	122	123	124	125	122	123	124	125	122	123	124	125
Flavodoxin/ Ferredoxin	*isiB*	flavodoxin	80,57	19,52	40,64	28,17	4,84	5,00	4,93	4,89	16,66	3,901	8,247	5,765	0,98	0,15	1,69	0,37	1,60	1,84	2,03	1,96	0,614	0,079	0,832	0,188
	petF (11)	ferredoxin	2,41	2,35	3,26	2,36	20,57	22,40	21,27	21,41	0,117	0,105	0,153	0,110	40,45	143,8	145,0	71,35	32,70	52,44	52,01	52,31	1,237	2,743	2,787	1,364
Plastocyanin/ cytochrome c6	*petE*	plastocyanin	8,87	10,16	20,14	12,84	5,95	5,69	6,30	6,47	1,492	1,785	3,198	1,985	8,68	21,94	64,97	40,76	2,36	3,80	3,69	3,65	3,684	5,771	17,60	11,177
	petJ	cytochrome c553 (c6)	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	n.a	n.a	n.a	n.a	0,00	1,24	1,99	1,88	0,00	5,54	4,86	4,18	n.a	0,223	0,410	0,450
FBAI/FBAII	*fbaA*	fructose-1,6-bisP aldolase class II	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	n.a	n.a	n.a	n.a	3,42	22,85	18,57	8,89	5,04	3,38	3,09	3,12	0,680	6,753	6,000	2,845
	fda	fructose-1,6-bisP aldolase class I	22,00	21,51	13,82	24,96	6,80	7,03	6,88	6,53	3,236	3,062	2,007	3,822	1,53	0,36	0,94	0,86	1,95	4,14	4,12	4,04	0,786	0,087	0,227	0,212
IsiA	*isiA*	iron stress-induced chl-binding protein	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	n.a	n.a	n.a	n.a	6,33	0,15	2,29	2,45	**14,3**	**2,63**	**2,87**	**4,34**	0,442	0,056	0,798	0,564
SOD	*sodN*	superoxide dismutase [Ni]	16,00	12,99	17,17	20,60	7,60	7,35	7,64	7,37	2,105	1,766	2,247	2,794	20,25	21,08	22,50	16,42	1,83	3,63	3,63	3,88	11,05	5,814	6,207	4,232
	sodC	superoxide dismutase [Cu-Zn]	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	n.a	n.a	n.a	n.a	0,00	1,94	1,92	2,06	0,00	4,11	4,30	4,40	n.a	0,472	0,447	0,469
Storage	*ftn*	ferritin	6,92	9,95	20,36	13,79	4,11	3,80	4,16	4,48	1,682	2,616	4,889	3,076	54,96	75,16	95,30	81,09	**11,8**	**1,22**	**1,20**	**1,29**	4,651	61,44	79,16	62,76
Transport/ Assimilation	*fepD*	Fe transport system, permease component	0,08	0,02	0,01	0,01	1,21	1,21	1,20	1,18	0,070	0,015	0,011	0,007	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	n.a	n.a	n.a	n.a
	sufC	ABC transporter involved in Fe-S cluster assembly, ATPase component	0,81	0,59	0,62	0,63	5,28	4,64	5,41	5,09	0,154	0,126	0,115	0,124	1,04	3,03	2,74	3,94	**7,91**	**4,43**	**4,57**	**4,44**	0,131	0,685	0,599	0,888
	sufD	ABC transporter involved in Fe-S cluster assembly, permease component	0,22	0,27	0,20	0,21	4,74	5,70	5,04	4,93	0,046	0,047	0,039	0,042	0,42	1,69	1,72	1,89	**5,17**	**3,51**	**3,61**	**4,07**	0,082	0,481	0,477	0,465
	futC	ABC transporter ATP binding component	0,23	0,05	0,16	0,24	4,10	4,48	4,31	4,23	0,056	0,011	0,037	0,057	0,09	0,64	1,08	1,24	2,70	3,56	3,71	4,13	0,035	0,179	0,292	0,301
	idiB	ABC-type Fe^{3+} transport system, membrane component	0,12	0,10	0,09	0,12	5,01	5,16	5,01	4,70	0,024	0,019	0,018	0,026	0,00	0,63	1,94	2,46	2,17	3,69	3,75	3,69	0,000	0,172	0,517	0,666
	feoB	ferrous iron transporter	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	n.a	n.a	n.a	n.a	0,00	0,00	0,01	0,00	**0,26**	**0,01**	**0,00**	**0,01**	0,000	0,324	8,141	0,000
		Mn^{2+}/Fe^{2+} transporter NRAMP family	0,31	0,11	0,16	0,14	6,07	5,49	6,23	5,95	0,052	0,020	0,026	0,024	0,00	0,33	0,48	0,62	**0,00**	**3,55**	**3,37**	**3,36**	n.a	0,092	0,142	0,184
Oxidative Stress	*dpsA*	stress-inducible DNA-binding protein	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	n.a	n.a	n.a	n.a	0,00	1,65	2,65	4,67	5,42	0,43	0,63	0,53	0,000	3,810	4,200	8,868
		Catalase	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	n.a	n.a	n.a	n.a	0,00	0,00	0,00	0,00	**0,00**	**0,01**	**0,00**	**0,00**	n.a	0,000	0,000	n.a
	(3)	SOD	16,00	12,99	17,17	20,60	7,60	7,35	7,64	7,37	2,105	1,766	2,247	2,794	20,25	23,02	24,42	18,48	**1,83**	**7,73**	**7,93**	**8,28**	11,05	2,977	3,082	2,232
	(5)	Peroxiredoxin	7,72	13,27	17,72	15,68	23,93	22,32	24,94	24,51	0,323	0,594	0,711	0,640	4,70	22,29	22,13	17,15	11,76	15,54	15,61	15,73	0,400	1,435	1,418	1,091
		Glutathione peroxidase	1,18	2,32	1,72	4,49	5,48	4,59	5,80	5,70	0,215	0,505	0,297	0,788	0,00	9,64	7,21	2,61	1,74	3,86	3,59	3,58	0,000	2,497	2,009	0,730
	(2)	FTR system	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	n.a	n.a	n.a	n.a	6,32	21,26	20,32	22,95	**4,20**	**10,9**	**10,5**	**10,1**	1,504	1,957	1,939	2,276
	(3)	Thioredoxin	8,21	6,70	18,37	11,22	21,23	20,58	21,25	20,37	0,387	0,326	0,865	0,551	0,43	8,68	12,11	9,61	**6,04**	**11,5**	**11,3**	**11,5**	0,071	0,755	1,070	0,838
	(3)	Rubredoxin	3,81	4,82	7,43	6,55	6,79	6,94	7,13	6,74	0,561	0,695	1,043	0,973	1,18	3,97	3,64	6,12	5,54	9,38	8,96	8,03	0,213	0,424	0,406	0,763
Fe-containing proteins (8)			6,59	14,95	20,73	13,66	35,5	33,6	35,1	36,0	0,186	0,445	0,590	0,380	9,58	7,80	14,57	8,83	**2,23**	**21,7**	**22,0**	**20,5**	4,297	0,359	0,661	0,431

[1] The number between brackets indicates the number of genes taken in consideration for the category

[2] Red numbers indicate a notable change in the mapping intensity of metagenomic reads

CHAPITRE III

ECOTYPIC VARIABILITY IN RESPONSE TO VARIOUS PHYSIOLOGICAL STRESSES

Context of the work

The work presented in the first two chapters highlighted the wide genetic diversity of marine picocyanobacteria in general, and *Synechococcus* in particular. Both the analysis of *in situ* distribution of its lineages and the study of the genomic content of representative strains revealed a striking need for a better characterization of the physiological diversity of this genus on the one hand to experimentally confirm the hypotheses built from correlation to environmental parameters, and on the other hand to relate genomic variations to experimentally characterized traits. Indeed, most of the current knowledge of *Synechococcus* physiology is based on the study of a few model strains, whose representativeness with regard to the whole genus can be questioned. As the niche partitioning of the lineages of *Synechococcus* started to be uncovered, a number of studies have compared model strains considered as representative of *coastal* versus *open ocean* environments, based on their isolation site (Palenik *et al.*, 2006; Stuart *et al.*, 2009; Tetu *et al.*, 2009, 2013; Stuart *et al.*, 2013; Mackey *et al.*, 2015). These studies mostly focused on the response to trace metals shock or starvation, and suggested that coastal strains would be more resilient than oceanic strains when facing a stress thanks to a more developed ability to sense and response to environmental variability. More recently, and mostly during the course of my PhD, new research started to focus more broadly on the physiological differences between lineages with regard to temperature (Mackey *et al.*, 2013; Pittera *et al.*, 2014; Varkey *et al.*, 2016) or light (Six *et al.*, 2007a; Mackey *et al.*, 2017).

In the context of the SAMOSA ANR project (2014-2018), which focuses on ecotypic variability in *Synechococcus* both *in situ* and in controlled experimental conditions, we aimed at making a few more steps in the characterization of phenotypic variability, by analyzing the light and temperature *preferenda* of a dozen strains as well as their lipid composition, and the physiological response of the model strain WH7803 as well as 4 strains more ecologically representative of the diversity existing in the natural environment.

As concerns WH7803, the physiological experiments aimed at studying the variability of the response to a large range of stresses (high light, UV, low/high temperature) and on the comparison of two light acclimation states (low *vs*. high light). Additionally, since as we highlighted in the introduction, response to light variations and to changes in other environmental parameters are highly entangled, we also compared the physiology of the WH7803 strain acclimated to two different temperatures in light conditions that are close to those observed *in situ*, i.e. a modulated light-dark cycle. The aim was thus to analyze how this strain dealt with distinct light-temperature combinations.

As for the four ecologically representative *Synechococcus* strains, they were submitted to a more limited set of stresses (high-light and low-temperature shocks), with two goals in mind: first, to identify a core stress response of this genus to stress in general and to high-light and low-temperature stresses in particular, and second to identify differences in stress response between strains that could be related to the distinct niches occupied by these lineages *in situ*.

Given the heavy experimental plan and the amount of data to process, several members of MaPP group were involved in sample harvesting and in their analysis. I personally contributed to

the experimental work by preparing cultures and leading sampling in tight collaboration with Ngoc-An Nguyen for stress experiments and with Morgane Ratin for light-dark experiments. I have more particularly led the experimental set up and sampling for the 4 ecologically representative *Synechococcus* strains and have been involved in the analysis of all the results presented in this chapter.

1. Experimental procedures

1.1 Stress experiments

1.1.1 Choice of strains

The rationale of comparing physiological differences between strains was to unveil potential differences in phenotypic plasticity between lineages that could explain the distribution of *Synechococcus* populations. With this goal in mind, we selected 4 strains of *Synechococcus* representative of most clades I, II, III and IV, which were considered to be the most abundant lineages in the oceans at the beginning of my PhD. We also added strain WH7803, representative of the less abundant clade V, as it is a model strain on which a large number of physiological studies have been performed (Kana and Glibert, 1987; Sweeney and Borgese, 1989; Scanlan *et al.*, 1993; Garczarek *et al.*, 2008; Blot *et al.*, 2011), allowing a comparison with previous work. As the objective was to unveil differences between clades, all strains were chosen with the same pigment type (type 3, see INTRODUCTION) to limit variability in response to light. All strains were also chosen among those having a sequenced genome, in order both to facilitate transcriptomics analyses and to be able to integrate physiological and genomic data. Finally, when several strains in a clade met these criteria, we chose one representative of the dominant ESTU in the clade. The selected strains are listed in (Table 2).

Table 2 General information regarding *Synechococcus* strains used in this study

Strain	*MVIR-18-1*	*A15-62*	*WH8102*	*BL107*	*WH7803*
Clade	I	II	III	IV	V
ESTU	IA	IIA	IIIA	IVA	VA
Pigment type	3a	3dB	3c	3dA	3a
Genome accession	-	-	BX548020	AATZ00000000	CT971583
Axenic	No	No	Yes	No	Yes
Isolation site	Southern Norwegian Sea	Atlantic Ocean	Carribean Sea	Mediterranean Sea	Sargasso Sea
Isolation latitude	61°00’ N	17°37’N	22°29’N	41°43’N	33°45’N
Isolation longitude	1°59’ E	20°57’W	65°36’W	3°33’E	67°30’W
Isolation date	23/07/2007	04/10/2004	15/03/1981	15/03/2001	01/07/1978
Isolation depth (m)	25	15	-	1800	25
Isolation temperature	13.98°C	26.02°C*	25.78°C*	13.89°C*	25.85°C*

*Temperatures were retrieved from (Pittera *et al.*, 2014)

1.1.2 Experimental conditions

All strains were grown in modified PCR-S11 medium (Rippka *et al.*, 2000) with 1mM sodium nitrate and 1µg/L vitamin B12, based on seawater reconstituted from Red Sea Salts (Red Sea Aquatics Ltd) and distilled water. All strains were acclimated for at least two weeks at 21°C and 20 µmol photons m^{-2} s^{-1} (µE m^{-2} s^{-1}) under LED lights (Alpheus, Montgeron, France) before being submitted to stress experiments.

Light stress corresponded to a shift from low-light acclimated conditions (hereafter LL, 20 µE m^{-2} s^{-1}) to 250 µE m^{-2} s^{-1} (hereafter HL), the temperature being maintained at 21°C. After this shift at T0, cultures were sampled after 20min (T0.3), 1h (T1), 3h (T3) and 6h (T6). At 6h, cultures were shifted back to initial conditions to assess their ability to recover from the stress, and were sampled 1 h later (R1 for recovery 1h) and 24h later (R24).

Low temperature (LT) stress corresponded to a shift from 21°C to 13°C in temperature-controlled chambers (Liebherr-Hausgeräte, Lienz, Austria), the light intensity being maintained at 20 µE m^{-2} s^{-1}. After this shift at T0, cultures were sampled after 12h (D0.5), 1 day (D1), 2 days (D2), 3 days (D3) and 4 days (D4).

All experiments were performed at least in 3 replicates.

1.1.3 Preparation of cultures

Given the high volume needed for sample analysis (see below), cultures were progressively brought from a volume of 50 mL to a volume of 10 L. To determine the optimal time to start the experiment, we followed several parameters along a growth cycle: i) *Synechococcus* cell concentration; ii) Percentage of contamination by heterotrophic bacteria; iii) Photosynthetic efficiency of photosystem II, as assessed by the ratio of F_V:F_M (Figure 41).

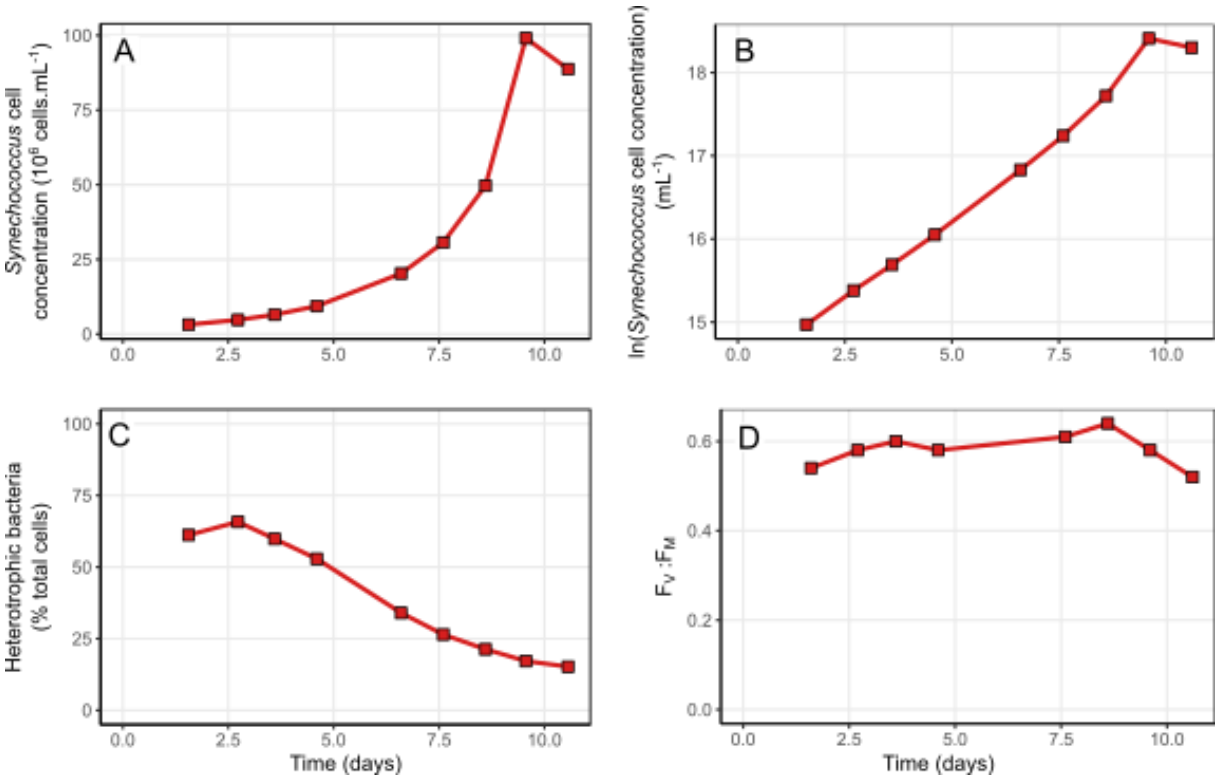


Figure 41 Example of preculture monitoring. Cell density of cultures was followed by flow cytometry (A, B) to calculate growth rates and the maximal concentration achieved in exponential phase. The relative concentration of heterotrophic contaminant (for non-axenic strains) decreased thoughout the exponential phase (C). Physiological status of the cells was assessed through fluorescence emission of phycoerythrin (not shown) and PSII photosynthetic efficiency (F_V:F_M, D). Curves shown on this figure were obtained with a preculture of the strain A15-62.

The growth curve obtained from cell density measurements were used to calculate the growth rate of each strain in acclimated condition and to determine the maximal density that the culture can reach in exponential phase. We then determined the optimal initial cell concentration at which stress experiments should be started in order to meet several conditions:

1- Cultures have to stay in exponential phase for at least 4 days of temperature stress;
2- Cell density has to be as high as possible in order to maximize cell concentration at the time of sampling, in particular for samples used for RNA extraction and transcriptomics;
3- Contamination by heterotrophic bacteria has to be minimal, as they add noise to downstream analyses, in particular for the analysis of lipid composition (see below) but also for RNAseq;
4- F_V:F_M (used as a proxy of physiological status) has to be as high as possible at the beginning of the stress experiment.

F_V:F_M was quite stable along the whole exponential phase and laid between 0.5 and 0.6 depending on the strain. As the contamination level decreased steadily during the exponential phase (Figure 41), the choice of initial concentration was a compromise between conditions 2 and 3 on one side, which favor a high concentration, and condition 1 on the other side, which favors a lower concentration. Initial concentrations that were chosen for each strain are summarized in Table 3.

Table 3 Initial cell concentration determined for each strain

Strain	Clade	Initial concentration (cells/mL)	Comment
MVIR-18-1	I	3.10×10^7	As this strain is expected to keep growing during cold stress at 13°C, the culture was diluted to 1.7×10^7 for cold stress
A15-62	II	3.80×10^7	
WH8102	III	9.50×10^6	This strain has large cells and thus a low maximal density in exponential phase (personal observation)
BL107	IV	2.80×10^7	This strain is expected to keep growing during cold stress at 13°C
WH7803	V	1.00×10^7	This strain has large cells and thus a low maximal density in exponential phase

1.1.4 Experimental procedures and sampling

In order to get maximal information on how cells responded to stress, we monitored cell physiology during stress, while sampling for a number of analyses. Cell physiology was monitored by following cell density, PSII quantum yield (F_V:F_M, used as a proxy of physiological status), the level of fluorescence emission by phycobiliproteins and the repair rate of the D1 protein.

Cell density measurements

Cell density was measured by flow cytometry using a FACSCanto II (Becton Dickinson, San Jose, CA, USA) with laser emission set at 488 nm and using distilled water as sheath fluid. Aliquots of cultures were sampled at each time point, fixed with 0.25% glutaraldehyde grade II (Sigma Aldrich, St Louis, MO, USA) and stored at -80°C until analysis (Marie et al., 1999).

In vivo *fluorescent measurements*

PSII quantum yield (F_V/F_M) was measured using a Pulse Amplitude Modulation fluorometer (Phyto-PAM, Walz, Effeltrich, Germany) connected to a chart recorder (Vernier, LabPro, Beaverton, OR, USA), upon excitation at 520 nm. 2mL of culture were acclimated to dark during 3 minutes before exposure to modulated light only, under which basal level of fluorescence (F_0) was measured. The PSII blocker 3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU) was then added at a concentration of 100µM while triggering saturating light pulses (655 nm; 5000 µE m^{-2} s^{-1}), and maximal fluorescence was measured in this new condition (F_M) (Six *et al.*, 2007a). F_V:F_M was calculated as:

$$F_V:F_M = (F_M-F_0)/F_M$$

Fluorescence emission spectra were also recorded upon excitation at 530 nm with a LS-50B spectrofluorometer (Perkin-Elmer, Waltham, MA, USA) as described in (Six *et al.*, 2007b). This allows to study the phycobiliprotein coupling (and thus to observe any dismantlement of phycobilisome rods) in particular by calculating the ratio for fluorescence emission of phycoerythrin (PE) to phycocyanin (PC), obtained by dividing the values measured at the PE maximum (between 565 and 575 nm) to the PC maximum (between 645 and 655 nm).

PSII repair rate measurements

To calculate PSII repair rate (R_{PSII}), two 50 mL-aliquots of culture were sampled, and one was used as a control (-linco), while lincomycin (a blocker of translation) was added to the other at a final concentration of 0.5 mg/mL (+linco). As lincomycin kills the cells if the exposure is too long, this measurement was only performed for HL stress experiments (6 h long) and not for temperature stress experiments (4 days long). Both – and + linco aliquots were placed in the same condition at the beginning of the stress (T0) (i.e. control condition or stress condition) and F_V:F_M was monitored on both cultures at each time point (T0 to T6). Exponential curves were fit to F_V:F_M measurements and the difference of exponential coefficients used as a proxy of R_{PSII} (Figure 42).

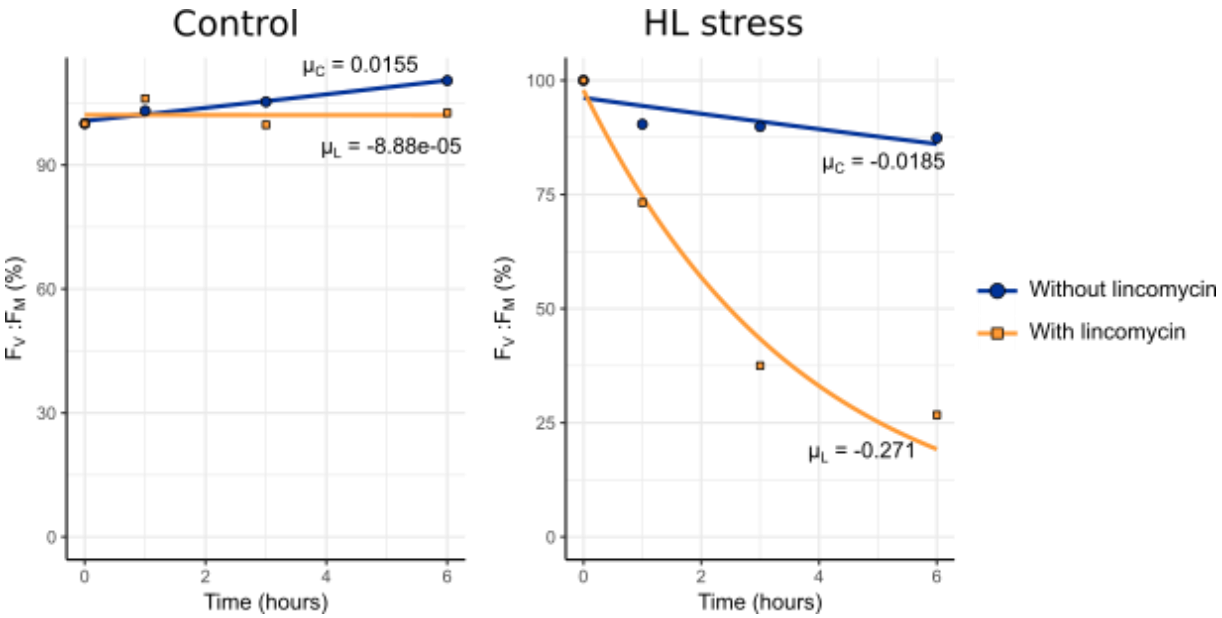


Figure 42 Calculation of PSII repair rate (R_{PSII}). After plotting F_V:F_M against time for cultures with without 0.5 mg ml^{-1} lincomycin, R_{PSII} is calculated as the difference between the coefficients of exponential fits on each curve. Examples are given for a control, low-light acclimated culture (R_{PSII} = 0.0155-(-0.00008) = 0.0156 h^{-1}) and for a culture under high-light stress (R_{PSII} = -0.0185-(-0.271) = 0.253 h^{-1}).

For each strain, we thus obtained one value or repair rate in control condition (equivalent of R_{PSII} in acclimated condition) and one value under HL stress.

Along with these direct measurements of cell physiology, samples were harvested for analysis of lipid composition, protein content, pigment content and RNA sequencing. This sampling is

detailed in Figure 43. While I participated to the sampling, samples for lipid and pigment content were processed and analyzed by Christophe Six and Solène Breton, and thus no further detail will be given in this thesis. For RNA sequencing, samples were harvested by centrifugation of 2 x 150 mL of culture with 0.005% Pluronic® (Poloxamer 188, Sigma Aldrich, St Louis, MO, USA) in a 5804R centrifuge (Eppendorf, Hamburg, Germany) at 4°C, 10 000 x g for 7 minutes. Cell pellets were resuspended in < 2mL then transferred to 2mL tubes and centrifugated again in a 5415D centrifuge (Eppendorf) at 16 000 x g for 1 minute. Cell pellets were resuspended in 1mL Qiazol (QIAGEN, Hilden, Germany), flash-frozen in liquid nitrogen and stored at -80°C until further analysis. All these steps were performed on ice to limit RNA degradation and in less than 20 minutes.

RNA extraction and library preparation were performed by Morgane Ratin and Ngoc-An Nguyen, both members of the MaPP team, and RNA samples were then sent for sequencing to the GenoToul sequencing platform in Toulouse using the Truseq stranded mRNA protocol (Illumina).

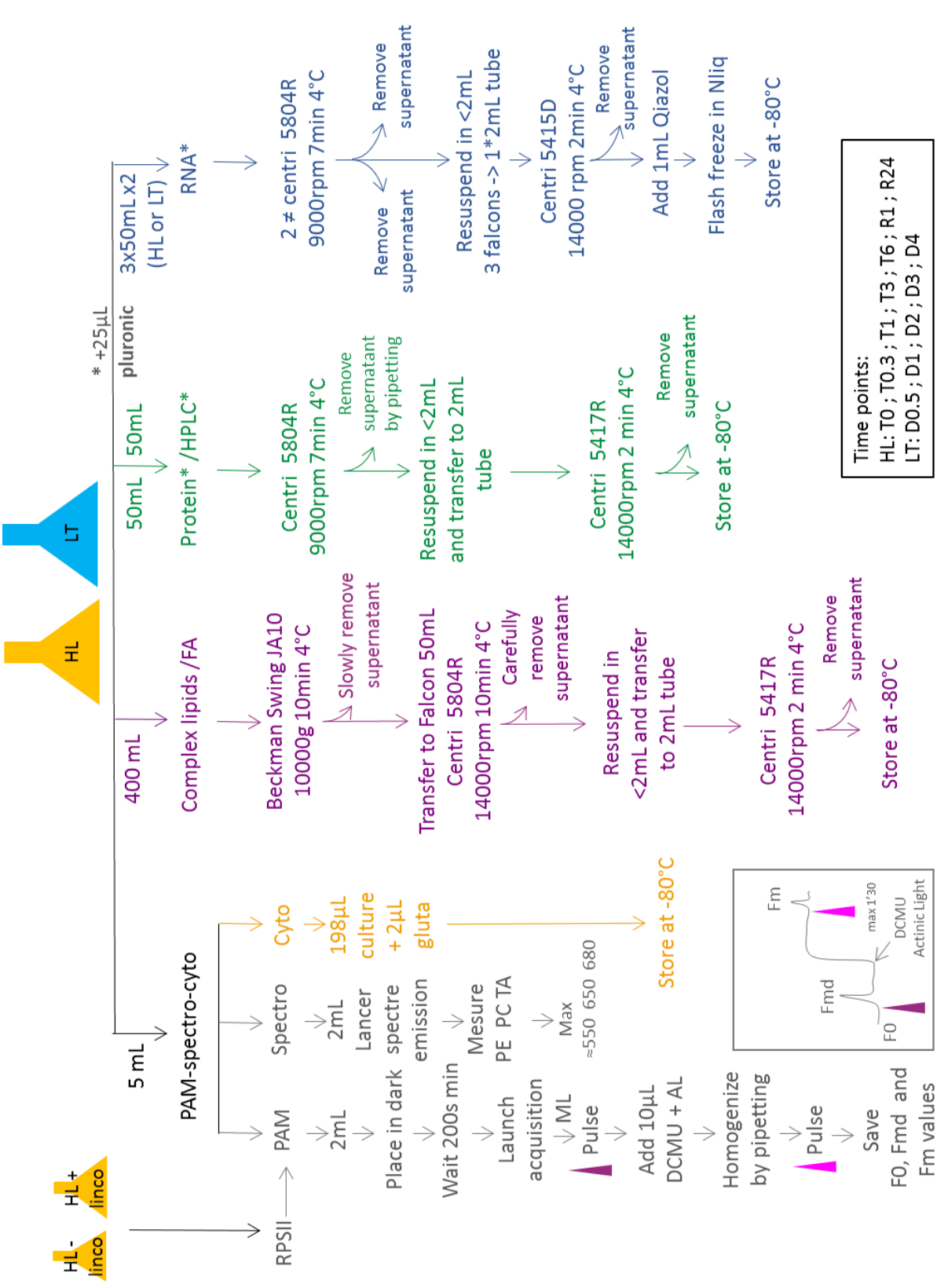


Figure 43 Experimental plan for high-light and low-temperature stresses. This plan was applied for each sampling point.

1.2 Light-dark cycle experiments at two temperatures

1.2.1 Choice of experimental conditions

Given the heavy logistics implied by a continuous culture and a regular sampling during a diurnal cycle, only one strain was used for this experiment. We chose WH7803, in order to ease comparison with previous studies of a diurnal cycle performed on this strain (Mella-Flores *et al.*, 2012) and because this is one of the few axenic strains of *Synechococcus* available. The strain was grown in 8L of PCR-S11 medium under varying light intensity mimicking a diurnal light cycle provided by LED ramps (Alpheus, France) (Figure 44) in temperature-controlled chambers (Liebherr-Hausgeräte, Lienz, Austria). The light phase started at 6h00, light intensity was maximal at 12h00 with 650 µE and the dark phase started at 18h00, giving a total irradiance of 13.4 mol photons m^{-2} d^{-1}.

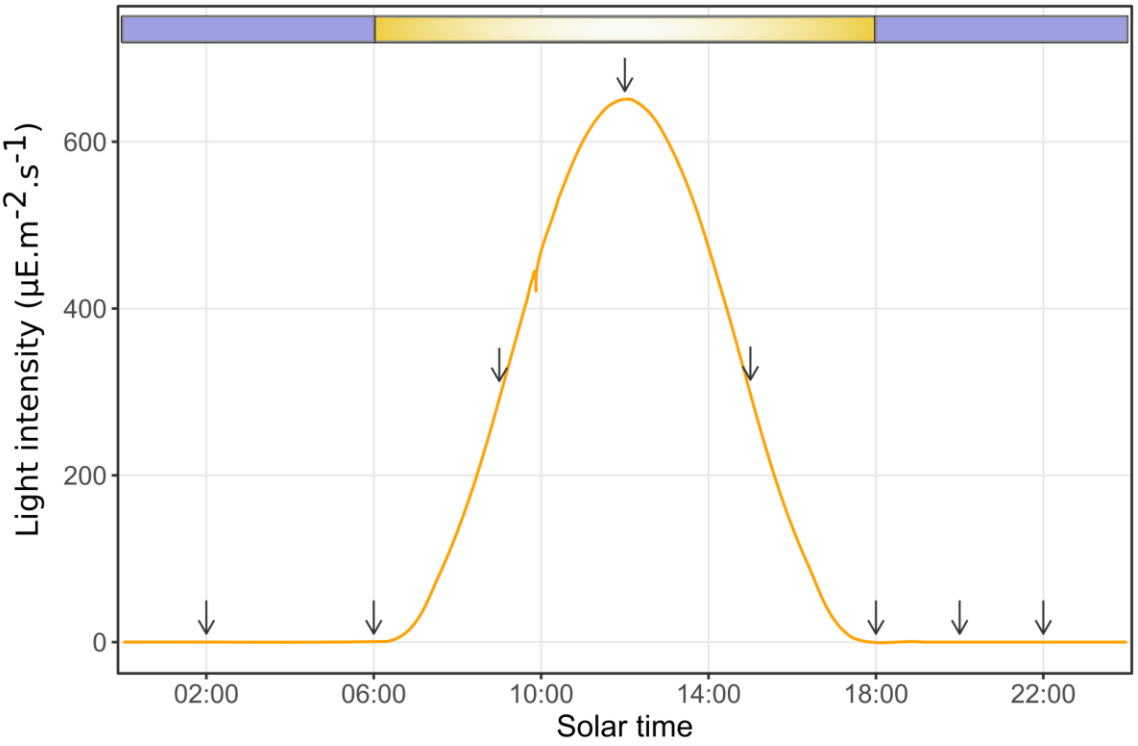


Figure 44 Light intensity delivered along a light:dark cycle. Arrows indicate sampling time points.

We maintained the culture in exponential growth phase by constraining cell density with a constant outflow of culture compensated by a constant inflow of fresh medium. If one wants to maintain a constant cell density in a constant volume of culture V_{cult} with a growth rate µ, a volume of $V = V_{cult} \times \mu$ must be replaced by time unit. For example, with a growth rate of 1.2 d^{-1} and a volume V_{cult} of 8L, one has to remove 1.2 x 8 = 9.6 L of culture per day and to add as much fresh medium. As we planned to sample a total volume of 5.2 L per day (see below), we had a constraint on the growth rate that had to be higher than 0.65 d^{-1}. As the objective was to compare one cold and one warm temperatures, we measured growth rate as a function of temperature in the same light-dark conditions but in batch cultures (Figure 45).

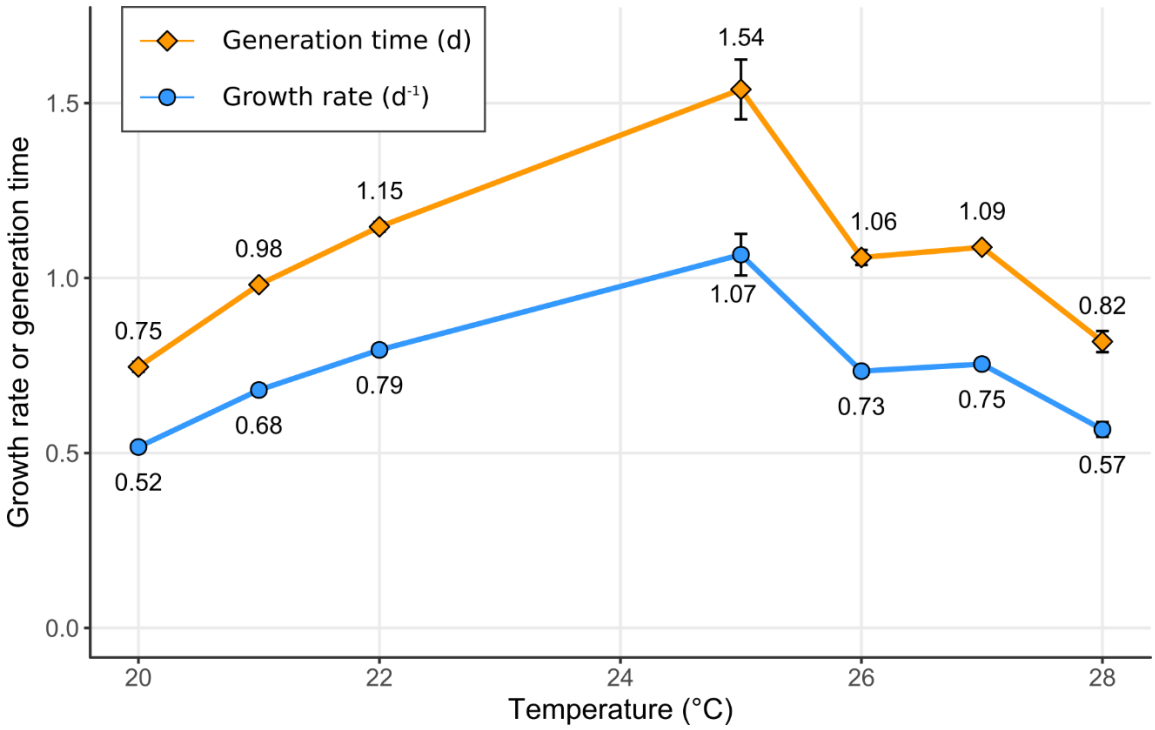


Figure 45 Growth rate of WH7803 as a function of temperature under a light:dark cycle. These preliminary experiments were used to select experimental temperatures (this preliminary work was performed by Ngoc-An Nguyen).

Interestingly, this curve is quite different from the one obtained in continuous light at 80 µE m^{-2} s^{-1} (Pittera *et al.*, 2014), where the optimal growth temperature was above 30°C. Here the optimal growth temperature is around 25°C, which is very close to the isolation temperature of this strain (25.85°C). In order to compare temperatures as extreme as possible while maintaining the growth rate above 0.65, we chose 21°C as the "cold" temperature and 27°C as the "warm" temperature.

1.2.2 Experimental setup and sampling scheme

Cultures were grown in 10L quartz flasks (Atelier Jean Premont, Bordeaux, France) filled up to 8 liters, with a continuous inflow of fresh medium and outflow of culture (Figure 46) and the flow rate was regularly adjusted to keep the cell concentration constant and the outflow higher or equal than the inflow. An automatic sampler allowed to sample small volumes every hour for flow cytometry, and an additional peristaltic pump allowed manual sampling. Samples were taken at 6h, 9h, 12h, 15h, 18h, 20h, 22h and 2h except for D1 repair that was measured only during day (from 6h to 18h) and for flow cytometry samples that were sampled every hour of day and night. The volume sampled for each type of sample is indicated in Table 4.

Table 4 Volumes sampled in light:dark cycle experiment.

Sample	Volume per sample (ml)	Nb of samples/day	Volume/day (mL)	Sampling
RNA	500	8	4000	2 x 250 ml (technical duplicates)
PAM	2	8	16	
Spectrofluorimetry	2	8	16	
R_{PSII}	30	5	150	2 x 15mL (+/- lincomycin) aliquoted at each time point
FCM	0,4	24	9,6	2 x 200µL (technical duplicates) with 0.25 % gluta
FCM Vésicules	0,4	24	9,6	2 x 200µL (technical duplicates) with 0.5 % gluta
Protein	50	8	400	
HPLC (pigments)	50	8	400	
Dead volume FCM	2	24	48	
Dead volume other	20	8	160	18 mL measured on actual tube length
Total	**656.8**		**5209.2**	

F_V:F_M, spectrofluorimetry and cell density were measured as described above. RNA samples were also treated as described above for stress experiments except for the centrifugation steps, including a first centrifugation of 250 mL in a Beckman centrifuge with a JA-10 rotor at 9000 rpm (8890 x g), 4°C during 7 minutes, followed by a second centrifugation in 2ml Eppendorf tubes in a 5415D centrifuge (Eppendorf) at 16,000 × g for 1 minute. For measurements of PSII repair rate, 2 aliquots of 15mL were sampled at the beginning of each time point during the light phase, and lincomycin was added to one of them at a final concentration of 0.5 mg/mL. Both aliquots were placed within the thermoregulated chambers in the same conditions as the main culture and F_V/F_M was monitored every 15min during 1h. R_{PSII} was calculated from these measurements as described above.

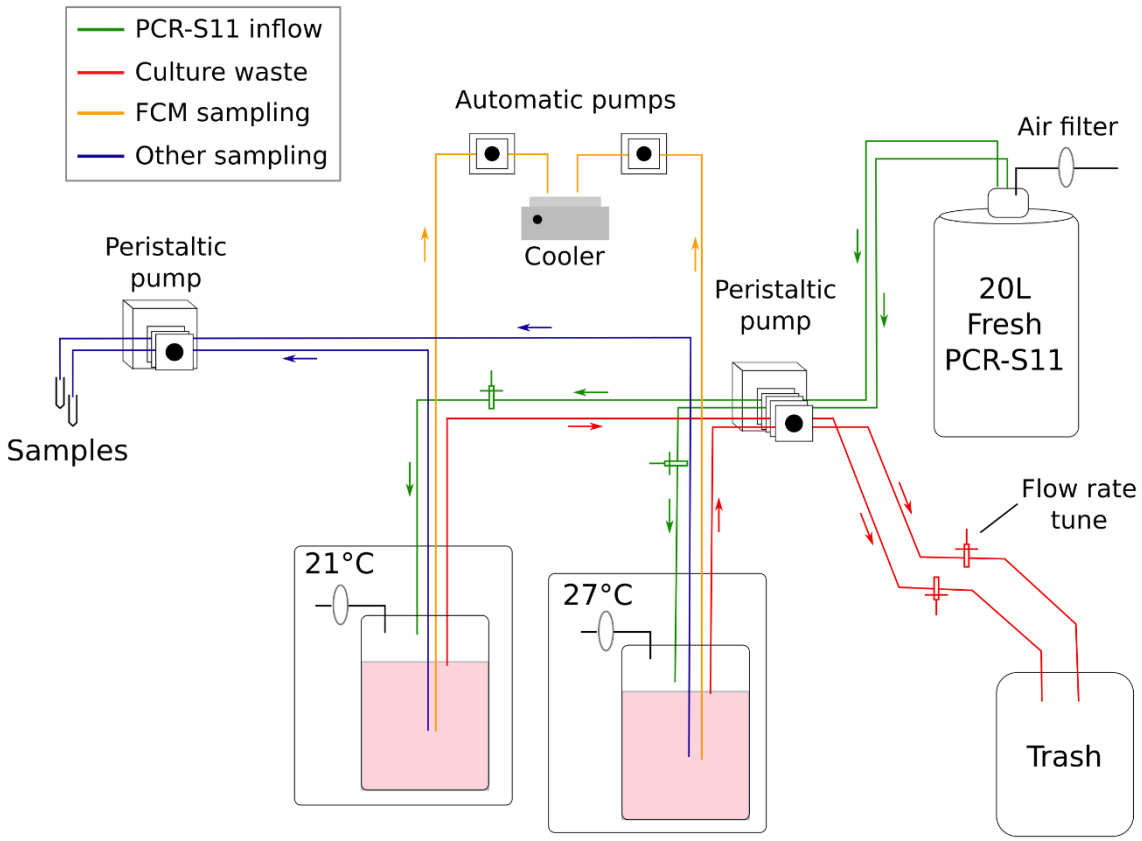


Figure 46 Experimental setup used for axenic, continuous growth of cultures under a light:dark cycle.

1.3 Analysis of transcriptomic data

RNA sequencing resulted in raw paired-end reads of 150 bp that went through an analysis workflow developed and applied by U. Guyet. Briefly, reads provided by the sequencing platform were submitted to a series of quality checks (data cleaning, selection based on read size and sequencing quality) before alignment against the reference genome using Bowtie 2 (Langmead and Salzberg, 2012). After checking for outliers, technical replicates were pooled and read counts were obtained from alignments using HTseq Count (Anders *et al.*, 2015). Finally, the pipeline implemented in SARTools (Varet *et al.*, 2016) was used to calculate differential expression relative to control condition (T0 for HL stress, J0 for LT stress, 6h for light:dark cycle). For this critical step, all samples of each strain were pooled together, in order to allow a global normalization of signal over all conditions. Results were then filtered to keep only genes showing a fold change higher than two with a false discovery rate below 0.05. While I have not contributed to these data-processing steps, I was involved in the analysis of all the results presented in this chapter.

2. Ecotypic variability in response to high-light and low-temperature stresses

2.1 Physiological response to high-light stress

The five strains of marine *Synechococcus* were submitted to a high-light (HL) stress triggered by shifting cultures acclimated to low-light (LL, 20 µE m^{-2} s^{-1}) to an intensity of 250 µE m^{-2} s^{-1}. The strains appeared to be differentially affected by HL stress, as shown by the variation of F_V:F_M, used as a proxy of physiological status (Figure 47). Responses ranged from a low impairment of PSII activity (10% decrease of F_V:F_M after 6h of stress) for strains MVIR-18-1 (clade I) and A15-62 (clade II) to a high sensitivity to HL for strains BL107 (clade IV, 40% decrease after 6h) and WH8102 (clade III, 60% decrease after 6h), the model strain WH7803 (clade V) showing an intermediary sensitivity. All strains were able to recover after shifting them back to initial LL conditions for 24h, with F_V:F_M values close to both control and pre-stress levels (Figure 47), showing that the stress was no lethal and was well induced by HL conditions.

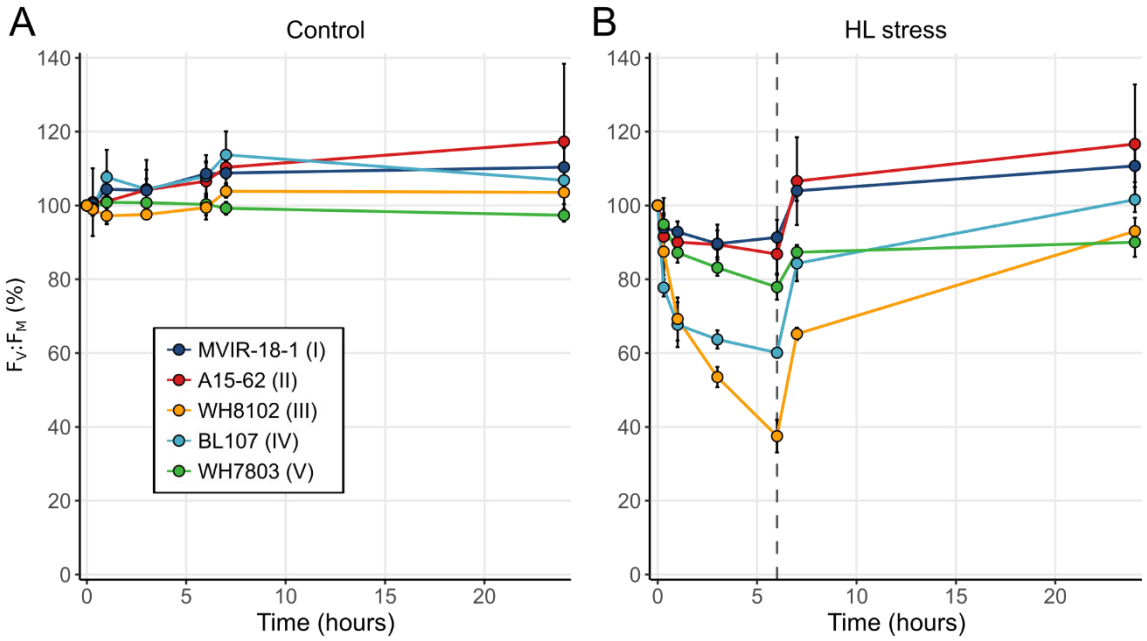


Figure 47 Variation of PSII quantum yield during the course of a high-light stress. A. Control condition (low-light, 20 µE m^{-2} s^{-1}). B. High-light (HL) stress condition. Cultures were shifted to HL (250 µE m^{-2} s^{-1}) at T0. The dashed line indicates the time (6h) at which cultures were shifted back to LL conditions for recovery. Strains and their corresponding clade are indicated in insert. F_V:F_M is expressed as percentage of T0 value.

Only two strains showed significant variations in the ratio of phycoerythrine (PE) phycocyanine (PC) in response to HL stress: this ratio decreased during stress for BL107 and increased for WH7803 (Figure 48). These fluorescence variations are likely due to a disconnection

of the distal PEII hexamers in WH7803 (as observed in WH8102 in response to UV by (Six *et al.*, 2007b)), and to a disconnection between PC and APC for BL107, i.e. at the basis of rods. Interestingly, degradation of phycobilisomes does not seem to be related to PSII degradation, since the most affected strain at PSII level (WH8102) did not show any variation of PE:PC ratio.

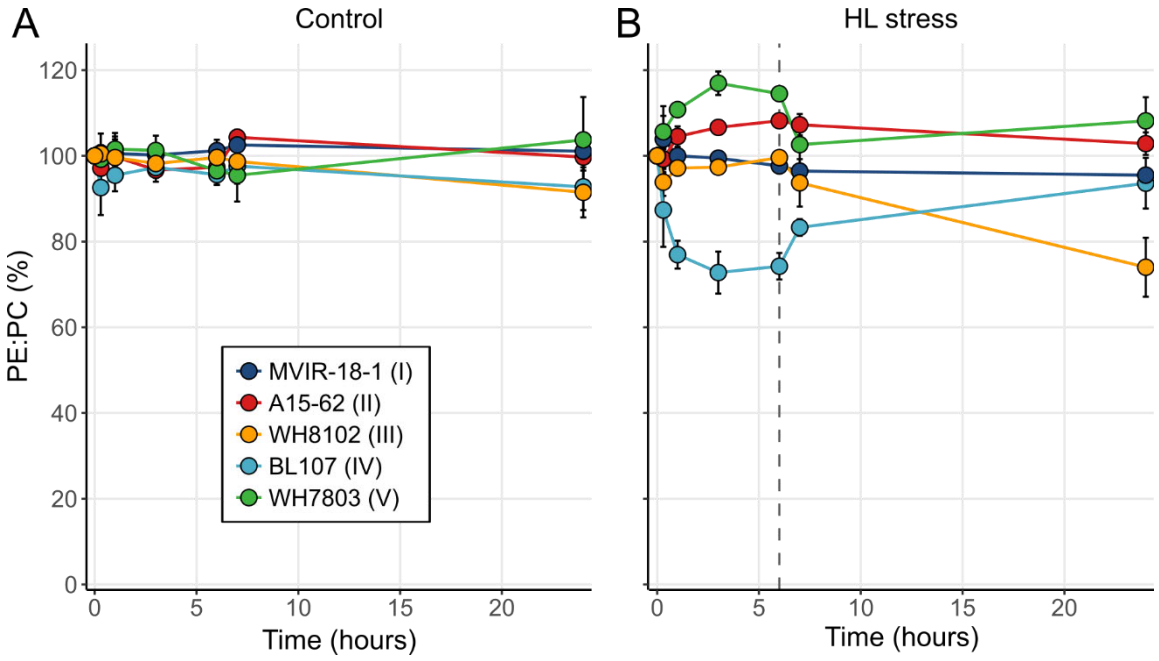


Figure 48 Variation of PE to PC ratio during HL stress. A. Control condition (LL, 20 µE m^{-2} s^{-1}). B High-ligh (HL) stress condition. Cultures were shifted to HL (250 µE m^{-2} s^{-1}) at T0. The dashed line indicates the time at which cultures were shifted back to LL for recovery (6h). Strains and their corresponding clade are indicated in insert. PE:PC ratio is expressed as percentage of T0 value.

Finally, we also calculated a proxy of PSII repair rate (R_{PSII}) integrated over the whole duration of the stress (see above). While in control conditions (LL), the 5 strains had a low R_{PSII} (below 0.1 h^{-1}), all of them were able to increase their R_{PSII} under HL stress, up to 0.34 h^{-1} in A15-62. (Figure 49). Interestingly, the two strains that were the least affected by HL stress, MVIR-18-1 and A15-62, showed no detectable R_{PSII} under LL, and thus probably underwent very little damages to PSII at this irradiance. Both strains also displayed the largest difference between HL and control LL conditions, suggesting that their higher resistance to HL stress might be due to their higher PSII repair efficiency. On the contrary, BL107, whose F_V:F_M was strongly affected by HL, showed the highest R_{PSII} in LL but the lowest difference of R_{PSII} between LL and HL. Thus, the D1 turnover was less responsive to HL stress in BL107 than in the other *Synechococcus* strains. Surprisingly, WH7803 showed the lowest R_{PSII} rate under HL but still managed to maintain its F_V:F_M at 80% of control level after 6h of stress, suggesting that this strain exhibited a smaller photoinactivation cross section (σ_i) in HL conditions than the other strains (Six *et al.*, 2007a).

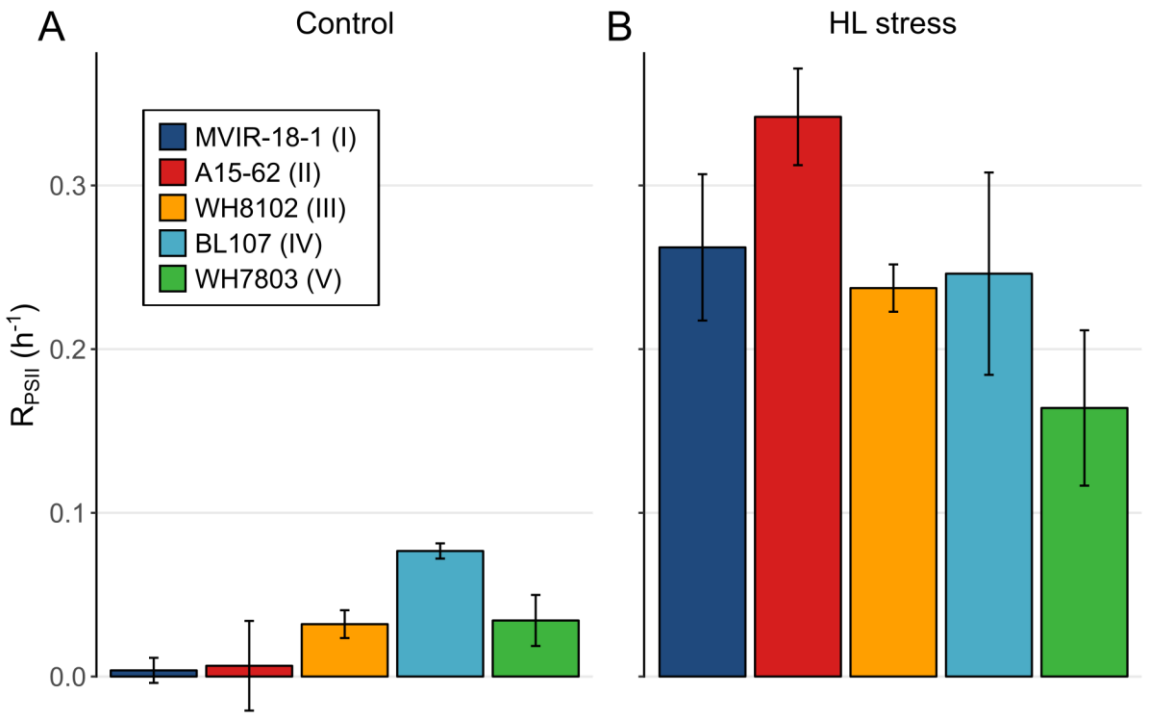


Figure 49 PSII repair rate under control and HL stress conditions. A. Control conditions (LL, 20µE.m-2.s-1). B. High-light stress condition (250 µE.m-2.s-1). Strains and their corresponding clade are indicated in insert.

2.2 Physiological response to cold stress

The same five strains were submitted to a cold stress by shifting cultures from 22°C to 13°C while maintaining LL intensity (20 $\mu E.m^{-2}.s^{-1}$). Here again, responses were contrasted between strains: MVIR-18-1 (clade I) was not affected by the temperature shift, maintaining a stable F_V:F_M over the course of the experiment (Figure 50A), while WH7803 (clade V) was rapidly affected with F_V:F_M values below 60% of initial level after only one day of stress. This decrease reflects either an increase of non-photochemical quenching, resulting in a decrease of F_M or an increase in PSII photoinactivation upon cold stress. The three other strains showed a progressive decrease in PSII photosynthetic efficiency, but only BL107 (clade IV) could maintain it above 80% of initial level. As expected, only the two strains representative of cold thermotypes (clades I and IV) were able to maintain growth at 13°C, though with a one-day latency for clade IV strain BL107 that was not observed for clade I strain MVIR-18-1 (Figure 50B). The other strains either died (WH7803 and A15-62) or maintained a constant cell density (WH8102).

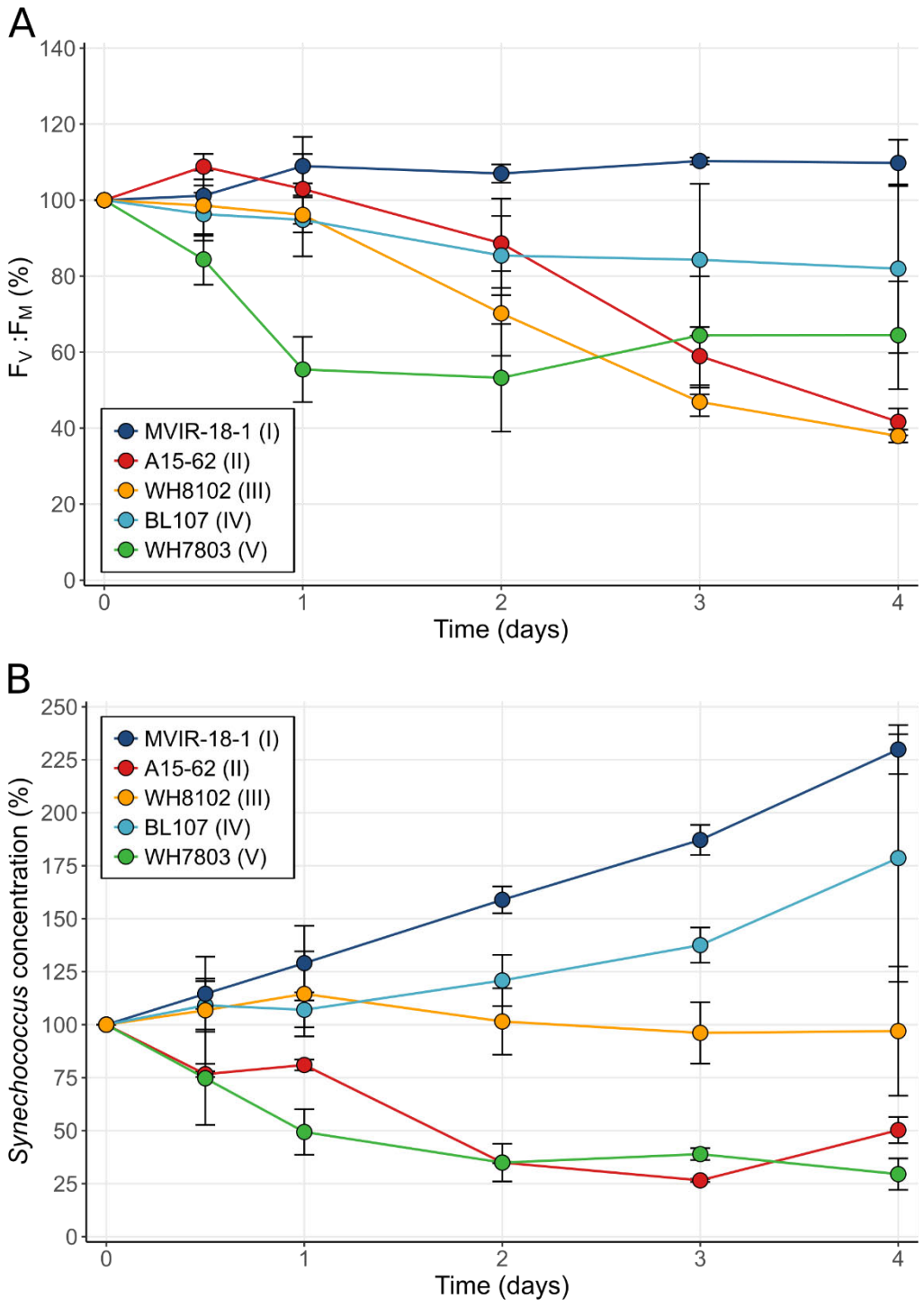


Figure 50 Effect of low temperature stress on PSII efficiency (A) and cell concentration (B). These parameters are expressed in percentage of the value measured at T0, time at which cultures where shifted from 22°C (acclimated condition) to 13°C (stress condition). Strains and their corresponding clade are indicated in insert.

All the strains that were affected by cold stress (i.e. all strains but MVIR-18-1) showed an increase in the PE:PC ratio (Figure 51), a phenomenon known to occur upon cold stress in tropical and temperate strains and indicative of an impairment of the phycobiliprotein coupling likely

leading to the disconnection of PE from phycobilisome rods (Pittera *et al.*, 2014; Six *et al.*, 2007b). This disconnection could be a way to limit the oxidative stress both by decreasing the amount of energy reaching PSII reaction center and because free PE might dissipate energy through its high fluorescence (Pittera *et al.*, 2014), but most certainly disturbs growth.

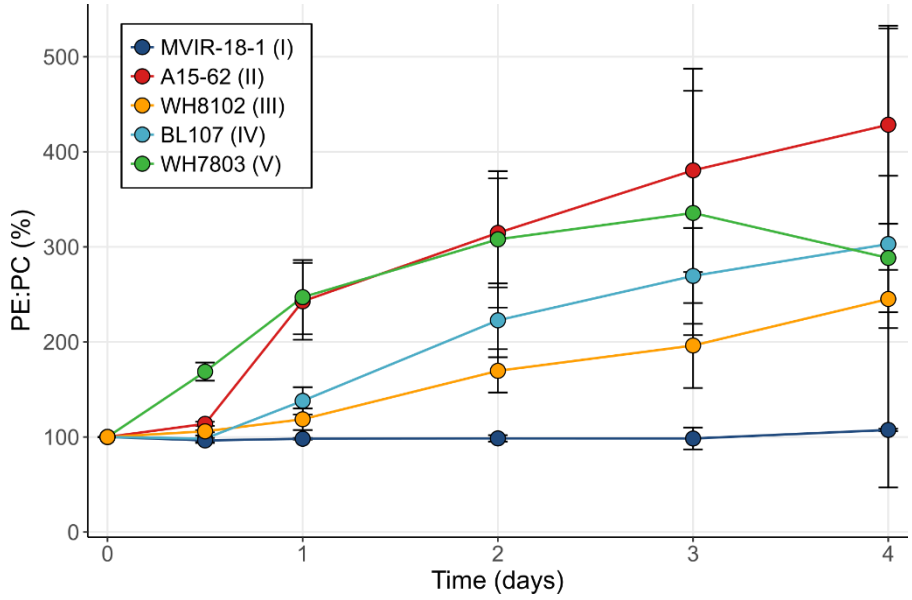


Figure 51 Variation of PE to PC ratio during low-temperature stress. The ratio is expressed as a percentage of T0 value. Strains and their corresponding clade are indicated in insert.

Strains representative of cold thermotypes clearly better stand cold stress than the three other strains, and the low photoinactivation of PSII (low decrease in F_V:F_M) suggests that they are likely able to better deal with the imbalance between the light energy absorption and its utilization in downstream reactions (Latifi *et al.*, 2009).

2.3 Transcriptomic response to stress: preliminary results.

Transcriptomic response was monitored during high-light (HL) and low-temperature (LT) stress by RNA sequencing after 20 min (T0.3), 1h (T1), 3h (T3) and 6h (T6) for each strain. Given the time needed for sequencing and data processing, only preliminary results are available at this stage.

2.3.1 Global transcriptomic response

A global analysis of gene expression during HL and LT stress reveals a quite similar response in both types of stress. Indeed, a majority of differentially expressed genes in HL stress show the same pattern of expression as in LT stress (Figure 52). However, the differential gene expression levels vary between stresses in a strain-dependent fashion, the cold-thermotypes, MVIR-18-1 (clade I) and BL107 (clade II), displaying much lower fold change values in response to the LT stress compared to HL stress in agreement with their ability to better stand cold conditions. . On

the contrary, the differential expression levels were higher during LT than HL stress for WH7803 (clade V) and in a lesser extent for A15-62 (clade II). It is also worth noting that a significant part of the differentially expressed genes share the same expression pattern between strains (e.g clusters 2, 4 and 5 on Figure 52), while some genes seem to be differentially expressed only in a specific strain. More specifically, all strains but BL107 display a group of genes downregulated in response to stress that shows little or no differential expression in other strains (cluster 6 for WH8102, cluster 13 for A15-62, cluster 3 for WH7803 and part of cluster 9 for MVIR-18-1). In contrast, strain-specific groups of genes upregulated in response to stress were observed mainly in WH8102 (cluster 7) and in MVIR-18-1 (cluster 15).

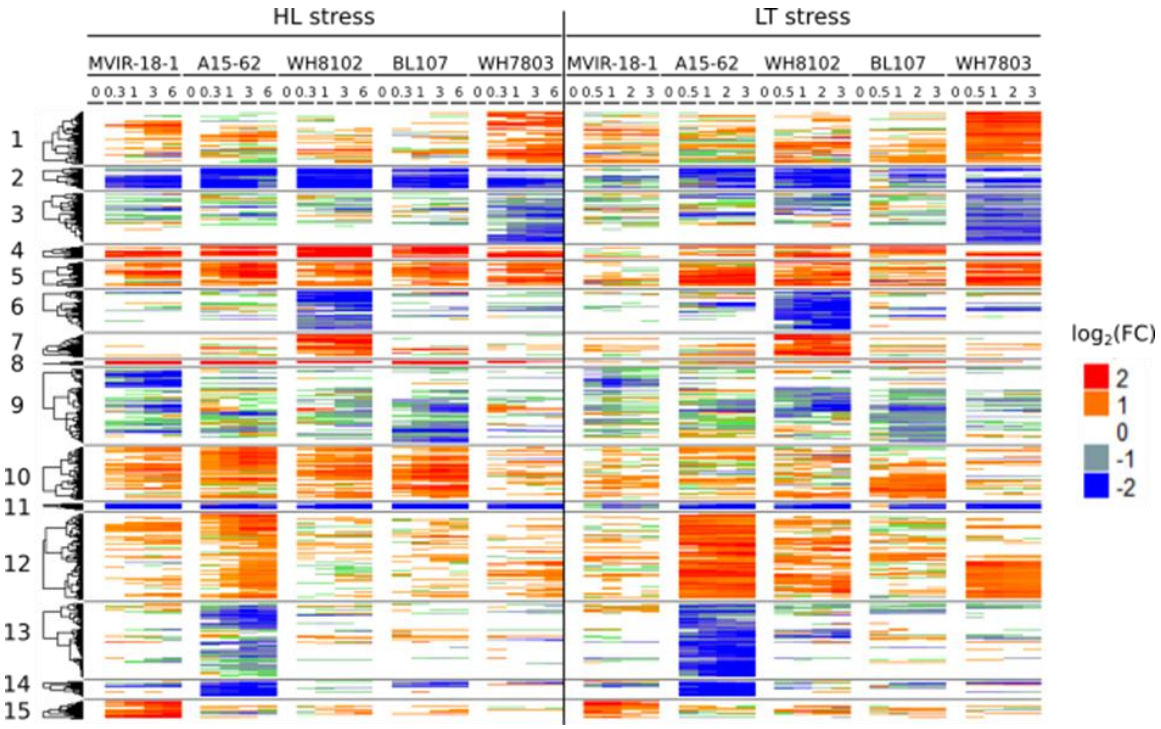


Figure 52 Global transcriptomic response to HL and LT stress in five *Synechococcus* strains. Gene expression values correspond to log2 of fold change as compared to T0. Any value over 2 is plotted in red, and any value below -2 is plotted in blue. Time points are expressed in hours for HL stress and in days for LT stress. Only the most differentially expressed genes are shown.

These differences between strains are thus quite promising with respect to our goal to unveil the molecular bases of the ecotypic variability of the *Synechococcus* response to stress. In the next paragraphs, I will detail the expression patterns of genes belonging to functional categories that have been shown in previous studies to play a key role in the metabolism and/or response to stress in marine *Synechococcus* or other cyanobacteria.

2.3.2 Expression of genes involved in photosynthesis

As photosynthesis is central to the metabolism of marine picocyanobacteria (see INTRODUCTION), we explored in more details the expression pattern of genes involved in this process. The genes encoding phycobiliproteins were massively downregulated in all strains upon HL stress (Figure 53), likely resulting in a reduction of the amount of energy reaching the PSII reaction center. In agreement with a previous study performed on WH8102, which showed that UV stress induced a partial dismantling of phycobilisome rods, particularly affecting the tips of the rods (Six *et al.*, 2007b), this downregulation was higher for genes encoding PEI and PEII (distal part of the rod, e.g. *mpeA*) than for genes encoding PC (e.g. *cpcA*), the lowest decrease being observed for APC (e.g. *apcA*). Interestingly, the expression patterns of the genes encoding these proteins upon LT stress were quite different between strains, with representatives of cold thermotypes (MVIR-18-1 and BL107) only showing a weak downregulation, mainly of PE-encoding genes, while representatives of warm thermotypes (WH8102 and A15-62) exhibited a downregulation of phycobiliprotein-encoding genes similar to that observed in HL stress.

The expression of PSI-encoding genes (e.g. *psaA/B/E/J*) followed a similar expression pattern as phycobiliprotein encoding genes, with a strong downregulation in all strains in response to HL stress, and mainly in warm thermotypes strains in response to LT stress (Figure 53). In contrast, most genes involved in PSII synthesis (e.g. *psbB*/*E/H*) showed no or little downregulation upon HL stress, and only a weak downregulation in warm-adapted strains A15-62 (clade II) and WH8102 (clade III) upon LT stress. Notable exceptions are *psbA* and *psbD*, encoding the two subunits of the PSII reaction center (D1 and D2). *psbD* and *psbA* gene copies encoding D2 and the D1:2 isoforms, respectively, were indeed quite strongly upregulated upon HL stress, while the *psbA* copy coding for the D1:1 isoform was downregulated in the same conditions. This likely reflects a degradation and an acceleration of the turnover of the D1 and D2 proteins in response to HL as well as the replacement of the D1.1 isoform, which would confer a higher PSII activity (Campbell *et al.*, 1996), by the more photostable D1.2 in HL conditions (Tichý *et al.*, 2003), as a way to limit PSII inactivation (Garcia-fernandez *et al.*, 1998; Garczarek *et al.*, 2008). Interestingly, this replacement is also observed in LT stress, indicating that cells also undergo PSII degradation during this stress. However, the effect is hardly visible in strains MVIR-18-1 (clade I) and BL107 (clade IV), which seem to undergo much less photoinactivation than the three other strains. Noteworthy, these expression patterns correlate very well with the differences in the decrease in F_V:F_M observed between strains (Figure 50A).

HL stress			MVIR-18-1					A15-62					WH8102					BL107					WH7803				
	Gene name	Product	T0	T0.3	T1	T3	T6	T0	T0.3	T1	T3	T6	T0	T0.3	T1	T3	T6	T0	T0.3	T1	T3	T6	T0	T0.3	T1	T3	T6
PBPs	*apcA*	APC α chain	0	-2	-2	-3	-3	0	-2	-2	-1	0	0	-3	-3	-2	-2	0	-2	-2	-2	-2	0	-2	-2	-2	-1
	cpcA	PC α chain	0	-2	-2	-3	-3	0	-1	-1	0	0	0	-3	-3	-3	-3	0	-3	-4	-4	-3	0	-2	-3	-3	-3
	mpeA	C-PE class II α chain	0	-2	-5	-6	-6	0	-2	-4	-5	-3	0	-1	-4	-7	-7	0	-2	-5	-6	-6	0	-2	-4	-5	-5
PSI	*psaA*	PSI core, PsaA	0	-2	-4	-4	-4	0	-2	-4	-3	-1	0	-3	-4	-4	-4	0	-3	-5	-5	-4	0	-1	-4	-3	-2
	psaB	PSI core, PsaB	0	-2	-4	-5	-5	0	-1	-4	-3	-1	0	-2	-4	-4	-4	0	-1	-4	-5	-4	0	-1	-3	-3	-2
	psaE	PSI, PsaE	0	-1	-1	-2	-2	0	-4	-4	-4	-1	0	-3	-3	-3	-3	0	-2	-2	-3	-2	0	-2	-2	-2	-2
	psaJ	PSI, PsaJ	0	-2	-2	-2	-3	0	-3	-3	-3	-1	0	-4	-4	-4	-4	0	-3	-3	-3	-3	0	-2	-2	-1	-1
PSII	*psbB*	PSII, CP47	0	-1	-1	-1	-1	0	0	0	0	0	0	-1	-1	-1	-1	0	-2	-2	-2	0	0	-1	0	0	0
	psbD	PSII, D2	0	0,5	1,2	1	1	0	1,1	1,3	1,6	1,2	0	1,3	1	1,2	2,1	0	1	0,9	1	0,8	0	1,2	1,3	1,2	1,2
	psbE	cytochrome b559 subunit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	psbH	PSII, PsbH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D1	*psbA*	PSII, D1.1	0	-3	-5	-6	-6	0	0,6	-2	-2	-1	0	0	-2	-5	-6	0	-2	-4	-5	-6	0	-2	-4	-6	-6
	psbA	PSII, D1.2 (average)	0	2,3	2,5	2,6	2,1	0	1,1	1,7	1,8	1,1	0	4,7	4,8	4,2	4,2	0	1,4	1,4	1,5	1,3	0	2,5	2,6	2,1	2,6
C fixation	*csoS1A*	carboxysome shell, CsoS1A	0	0	0	-1	-2	0	1,5	0	0	0	0	2,4	2	0	0	0	0	-1	-2	-3	0	1,2	0	0	0
	csoSCA	carboxysomal carbonic anhydrase	0	0	0	-2	-2	0	1,1	0	0	0	0	1,9	1,2	0	-2	0	0	-1	-3	-3	0	1,6	1,5	1,4	1,4
	rbcL	RuBisCo large subunit	0	0	0	-2	-3	0	1,5	1,1	0	0	0	2,3	1,8	0	-2	0	0	-1	-3	-3	0	1,3	1,2	1	0

LT stress			MVIR-18-1					A15-62					WH8102					BL107					WH7803				
	Gene name	Product	J0	J0.5	J1	J2	J3	J0	J0.5	J1	J2	J3	J0	J0.5	J1	J2	J3	J0	J0.5	J1	J2	J3	J0	J0.5	J1	J2	J3
PBPs	*apcA*	APC α chain	0	-1	-2	-1	0	0	-1	-2	-2	-3	0	-2	-3	-4	-5	0	0	-1	-2	-2	0	-1	-2	-2	-2
	cpcA	PC α chain	0	0	0	0	0	0	-2	-3	-3	-4	0	-2	-4	-5	-6	0	0	-2	-2	-2	0	-2	-3	-3	-3
	mpeA	C-PE class II α chain	0	0	-2	-2	0	0	-2	-3	-4	-5	0	-2	-4	-7	-8	0	0	-2	-3	-3	0	-2	-3	-3	-3
PSI	*psaA*	PSI core, PsaA	0	0	-1	-1	0	0	-1	-2	-3	-3	0	0	-2	-4	-5	0	0	-1	-1	-1	0	0	-1	-2	-2
	psaB	PSI core, PsaB	0	0	0	-1	0	0	-2	-2	-3	-3	0	0	-2	-4	-5	0	0	-1	-1	-1	0	-1	-2	-2	-2
	psaE	PSI, PsaE	0	0	0	0	0	0	-2	-2	-2	-3	0	-1	-2	-3	-4	0	0	0	0	0	0	-1	-2	-2	-2
	psaJ	PSI, PsaJ	0	0	0	0	0	0	0	0	-1	-2	0	0	-1	-3	-3	0	0	0	0	0	0	0	0	-1	-1
PSII	*psbB*	PSII, CP47	0	0	0	0	0	0	0	-1	-1	-1	0	-1	-2	-3	-4	0	0	0	0	0	0	0	-1	-1	-1
	psbD	PSII, D2	0	0	-1	-1	0	0	-1	-1	-1	-1	0	-1	-1	-1	-2	0	0	-1	0	0	0	0	0	0	0
	psbE	cytochrome *b559* subunit	0	0	0	0	0	0	0	-1	-2	-2	0	0	0	0	-2	0	0	0	0	0	0	0	0	0	0
	psbH	PSII, PsbH	0	0	0	0	0	0	0	-1	-1	-1	0	0	0	0	-2	0	0	0	0	0	0	0	0	0	0
D1	*psbA*	PSII, D1.1	0	0	-1	-2	-1	0	-1	-2	-2	-3	0	0	-1	-4	-6	0	0	-2	-3	-2	0	-3	-4	-4	-4
	psbA	PSII, D1.2 (average)	0	0,5	0	0	0	0	1,2	1,4	2,2	2,5	0	0,9	1,6	1,5	1,3	0	0,3	0,3	0,3	0,7	0	2,4	2,6	2,3	2,2
C fixation	*csoS1A*	carboxysome shell, CsoS1A	0	0	0	0	0	0	0	-2	-2	-2	0	-2	-3	-3	-4	0	0	-2	-2	-2	0	0	-1	-2	-2
	csoSCA	carboxysomal carbonic anhydrase	0	-1	-1	0	0	0	0	0	0	0	0	-1	-2	-2	-1	0	0	-3	-3	-3	0	0	0	0	0
	rbcL	RuBisCo large subunit	0	-1	-1	0	0	0	-1	-2	-2	-2	0	-2	-3	-4	-5	0	0	-2	-2	-2	0	0	-1	-2	-2

Figure 53 Expression of genes involved in photosynthesis. Representative genes were chosen for each category. Values correspond to rounded $\log_2$ of fold change compared to T0 (control condition). Any value between -1 and 1, corresponding to a fold change of 2, was considered as not differentially expressed. The upper panel shows response to HL stress, while the lower panel shows response to LT stress.

As concerns carbon fixation, gene expression of *csoCA,* encoding carbonic anhydrase, and *rbcL,* encoding the RuBisCO large subunit, vary between strains during HL stress. While in MVIR-18-1 (clade I) and BL107 (clade IV), both genes are downregulated after T1 or T3, they show a

peak of expression at T0.3 in the three other strains. In LT stress, the expression of these genes follows that of photosystem I- and phycobiliprotein-encoding genes, with a more significant decrease in warm-adapted strains.

Overall, the response of genes involved in photosynthesis is quite similar for all strains in HL stress (except for genes involved in the carbon fixation process), while clear differences can be observed between thermotypes in LT stress.

2.3.3 Expression of genes involved in photoprotection and response to oxidative stress

Exposure to HL and LT has been previously shown to trigger the production of reactive oxygen species (ROS), due to the excess of energy arriving on the PSII reaction center (Latifi *et al.*, 2009; Blot *et al.*, 2011; Pittera *et al.*, 2014). We thus explored the expression of genes involved in the dissipation of excess energy and the detoxification of ROS. As discussed in the INTRODUCTION, excess energy can be dissipated by heat, mainly by carotenoids localized close to the reaction center and by the orange carotenoid protein (for strains possessing it). Both carotenoid biosynthesis genes and the *ocp* and *frp* genes (involved in OCP-mediated NPQ) were upregulated in response to HL stress, and to a lesser extent in LT stress (Figure 54). Surprisingly, *ocp* was not differentially expressed in BL107 and WH7803 in LT stress, though this gene was upregulated in both strains in HL stress. Some of the genes encoding HLIPs (high-light inducible proteins), which are also believed to be involved in photoprotection, were also strongly upregulated in HL stress and to a lesser extent in LT stress. Thus, while photoprotection is, as expected, efficiently triggered by HL stress, its possible involvement in the response to LT is less clear, at least in the conditions of our experiment in which light intensity was quite low (20 µE m^{-2} s^{-1}).

HL stress			MVIR-18-1					A15-62					WH8102					BL107					WH7803				
	Gene name	Product	T0	T0.3	T1	T3	T6	T0	T0.3	T1	T3	T6	T0	T0.3	T1	T3	T6	T0	T0.3	T1	T3	T6	T0	T0.3	T1	T3	T6
Photo-protection	*crtW*	beta-carotene ketolase	0	3	2,8	2,1	1,6	-	-	-	-	-	0	5,7	5,4	5,2	5	0	4,6	4,2	3,8	3	0	5,8	5,7	4,4	3,7
	pds	phytoene dehydrogenase	0	0	1	1,5	1	0	1,7	2,5	2,8	2,2	0	2,2	2,8	3	3,2	0	1,2	1,9	2,5	2,3	0	1	1,4	1,4	1,1
	ocp	OCP	0	2	1,8	1,3	0	-	-	-	-	-	0	5,8	5,5	5,2	4,8	0	3,6	3,3	3,1	2,3	0	4,6	4,5	3,4	2,8
	frp	fluorescence recovery protein	0	2,2	1,9	1,6	1,2	-	-	-	-	-	0	4,1	4,2	4,3	4,1	0	4	4,1	4	3,4	0	3,7	3,6	2,6	2
	hli	HLIP	0	4,3	4,8	5,1	4,9	0	5,8	5,3	3,4	1,6	0	5,7	5,7	5,2	4,9	0	5,3	5,1	4,7	3,5	0	3	2,5	0	0
	hli	HLIP	0	5,9	6,4	6,4	5,6	0	7,1	6,6	5	3,5	0	6,9	6,7	6	5,6	0	7	7	6,6	5,8	0	4,4	3,9	3,3	2,9
	hli	HLIP	0	3,8	4,7	4,8	4	0	5,4	5,8	4,8	3,7	0	6,5	6,1	5,2	5,2	0	1,8	4,1	4,2	3,4	0	3,4	3,2	2,3	1,6
Oxidative stress	*trxB*	thioredoxin reductase (NTRC)	0	0	0	1,2	1,6	0	0	1,4	1,8	1,3	0	0	1,1	1,1	1,2	0	0	0	2,2	2,7	0	0	1	1,8	2,2
	prxQ	peroxiredoxin	0	0	1,9	2,6	2,4	0	1,4	2,6	3,2	2,8	0	1,2	1,8	2,1	2,4	0	1,1	2,3	2,8	2,6	0	1,1	1,3	1,2	1
	sodC	SOD [Cu-Zn]	0	0	1	1,4	1,5	0	0	0	0	0	-	-	-	-	-	0	0	0	1,3	1,3	0	0	0	0	0
	ahpC	2-Cys peroxiredoxin	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1,9	1,3	1,2	1,2	0	0	1,6	2,7	2,6
	sodX	Ni-type SOD maturation protease	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	-	-	-	-
	flv1	flavoproteins (Mehler reaction)	0	0	0	0	0	0	0	0	1	1,1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	flv3		0	0	0	0	0	0	0	0	1,1	1,1	0	0	1	1,3	1,6	0	0	1,3	1,9	1,8	0	0	0	0	0

LT stress			MVIR-18-1					A15-62					WH8102					BL107					WH7803				
	Gene name	Product	J0	J0.5	J1	J2	J3	J0	J0.5	J1	J2	J3	J0	J0.5	J1	J2	J3	J0	J0.5	J1	J2	J3	J0	J0.5	J1	J2	J3
Photo-protection	*crtW*	beta-carotene ketolase	0	2,5	2,5	1,9	1,7	0	-	-	-	-	0	1,6	2,9	3,1	2,5	0	1	0	0	0	0	1,1	1,5	1,3	1,3
	pds	phytoene dehydroge-se	0	0	0	0	0	0	1	0	0	0	0	1,1	1,1	0	0	0	1,3	1,9	1,8	1,6	0	0	0	0	0
	ocp	OCP	0	1,6	1,6	1,2	1,1	-	-	-	-	-	0	1,1	2,1	2,3	1,8	0	0	0	0	0	0	0	0	0	0
	frp	fluorescence recovery protein	0	1,6	1,7	1,2	0	-	-	-	-	-	0	0	1,9	2,9	2,4	0	0	0	0	0	0	0	0	0	0
	hli	HLIP	0	0	0	0	0	0	0	0	-1,2	0	0	0	1,6	1,4	1,4	0	1,1	1,2	0	0	0	0	0	0	0
	hli	HLIP	0	1,2	0	0	1,1	0	1	1,1	1,3	1,6	0	1,1	1,9	2,4	2,2	0	2,3	2,8	2,9	2,6	0	2,1	2,4	2,2	2,2
	hli	HLIP	0	1	0,4	0	0	0	0	0	0	0	0	1,8	2,6	2,3	2,3	0	0,8	1,2	1,5	1,4	0	0	0	0	0
Oxidative stress	*trxB*	thioredoxin reductase (NTRC)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	1,1	1,1	0	1,8	2,1	2	1,9
	prxQ	peroxiredoxin	0	0	1	1,1	1	0	0	0	0	0	0	0	0	0	1,4	0	0	1,3	1,5	1,3	0	1	1,1	1,1	1,2
	sodC	SOD [Cu-Zn]	0	0	0	0	0	0	0	-1,1	-1,3	-1,3	-	-	-	-	-	0	0	0	0	0	0	0	0	0	0
	ahpC	2-Cys peroxiredoxin	0	0	1,4	1,3	1	0	-1,5	-1,9	-2,2	-2,4	0	0	0	-1,9	-3,2	0	0	0	0	0	0	0	0	0	0
	sodX	Ni-type SOD maturation protease	0	0	1,3	1	1,1	0	2,3	2	1,6	1,9	0	0	0	0	0	0	0	0	0	0	-	-	-	-	-
	flv1	flavoproteins (Mehler reaction)	0	0	0	0	0	0	1,3	1,1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	flv3		0	0	0	0	0	0	1,3	1,7	2	2,1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 54 Expression of genes involved in photoprotection and response to oxidative stress. Representative genes were chosen for each category. Values correspond to rounded $\log_2$ of fold change compared to T0 (control condition). Any value between -1 and 1, corresponding to a fold change of 2, was considered as not differentially expressed. The upper panel shows response to HL stress, while the lower panel shows response to LT stress. Dashes were used when a gene was absent from a genome.

In HL stress, genes involved in the response to oxidative stress were upregulated only after 1h of stress, suggesting that generation of ROS is not immediately triggered by abrupt changes in light intensity (Figure 54). This response mainly involved peroxiredoxins (*prxQ*, *ahpC*), thioredoxin (*trxB*, NTRC complex), flavoproteins (*flv1* and *flv3*) and a Cu-Zn superoxide-dismutase (SOD,

sodC). The response was lower in LT stress and may rather involve the Ni-SOD since the Ni-type SOD maturation protein encoding gene *sodX* was upregulated in both MVIR-18-1 and A15-62. WH8102 displayed no differential expression of these genes but showed the highest OCP response. Thus, energy dissipation might be sufficient to avoid the generation of ROS in this strain.

2.3.4 Expression of genes from other functional categories

Cold temperatures are known to impact membrane rigidity, and bacteria are able to modify membrane fluidity as a response to changes in temperature, in particular by modifying the degree of unsaturation of membrane lipids (Varkey *et al.*, 2016; Pittera *et al.*). Surprisingly, we observed only little differential expression of genes involved in the synthesis of membrane lipids, except in strain A15-62 in which a number of these genes were upregulated (Figure 55). Two fatty-acid desaturases (*desA3* and *desC4*) were upregulated in WH7803 and one (*desA2*) in WH8102, while none of them was differentially expressed in other strains.

HL stress			MVIR-18-1					A15-62					WH8102					BL107					WH7803				
	Gene name	Product	T0	T0.3	T1	T3	T6	T0	T0.3	T1	T3	T6	T0	T0.3	T1	T3	T6	T0	T0.3	T1	T3	T6	T0	T0.3	T1	T3	T6
Lipid Desaturases	*desA2*	delta-12 fatty acid desaturase	-	-	-	-	-	0	0	0	1,2	0	0	0	0	1,5	1,8	-	-	-	-	-	0	0	0	0	0
	desC4	delta-9 fatty-acid desaturase	0	0	0	0	0	-	-	-	-	-	-	-	-	-	-	0	0	0	0	0	0	3,6	3,2	2,6	2,6
	desA3	delta-12 fatty acid desaturase	0	1,2	1,4	1,8	1,6	-	-	-	-	-	0	0	0	0	1,2	0	0	0	0	0	0	1,1	0	0	0
Chaperones	*dnaK3*	chaperone DnaK	0	0	0	-1,3	-1,7	0	1,5	0	0	0	0	2,1	2,1	1,7	1,7	0	0	2,3	1,6	1,1	0	2,2	2,6	1,8	2,1
	grpE	chaperone GrpE	0	0	0	0	0	0	1,2	0	0	0	0	1,8	1,8	1,7	2	0	0	1,8	1,2	0	0	1,7	2,1	1,8	2
	clpB1	ATP-dependent Clp protease subunit ClpB	0	0	0	0	0	0	0	0	0	0	0	1,7	1,9	1,3	1,4	0	0	2,7	1,5	0	0	1,7	2,2	0	1,8
	0	small hsp (HSP20) family	0	0	0	0	-1,1	0	0	0	0	0	0	1	1,3	1,1	1,5	0	0	3	1,7	1,2	0	1,4	2,9	1,6	2,2
	dnaJ	chaperone DnaJ	0	0	0	0	0	0	1,3	0	0	1,3	0	1,6	1,7	1,5	1,6	0	0	1,8	1,3	0	0	1,2	1,8	1,5	1,8
	groL2	chaperonin GroEL	0	0	0	-1,5	-1,9	0	0	0	0	0	0	1,4	1,3	0	1	0	0	1,5	1,2	1	0	1,9	2,3	1,7	2,2

LT stress			MVIR-18-1					A15-62					WH8102					BL107					WH7803				
	Gene name	Product	J0	J0.5	J1	J2	J3	J0	J0.5	J1	J2	J3	J0	J0.5	J1	J2	J3	J0	J0.5	J1	J2	J3	J0	J0.5	J1	J2	J3
Lipid Desaturases	*desA2*	delta-12 fatty acid desaturase	-	-	-	-	-	0	1	0	0	0	0	2	1,6	1,4	0	0	-	-	-	-	0	0	0	0	0
	desC4	delta-9 fatty-acid desaturase	0	0	-1	0	0	0	-	-	-	-	-	-	-	-	-	0	0	1,1	1,1	1,1	0	4,5	3,7	3,1	2,9
	desA3	delta-12 fatty acid desaturase	0	0	0	0	0	-	-	-	-	-	0	1,1	0	1	0	0	0	0	0	0	0	1,3	1,4	1,3	1
Chaperones	*dnaK3*	chaperone DnaK	0	-3,4	-2,1	-1,4	-1,2	0	0	0	0	0	0	1,8	1,7	0	0	0	-1,8	0	0	0	0	1,9	1,9	1,7	1,7
	grpE	chaperone GrpE	0	-1,2	0	0	0	0	0	1,2	1,3	1,4	0	1,9	2,4	2	0	0	0	0	0	0	0	1,6	1,6	1,5	1,4
	clpB1	ATP-dependent Clp protease subunit ClpB	0	0	0	0	0	0	0	1,4	2	2,3	0	0	0	1,2	0	0	0	0	1,1	1,4	0	2,4	2,9	2,4	2,2
	-	small hsp (HSP20) family	0	0	-1,1	-1	0	0	0	0	1,4	1,9	0	1,3	2,6	3,4	1,5	0	0	0	0	0	0	1,5	2,2	1,8	1,7
	dnaJ	chaperone DnaJ	0	-1,2	0	0	0	0	1,8	1,9	1,9	2	0	1,8	2,1	1,6	0	0	-1,6	0	0	0	0	1,9	2	1,8	1,8
	groL2	chaperonin GroEL	0	-2,4	-1,7	-1,1	0	0	0	0	0	0	0	1,1	1,2	0	0	0	-1,2	0	0	0	0	1,4	1,6	1,4	1,3

Figure 55 Expression of genes coding for lipid desaturases and chaperones. Representative genes were chosen for each category. Values correspond to rounded $\log_2$ of fold change compared to T0 (control condition). Any value between -1 and 1, corresponding to a fold change of 2, was considered as not differentially expressed. The upper panel shows response to HL stress, while the lower panel shows response to LT stress. Dashes were used when a gene was absent from a genome.

Genes encoding for chaperonins, proteins involved in the folding and/or stabilization of other proteins, also showed some interesting patterns. They were highly upregulated in response to HL stress in WH8102, BL107 and WH7803, i.e. the three strains that were the most impacted by this stress (according to F_V:F_M variation, see above), while they were not in the two other strains. In response to LT stress, they were upregulated in strains representative of warm thermotypes (A15-62, WH8102 and WH7803) but not in cold-adapted ones, in which they were often downregulated (Figure 55). Thus, the expression of these chaperone proteins is highly representative of the physiological status of the cells in response to these stresses.

2.4 Discussion

2.4.1 Physiological response to stress

Synechococcus cultures were submitted to stresses while they were in exponential growth phase, in a nutrient-rich medium, so they are not expected to be stressed by nutrient availability. Thus, the only factors potentially perturbing photosynthesis are light and/or temperature. HL stress experiments were performed at 22°C, a temperature at which all strains used here grow well according to previous work (Pittera, 2015; Varkey *et al.*, 2016; see also chapter I part 3), and thus only light is expected to play a role in the observed stress. Similarly, LT stress was performed under LL conditions, in which light stress is weak (see control on Figure 49) and only temperature is expected to play a role in this stress.

Strains were chosen in order to have representatives of the most abundant clades in the oceans, though only one strain within each clade was studied. While we know from previous work that differences exist between clades in their adaptation and ability to acclimate with regards to temperature (Zwirglmaier *et al.*, 2008; Mackey *et al.*, 2013; Pittera *et al.*, 2014; Varkey *et al.*, 2016), little is known on the differences in light physiology between *Synechococcus* clades (see INTRODUCTION), since pigmentation is not related to clades. As no clear depth partitioning has been described in this genus between clades or between pigment types, we made no *a priori* assumptions on whether some strains would be more sensitive to HL stress than others.

In this context, we observed a quite variable response to HL stress between strains, with WH8102 (clade III) being the most sensitive strain and A15-62 (clade II) and MVIR-18-1 (clade I) the most resistant strains. This better ability to stand large variation of light intensity can be related to the higher PSII repair rate measured in these strains. With only one strain per clade, it is too early to relate these differences to clade ecology, but this warrants further studies on variability in HL stress response both between clades and between pigment types. In the past decade, a number of studies focused on the differences between an “open-ocean” representative (usually clade III strain WH8102) and a “coastal” representative (usually clade I strain CC9311 or the clade IV strain CC9902) based on their isolation site (Palenik *et al.*, 2006; Stuart *et al.*, 2009; Tetu *et al.*, 2013; Stuart *et al.*, 2013). Here, by looking simultaneously at five strains we observed a range of stress responses that cannot be interpreted by a mere coastal/open-ocean dichotomy.

In contrast, the differences in physiological response to LT stress can clearly be interpreted as differences between clades, in light of previous knowledge on their distinct thermal adaptation. The members of the two cold thermotypes (clades I and IV) were the least impacted by LT and could maintain growth, contrary to members of warm thermotypes (clades II and III). While strain MVIR-18-1 (clade I) was known from previous work to be able to grow at 13°C and to stand cold stress (Pittera *et al.*, 2014), the data presented in this chapter and those in chapter I confirm that clade IV, for which no growth data was available below 18°C (Varkey *et al.*, 2016), can be considered as a cold thermotype. Interestingly, the clade IV strain was more affected by cold stress than the clade I strain, which is consistent with our finding that clade I predominates at higher latitudes that clade IV *in situ* (see Chapter 1), though one must be cautious when extrapolating data from a single strain to a whole clade. The fact that A15-62 (clade II), WH8102 (III) and WH7803 (V) cannot maintain growth at 13°C suggests that this temperature is at or below the lower limit of the range of temperatures at which these strains can grow, as expected from the temperature preferenda and boundary limits for growth previously determined for other strains of the same clades (Pittera *et al.*, 2014, see also chapter I part 3). Finally, all strains but MVIR-18-1 showed a dismantling of their phycobilisomes in response to cold stress. While this specificity of the clade I strain could be related to the higher flexibility of phycocyanin demonstrated in (Pittera *et al.*, 2016), it must be noted that the latter study also predicted a higher flexibility of clade IV phycocyanin, which is not reflected in our results.

2.4.2 Transcriptomic response to stress

Transcriptomic responses to HL and LT stresses share a number of similarities, as revealed by the number of genes displaying the same differential expression in both conditions. As far as we know, this is the first analysis of the global transcriptome comparing these two conditions in marine *Synechococcus*, and this result is quite consistent with the similarities in terms of physiology observed between acclimation to HL and LT (Mackey *et al.*, 2013; Pittera *et al.*, 2014; Mackey *et al.*, 2017), which both notably involve a strong downregulation of photosynthesis-related genes in both conditions (Six *et al.*, 2007b; Mackey *et al.*, 2013; Varkey *et al.*, 2016).

Only few differences were observed between strains at the transcriptomic level in response to HL stress in the functional categories we explored, except for chaperonins that were more expressed in the most affected strains. The response was much more contrasted in response to temperature, members of the warm thermotypes showing a much stronger stress response. In contrast, cold-adapted strains exhibited only a weak downregulation of photosynthesis-related genes, which shows that they were able to maintain the balance between the absorption of light energy and its use in downward processes. More specifically, the low temperature likely induces a lower metabolic rate that requires an adjustment of electron flux to match the lower demand and prevent PS damage and oxidative stress, and only strains of clade I and IV are able to make this adjustment. Alternatively, the metabolic rate could be less affected in strains of the latter clades due to a global adaptation of the metabolism to cold temperatures. Our results also suggest a lower photoinactivation of PSII in these strains, which would thus not need to reduce the amount of

energy reaching PSII reaction center. (Pittera *et al.*, 2014) hypothesized that this ability of cold thermotypes to maintain energetic balance was due to a more efficient non-photochemical quenching of fluorescence. Expression of genes involved in this process does not show any clear difference between thermotypes, but we cannot rule out the occurrence of post-transcriptional differences, in particular in the OCP recycling process.

In a comparison of proteomes of cultures acclimated to 18°C and 22°C, (Varkey *et al.*, 2016) suggested that the better resistance to cold of the clade IV strain BL107 was in large part due to its ability to maintain a high expression of proteins involved in translation. In our analysis, we did not see any difference between thermotypes concerning these genes (not shown), but noticed a strong difference in the expression of chaperones. Their overexpression is observed in response to a large variety of stresses (Tetu *et al.*, 2013), but is of particular interest in the context of temperature stress. Since temperature influences protein flexibility and folding, the upregulation of chaperone-encoding genes in warm-adapted strains and their downregulation in cold-adapted ones during LT stress could indicate that protein stability is disturbed at 13°C for the former, and at 22°C for the latter.

In conclusion, this preliminary analysis of the ecotypic variability of the stress response lays the ground for a deeper exploration of the mechanisms involved. Physiological analyses will benefit from the on-going characterization of pigment content and lipid composition during stress (performed by C. Six and S. Breton), while the analysis of the expression of a number of genes belonging to functional categories that were not analyzed here, in particular those displaying a strain-specific response, will certainly reveal further differences both between strains and between stresses.

3. Effect of temperature on growth in light:dark cycle

3.1 Shift in cell cycle timing

Growth of the model strain WH7803 was monitored along a light-dark cycle (12h:12h) at two temperatures, 21°C and 27°C. Flow cytometry was used to identify the proportion of cells in the different phases of the cell cycle. In the experimental conditions tested, this cycle is quite complex since populations with 1 (1C), 2 (2C), 3 (3C) and 4 (4C) copies of the chromosome coexist. Cell-sorting by flow cytometry coupled with microscopic observations showed that 3C and 4C populations are not composed of single cells with 3 or 4 chromosomes but instead of cells attached by pairs, containing either 1 or 2 chromosomes (Figure 56). These cells could not be separated by sonication, and thus are probably not fully individualized.

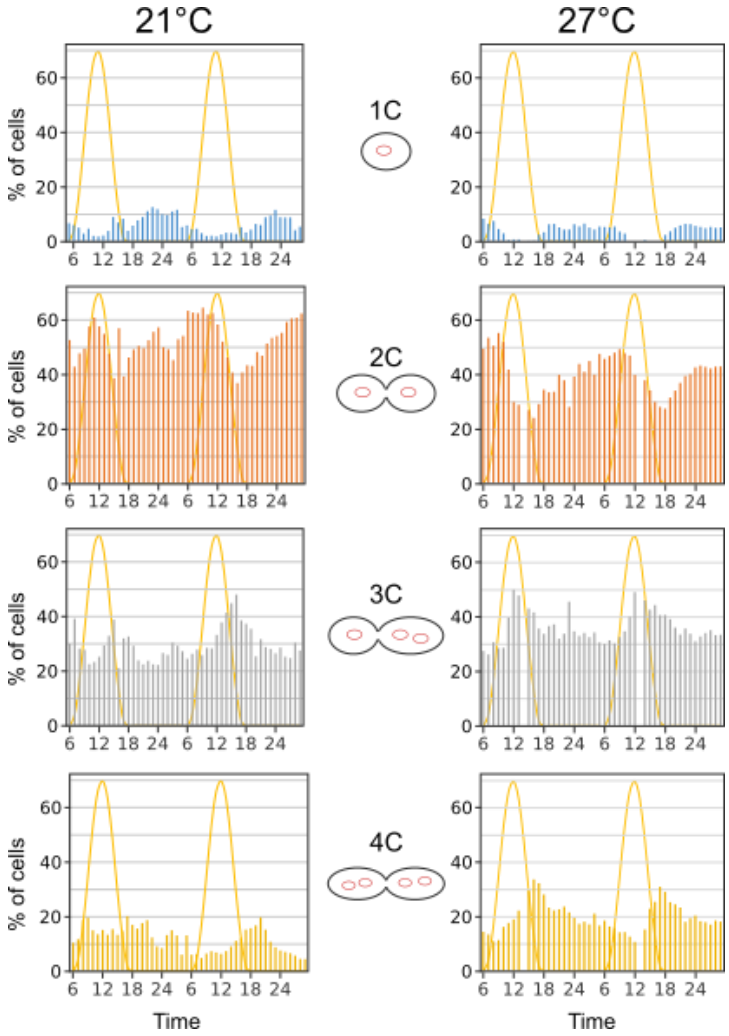


Figure 56 WH7803 cell cycle over a light :dark cycle at 21°C and 27°C. The proportion of each population is plotted for two consecutive days, overlayed on the curve of the light intensity. A drawing indicates the shape of cells for each population. 1C, 2C, 3C, 4C: 1, 2, 3 or 4 chromosomes.

The variation of the proportion of each population (Figure 56) along one light:dark cycle allowed us to propose a model of cell cycle dynamics. At the end of the dark period, 1C cells begin to synthesize DNA, increasing the proportion of 2C cells. Later during daytime, 2C cells synthesize new chromosomes, increasing the proportion of 3C and potentially 4C cells. Finally, in the late afternoon 3C cells synthesize a fourth chromosome, increasing the proportion of 4C cells. These different synthesis phases are overlapping, since not all the cells in each population have the exact same timing. At the light-to-dark transition, cell division begins, with 4C cells dividing into two 2C cells (and/or potentially into 1C cells), and 2C cells dividing into 1C cells, while 3C cells probably divide into 2C and 1C cells. This division phase lasts until midnight, and is followed by a phase in which the proportion of the different populations stays quite stable, most cells probably being in G1 phase. It is however worth noting that a significant proportion stays in an asymmetrical 3C state and to a lesser extent in a 4C state throughout the light:dark cycle, which could correspond either to unsynchronized cells or to cells having a different timing (e.g. cycling on two consecutive days).

The expression of genes involved in cell cycle and cell division is quite consistent with this succession of steps. The *ftsZ* gene, which codes for a protein forming a ring at the future septum site and is usually expressed during the S phase (Holtzendorff *et al.*, 2001; Zinser *et al.*, 2009), shows a maximal expression at 12h00 and 15h00 (Figure 57), corresponding to the increase in the proportion of 3C and 4C cells. Genes involved in DNA replication (e.g. dnaA) show only little differential expression along the light:dark cycle, but it is also consistent with an early S phase, with only a small peak between 9h00 and 12h00 (Figure 57). The sepF and ftsI genes, which are involved in cell separation, show a peak at 18h with a significant expression maintained until 22h00, corresponding to the timing of cell division observed by flow cytometry.

Interestingly, while the same sequence was observed at both temperatures, the DNA synthesis phase (as predicted from flow cytometry data) was shifted at 21°C compared to 27°C. The transition from 1C to 2C started earlier (around 4 am against 6 am at 27°C), while the other synthesis steps occurred 2 hours later compared to 27°C. Despite this shift, cell division started at the light-to-dark transition at both temperatures, but there was likely an overlap between the end of S phase (2C and 3C to 4C) and the division phase at 21°C, as shown by the fact that the proportion of 4C cells still increases between 18h00 and 20h00 at this temperature. Finally, we also observed a difference in the proportion of each population between the two temperatures: there were in particular less 4C and more 2C cells at 21°C, which is consistent with a lag of S phase preventing most 2C/3C cells to synthesize new chromosomes before division at the light-to-dark transition. Thus at colder temperature, we observed a lag in cell growth and DNA synthesis, which is likely due to a slower metabolism, and might explain the slightly lower growth rate measured at this temperature (see Figure 45).

Overall, the cell cycle of WH7803 seems quite complex. Previous studies on this strain have reported either a bimodal DNA distribution (Mella-Flores *et al.*, 2012) or a multimodal distribution similar to the one we observed (Binder and Chisholm, 1995), but the determinant of these modes has not been identified yet. Moreover, the timing of WH7803 cell division differed between studies,

occurring at the beginning of the afternoon, in the middle of the afternoon or just after light-to-dark transition (see (Scanlan, 2003) for review). Thus, additional studies are needed to identify the determinants of both division mode and division timing.

	Gene name	Product	21°C 6h	9h	12h	15h	18h	20h	22h	2h	27°C 6h	9h	12h	15h	18h	20h	22h	2h
Cell division	*ftsZ*	cell division protein FtsZ	0	0	1,18	1,66	0	0	0	0	0	1,17	1,33	1,94	0	0	0	0
	minC	septum formation inhibitor	0	0	0	0	1,46	0	0	0	0	0	0	0	1,17	0	0	0
	minD	septum site-determining protein	0	0	0	1,22	0	-1,1	-1,1	0	0	0	0	1,76	0	0	0	0
	sepF	cell division protein SepF	0	0	0	0	2,21	1,34	0	0	0	0	0	0	2,31	1,05	0	0
	ftsI	cell division peptidoglycan synthetase	0	0	0	0	2,44	2,21	1,23	0	0	0	0	1,54	3,25	2,17	1,66	0
DNA replication	*dnaA*	chromosomal replication initiator DnaA	0	1,07	0	0	-1,7	0	0	0	0	1,58	1,62	0	0	0	0	0
	dnaB	replicative DNA helicase	0	0	0	0	0	0	0	0	0	0	-1,1	0	0	0	0	0
	polA	DNA polymerase I	0	1,04	0	0	0	0	0	0	0	1,06	0	0	-1	0	0	0

Figure 57 Expression of genes involved in cell division and DNA replication over a light:dark cycle at 21°C and 27°C. Representative genes were chosen within each category. Values of expression correspond to $\log_2$ of fold change compared to 6h. Any value between -1 and 1, corresponding to a fold change of 2, was considered as not differentially expressed.

3.2 Photophysiology along a light:dark cycle

In light:dark cycle, light physiology of *Synechococcus* is strongly influenced by the daily variations of light intensity (Mella-Flores *et al.*, 2012). As expected, at both temperatures PSII efficiency decreased in the morning and increased in the afternoon, mirroring light intensity (Figure 58). Several differences however arose when comparing the two temperatures. First, FV/FM was higher at the dark-to-light transition at 21°C than at 27°C (mean: 0.70 vs 0.60, p=0.0143 in one-sided Wilcoxon test, not visible on Figure 58 because of normalization). Second, the minimal PSII efficiency was observed from 9h to 12h at 27°C, temperature at which the decrease was sharper, while it was observed at 12h at 21°C. Third, a peak of F_V:F_M was observed at 18h at 27°C only, followed by a continuous decrease over night.

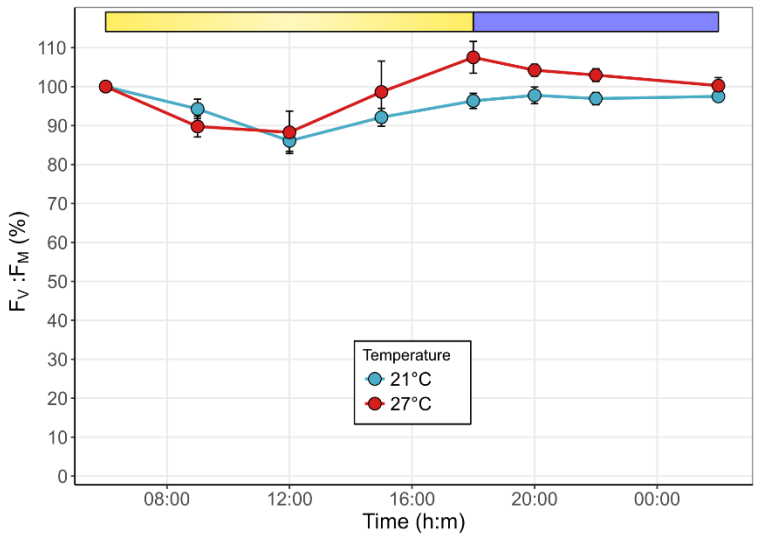


Figure 58 Variation of photosystem II efficiency along a light:dark cycle at 21°C and 27°C. Error bars indicate standard deviation.

Although the decrease in PSII photosynthetic efficiency is similar in both strains, looking closely at R_{PSII} shows that photoinactivation is actually lower at 21°C than at 27°C (the decrease in $F_V:F_M$ in presence of lincomycin is steeper at 27°C, Figure 59A) but R_{PSII} is higher at 27°C (Figure 59B). Thus, the rate of PSII repair matches the rate of photoinactivation and allows in both cases to maintain PSII efficiency within 80% of the initial value (Figure 58). At 27°C, this compensation is already reached at 9h00, while at 21°C it is only 12h00 from 12h. When light decreases during the afternoon, R_{PSII} also decreases to reach a near-zero value at the light-to-dark transition, a minimal level that was previously shown to be maintained during the whole dark period (Mella-Flores *et al.*, 2012).

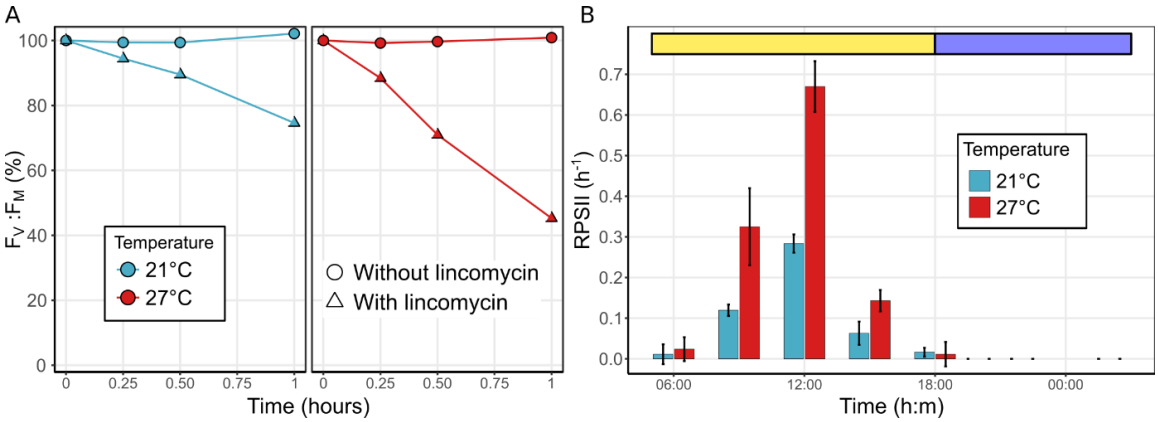


Figure 59 PSII repair rate along a light:dark cycle at 21°C and 27°C. A. Examples of PSII inactivation with (triangles) and without (circles) lincomycin at 21°C and 27°C, obtained at 12h00. B. R_{PSII} measured at each time point along the light:dark cycle. R_{PSII} was not measured at night as it is known to be close to zero (Mella-Flores *et al.*, 2012).

The expression pattern of genes involved is this process is very consistent with physiological measurements: the gene encoding the D1.1 isoform is downregulated during the day with a minimal expression at 12h, and a maximal expression during the whole dark period, while the expression of D1.2 isoform-encoding genes shows the exact reverse pattern (Figure 60).

			21°C								27°C							
	Gene name	Product	6h	9h	12h	15h	18h	20h	22h	2h	6h	9h	12h	15h	18h	20h	22h	2h
D1 repair	*psbA*	PSII protein D1.2	0	2,6	3,19	2,74	0	0	0	0	0	0	0	0	0	-1,1	0	0
	psbA	PSII protein D1.2	0	1,88	2,49	1,46	0	-1,5	-1,2	0	0	0	1,19	1,05	0	0	0	0
	psbA	PSII protein D1.2	0	2,86	3,62	3,29	0	0	0	0	0	1,14	1,49	1,47	-1,1	0	0	0
	ftsH1	possible PSII D1 repair protein	0	0	1,18	0	0	0	0	0	0	0	0	0	0	0	0	0
	ftsH2	PSII D1 repair protein	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 60 Expression of genes involved in D1 repair over a light:dark cycle at 21°C and 27°C. Values of expression correspond to log2 of fold change compared to 6h. Any value between -1 and 1, corresponding to a fold change of 2, was considered as not differentially expressed.

Our observation that the PSII repair rate is lower at 21°C could be explained both by the lower photoinactivation rate of PSII and by the effect of temperature, which slows down the translation of the D1 protein (Murata *et al.*, 2007). Interestingly, despite a higher photoinactivation at 27°C, changes in the expression of D1-encoding genes were higher at 21°C (Figure 60). Thus it is possible that at 21°C, WH7803 has to compensate the slower translation of D1 by an increase in transcription of the gene, though another explanation could be a continuous higher transcription rate of D1 genes at 27°C throughout the light:dark cycle (hidden by the normalization at 6h00).

Aside from the specific process of D1 repair, all genes involved in light energy harvesting and transmission follow more or less the same expression pattern (Figure 61): the expression of genes involved in PSII, PSI and phycobilisome biosynthesis as well as those involved in electron transport increases during the day to reach a maximum around 15h. This is opposite to what is observed during a high-light stress (see above) and rather suggests an optimization of light harvesting during the day. However, the increasing light intensity still promotes the production of ROS, as revealed by the triggering of the expression of photoprotection and ROS detoxification genes, which is maximal between 12h and 15h (Figure 61).

	Gene name	Product	21°C 6h	9h	12h	15h	18h	20h	22h	2h	27°C 6h	9h	12h	15h	18h	20h	22h	2h
PSII	*psbB*	PSII protein CP47	0	0	0	0	0	-1	-1	0	0	0	1,03	1,88	0	0	0	0
	psbC	PSII CP43 protein	0	0	0	1,31	1,03	0	0	0	0	0	1,14	1,85	0	0	0	0
	psbD	PSII D2 protein	0	0	1,17	1,85	0	-1,1	-1,6	0	0	0	1,43	2,28	0	0	0	0
PSI	*psaA*	PSI core protein PsaA	0	0	0	1,49	1,74	0	0	0	0	0	0	2,17	1,55	0	0	0
	psaB	PSI core protein PsaB	0	0	0	1,4	1,75	0	0	0	0	0	0	2,1	1,65	0	0	0
	psaC	PSI Fe-S center subunit VII	0	0	0	1,21	0	-1,3	-1,5	0	0	0	1,2	1,93	1,14	0	0	0
ETP	*ndhD2*	NADH dehydrogenase I subunit	0	1,54	1,81	0	-1,7	0	0	0	0	1,49	1,17	0	-1,2	0	0	0
	petJ	cytochrome c553 (c6)	0	0	0	1,39	2,13	0	0	0	0	0	0	0	1,23	0	0	0
	petB	cytochrome b6	0	0	0	0	0	-1,3	-1,2	0	0	0	0	1,43	0	0	0	0
	petF	2Fe-2S ferredoxin	0	2,23	2,59	2,26	0	0	0	0	0	2,52	2,95	2,46	0	0	0	0
	atp1	ATP synthase protein I	0	1,87	1,71	1,63	-1,5	-1,6	0	0	0	2,91	2,86	2,68	0	0	0	0
	atpA	ATP synthase alpha chain	0	2,06	1,88	1,65	0	0	0	0	0	3,38	3,2	2,88	0	0	0	0
PBPs & lyases	*apcF*	APC beta-18 chain	0	0	0	1,68	0	-1,5	-1,4	0	0	0	1,55	2,37	0	0	0	0
	cpcB	PC beta chain	0	0	0	2,16	1,12	0	-1,2	0	0	0	1,47	2,98	1,07	0	0	0
	cpeA	C-PE class I. alpha subunit	0	0	0	2,62	2,71	0	0	0	0	0	1,05	3,2	1,24	0	0	0
	mpeA	C-PE class II. alpha chain	0	0	-1	2,41	2,44	0	0	0	0	0	0	3,68	1,78	1,16	0	0
	cpeS	PEB:Cys-82 beta-PE lyase	0	0	0	2,09	0	0	0	0	0	0	0	2,25	1,14	0	0	0
	cpeZ	PEB:Cys-82 alpha-PE lyase	0	0	0	2,01	0	-1,2	0	0	0	0	1,15	2,52	0	0	0	0
Photoprotection & ROS detoxification	*grx3*	monothiol glutaredoxin	0	1,15	1,38	1,06	0	0	0	0	0	0	0	0	0	0	0	0
	gst	glutathione S-transferase	0	1,02	2,12	2,48	0	0	0	0	0	0	1,74	2,17	0	0	0	0
	ocp	OCP	0	2,41	2,4	0	-1,2	0	0	0	0	1,05	0	-1,2	-1,4	-1,2	0	0
	frp	fluorescence recovery protein	0	1,06	1,16	0	0	0	0	0	0	0	0	0	0	0	0	0
	katG	catalase/ peroxidase	0	0	0	1,99	2,87	1,07	0	0	0	0	0	2,48	2,46	0	0	0
	prxQ	peroxiredoxin	0	1,05	1,69	1,96	0	0	0	0	0	1,17	1,88	1,78	0	0	0	0
	sodB	SOD [Fe]	0	0	1,86	1,64	0	0	0	0	0	0	1,35	1,06	0	0	0	0
	sodC	SOD [Cu-Zn]	0	0	0	-1,5	-2,1	0	0	0	0	0	-1,2	-1,9	-1,8	-1,1	0	0
	ahpC	2-Cys peroxiredoxin	0	3,28	4,96	3,94	0	0	0	0	0	2,51	3,61	1,97	0	0	0	0

Figure 61 Expression of genes involved in absorption and transfer of light energy over a light:dark cycle at 21°C and 27°C. Values of expression correspond to log2 of fold change compared to 6h. Any value between -1 and 1, corresponding to a fold change of 2, was considered as not differentially expressed. ETP: electron transport. The NADH dehydrogenase subunit D2 isoform is involved in photosynthesis while the isoform D1 is involved in respiration. The two ATP synthase genes presented here seem to be expressed primarily for photosynthesis, though they could also be involved in respiration.

3.3 Metabolic timing along the light:dark cycle

Along with the regulation of photophysiology, the analysis of the transcriptome reveals that the whole metabolism is synchronized with the modulation of the light intensity. While genes involved in light harvesting are upregulated along the day, the expression of genes involved in carbon fixation peaks at 9h and stays high until 15h (Figure 62). A similar pattern was observed in previous experiments on *Synechococcus* and *Prochlorococcus* (Zinser *et al.*, 2009; Mella-Flores *et al.*, 2012) and was interpreted as a preparation of cells to an increase in energy supply. This early expression of carbon fixation genes allows to maintain the balance between energy supply and energy demand, and probably explains why light-dark acclimated cultures did not behave as in HL stress when exposed to light at 6h00.

	Gene name	Product	21°C								27°C							
			6h	9h	12h	15h	18h	20h	22h	2h	6h	9h	12h	15h	18h	20h	22h	2h
Carbon fixation	*csoS1A*	carboxysome shell protein CsoS1A	0	2,08	1,86	0	-2,7	-2,1	-1,1	0	0	2,53	2,29	1,37	-1,4	0	0	0
	cupB	CO2 hydration protein	0	0	0	0	0	0	0	0	0	2,01	2,1	2,53	0	0	0	0
	rbcL	RuBisCO large subunit	0	2,24	2,3	1,08	-2,1	-1,8	0	0	0	2,93	2,92	1,75	0	0	0	0
Glycogen synthesis	*glgA*	glycogen synthase	0	1,39	1,54	2,04	-1,6	-1,8	-1,6	0	0	1,59	2,26	2,72	0	0	0	0
Carbon catabolism	*glgP*	glycogen phosphorylase	0	0	0	1,79	2,49	1,48	0	0	0	0	0	1,34	2,03	0	0	0
	ctaA	cytochrome oxidase biogenesis protein	0	-1,9	-2,1	2,77	6,66	4,02	1,5	0	0	-3,2	-3,9	1,25	2,53	0	0	0
	ctaB	protoheme IX farnesyltransferase	0	-1,4	-1,6	2,16	6,73	3,93	1,5	0	0	-2,9	-3,5	1,21	2,54	0	0	0
	ctaCI	cytochrome c oxidase subunit II	0	-2	-1,2	2,44	7,12	5,81	3,95	1,04	0	-3	-2,6	1,65	4,1	2,53	2,31	0
	ctaD1	cytochrome c oxidase subunit I	0	-1,5	-2	1,45	6,64	5,55	3,82	1,05	0	-2,9	-3,6	0	3,68	2,19	1,94	0
	ndhD2	NADH dehydrogenase I subunit	0	1,54	1,81	0	-1,7	0	0	0	0	1,49	1,17	0	-1,2	0	0	0

Figure 62 Expression of genes involved in carbon metabolism over a light:dark cycle at 21°C and 27°C. Values of expression correspond to log2 of fold change compared to 6h. Any value between -1 and 1, corresponding to a fold change of 2, was considered as not differentially expressed. The NADH dehydrogenase subunit D2 isoform is involved in photosynthesis while the isoform D1 is involved in respiration.

In opposition to the morning-centered carbon fixation, respiration occurs mainly at the light-to-dark transition (18h00), when photosynthesis cannot supply the cell with the energy necessary for cell division, as shown by the expression pattern of genes involved in glycogen degradation and respiratory electron transport (Figure 62). This is consistent with what was observed in *Prochlorococcus,* with carbon fixation and storage being enhanced during the day, and carbon reserves being used at night (Zinser *et al.*, 2009).

We also analyzed the expression of genes involved in nutrient acquisition, which reflects nutrient demand for the biosynthesis of cell components. The expression of phosphate-acquisition genes shows two peaks, the first between 12h and 15h and the second between 20h and 22h, the minimum being observed at the end of dark period (2h00 to 6h00) (Figure 63). The first peak is likely related to the P demand for DNA synthesis, while the second follows division and could be linked to a number of other biosynthesis pathways. Concerning nitrogen, the expression of the ammonium transporter, the source of N preferentially used by *Synechococcus* cells, increases during the day when C fixation allows to build cell biomass, is maximal at the light-to-dark transition when division occurs, and decreases to reach a minimum at the end of the night. Surprisingly, genes involved in the use of other sources of N, including cyanate and nitrate, were also strongly upregulated from 9h00 to 18h00 (Figure 63). This was not expected as additional sources of N are not believed to be used when ammonium is present (Moore et al., 2002), which was always the case in our experiment due to the continuous input of fresh medium. Two complementary explanations can be given to this observation: first, a change is the balance between N and C availability, which is sensed by the cell (see INTRODUCTION), could trigger an N stress-like response as suggested by the overexpression of the regulator ntcA. Second, nitrate reductase and nitrite reductase might be used to oxidize ferredoxin and thus divert excess energy from photosynthesis (Flores *et al.*, 2005).

	Gene name	Product	21°C 6h	9h	12h	15h	18h	20h	22h	2h	27°C 6h	9h	12h	15h	18h	20h	22h	2h
Phosphate	*pstA*	ABC-type phosphate transport system permease component	0	0	0	0	0	0	0	0	0	2,29	2,1	2,11	1,81	1,08	1,03	0
	pstS	phosphate ABC transporter, phosphate-binding protein PstS	0	1,83	3,2	3,27	1,87	3,09	2,63	1,03	0	0	2,26	2,92	1,74	2,28	1,57	0
	sphX	ABC-type phosphate transporter, substrate binding component, cyanobacterial-specific	0	0	0	1,63	5,15	3,78	2,36	0	0	-2	-1,9	0	2,31	1,45	1,2	0
	phoH	phosphate starvation-inducible protein	0	1,34	1,25	0	0	1,68	1,89	0	0	0	0	0	0	0	0	0
	phoD	phosphodiesterase/ alkaline phosphatase D	0	0	1,2	1,37	1,98	0	0	0	0	0	0	1,18	1,04	0	0	0
	ptrA	transcriptional phosphate regulator	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Nitrogen	*amt1*	ammonium transporter	0	2,13	2,1	2,95	4,59	3,15	2,42	0	0	1,39	1,06	1,46	1,64	0	0	0
	cynS	cyanate hydratase	0	4,57	3,7	3,99	3,72	1,74	0	0	0	2,79	2,9	2,18	0	0	0	0
	nrtP	nitrate transporter	0	4,36	3,78	3,86	2,29	0	1,12	0	0	3,42	3,18	2,28	0	0	0	0
	nirA	ferredoxin-nitrite reductase	0	4,35	4,06	4,08	3,3	1,11	0	0	0	3,29	3,28	2,75	0	0	0	0
	glsF	ferredoxin-dependent glutamate synthase	0	0	0	-1	-1,2	0	0	0	0	0	0	-1,6	-1,6	0	0	0
	glnA	glutamine synthetase type I	0	1,78	1,91	2,23	1,48	0	0	0	0	2,33	2,16	1,76	0	0	0	0
	glnB	nitrogen regulatory protein P-II	0	0	0	1,93	1,53	0	0	0	0	0	0	0	0	0	0	0
	ntcA	global nitrogen regulatory protein	0	2,7	2,67	2,03	0	0	0	2,37	0	2,55	1,85	0	0	0	0	0

Figure 63 Expression of genes involved in N and P assimilation over a light:dark cycle at 21°C and 27°C. Values of expression correspond to log2 of fold change compared to 6h. Any value between -1 and 1, corresponding to a fold change of 2, was considered as not differentially expressed.

3.4 Conclusion and perspectives

This on-going analysis of physiological and transcriptomic response along a light-dark cycle allowed us to describe the physiology of WH7803 in light-dark experiments at two distinct temperature, in terms of cell cycle, response to light and global metabolism. This should allow us to decipher whether growth below and above temperature optimum (i.e. 25 °C, see Figure 45) implies different constraints. Preliminary results seem to show that the differences between these two conditions mainly happens at night and between transitions between light and dark periods (not shown), but the genes involved in these differences have not been analyzed in detail yet. Their analysis and a comparison with the transcriptomic data obtained in response to abrupt shifts in light and temperature will certainly bring new insights on how this model strain acclimates to varying environmental conditions.

CONCLUSION

&

PERSPECTIVES

« Qui n'avance pas recule »

Jacques Brel – A jeun

1. Diversity level and its link to niche adaptation

1.1 What is the link between ecological niche occupancy and genetic diversity?

One of the major results of the analyses of the environmental distribution of picocyanobacterial lineages is the identification of an important microdiversity within both *Synechococcus* and *Prochlorococcus* genera, and more importantly the fact that this diversity is ecologically relevant, in the sense that it can be related to functional traits, such as the capacity to thrive in Fe-depleted areas. The use of the term "microdiversity" is somewhat confusing, since it does not have a clear definition. It is usually defined as the diversity encountered within closely related taxonomic units (e.g. > 97% similar 16S rRNA gene; (Larkin and Martiny, 2017)). Here, we use this term to designate taxonomical units defined within each clade based on differences in their global distribution patterns, probably reflecting different ecological niche *preferenda*. This method, based on the high resolution marker *petB,* allowed us to delineate niche boundaries and identify the main physico-chemical parameters influencing the distribution of picocyanobacteria more precisely than if units were defined solely on the base of an arbitrary identity cut-off or by their position in a phylogenetic tree. Although ESTUs were defined using a single marker gene, the fact that each ESTU seems to have its own gene repertoire *in situ*, in particular for *Synechococcus*, tends to suggest that this link between genetic and ecological units is relevant. Still, the definition of ESTUs and the appraisal of their corresponding niches remain perfectible, in particular for clades that were present in low abundance in the *Tara* Oceans dataset (e.g. clades thriving at high latitude or in coastal areas). ESTU classification will thus need to evolve as new data from a variety of environments become available.

1.2 Where are we on the path to niche delineation?

The datasets of *petB* metabarcodes or metagenomes that were analyzed during this PhD cover a wide variety of environments, extending until 80°N and encompassing both open-ocean and coastal areas (Figure 64A). This coverage allowed us to get a global view of the biogeography of marine picocyanobacteria and brought new insights into the ecology of some clades. This is particularly true for *Synechococcus*, for which many clades had a poorly known distribution due to the use of markers of low resolution: for example, clades V/VI/VII/CRD1 could not be differentiated using 16S oligonucleotide probes used in previous studies (Zwirglmaier *et al.*, 2008; Mella-Flores *et al.*, 2011). The global picture of the distribution of picocyanobacterial lineages we obtained (Figure 64B-C) clearly shows that different populations occupy distinct niches. Still, determining the precise boundaries of each niche, as defined as a multidimensional space whose dimensions comprise all sorts of abiotic and biotic parameters (see the introduction of CHAPITRE I), remains quite challenging.

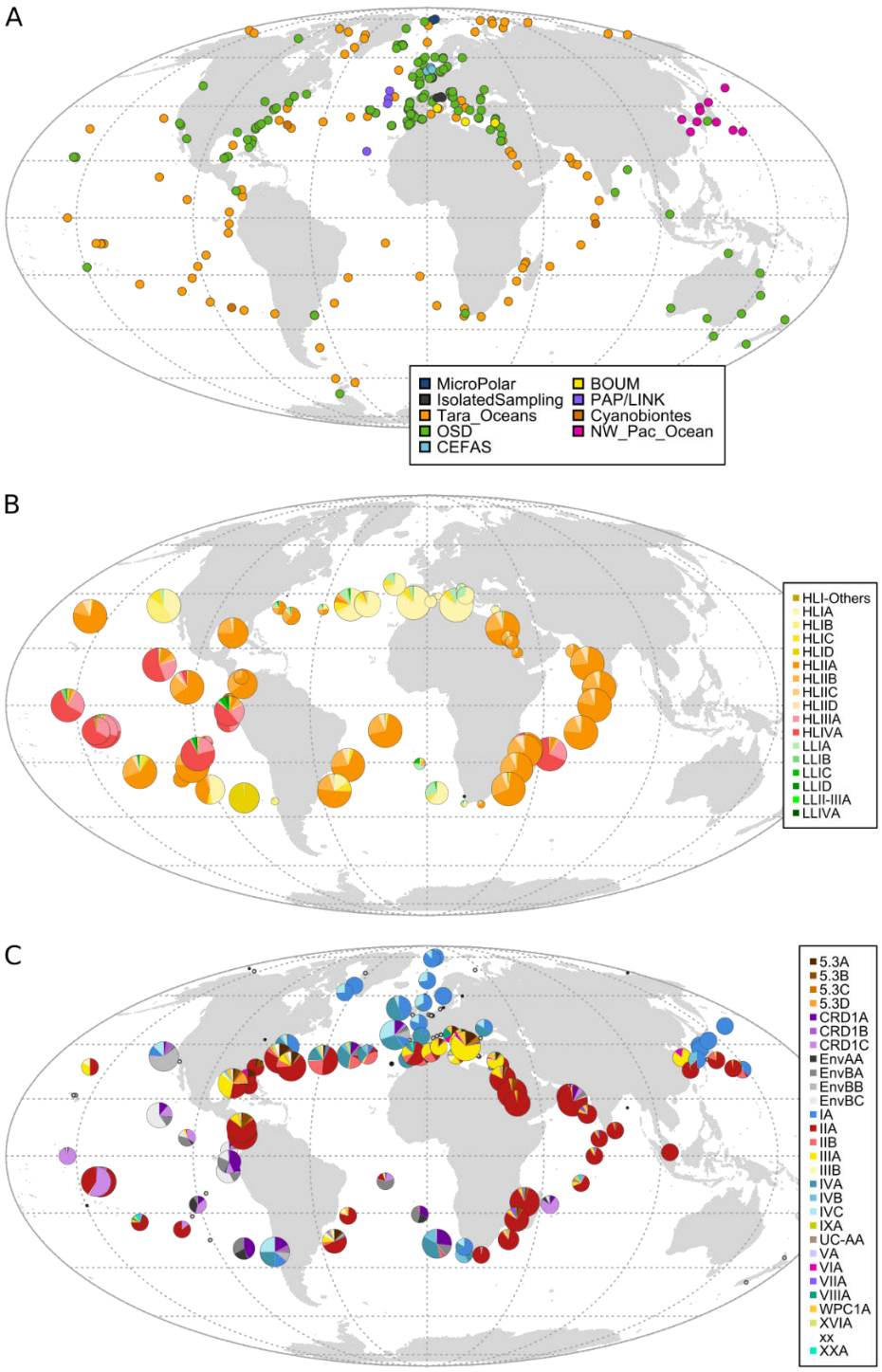


Figure 64 Worldwide distribution of marine picocyanobacterial taxa as assessed using *petB* gene. A. Transects of all the cruises and sampling campaign for which *petB* metabarcodes or *petB* miTags were analysed during this phD. B. Global distribution of *Prochlorococcus* ESTUs. The size of circles is proportional to the number of *petB* reads. Only *Tara* Oceans transect is shown, as *Prochlorococcus petB* is not well amplified by primers used for metabarcodes, and this genus was almost absent from OSD dataset. C. Global distribution of *Synechococcus* ESTUs.

Here, we managed to determine some broad characteristics of the niches of the most abundant ESTUs and some of the less abundant ones, based on the environmental parameters that were available in our datasets. However, we are still quite far from having a complete picture of niche limits, in part because correlation analyses cannot detect non-linear relationships, but also (and maybe mostly) because we lack information on many other environmental parameters. This is particularly the case for the effects of biotic interactions on picocyanobacterial distribution. Predators, be it grazers or phages, have a major impact on survival *in situ* since their predatory activity is sufficient to balance cellular growth of picocyanobacteria (Partensky *et al.*, 1999a; Hunter-Cevera *et al.*, 2014). They are also believed to constitute a strong selective agent, as revealed by the preponderance of cell-surface modification genes in strain-specific gene sets and the differential geographic distribution of these genes (see CHAPITRE II), thus being one of the drivers of genetic diversification (both in terms of gene content and of sequence modification (Lindell *et al.*, 2004; Avrani *et al.*, 2011)). Finally, phages carry genes whose expression can affect host metabolism (e.g. genes involved in photosynthesis (Sullivan *et al.*, 2005, 2006; Lindell *et al.*, 2005a; Fridman *et al.*, 2017)) and how this interaction affects the niche of the host and of the phage is unclear. More generally, as long as the extent of biotic interactions and the identification of picocyanobacterial partners and their respective niches are unknown, it will remain difficult to determine what part of the observed distribution is due to a top-down control and what part to a bottom-up control. It is thus likely that a better knowledge of biotic interactions would explain the part of the variance that remains unexplained with abiotic parameters.

Another important aspect of our approach, in which niches are described from distribution patterns (i.e. environmental realized niches), is that niche limits correspond to conditions in which each unit was found, not necessarily coinciding with conditions in which it can actively grow since cells can for example be taken outside of their niche by currents. Thus, while the niches we determined reflect an environmental reality, experimental work will be needed to evaluate whether their limits correspond to the physiology of cells (i.e. their fundamental niche). Still, even though niche boundaries remain to be refined, the work performed in the framework of this PhD improved our knowledge of ecological preferences of the different picocyanobacterial lineages, and brought invaluable information to interpret both genomic and physiological data in terms of adaptation, as we did in CHAPITRE II andCHAPITRE III.

Another important goal of niche-partitioning analyses is to delineate population units with well constrained niches, on which selection primarily acts. *Prochlorococcus* and *Synechococcus* have distinct, though overlapping niches in the ocean, whose boundaries are defined by the limits of the niches of large taxonomical units (e.g., clades), themselves comprising a number of sub-units with different niches (Kashtan *et al.*, 2014; Larkin *et al.*, 2016; Farrant *et al.*, 2016). Thus, the global niche of the *Synechococcus* genus can be seen as a series of "Russian dolls", with nested units of decreasing sizes exhibiting narrower and narrower ecological niches (Figure 65). Many advances have been made in the study of marine picocyanobacteria by defining ever smaller and more closely related units with distinct environmental niches, but the question remains open whether the ultimate population unit will one day be identified. Alongside with niche delineation, this would also greatly improve our understanding of evolutionary dynamics of these units over short time scales.

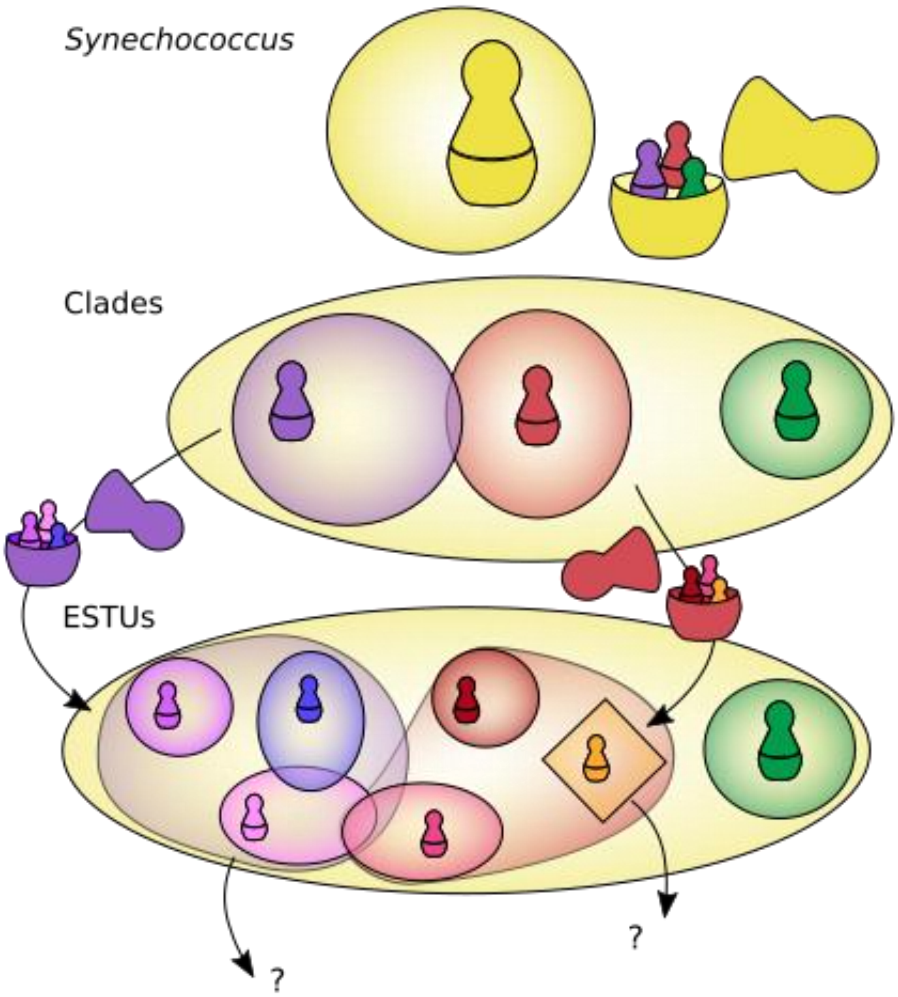


Figure 65 A simple model for niche partitioning in *Synechococcus*: a series of Russian dolls. Each taxonomic unit is represented as a Russian doll, surrounded by a halo corresponding to its ecological niche. *Synechococcus* comprises many, and its niche corresponds to the union of the niches of every clades. Similarly, each clade can comprise several ESTUs, each having its own niche. Note that niches can partially overlap. A higher taxonomic resolution might unveil smaller taxonomic units, which is symbolized by question marks.

2. Models of evolution of marine picocyanobacterial lineages

2.1 Mechanisms of ecotype formation

Prochlorococcus and *Synechococcus* seem to partition into genetic units with different ecological *preferenda* that could be called ecotypes. What evolutionary sequence could have led to the formation of such populations, which differ both genetically and ecologically? This question has been widely explored in particular by F. Cohan and his colleagues who established a number of theoretical models (up to 10, Cohan and Perry, 2007) explaining ecotype formation depending on the forces driving the evolution of the microbial group studied. Before describing some of these models, it is important to give the definition they use for the term "ecotype": (Cohan and Kane, 2001) defined it as a population "using the same or similar ecological resources, such that an adaptive mutant from within the ecotype out-competes to extinction all other strains of the same ecotype; an adaptive mutant does not, however, drive to extinction strains from other ecotypes". In terms of ecological niche, an ecotype thus comprises all lineages sharing the exact same niche. I will present here only some of the models, which I think are of interest with regard to picocyanobacterial diversification (Figure 66). In the "Stable ecotype" model (Figure 66), the emergence of new ecotypes (i.e. colonization of a new niche) is rare compared to the rate of periodic selection events (selective sweeps) that purge the genetic diversity within an ecotype. As a result, these selective sweeps promote the genetic divergence of ecotypes, and they form distinct genetic clusters. In the "Fractal-like" model (Figure 66, Cohan and Koeppel, 2008), new ecotypes are formed much more rapidly than the rate of evolution of the whole genome, so ecotypes cannot be differentiated as genetically distinct clusters using classical markers (such as 16S rDNA). Other markers can reveal ecologically distinct clusters that appear to be ecotypes, and even more rapidly evolving markers can reveal even more clusters that are ecologically distinct. Finally, in the "Nano-niche" model (Figure 66, Cohan and Perry, 2007), several ecotypes share the same set of resources but differ in their use of these resources (e.g. they use them in different proportions). In this model, a particularly adaptive mutation in one ecotype can drive extinct the others, since they use the same set of resources. As subtle differences can be sufficient to colonize a new nano-niche, ecotypes are not necessarily distinguished based on a classical marker genes, but can be when using a fast-evolving gene.

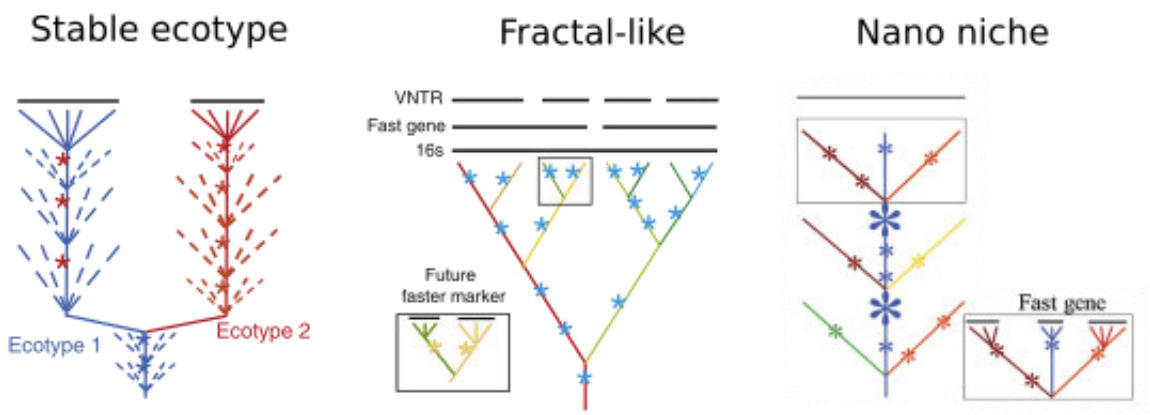


Figure 66 Models of ecotype formation. These models are based on a constant diversification interrupted by events of periodic selection (corresponding to selective sweeps). Dotted lines represent extinct lineages, and asterisks represent selection events. Genetically consistent taxa are identified by a black trait at the tips of the trees. In the stable ecotype model, ecotype divergence is rare compared to periodic selection events, and each ecotypes are genetically distinct. In the fractal-like model, ecotype formation is fast, too fast to identify them with well-conserved markers such as 16S rDNA. The use of faster-evolving markers can however distinguish ecotypes. In the nano-niche model, ecotypes differ slightly in their use of a common set of resources, thus a particularly adaptive variant of one ecotype can drive extinct other ecotypes in strong selective sweeps (larger asterisks). As ecotypes are only slightly different in their niches, they are not well differentiated with a conserved marker. Figures were taken from (Cohan and Perry, 2007) and (Cohan and Koeppel, 2008).

In our genomic analysis, we observed a slow divergence of lineages with a strong clonality, which is expected to happen when recombination is infrequent compared to mutation (Spratt *et al.*, 2001). Thus, marine picocyanobacteria are good candidates to apply Cohan's models, which rely on a hypothesis of low recombination (genetic homogeneity is maintained by periodic selective sweeps rather than recombination).

The fact that the use of a high-resolution marker enables to unveil new "ecotypes" that are not identified with a conserved, low-resolution marker, as we found here using *petB* and consistently with previous studies that used other markers (Tai *et al.*, 2011; Mazard *et al.*, 2012a; Kashtan *et al.*, 2014; Larkin *et al.*, 2016; Kashtan *et al.*, 2017), fits well with two of the models described above, the "Nano-niche" and the "Fractal-like" models. However the "Fractal-like" model suggests that ecological differentiation is too fast to be reflected by genomic differentiation, while our results and those of others suggest that picocyanobacterial ecotypes can be distinguished by differences in gene content and/or genomic sequence (Kashtan *et al.*, 2014; Kent *et al.*, 2016; Farrant *et al.*, 2016). Thus the "Nano-niche" model seems to be one of the best suited to describe picocyanobacterial evolution among those presented, though the actual differences between ecotypes remain to be fully described. If this model was described in a "bottom-up" conception of ecotypes (i.e. driven by access to resources), it can also be easily adapted to "top-down" components of the niche, ecotypes being differentiated by their ability to resist or escape predators.

2.2 Integration of evolution models with picocyanobacterial lifestyle

Despite their utility in defining a conceptual framework, none of the models presented above is entirely satisfying to describe picocyanobacterial evolution, mainly because they are static both spatially and temporally. The timescale and the frequency at which the processes of ecotype formation and selective sweeps happen are not considered, which limits interpretation. The incorporation of a timescale in our analysis of mechanisms driving genome diversification showed that they can be quite slow, happening in a time when environmental conditions, and thus selective pressures, can change. For now, little is known about the dynamics of formation of new populations in marine picocyanobacteria on a short-term. Such early events involved in ecological differentiation have been explored in other marine bacteria, such as *Vibrio cyclitrophicus*, in which genomic divergence of populations recently adapted to particles of different sizes was analyzed (Shapiro *et al.*, 2012). In this study, they found that diverging populations differed by a set of SNPs localized in adaptive genes as well as in their flexible gene content, while they did not (yet) differ in their core genome, due to gene-specific selective sweeps enabled by frequent recombination. However, this kind of short-term dynamics is not necessarily happening in marine picocyanobacteria, which have a very different lifestyle (Polz *et al.*, 2006). Indeed, while these *Vibrio* swim towards marine particles where they aggregate, picocyanobacteria have adopted a free-living lifestyle in which there is little chance of encounter between cells: they live in a dilute environment and most of them are not motile. Recent studies estimated that a *Prochlorococcus* cell is in average 200 cell lengths away from the closest *Prochlorococcus* cell, and that a *Prochlorococcus* daughter cell will be centimeters apart from its parent within minutes, tens of meters apart within an hour, and a few kilometers apart within a week (Kashtan *et al.*, 2014; Biller *et al.*, 2014a). Thus, we can expect that genetic exchange with relatives is rare for a given cell, and picocyanobacterial cells could be more likely to get genes either through the mediation of phages or vesicles, or from more distantly related organisms (thus increasing the size of the pangenome). One can then wonder what cohesive force maintains homogeneity within a taxon or an ecotype: is it selection only or does homologous recombination also play a significant role?

Altogether, we are lacking fundamental information about the recombination rate and the encounter rate between cells of a taxon and between taxa, the level of genetic homogeneity within a population, and more generally about the role played by dispersion in picocyanobacterial ecology, to further understand how new populations are formed. From our current knowledge, we can only speculate a model of evolution, in which new populations emerge and rise locally in response to a selective pressure (e.g. mutations or acquisition of genes allowing to escape viruses and grazers) and then slowly diverge from other populations. The conditions necessary for this new genetically distinct population to colonize the whole ocean and for its genomic backbone (or specific genes of variants) to sweep the whole *Synechococcus* and *Prochlorococcus* are not yet understood, and this process is likely to occur at a timescale of several millions years.

3. Towards a better understanding of picocyanobacterial life in the ocean

3.1 Unknown taxa and their ecological importance

While 30 years of studies on marine picocyanobacteria built a quite good knowledge about the ecology of long-known clades, some others have been overlooked and some were identified only recently – and might even have been identified several times and given different names due to the use of different markers (see INTRODUCTION). These poorly known clades include most *Prochlorococcus* LL clades (whose genetic diversity is likely under-sampled, and whose global distribution has not been documented yet) as well as a number of *Synechococcus* clades, among which EnvA and EnvB are good examples. These clades were identified by amplification of *petB* gene *in situ* (Mazard *et al.*, 2012a) and we found them to dominate or co-dominate *Synechococcus* communities in various places in the ocean (Farrant *et al.*, 2016) (Figure 64). The absence of isolated strains prevents us from testing any hypothesis about their ecological preferences, but our global mapping of *Synechococcus* communities will allow us to target specific regions of the world to try and isolate them. Another mystery clade has been rising since its identification as a phylogenetic group in 2005 (Saito *et al.*, 2005), namely clade CRD1, whose prevalence was unknown when I started my PhD, but is now recognized as dominating large expenses of the ocean (Sohm *et al.*, 2015; Farrant *et al.*, 2016). We further showed that it could be separated in several genetic populations (ESTUs) showing differences in their distribution, which is a good example of how large-scale analyses can generate new hypotheses testable in the lab. Indeed, based on these results, and because we were lucky enough to have one representative of each ESTU in culture, we tested the hypothesis that they had different thermal *preferenda*. This preliminary work has been performed by Théo Sciandra, a master student who I co-supervised, and by Garance Monier (Figure 67). The results are encouraging, since the strain representative of CRD1B ESTU, found to have a colder niche (its highest abundance being in upwelling areas), is able to grow at lower temperatures than the CRD1C strain, an ESTU dominating in warm areas. Finally, the strain representative of ESTU CRD1A, for which we hypothesized a very wide temperature tolerance from its distribution in the oceans, is indeed able to grow over a large range of temperatures. As the clade as a whole was found to dominate in iron-deplete regions, future plans of the team in the framework of the forthcoming ANR project CINNAMON (2018-2020), also include a comparison of the different ESTUs within the CRD1 clade with other *Synechococcus* clades/ESTUs in their ability to cope with various concentrations of iron.

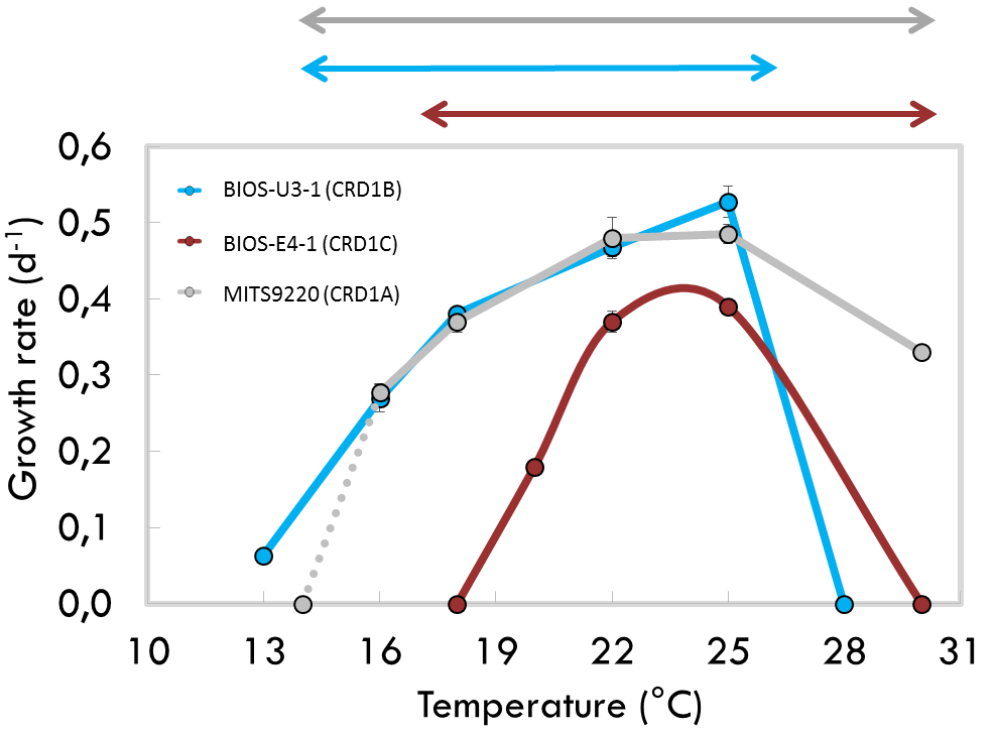


Figure 67 Thermal preferenda of three CRD1 strains representative of three ESTUs. Growth rate as a function of temperature, measured at a light intensity of 20 µE.m^{-2}.s^{-1}. The ESTU to which each strain belongs to is indicated is the legend. Arrows drawn above the graph indicate the range of temperatures at which theses ESTUs were found in the analysis of the *Tara* Oceans dataset (Farrant et al., 2016). These preliminary experiments were performed by Théo Sciandra.

3.2 What interactions occur between clades or ecotypes *in situ*?

One striking result from our global survey of picocyanobacterial distribution, which is also evident on Figure 64, is the coexistence of multiple clades or ESTUs in every samples. While we focused on where each taxon dominates with the idea of delineating its niche, it is important to note that a given taxon is also present in other environments at low abundance, where another one dominates. This coexistence seems to violate the niche exclusion principle we evoked in the INTRODUCTION, and meets the questions asked by G. E. Hutchinson when he enunciated the "paradox of the plankton" (Hutchinson, 1961). In his paper, he formulates a number of hypotheses to explain the coexistence of species in a same niche, which include niche heterogeneity in space (i.e. co-existence of several niches in a same place) and time (i.e. the environment changes faster than the time needed for a population to take over all other populations) and the influence of biotic interactions. Both types of hypotheses can be applied to marine picocyanobacteria. In particular, the role of cyanophages in maintaining diversity by frequency-dependent selection has been proposed several times (Rodriguez-Valera *et al.*, 2009; Avrani *et al.*, 2011, 2012; Biller *et al.*, 2014a), and this applies to grazers as well. Besides antagonistic interactions, positive interactions (symbiosis), such as the one observed between *Prochlorococcus* and *Pelagibacter* that were shown to be interdependent (see the Black Queen Hypothesis in INTRODUCTION (Morris *et al.*, 2012;

Biller *et al.*, 2014a)), can influence diversity as well: tight links with different partners might favor the maintenance of several populations. Both kinds of interactions thus most certainly affect patterns of diversity, but the level of diversity impacted by frequency-dependent selection is not clear: does this play a role in coexistence of closely-related populations, of long-diverged populations like clades, or both?

One possibility that was not in the scope of Hutchinson's paper is that co-occurring populations occupy similar or overlapping niches but are *not* in competition for resources or resistance to predators because cells are too far away from each other. As we said above, two *Prochlorococcus* cells are in average 200 cell lengths away (Biller *et al.*, 2014a) and rely for nutrients acquisition on diffusion taking place at smaller distances (Zehr *et al.*, 2017). In this case, relative abundance of different populations would be dictated by intrinsic metabolic characteristics of each population, and not by competitive exclusion between them. This specific question can be addressed by co-culturing different ecotypes in varying experimental conditions, which has rarely been done with marine picocyanobacteria.

Finally, a last explanation for co-occurrence of several populations can be a simple effect of dispersion: even if populations have clearly distinct niches, they can co-occur when one is taken outside of its environment due to passive dispersion. This implies that part of the observed cells in a given sample (e.g. HLI cells in a sample largely dominated by HLII cells) might be incapable of growing fast enough to maintain a stable population size, but that the population size remains constant thanks to a continuous input from a more or less distant reservoir. The question then remains to understand the physiological state of these cells brought outside of their niche: if not dividing, are they surviving in a kind of dormancy?

All these hypotheses are not mutually exclusive and can influence at the same time the local diversity of picocyanobacterial populations, but the relative weight of each process and the conditions in which they are favored remain to be assessed.

3.3 A step forward: measuring the metabolic state of cells in the ocean

The hypothesis we just mentioned concerning a potential cell dormancy outside the ecological niche can now be addressed in the field. Many studies have measured physiological parameters of picocyanobacteria *in situ*, such as photosynthesis, mortality or uptake of nutrients (see INTRODUCTION) usually by sorting them from other plankton by filtration or flow cytometry. However, these methods cannot differentiate between distinct populations of *Synechococcus* or *Prochlorococcus* mixed in the same sample, and measurements likely reflect only the physiology of the dominant one. In the goal of assessing the metabolic state of cells, sequencing data are imperfect in the sense that they do not measure the phenotype directly, but they allow to analyze independently the response of co-occurring taxonomic units. In that prospect, metatranscriptomic data that are becoming available promise new discoveries, in particular in combination with metagenomic data. Our analysis of a regional response to iron fertilization gives an example of

how these data can help differentiate between the adaptive and acclimation responses of natural populations, and will get one step further by differentiating the response of each ESTU.

In this context, we need a solid reference framework to interpret patterns of gene expression, which can only be obtained by transcriptomic experiments in controlled conditions. We brought our brick to this wall with analyses of both the temporal gene expression pattern along a day and the response to various stresses. For example, we found an overexpression of genes related to nitrogen stress in the morning during the light-dark cycle experiment, despite the use of an N-replete medium. Without this piece of information, such a pattern of expression would have been misinterpreted as an N-stress if one had measured gene expression *in situ* only in the morning. Comparison of expression profiles with those obtained in controlled conditions (e.g. cold stress) can moreover give insights into the metabolic state of cells even when genes have no or only a putative function. One of the caveats of such omics analyses is however their dependence on genomic references: our current database of genomes, as we showed, is far from representing the whole pangenome of either *Prochlorococcus* or *Synechococcus*, and we are probably lacking the distribution and expression patterns of some genes of environmental importance. One of the challenges in the forthcoming years will thus be to find a way to access these unknown gene repertoires, which will probably be allowed by newly developed sequencing techniques (e.g MinIon technology (Lu *et al.*, 2016)) facilitating assembly in combination with currently used technologies.

Conclusion – a matter of scales

The question of scales has been an underlying leitmotiv in all the considerations about picocyanobacterial ecology explored in this thesis: spatial scales, temporal scales, diversity scales… The next step will be to bridge the gaps between the different levels of analysis: how to reconcile what is observed in a few generations in the lab with what happens in a single year, or in tens, hundreds or millions years? How to link observations made on single cells with those made on cultures and those made *in situ*, and how to link local variations with global patterns of distribution? What an exciting future!

References

A

Abby S, Daubin V. (2007). Comparative genomics and the evolution of prokaryotes. *Trends Microbiol* **15**: 135–141.

Ahlgren N a., Noble A, Patton AP, Roache-Johnson K, Jackson L, Robinson D, *et al.* (2014). The unique trace metal and mixed layer conditions of the Costa Rica upwelling dome support a distinct and dense community of *Synechococcus*. *Limnol Oceanogr* **59**: 2166–2184.

Ahlgren N a, Rocap G. (2006). Culture isolation and culture-independent clone libraries reveal new marine *Synechococcus* ecotypes with distinctive light and N physiologies. *Appl Environ Microbiol* **72**: 7193–204.

Ahlgren NA, Rocap G. (2012). Diversity and Distribution of Marine *Synechococcus*: Multiple Gene Phylogenies for Consensus Classification and Development of qPCR Assays for Sensitive Measurement of Clades in the Ocean. *Front Microbiol* **3**: 213.

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

Anders S, Pyl PT, Huber W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166–169.

Avrani S, Schwartz DA, Lindell D. (2012). Virus-host swinging party in the oceans. *Mob Genet Elements* **2**: 88–95.

Avrani S, Wurtzel O, Sharon I, Sorek R, Lindell D. (2011). Genomic island variability facilitates *Prochlorococcus*-virus coexistence. *Nature* **474**: 604–608.

Axelsen KB, Palmgren MG. (1998). Evolution of substrate specificities in the P-type ATPase superfamily. *J Mol Evol* **46**: 84–101.

Azam F, Fenchel T, Field J, Gray J, Meyer-Reil L, Thingstad F. (1983). The Ecological Role of Water-Column Microbes in the Sea. *Mar Ecol Prog Ser* **10**: 257–263.

B

Bailey S, Mann NH, Robinson C, Scanlan DJ. (2005). The occurrence of rapidly reversible non-photochemical quenching of chlorophyll a fluorescence in cyanobacteria. *FEBS Lett* **579**: 275–280.

Bailey S, Melis A, Mackey KRM, Cardol P, Finazzi G, van Dijken G, *et al.* (2008). Alternative photosynthetic electron flow to oxygen in marine *Synechococcus*. *Biochim Biophys Acta - Bioenerg* **1777**: 269–276.

Becking BLGM. (1934). Geobiologie of Inleiding Tot De Milieukunde. La Haye.

Behrenfeld MJ, Halsey KH, Milligan AJ. (2008). Evolved physiological responses of phytoplankton to their integrated growth environment. *Philos Trans R Soc B Biol Sci* **363**: 2687–2703.

Behrenfeld MJ, Kolber ZS. (1999). Widespread iron limitation of phytoplankton in the south Pacific Ocean. *Science* **283**: 840–843.

Behrenfeld MJ, Westberry TK, Boss ES, O'Malley RT, Siegel D a., Wiggert JD, *et al.* (2009). Satellite-detected fluorescence

reveals global physiology of ocean phytoplankton. *Biogeosciences Discuss* **5**: 4235–4270.

Berg G, Shrager J, van Dijken G, Mills M, Arrigo K, Grossman A. (2011). Responses of psbA, hli and ptox genes to changes in irradiance in marine *Synechococcus* and *Prochlorococcus*. *Aquat Microb Ecol* **65**: 1–14.

Bertilsson S, Berglund O, Karl DM, Chisholm SW. (2003). Elemental composition of marine *Prochlorococcus* and *Synechococcus*: Implications for the ecological stoichiometry of the sea. *Limnol Oceanogr* **48**: 1721–1731.

Berube PM, Biller SJ, Kent AG, Berta-Thompson JW, Roggensack SE, Roache-Johnson KH, *et al.* (2014). Physiology and evolution of nitrate acquisition in *Prochlorococcus*. *ISME J* **9**: 1195–1207.

Biard T. (2015). Diversité, biogéographie et écologie des Collodaires (Radiolaires) dans l'océan mondial. Université Pierre et Marie Curie - Paris VI.

Bibby TS, Mary I, Nield J, Partensky F, Barber J. (2003). Low-light-adapted *Prochlorococcus* species possess specific antennae for each photosystem. *Nature* **424**: 1051–1054.

Bibby TS, Nield J, Barber J. (2001). Iron deficiency induces the formation of an antenna ring around trimeric photosystem I in cyanobacteria. *Nature* **412**: 743–745.

Biller SJ, Berube PM, Lindell D, Chisholm SW. (2014a). *Prochlorococcus*: the structure and function of collective diversity. *Nat Rev Microbiol* **13**: 13–27.

Biller SJ, Schubotz F, Roggensack SE, Thompson AW, Summons RE, Chisholm SW. (2014b). Bacterial vesicles in marine ecosystems. *Science* **343**: 183–6.

Binder BJ, Chisholm SW. (1995). Cell Cycle Regulation in Marine *Synechococcus* sp. Strains. *Appl Environ Microbiol* **61**: 708–717.

Bird C, Wyman M. (2003). Nitrate/Nitrite Assimilation System of the Marine Picoplanktonic Cyanobacterium *Synechococcus* sp . Strain WH8103: Effect of Nitrogen Source and Availability on Gene Expression. *Appl Environ Microbiol* **69**: 7009–7018.

Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, *et al.* (1997). The Complete Genome Sequence of *Escherichia coli* K-12. *Science* **277**: 1453–1462.

Blot N, Mella-Flores D, Six C, Le Corguillé G, Boutte C, Peyrat A, *et al.* (2011). Light history influences the response of the marine cyanobacterium *Synechococcus* sp. WH7803 to oxidative stress. *Plant Physiol* **156**: 1934–54.

Boekema EJ, Hifney A, Yakushevska AE, Piotrowski M, Keegstra W, Berry S, *et al.* (2001). A giant chlorophyll–protein complex induced by iron deficiency in cyanobacteria. *Nature* **412**: 745–748.

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

Bouman HA, Ulloa O, Scanlan DJ, Zwirglmaier K, Li WKW, Platt T, *et al.* (2006). Oceanographic basis of the global surface distribution of *Prochlorococcus* ecotypes. *Science* **312**: 918–921.

Braakman R, Follows MJ, Chisholm SW. (2017). Metabolic evolution and the self-organization of ecosystems. *Proc Natl Acad Sci* **114**: E3091–E3100.

Brahamsha B. (1999). Non-flagellar swimming in marine *Synechococcus*. *J Mol Microbiol Biotechnol* **1**: 59–62.

Brierley AS. (2017). Plankton. *Curr Biol* **27**: R478–R483.

Buchan A, LeCleir GR, Gulvik CA, González JM. (2014). Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nat Rev Microbiol* **12**: 686–698.

Burnap RL, Troyan T, Sherman L a. (1993). The highly abundant chlorophyll-protein complex of iron-deficient *Synechococcus* sp. PCC7942 (CP43') is encoded by the isiA gene. *Plant Physiol* **103**: 893–902.

c

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

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

Campbell D, Bruce D, Carpenter C, Gustafsson P, Oquist G. (1996). Two forms of the Photosystem II D1 protein alter energy dissipation and state transitions in the cyanobacterium *Synechococcus* sp. PCC 7942. *Photosynth Res* **47**: 131–44.

Campbell L, Vaulot D. (1993). Photosynthetic picoplankton community structure in the subtropical North Pacific Ocean near Hawaii (station ALOHA). *Deep Res Part I* **40**: 2043–2060.

Casey JR, Lomas MW, Mandecki J, Walker DE. (2007). *Prochlorococcus* contributes to new production in the Sargasso Sea deep chlorophyll maximum. *Geophys Res Lett* **34**: 1–5.

Casey JR, Mardinoglu A, Nielsen J, Karl DM. (2016). Adaptive Evolution of Phosphorus Metabolism in *Prochlorococcus*. *mSystems* **1**: 1–15.

Chandler JW, Lin Y, Gainer PJ, Post AF, Johnson ZI, Zinser ER. (2016). Variable but persistent coexistence of *Prochlorococcus* ecotypes along temperature gradients in the ocean's surface mixed layer. *Environ Microbiol Rep* **4**: n/a-n/a.

Chen F, Wang K, Kan J, Suzuki MT, Wommack KE. (2006). Diverse and Unique Picocyanobacteria in Chesapeake Bay , Revealed by 16S-23S rRNA Internal Transcribed Spacer Sequences. *Appl Environ Microbiol* **72**: 2239–2243.

Chen PE, Shapiro BJ. (2015). The advent of genome-wide association studies for bacteria. *Curr Opin Microbiol* **25**: 17–24.

Chisholm SW, Frankel SL, Goericke R, Olson RJ, Palenik B, Waterbury JB, *et al.* (1992). *Prochlorococcus* marinus nov. gen. nov. sp.: an oxyphototrophic marine prokaryote containing divinyl chlorophyll a and b. *Arch Microbiol* **157**: 297–300.

Chisholm SW, Olson RJ, Zettler ER, Goericke R, Waterbury JB, Welschmeyer NA. (1988). A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* **334**: 340–343.

Choi DH, Noh JH, An SM, Choi YR, Lee H, Ra K, *et al.* (2016). Spatial distribution of cold-adapted *Synechococcus* during spring in seas adjacent to Korea. *Algae* **31**: 231–241.

Choi DH, Noh JH, Shim J. (2013). Seasonal changes in picocyanobacterial diversity as revealed by pyrosequencing in temperate waters of the East China Sea and the East Sea. *Aquat Microb Ecol* **71**: 75–90.

Christaki U, Jacquet S, Dolan JR, Vaulot D, Rassoulzadegan F. (1999). Growth and grazing on *Prochlorococcus* and *Synechococcus* by two marine ciliates. *Limnol Oceanogr* **44**: 52–61.

Cohan FM, Kane M. (2001). Bacterial Species and Speciation. *Syst Biol* **50**: 513–524.

Cohan FM, Koeppel AF. (2008). The Origins of Ecological Diversity in Prokaryotes. *Curr Biol* **18**: 1024–1034.

Cohan FM, Perry EB. (2007). A systematics for discovering the fundamental units of bacterial diversity. *Curr Biol* **17**: 373–386.

Coleman ML, Chisholm SW. (2007). Code and context: *Prochlorococcus* as a model for cross-scale biology. *Trends Microbiol* **15**: 398–407.

Coleman ML, Chisholm SW. (2010). Ecosystem-specific selection pressures revealed through comparative population genomics. *Proc Natl Acad Sci Unit States Am* **107**: 18634–18639.

Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, Delong EF, *et al.* (2006). Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311**: 1768–1770.

Collier JL, Brahamsha B, Palenik B.

(1999). The marine cyanobacterium *Synechococcus* sp. WH7805 requires urease (urea amidohydrolase, EC 3.5.1.5) to utilize urea as a nitrogen source: molecular-genetic and biochemical analysis of the enzyme. *Microbiolo* **145**: 447–459.

Cordero OX, Polz MF. (2014). Explaining microbial genomic diversity in light of evolutionary ecology. *Nat Rev Microbiol* **12**: 263–273.

Csurös M. (2010). Count: Evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* **26**: 1910–1912.

Cuhel RL, Waterbury JB. (1984). Biochemical composition and short term nutrient incorporation patterns in a unicellular marine cyanobacterium, *Synechococcus* (WH7803). *Limnol Oceanogr* **29**: 370–374.

D

Daubin V, Moran NA, Ochman H. (2003). Phylogenetics and the cohesion of bacterial genomes. *Science* **301**: 829–32.

Donald KM, Scanlan DJ, Carr NG, Mann NH, Joint I. (1997). Comparative phosphorus nutrition of the marine cyanobacterium *Synechococcus* WH7803 and the marine diatom *Thalassiosira weissflogii*. *J Plankton Res* **19**: 1793–1813.

Doolittle WF. (1999). Phylogenetic classification and the universal tree. *Science* **284**: 2124–2129.

Doolittle WF. (2012). Population genomics: How bacterial species form and why they don't exist. *Curr Biol* **22**: R451–R453.

Dufresne A, Garczarek L, Partensky F. (2005). Accelerated evolution associated with genome reduction in a free-living prokaryote. *Genome Biol* **6**: R14.1-10.

Dufresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik BP, *et al.* (2008). Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol* **9**: R90.

E

Eddy SR. (2011). Accelerated profile HMM searches. *PLoS Comput Biol* **7**. e-pub ahead of print, doi: 10.1371/journal.pcbi.1002195.

Edwards KF, Thomas MK, Klausmeier CA, Litchman E. (2016). Phytoplankton growth and the interaction of light and temperature : A synthesis at the species and community level. *Limnol Oceanogr* **61**: 1232–1244.

Erdner DDL, Anderson DMD. (1999). Ferredoxin and flavodoxin as biochemical indicators of iron limitation during open-ocean iron enrichment. *Limnol Oceanogr* **44**: 1609–1615.

F

Falkowski PG. (2004). The Evolution of Modern Eukaryotic Phytoplankton. *Science* **305**: 354–360.

Farrant GK, Doré H, Cornejo-castillo FM, Partensky F, Ratin M, Garczarek L. (2016). Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proc Natl Acad Sci* **113**: E3365-74.

Farrant GK, Hoebeke M, Partensky F, Andres G, Corre E, Garczarek L. (2015). WiseScaffolder: an algorithm for the semi-automatic scaffolding of Next Generation Sequencing data. *BMC Bioinformatics* **16**: 281.

Feingersch R, Philosof A, Mejuch T, Glaser F, Alalouf O, Shoham Y, *et al.* (2012). Potential for phosphite and phosphonate utilization by *Prochlorococcus*. *ISME J* **6**: 827–834.

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

Ferris MJ, Palenik B. (1998). Niche adaptation in ocean cyanobacteria. *Nature* **396**: 226–228.

Field CB. (1998). Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science* **281**: 237–240.

Fleischmann R, Adams M, White O, Clayton R, Kirkness E, Kerlavage A, *et al.* (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* **269**: 496–512.

Flombaum P, Gallegos JL, Gordillo R a, Rincón J, Zabala LL, Jiao N, *et al.* (2013). Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc Natl Acad Sci U S A* **110**: 9824–9.

Flores E, Frias JE, Rubio LM, Herrero A. (2005). Photosynthetic nitrate assimilation in cyanobacteria. *Photosynth Res* **83**: 117–133.

Follows MJ, Dutkiewicz S, Grant S, Chisholm SW. (2007). Emergent Biogeography of Microbial Communities in a Model Ocean. *Science* **315**: 1843–1846.

Frias-Lopez J, Thompson A, Waldbauer J, Chisholm SW. (2009). Use of stable isotope-labelled cells to identify active grazers of picocyanobacteria in ocean surface waters. *Environ Microbiol* **11**: 512–525.

Fridman S, Flores-Uribe J, Larom S, Alalouf O, Liran O, Yacoby I, *et al.* (2017). A myovirus encoding both photosystem I and II proteins enhances cyclic electron flow in infected *Prochlorococcus* cells. *Nat Microbiol*. e-pub ahead of print, doi: 10.1038/s41564-017-0002-9.

Fu F, Zhang Y, Feng Y, Hutchins DA. (2006). Phosphate and ATP uptake and growth kinetics in axenic cultures of the cyanobacterium *Synechococcus* CCMP 1334. *Eur J Phycol* **41**: 15–28.

Fuller NJ, Marie D, Partensky F, Vaulot D, Post AF, Scanlan DJ. (2003). Clade-Specific 16S Ribosomal DNA Oligonucleotides Reveal the Predominance of a Single Marine *Synechococcus* Clade throughout a Stratified Water Column in the Red Sea. *Appl Environ Microbiol* **69**: 2430–2443.

Fuller NJ, Tarran G a., Cummings DG, Woodward EMS, Orcutt KM, Yallop M, *et al.* (2006). Molecular analysis of photosynthetic picoeukaryote community structure along an Arabian Sea transect. *Limnol Oceanogr* **51**: 2502–2514.

Fuller NJ, West NJ, Marie D, Yallop M, Rivlin T, Post AF, *et al.* (2005). Dynamics of community structure and phosphate status of picocyanobacterial populations in the Gulf of Aqaba, Red Sea. *Limnol Oceanogr* **50**: 363–375.

G

Galens K, Orvis J, Daugherty S, Creasy HH, Angiuoli S, White O, *et al.* (2011). The IGS Standard Operating Procedure for Automated Prokaryotic Annotation. *Stand Genomic Sci* **4**: 244–251.

Galili T. (2015). dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **31**: 3718–3720.

Garcia-fernandez JM, Hess WR, Houmard J, Partensky F. (1998). Expression of the psbA Gene in the Marine Oxyphotobacteria *Prochlorococcus* spp. *Arch Biochem Biophys* **359**: 17–23.

Garczarek L, Dufresne A, Blot N, Cockshutt AM, Peyrat A, Campbell DA, *et al.* (2008). Function and evolution of the psbA gene family in marine *Synechococcus*: *Synechococcus* sp. WH7803 as a case study. *ISME J* **2**: 937–953.

Geider R, La Roche J. (2002). Redfield revisited: variability of C:N:P in marine microalgae and its biochemical basis. *Eur J Phycol* **37**: 1–17.

Genevaux P, Georgopoulos C, Kelley WL. (2007). The Hsp70 chaperone machines of *Escherichia coli*: A paradigm for the repartition of chaperone functions. *Mol Microbiol* **66**: 840–857.

Gierga G, Voss B, Hess WR. (2012). Non-coding RNAs in marine *Synechococcus* and their regulation under environmentally relevant stress conditions. *ISME J* **6**: 1544–57.

Giovannoni SJ. (2017). SAR11 Bacteria: The Most Abundant Plankton in the Oceans. *Ann Rev Mar Sci* **9**: 231–255.

Giovannoni SJ, Cameron Thrash J, Temperton B. (2014). Implications of streamlining theory for microbial ecology. *ISME J* **8**: 1553–1565.

Goericke R, Repeta DJ. (1992). The pigments of *Prochlorococcus* marinus: The presence of divinylchlorophyll a and b in a marine procaryote. *Limnol Oceanogr* **37**: 425–433.

Gogarten JP, Doolittle WF, Lawrence JG. (2002). Prokaryotic evolution in light of gene transfer. *Mol Biol Evol* **19**: 2226–2238.

Goris J, Konstantinidis KT, Klappenbach J a., Coenye T, Vandamme P, Tiedje JM. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57**: 81–91.

Grébert T, Doré H, Partensky F, Farrant GK, Emmanuel B, Picheral M, *et al.* Light color acclimation: a key process in the global ocean distribution of *Synechococcus* cyanobacteria. *Proc Natl Acad Sci* **in revisio**.

Guidi L, Chaffron S, Bittner L, Eveillard D. (2016). Plankton networks driving carbon export in the oligotrophic ocean. *Nature* **in press**.

Guillou L, Jacquet S, Chrétiennot-Dinet M-J, Vaulot D. (2001). Grazing impact of two small heterotrophic flagellates on *Prochlorococcus* and *Synechococcus*. *Aquat Microb Ecol* **26**: 201–207.

Guindon S, Gascuel O. (2003). A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Syst Biol* **52**: 696–704.

Guttierrez-Rodriguez A, Slack G, Daniels EF, Selph KE, Palenik B, Landry MR. (2014). Fine spatial structure of genetically distinct picocyanobacterial populations across environmental gradients in the Costa Rica Dome. *Limnol Oceanogr* **59**: 705–723.

H

Haeckel E. (1904). Kunstformen der Natur. Bibliograp. Verlag des Bibliographischen Instituts,: Leipzig und Wien.

Hanage WP, Fraser C, Spratt BG. (2006). Sequences, sequence clusters and bacterial species. *Philos Trans R Soc B Biol Sci* **361**: 1917–1927.

Hao W, Golding G. (2006). The fate of laterally transferred genes: life in the fast lane to adaptation or death. *Genome Res* 636–643.

Hardin G. (1960). The Competitive Exclusion Principle. *Science* **131**: 1292–1297.

Harris SR, Feil EJ, Holden MTG, Quail MA, Nickerson EK, Chantratita N, *et al.* (2010). Evolution of MRSA During Hospital Transmission and Intercontinental Spread. *Science (80-)* **327**: 469 LP-474.

Haswell ES, Phillips R, Rees DC. (2011). Mechanosensitive channels: What can they do and how do they do it? *Structure* **19**: 1356–1369.

Havaux M, Guedeney G, He Q, Grossman AR. (2003). Elimination of high-light-inducible polypeptides related to eukaryotic chlorophyll a/b-binding proteins results in aberrant photoacclimation in *Synechocystis* PCC6803. *Biochim Biophys Acta - Bioenerg* **1557**: 21–33.

Havaux M, Guedenney G, Hagemann M, Yeremenko N, Matthijs HCP, Jeanjean R. (2005). The chlorophyll-binding protein IsiA is inducible by high light and protects the cyanobacterium *Synechocystis* PCC6803 from photooxidative stress. *FEBS Lett* **579**: 2289–2293.

He Q, Dolganov N, Bjo O, Grossman AR, Natl P, Sci A. (2001). The High Light-inducible Polypeptides in *Synechocystis* PCC6803. *J Biol Chem* **276**: 306–314.

Heldal M, Norland S, Scanlan DJ, Mann NH. (2003). Elemental composition of single cells of various strains of marine *Prochlorococcus* and *Synechococcus* using X-ray microanalysis. *Limnol Oceanogr* **48**:

1732–1743.

Helman Y, Tchernov D, Reinhold L, Shibata M, Ogawa T, Schwarz R, *et al.* (2003). Genes Encoding A-Type Flavoproteins Are Essential for Photoreduction of O2 in Cyanobacteria. *Curr Biol* **13**: 230–235.

Holtzendorff J, Partensky F, Jacquet S, Bruyant F, Marie D, Garczarek L, *et al.* (2001). Diel expression of cell cycle-related genes in synchronized cultures of *Prochlorococcus* sp strain PCC 9511. *J Bacteriol* **183**: 915–920.

Ten Hoopen P, Cochrane G, Micro B3Consortium. (2014). Ocean Sampling Day Handbook - Version of June 2014. http://www.microb3.eu/ sites/default/files/osd/OSD_Handbook_v2.0 .pdf. 2014.

Hopkinson BM, Morel FMM. (2009). The role of siderophores in iron acquisition by photosynthetic marine microorganisms. *BioMetals* **22**: 659–669.

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

Huerta-Cepas J, Dopazo J, Gabaldón T. (2010). ETE: a python Environment for Tree Exploration. *BMC Bioinformatics* **11**: 24.

Hull PM. (2017). Emergence of modern marine ecosystems. *Curr Biol* **27**: R466–R469.

Humily F, Partensky F, Six C, Farrant GK, Ratin M, Marie D, *et al.* (2013). A gene island with two possible configurations is involved in chromatic acclimation in marine *Synechococcus*. *PLoS One* **8**: e84459.

Hunt DE, Lin Y, Church MJ, Karl DM, Tringe SG, Izzo LK, *et al.* (2013). Relationship between Abundance and Specific Activity of Bacterioplankton in Open Ocean Surface Waters. *Appl Environ Microbiol* **79**: 177–184.

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

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

Hutchinson GE. (1957). Concluding Remarks. *Cold Spring Harb Symp Quant Biol* **22**: 415–427.

Hutchinson GE. (1961). The paradox of the Plankton. *Am Nat* **95**: 137–145.

I

Ikeya T, Ohki K, Takahashi M, Fujita Y. (1997). Study on phosphate uptake of the marine cyanophyte *Synechococcus* sp . NIBB 1071 in relation to oligotrophic environments in the open ocean. *Mar Biol* **129**: 195–202.

J

Johnson PW, Sieburth JM. (1979). Chroococcoid cyanobacteria in the sea: A ubiquitous and diverse phototrophic biomass. *Limnol Oceanogr* **24**: 928–935.

Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward EMS, Chisholm SW. (2006). Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* **311**: 1737–1740.

Jones DT, Taylor WR, Thornton JM. (1992). The rapid generation of mutation data matricies from protein sequences. *Comput Appl Biosci* **8**: 275–282.

K

Kamennaya NA, Chernihovsky M, Marine HS, Beach C, Post AF. (2008). The cyanate utilization capacity of marine unicellular Cyanobacteria. *Limnol Oceanogr* **53**: 2485–2494.

Kamennaya NA, Post AF. (2011). Characterization of cyanate metabolism in marine *Synechococcus* and *Prochlorococcus* spp. *Appl Environ Microbiol* **77**: 291–301.

Kana TM, Glibert PM. (1987). Effect of irradiances up to 2000 μE m-2 s-1 on marine *Synechococcus* WH7803 -- I. Growth, pigmentation, and cell composition. *Deep - Sea Res Part A Oceanogr Res Pap* **34**: 479–495.

Karsenti E, Acinas SG, Bork P, Bowler C, De Vargas C, Raes J, *et al.* (2011). A Holistic Approach to Marine Eco-Systems Biology. *PLoS Biol* **9**: e1001177.

Kashtan N, Roggensack SE, Berta-Thompson JW, Grinberg M, Stepanauskas R, Chisholm SW. (2017). Fundamental differences in diversity and genomic population structure between Atlantic and Pacific *Prochlorococcus*. *ISME J* 1–15.

Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A, *et al.* (2014). Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* **344**: 416–20.

Katoh K, Standley DM. (2014). MAFFT: Iterative refinement and additional methods. In: D. R (ed) Vol. 1079. *Methods in Molecular Biology (Methods and Protocols)*. Humana Press: Totowa, NJ, pp 131–146.

Katsanevakis S, Coll M, Piroddi C, Steenbeek J, Ben Rais Lasram F, Zenetos A, *et al.* (2014). Invading the Mediterranean Sea: biodiversity patterns shaped by human activities. *Front Mar Sci* **1**: 1–11.

Kennemann L, Didelot X, Aebischer T, Kuhn S, Drescher B, Droege M, *et al.* (2011). Helicobacter pylori genome evolution during human infection. *Proc Natl Acad Sci* **108**: 5033–5038.

Kent AG, Dupont CL, Yooseph S, Martiny AC. (2016). Global biogeography of *Prochlorococcus* genome diversity in the surface ocean. *ISME J* **10**: 1856–1865.

Kerfeld CA, Melnicki MR, Sutter M, Dominguez-martin MA. (2017). Structure, function and evolution of the cyanobacterial orange carotenoid protein and its homologs. *New Phytol* **215**: 937–951.

Kettler GC, Martiny AC, Huang K, Zucker J, Coleman ML, Rodrigue S, *et al.* (2007). Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet* **3**: e231.

Kirilovsky D, Kaňa R, Prášil O. (2014). Mechanisms Modulating Energy Arriving at Reaction Centers in Cyanobacteria. In: Demmig-Adams B, Garab G, Adams III W, Govindjee (eds). *Non-Photochemical Quenching and Energy Dissipation in Plants, Algae and Cyanobacteria*. Springer Netherlands: Dordrecht, pp 471–501.

Konstantinidis KT, Tiedje JM. (2005a). Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A* **102**: 2567–2572.

Konstantinidis KT, Tiedje JM. (2005b). Towards a genome-based taxonomy for prokaryotes. *J Bacteriol* **187**: 6258–6264.

Kopf A, Bicak M, Kottmann R, Schnetzer J, Kostadinov I, Lehmann K, *et al.* (2015). The ocean sampling day consortium. *Gigascience* **4**: 27.

Kopf M, Hess WR. (2015). Regulatory RNAs in photosynthetic cyanobacteria. *FEMS Microbiol Rev* **39**: 301–315.

Krumhardt KM, Callnan K, Roache-Johnson K, Swett T, Robinson D, Reistetter EN, *et al.* (2013). Effects of phosphorus starvation versus limitation on the marine cyanobacterium *Prochlorococcus* MED4 I: Uptake physiology. *Environ Microbiol* **15**: 2114–2128.

Kumar S, Stecher G, Tamura K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* **33**: 1870–1874.

Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V, *et al.* (1997). The complete genome sequence of the Gram-positive bacterium Bacillus subtilis. *Nature* **390**: 249–256.

L

Lan R, Reeves PR. (2000). Intraspecies variation in bacterial genomes: The need for a species genome concept. *Trends Microbiol* **8**: 396–401.

Langfelder P, Horvath S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC*

Bioinformatics **9**: 559.

Langmead B, Salzberg SL. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–9.

Lantoine F, Neveux J. (1997). Spatial and seasonal variations in abundance and spectral characteristics of phycoerithrins in the tropical northeastern Atlantic Ocean. *Deep Res I* **44**: 223–246.

Larkin AA, Blinebry SK, Howes C, Lin Y, Loftus SE, Schmaus CA, *et al.* (2016). Niche partitioning and biogeography of high light adapted *Prochlorococcus* across taxonomic ranks in the North Pacific. *ISME J* 1–13.

Larkin AA, Martiny AC. (2017). Microdiversity shapes the traits, niche space, and biogeography of microbial taxa. *Environ Microbiol Rep* **9**: 55–70.

Larsson J, Celepli N, Ininbergs K, Dupont CL, Yooseph S, Bergman B, *et al.* (2014). Picocyanobacteria containing a novel pigment gene cluster dominate the brackish water Baltic Sea. *ISME J* **8**: 1892–1903.

Latifi A, Jeanjean R, Lemeille S, Havaux M. (2005). Iron Starvation Leads to Oxidative Stress in Iron Starvation Leads to Oxidative Stress in Anabaena sp . Strain PCC 7120. *J Bacteriol* **187**: 1–4.

Latifi A, Ruiz M, Zhang CC. (2009). Oxidative stress in cyanobacteria. *FEMS Microbiol Rev* **33**: 258–278.

Lavin P, González B, Santibáñez JF, Scanlan DJ, Ulloa O. (2010). Novel lineages of *Prochlorococcus* thrive within the oxygen minimum zone of the eastern tropical South Pacific. *Environ Microbiol Rep* **2**: 728–738.

Lawrence JG. (2002). Gene transfer in bacteria: speciation without species? *Theor Popul Biol* **61**: 449–460.

Lawrence JG, Ochman H. (1998). Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci U S A* **95**: 9413–9417.

Lenton TM, Daines SJ. (2017). Biogeochemical Transformations in the History of the Ocean. *Ann Rev Mar Sci* **9**: 31–58.

Lerat E, Daubin V, Ochman H, Moran NA. (2005). Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol* **3**: 0807–0814.

Li L, Stoeckert CJJ, Roos DS. (2003). OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes -- Li et al. 13 (9): 2178 -- Genome Research. *Genome Res* **13**: 2178–2189.

Lichtlé C, Thomas JC, Spilar A, Partensky F. (1995). Immunological and ulstrastructural characterization of the photosynthetic complexes of the prochlorophyte *Prochlorococcus* (Oxychlorobacteria). *J Phycol* **31**: 934–941.

Lindell D, Erdner D, Marie D, Prásil O, Koblíjek M, Le Gall F, *et al.* (2002). Nitrogen Stress Response of *Prochlorococcus* Strain PCC9511 (Oxyphotobacteria) Involves Contrasting Regulation of ntcA and amt1. *J Phycol* **38**: 1113–1124.

Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW. (2005a). Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438**: 86–89.

Lindell D, Padan E. (1998). Regulation of ntcA Expression and Nitrite Uptake in the Marine *Synechococcus* sp . Strain WH7803. *J Bacteriol* **180**: 1878–1886.

Lindell D, Penno S, Al-qutob M, David E, Rivlin T, Lazar B, *et al.* (2005b). Expression of the nitrogen stress response gene ntcA reveals nitrogen-sufficient *Synechococcus* populations in the oligotrophic northern Red Sea. *Limnol Oceanogr* **50**: 1932–1944.

Lindell D, Post AF. (2001). Ecological Aspects of ntcA Gene Expression and Its Use as an Indicator of the Nitrogen Status of Marine *Synechococcus* spp . *Appl Environ Microbiol* **67**: 3340–3349.

Lindell D, Post AF. (1995). Ultraphytoplankton succession is triggered by deep winter mixing in the Gulf of Aqaba (Eilat), Red Sea. *Limnol Ocean* **40**: 1130–1141.

Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F, Chisholm SW. (2004). Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc Natl*

Acad Sci U S A **101**: 11013–11018.

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

Liu S, Cheng Y, Zhang X, Guan Q, Nishiuchi S, Hase K, *et al.* (2007). Expression of an NADP-malic enzyme gene in rice (Oryza sativa. L) is induced by environmental stresses; over-expression of the gene in Arabidopsis confers salt and osmotic stress tolerance. *Plant Mol Biol* **64**: 49–58.

Logares R, Sunagawa S, Salazar G, Cornejo-Castillo FM, Ferrera I, Sarmento H, *et al.* (2013). Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environ Microbiol* **16**: 2659–2671.

Loman NJ, Pallen MJ. (2015). Twenty years of bacterial genome sequencing. *Nat Rev Microbiol* **13**: 1–9.

Lommer M, Specht M, Roy A-S, Kraemer L, Andreson R, Gutowska MA, *et al.* (2012). Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biol* **13**: R66.

Lourenço SO, Barbarino E, Marquez UML, Aidar E. (1998). Distribution of Intracellular Nitrogen in Marine Microalgae: Basis for the Calculation of Specific Nitrogen-to-Protein Conversion Factors. *J Phycol* **34**: 798–811.

Lu H, Giordano F, Ning Z. (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics Bioinforma* **14**: 265–279.

M

Mackey KR, Post AF, McIlvin MR, Saito MA. (2017). Physiological and proteomic characterization of light adaptations in marine *Synechococcus*. *Environ Microbiol* **19**: 2348–2365.

Mackey KRM, Paytan A, Caldeira K, Grossman AR, Moran D, McIlvin M, *et al.* (2013). Effect of Temperature on Photosynthesis and Growth in Marine *Synechococcus* spp. *Plant Physiol* **163**: 815–829.

Mackey KRM, Post AF, McIlvin MR, Cutter G a., John SG, Saito M a. (2015). Divergent responses of Atlantic coastal and oceanic *Synechococcus* to iron limitation. *Proc Natl Acad Sci* **112**: 9944–9949.

Maechler M, Peter R, Struyf A, Hubert M, Hornik K. (2013). cluster: Cluster Analysis Basics and Extensions.

Magoč T, Salzberg SL. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**: 2957–2963.

Malmstrom RR, Coe A, Kettler GC, Martiny AC, Frias-Lopez J, Zinser ER, *et al.* (2010). Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific oceans. *ISME J* **4**: 1252–64.

Malmstrom RR, Rodrigue S, Huang KH, Kelly L, Kern SE, Thompson A, *et al.* (2012). Ecology of uncultured *Prochlorococcus* clades revealed through single-cell genomics and biogeographic analysis. *ISME J* **7**: 184–98.

Mann EL, Ahlgren N, Moffett JW, Chisholm SW. (2002). Copper toxicity and cyanobacteria ecology in the Sargasso Sea. *Limnol Ocean* **47**: 976–988.

Mann EL, Chisholm SW. (2000). Iron limits the cell division rate of *Prochlorococcus* in the eastern equatorial Pacific. *Limnol Oceanogr* **45**: 1067–1076.

Marie D, Brussaard CPD, Partensky F, Vaulot D. (1999). Flow cytometric analysis of phytoplankton, bacteria and viruses. In: Robinson JP (ed) Vol. 11.11. *Current protocols in cytometry*. John Wiley & Sons: New York, pp 1–15.

Martin JH, Coale KH, Johnson KS, Fitzwater SE, Gordon RM, Tanner SJ, *et al.* (1994). Testing the iron hypothesis in ecosystems of the equatorial Pacific Ocean. *Nature* **371**: 123–129.

Martinez E, Maamaatuaiahutapu K. (2004). Island mass effect in the Marquesas Islands: Time variation. *Geophys Res Lett*

31: 1–4.

Martiny A, Satish K, Berube PM. (2009a). Widespread metabolic potential for nitrite and nitrate assimilation among *Prochlorococcus* ecotypes. *Proc Natl Acad Sci U S A* **106**: 10787–10792.

Martiny AC, Coleman ML, Chisholm SW. (2006). Phosphate acquisition genes in *Prochlorococcus* ecotypes: Evidence for genome-wide adaptation. *Proc Natl Acad Sci U S A* **103**: 12552–12557.

Martiny AC, Huang Y, Li W. (2009b). Occurrence of phosphate acquisition genes in *Prochlorococcus* cells from different ocean regions. *Environ Microbiol* **11**: 1340–7.

Martiny AC, Tai APK, Veneziano D, Primeau F, Chisholm SW. (2009c). Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. *Environ Microbiol* **11**: 823–32.

Mary I, Tarran GA, Warwick PE, Terry MJ, Scanlan DJ, Burkill PH, *et al.* (2007). Light enhanced amino acid uptake by dominant bacterioplankton groups in surface waters of the Atlantic Ocean. *FEMS Microb Ecol* **63**: 36–45.

Mather AE, Reid SWJ, Maskell DJ, Parkhill J, Fookes MC, Harris SR, *et al.* (2013). Distinguishable Epidemics of Multidrug-Resistant Salmonella Typhimurium DT104 in Different Hosts. *Science (80-)* **341**: 1514–1517.

Mazard S, Ostrowski M, Partensky F, Scanlan DJ. (2012a). Multi-locus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*. *Environ Microbiol* **14**: 372–86.

Mazard S, Wilson WH, Scanlan DJ. (2012b). Dissecting the Physiological Response To Phosphorus Stress in Marine *Synechococcus* Isolates (Cyanophyceae). *J Phycol* **48**: 94–105.

McDonald BR, Currie CR. (2017). Lateral gene transfer dynamics in the ancient bacterial genus Streptomyces. *MBio* **8**: e00644-17.

Mella-Flores D, Mazard S, Humily F, Partensky F, Mahé F, Bariat L, *et al.* (2011). Is the distribution of *Prochlorococcus* and *Synechococcus* ecotypes in the Mediterranean Sea affected by global warming? *Biogeosciences* **8**: 2785–2804.

Mella-Flores D, Six C, Ratin M, Partensky F, Boutte C, Le Corguillé G, *et al.* (2012). *Prochlorococcus* and *Synechococcus* have evolved different adaptive mechanisms to cope with light and UV stress. *Front Microbiol* **3**: 1–20.

Michelle Wood A, Phinney DA, Yentsch CS. (1998). Water column transparency and the distribution of spectrally distinct forms of phycoerythrin-containing organisms. *Mar Ecol Prog Ser* **162**: 25–31.

Michelou VK, Cottrell MT, Kirchman DL. (2007). Light-Stimulated Bacterial Production and Amino Acid Assimilation by Cyanobacteria and Other Microbes in the North Atlantic Ocean. *Appl Environ Microbiol* **73**: 5539–5546.

Moore CM, Mills MM, Arrigo KR, Berman-Frank I, Bopp L, Boyd PW, *et al.* (2013). Processes and patterns of oceanic nutrient limitation. *Nat Geosci* **6**: 701–710.

Moore LR, Chisholm SW. (1999). Photophysiology of the marine cyanobacterium *Prochlorococcus* : Ecotypic differences among cultured isolates. *Limnol Oceanogr* **44**: 628–638.

Moore LR, Goericke R, Chisholm SW. (1995). Comparative physiology of *Synechococcus* and *Prochlorococcus*: influence of light and temperature on growth, pigments, fluorescence and absorptive properties. *Mar Ecol Prog Ser* **116**: 259–275.

Moore LR, Ostrowski M, Scanlan DJ, Feren K, Sweetsir T. (2005). Ecotypic variation in phosphorus-acquisition mechanisms within marine picocyanobacteria. *Aquat Microb Ecol* **39**: 257–269.

Moore LR, Post AF, Rocap G, Chisholm SW. (2002). Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Limnol Oceanogr* **47**: 989–996.

Moore LR, Rocap G, Chisholm SW. (1998). Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393**: 464–467.

Van Mooy BAS, Devol AH. (2008). Assessing nutrient limitation of *Prochlorococcus* in the North Pacific subtropical gyre by using an RNA capture method. *Limnol Oceanogr* **53**: 78–88.

Van Mooy BAS, Fredricks HF, Pedler BE, Dyhrman ST, Karl DM, Koblizek M, *et al.* (2009). Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. *Nature* **458**: 69–72.

Van Mooy BAS, Rocap G, Fredricks HF, Evans CT, Devol AH. (2006). Sulfolipids dramatically decrease phosphorus demand by picocyanobacteria in oligotrophic marine environments. *Proc Natl Acad Sci U S A* **103**: 8607–8612.

Morel A, Ahn Y-H, Partensky F, Vaulot D, Claustre H. (1993). *Prochlorococcus* and *Synechococcus* - a Comparative-Study of Their Optical-Properties in Relation To Their Size and Pigmentation. *J Mar Res* **51**: 617–649.

Morris JJ, Lenski RE, Zinser ER. (2012). The Black Queen Hypothesis : Evolution of Dependencies through Adaptative Gene Loss. *MBio* **3**: 1–7.

Morrissey J, Bowler C. (2012). Iron utilization in marine cyanobacteria and eukaryotic algae. *Front Microbiol* **3**: 1–13.

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

Mullineaux CW. (2014). Electron transport and light-harvesting switches in cyanobacteria. *Front Plant Sci* **5**: 1–6.

Murata N, Takahashi S, Nishiyama Y, Allakhverdiev SI. (2007). Photoinhibition of photosystem II under environmental stress. *Biochim Biophys Acta - Bioenerg* **1767**: 414–421.

Muro-pastor MI, Reyes JC, Florencio FJ. (2001). Cyanobacteria Perceive Nitrogen Status by Sensing Intracellular 2-Oxoglutarate Levels. *J Bacteriol* **276**: 38320–38328.

N

Needham DM, Sachdeva R, Fuhrman JA. (2017). Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *ISME J* **11**: 1614–1629.

Nishiyama Y, Allakhverdiev SI, Murata N. (2006). A new paradigm for the action of reactive oxygen species in the photoinhibition of photosystem II. *Biochim Biophys Acta - Bioenerg* **1757**: 742–749.

Nowell RW, Green S, Laue BE, Sharp PM. (2014). The extent of genome flux and its role in the differentiation of bacterial lineages. *Genome Biol Evol* **6**: 1514–1529.

O

Ocean Sampling Day Consortium P. (2015). Registry of samples and environmental context from the Ocean Sampling Day 2014. e-pub ahead of print, doi: 10.1594/PANGAEA.854419.

Ochman H. (2001). Genes Lost and Genes Found: Evolution of Bacterial Pathogenesis and Symbiosis. *Science* **292**: 1096–1099.

Ochman H, Lerat E, Daubin V. (2005). Examining bacterial species under the specter of gene transfer and exchange. *Proc Natl Acad Sci* **102**: 6595–6599.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, *et al.* (2015). vegan: Community Ecology Package. http://cran.r-project.org/package=vegan.

Olson R, Chisholm S. (1990). Pigments, size, and distribution of *Synechococcus* in the North Atlantic and Pacific Oceans. *Limnol Oceanogr* **35**: 45–58.

Olson RJ, Chisholm SW, Zettler ER, Altabet MA, Dusenberry JA. (1990). Spatial and temporal distributions of prochlorophyte picoplankton in the North Atlantic Ocean. *Deep Sea Res Part A, Oceanogr Res Pap* **37**: 1033–1051.

Olson RJ, Chisholm SW, Zettler ER, Armbrust E V. (1988). Analysis of *Synechococcus* pigment types in the sea using single and dual beam flow cytometry. *Deep Sea Res Part A, Oceanogr Res Pap* **35**: 425–440.

Osburne MS, Holmbeck BM, Coe A, Chisholm SW. (2011). The spontaneous mutation frequencies of *Prochlorococcus* strains are commensurate with those of other bacteria. *Environ Microbiol Rep* **3**: 744–749.

Osburne MS, Holmbeck BM, Frias-Lopez J, Steen R, Huang K, Kelly L, *et al.* (2010). UV hyper-resistance in *Prochlorococcus* MED4 results from a single base pair deletion just upstream of an operon encoding nudix hydrolase and photolyase. *Environ Microbiol* **12**: 1978–1988.

Ostrowski M, Mazard S, Tetu SG, Phillippy K, Johnson A, Palenik B, *et al.* (2010). PtrA is required for coordinate regulation of gene expression during phosphate stress in a marine *Synechococcus*. *ISME J* **4**: 908–921.

P

Paerl HW. (1991). Ecophysiological and Trophic Implications of Light-Stimulated Amino Acid Utilization in Marine Picoplankton. *Appl Environ Microbiol* **57**: 473–479.

Palenik B. (2001). Chromatic Adaptation in Marine *Synechococcus* Strains. *Appl Environ Microbiol* **67**: 991–994.

Palenik B. (2015). Molecular Mechanisms by Which Marine Phytoplankton Respond to Their Dynamic Chemical Environment. *Ann Rev Mar Sci* **7**: 325–340.

Palenik B, Brahamsha B, Larimer FW, Land M, Hauser L, Chain P, *et al.* (2003). The genome of a motile marine *Synechococcus*. *Nature* **424**: 1037–1042.

Palenik B, Ren Q, Dupont CL, Myers GS, Heidelberg JF, Badger JH, *et al.* (2006). Genome sequence of *Synechococcus* CC9311: Insights into adaptation to a coastal environment. *Proc Natl Acad Sci* **103**: 13555–13559.

Palenik B, Toledo G, Ferris M. (1997). Cyanobacterial diversity in marine ecosystems as seen by RNA polymerase (rpoC1) gene sequences. In: *International Symposium on Marine Cyanobacteria and Related Organisms*. Musée océanographique, pp 101–105.

Palinska KA, Jahns T, Rippka R, Tandeau De Marsac N. (2000). *Prochlorococcus* marinus strain PCC 9511, a picoplanktonic cyanobacterium, synthesizes the smallest urease. *Microbiology* **146**: 3099–3107.

Paradis E, Claude J, Strimmer K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**: 289–290.

Partensky F, Blanchot J, Lantoine F, Neveux J, Marie D. (1996). Vertical structure of picophytoplankton at different trophic sites of the tropical northeastern Atlantic Ocean. *Deep - Sea Res Part I - Oceanogr Res Pap* **43**: 1191–1213.

Partensky F, Blanchot J, Vaulot D. (1999a). Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters: a review. *Bull l'institut océanographique* **19**: 457–475.

Partensky F, Garczarek L. (2010). *Prochlorococcus*: advantages and limits of minimalism. *Ann Rev Mar Sci* **2**: 305–31.

Partensky F, Hess WR, Vaulot D. (1999b). *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* **63**: 106–127.

Partensky F, Hoepffner N, W K, Ulloa O, Vaulot D. (1993). Photoacclimation of *Prochlorococcus* sp. (Prochlorophyta) Strains lsolated from the North Atlantic and the Mediterranean Sea. *Plant Physiol* **101**: 285–296.

Paulsen ML, Doré H, Garczarek L, Seuthe L, Müller O, Sandaa R-A, *et al.* (2016). *Synechococcus* in the Atlantic Gateway to the Arctic Ocean. *Front Mar Sci* **3**. e-pub ahead of print, doi: 10.3389/fmars.2016.00191.

Paz-Yepes J, Brahamsha B, Palenik B. (2013). Role of a microcin-C-like biosynthetic gene cluster in allelopathic interactions in marine *Synechococcus*. *Proc Natl Acad Sci U S A* **110**: 12030–5.

Pearman PB, Guisan A, Broennimann O, Randin CF. (2008). Niche dynamics in space and time. *Trends Ecol Evol* **23**: 149–158.

Penno S, Lindell D, Post AF. (2006). Diversity of *Synechococcus* and *Prochlorococcus* populations determined from DNA sequences of the N-regulatory gene ntcA. *Environ Microbiol* **8**: 1200–1211.

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

Pesant S, Not F, Picheral M, Kandels-Lewis S, Le Bescot N, Gorsky G, *et al.* (2015). Open science resources for the discovery and analysis of *Tara* Oceans data. *Sci Data* **2**: 150023.

Pierella Karlusich JJ, Ceccoli RD, Graña M, Romero H, Carrillo N. (2015). Environmental selection pressures related to iron utilization are involved in the loss of the flavodoxin gene from the plant genome. *Genome Biol Evol* **7**: 750–767.

Pitt FD, Mazard S, Humphreys L, Scanlan DJ. (2010). Functional Characterization of *Synechocystis* sp. Strain PCC 6803 pst1 and pst2 Gene Clusters Reveals a Novel Strategy for Phosphate Uptake in a Freshwater Cyanobacterium. *J Bacteriol* **192**: 3512–3523.

Pittera J. (2015). Adaptation des cyanobactéries marines du genre *Synechococcus* au gradient latitudinal de température. Université Pierre et Marie Curie. https://tel.archives-ouvertes.fr/tel-01287967.

Pittera J, Humily F, Thorel M, Grulois D, Garczarek L, Six C. (2014). Connecting thermal physiology and latitudinal niche partitioning in marine *Synechococcus*. *ISME J* **8**: 1221–1236.

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* picocyanobacteria. *Revis*.

Pittera J, Partensky F, Six C. (2016). Adaptive thermostability of light-harvesting complexes in marine picocyanobacteria. *ISME J* 1–13.

Polz MF, Alm EJ, Hanage WP. (2013). Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends Genet* **29**: 170–175.

Polz MF, Hunt DE, Preheim SP, Weinreich DM. (2006). Patterns and mechanisms of genetic and phenotypic differentiation in marine microbes. *Philos Trans R Soc B Biol Sci* **361**: 2009–2021.

Post AF, Penno S, Zandbank K, Paytan A, Huse SM, Welch DM. (2011). Long term seasonal dynamics of *Synechococcus* population structure in the Gulf of Aqaba, northern Red Sea. *Front Microbiol* **2**: 1–12.

R

Raven JA, Evans MCW, Korb RE. (1999). The role of trace metals in photosynthetic electron transport in O2-evolving organisms. *Photosynth Res* **60**: 111–149.

Read RW, Berube PM, Biller SJ, Neveux I, Cubillos-Ruiz A, Chisholm SW, *et al.* (2017). Nitrogen cost minimization is promoted by structural changes in the transcriptome of N-deprived *Prochlorococcus* cells. *ISME J* **11**: 2267–2278.

Read TD, Massey RC. (2014). Characterizing the genetic basis of bacterial phenotypes using genome-wide association studies : a new direction for bacteriology. 1–11.

Rippka R, Coursin T, Hess W, Lichtlé C, Scanlan DJ, Palinska KA, *et al.* (2000). *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* **50**: 1833–1847.

Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY. (1979). Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria. *Microbiology* **111**: 1–61.

Rivers AR, Jakuba RW, Webb EA. (2009). Iron stress genes in marine *Synechococcus* and the development of a flow cytometric iron stress assay. *Environ Microbiol* **11**: 382–396.

Rocap G, Distel DL, Waterbury JB, Chisholm SW. (2002). Resolution of *Prochlorococcus* and *Synechococcus* Ecotypes by Using 16S-23S Ribosomal DNA Internal Transcribed Spacer Sequences. *Appl Environ Microbiol* **68**: 1180–1191.

Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA, *et al.* (2003). Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**: 1042–1047.

Rocha EP. (2008). Evolutionary patterns in prokaryotic genomes. *Curr Opin Microbiol* **11**: 454–460.

Rodriguez-r LM, Konstantinidis KT. (2014). Identify Bacterial Species. Microbe.

Rodriguez-Valera F, Martin-Cuadrado A-B, Rodriguez-Brito B, Pašić L, Thingstad TF, Rohwer F, *et al.* (2009). Explaining microbial population genomics through phage predation. *Nat Rev Microbiol* **7**: 828–836.

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

S

Saito M a., Rocap G, Moffett JW. (2005). Production of cobalt binding ligands in a *Synechococcus* feature at the Costa Rica upwelling dome. *Limnol Oceanogr* **50**: 279–290.

Saito MA, McIlvin MR, Moran DM, Goepfert TJ, DiTullio GR, Post AF, *et al.* (2014). Multiple nutrient stresses at intersecting Pacific Ocean biomes detected by protein biomarkers. *Science* **345**: 1173–7.

Sánchez-Baracaldo P. (2015). Origin of marine planktonic cyanobacteria. *Sci Rep* **5**: 17418.

Scanlan DJ. (2012). Marine Picocyanobacteria. In: Whitton BA (ed). *Ecology of Cyanobacteria II: Their Diversity in Space and Time*. Springer Netherlands: Dordrecht, pp 503–533.

Scanlan DJ. (2003). Physiological diversity and niche adaptation in marine *Synechococcus*. *Adv Microb Physiol* **47**: 1–64.

Scanlan DJ, Mann NH, Carr NG. (1993). The response of the picoplanktonic marine cyanobacterium *Synechococcus* species WH7803 to phosphate starvation involves a protein homologous to the periplasmic phosphate-binding protein of *Escherichia coli*. *Mol Microbiol* **10**: 181–191.

Scanlan DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR, *et al.* (2009). Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol Rev* **73**: 249–99.

Schirrmeister BE, Gugger M, Donoghue PCJ. (2015). Cyanobacteria and the great oxidation event: Evidence from genes and fossils. *Palaeontology* **58**: 769–785.

Scybert S, Pechous R, Sitthisak S, Nadakavukaren MJ, Wilkinson BJ, Jayaswal RK. (2003). NaCl-sensitive mutant of Staphylococcus aureus has a Tn917-lacZ insertion in its ars operon. *FEMS Microbiol Lett* **222**: 171–176.

Shapiro BJ, Friedman J, Cordero OX, Preheim SP, Timberlake SC, Szabo G, *et al.* (2012). Population Genomics of Early Events in the Ecological Differentiation of Bacteria. *Science (80-)* **336**: 48–51.

Shih PM, Wu D, Latifi A, Axen SD, Fewer DP, Talla E, *et al.* (2013). Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci U S A* **110**: 1053–8.

Shilova IN, Mills MM, Robidart JC, Turk-Kubo KA, Björkman KM, Kolber Z, *et*

al. (2017). Differential effects of nitrate, ammonium, and urea as N sources for microbial communities in the North Pacific Ocean. *Limnol Oceanogr*. e-pub ahead of print, doi: 10.1002/lno.10590.

Sieburth JM, Smetacek V, Lenz J. (1978). Pelagic ecosystem structure: Heterotrophic compartments of the plankton and their relationship to plankton size fractions. *Limnol Oceanogr* **23**: 1256–1263.

Simm S, Keller M, Selymesi M, Schleiff E. (2015). The composition of the global and feature specific cyanobacterial core-genomes. *Front Microbiol* **6**: 1–21.

Six C, Finkel Z V., Irwin AJ, Campbell DA. (2007a). Light variability illuminates niche-partitioning among marine Picocyanobacteria. *PLoS One* **2**: e1341.

Six C, Joubin L, Partensky F, Holtzendorff J, Garczarek L. (2007b). UV-induced phycobilisome dismantling in the marine picocyanobacterium *Synechococcus* sp. WH8102. *Photosynth Res* **92**: 75–86.

Six C, Thomas J-C, Garczarek L, Ostrowski M, Dufresne A, Blot N, *et al.* (2007c). Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.: a comparative genomics study. *Genome Biol* **8**: R259.

Six C, Thomas JC, Brahamsha B, Lemoine Y, Partensky F. (2004). Photophysiology of the marine cyanobacterium *Synechococcus* sp. WH8102, a new model organism. *Aquat Microb Ecol* **35**: 17–29.

Snel B, Bork P, Huynen M a. (1999). Genome phylogeny based on gene content. *Nat Genet* **21**: 108–110.

Snyder DS, Brahamsha B, Azadi P, Palenik B. (2009). Structure of Compositionally Simple Lipopolysaccharide from Marine *Synechococcus*. *J Bacteriol* **191**: 5499–5509.

Sohm J a, Ahlgren N a, Thomson ZJ, Williams C, Moffett JW, Saito M a, *et al.* (2015). Co-occurring *Synechococcus* ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. *ISME J* **10**: 1–13.

Spratt BG, Hanage WP, Feil EJ. (2001). The relative contributions of recombination and point mutation to the diversification of bacterial clones. *Curr Opin Microbiol* **4**: 602–606.

Steglich C, Behrenfeld M, Koblizek M, Claustre H, Penno S, Prasil O, *et al.* (2001). Nitrogen deprivation strongly affects Photosystem II but not phycoerythrin level in the divinyl-chlorophyll b -containing cyanobacterium *Prochlorococcus* marinus. *Biochim Biophys Acta* **1503**: 341–349.

Steglich C, Futschik M, Rector T, Steen R, Chisholm SW. (2006). Genome-wide analysis of light sensing in *Prochlorococcus*. *J Bacteriol* **188**: 7796–7806.

Steglich C, Futschik ME, Lindell D, Voss B, Chisholm SW, Hess WR. (2008). The challenge of regulation in a minimal photoautotroph: Non-coding RNAs in *Prochlorococcus*. *PLoS Genet* **4**: e1000173.

Steinberg DK, Carlson CA, Bates NR, Johnson RJ, Michaels AF, Knap AH. (2001). Overview of the US JGOFS Bermuda Atlantic Time-series Study (BATS): A decade-scale look at ocean biology and biogeochemistry. *Deep Res Part II Top Stud Oceanogr* **48**: 1405–1447.

Stoecker DK, Hansen PJ, Caron DA, Mitra A. (2017). Mixotrophy in the Marine Plankton. *Ann Rev Mar Sci* **9**: 311–335.

Stomp M, Huisman J, Stal LJ, Matthijs HCP. (2007). Colorful niches of phototrophic microorganisms shaped by vibrations of the water molecule. *ISME J* **1**: 271–82.

Stuart RK, Brahamsha B, Busby K, Palenik B. (2013). Genomic island genes in a coastal marine *Synechococcus* strain confer enhanced tolerance to copper and oxidative stress. *ISME J* **7**: 1139–1149.

Stuart RK, Dupont CL, Johnson DA, Paulsen IT, Palenik B. (2009). Coastal strains of marine *Synechococcus* species exhibit increased tolerance to copper shock and a distinctive transcriptional response relative to those of open-ocean strains. *Appl Environ Microbiol* **75**: 5047–5057.

Su Z, Dam P, Chen X, Olman V, Jiang T, Palenik B, *et al.* (2003). Computational Inference of Regulatory Pathways in

Microbes: an Application to Phosphorus Assimilation Pathways in *Synechococcus* sp. WH8102. *Genome Informatics* **14**: 3–13.

Su Z, Mao F, Dam P, Wu H, Olman V, Paulsen IT, *et al.* (2006). Computational inference and experimental validation of the nitrogen assimilation regulatory network in cyanobacterium *Synechococcus* sp . WH 8102. *Nucleic Acids Res* **34**: 1050–1065.

Su Z, Olman V, Mao F, Xu Y. (2005). Comparative genomics analysis of NtcA regulons in cyanobacteria: regulation of nitrogen assimilation and its coupling to photosynthesis. *Nucleic Acids Res* **33**: 5156–5171.

Su Z, Olman V, Xu Y. (2007). Computational prediction of Pho regulons in cyanobacteria. *BMC Genomics* **8**: 156.

Sullivan MB, Coleman ML, Weigele P, Rohwer F, Chisholm SW. (2005). Three *Prochlorococcus* cyanophage genomes: Signature features and ecological interpretations. *PLoS Biol* **3**: 0790–0806.

Sullivan MB, Lindell D, Lee JA, Thompson LR, Bielawski JP, Chisholm SW. (2006). Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol* **4**: 1344–1357.

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

Swan BK, Tupper B, Sczyrba A, Lauro FM, Martinez-Garcia M, González JM, *et al.* (2013). Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc Natl Acad Sci U S A* **110**: 11463–8.

Sweeney B, Borgese M. (1989). A circadian rhythm in cell division in a prokaryote, the cyanobacterium *Synechococcus* WH78031. *J Phycol* **25**: 183–186.

T

Tagliabue A, Bowie AR, Boyd PW, Buck KN, Johnson KS, Saito MA. (2017). The integral role of iron in ocean biogeochemistry. *Nature* **543**: 51–59.

Tai V, Burton R, Palenik B. (2011). Temporal and spatial distributions of marine *Synechococcus* in the Southern California Bight assessed by hybridization to bead-arrays. *Mar Ecol Prog Ser* **426**: 133–147.

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

Talarmin A, Van Wambeke F, Lebaron P, Moutin T. (2015). Vertical partitioning of phosphate uptake among picoplankton groups in the low Pi Mediterranean Sea. *Biogeosciences* **12**: 1237–1247.

Tamura K, Battistuzzi FU, Billing-Ross P, Murillo O, Filipski A, Kumar S. (2012). Estimating divergence times in large molecular phylogenies. *Proc Natl Acad Sci U S A* **109**: 19333–8.

Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, *et al.* (2005). Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial 'pan-genome'. *Proc Natl Acad Sci U S A* **102**: 13950–5.

Tetu SG, Brahamsha B, Johnson DA, Tai V, Phillippy K, Palenik B, *et al.* (2009). Microarray analysis of phosphate regulation in the marine cyanobacterium *Synechococcus* sp. WH8102. *ISME J* **3**: 835–849.

Tetu SG, Johnson DA, Varkey D, Phillippy K, Stuart RK, Dupont CL, *et al.* (2013). Impact of DNA damaging agents on genome-wide transcriptional profiles in two marine *Synechococcus* species. *Front Microbiol* **4**: 1–15.

Thompson AW, Huang K, Saito MA, Chisholm SW. (2011). Transcriptome response of high- and low-light-adapted *Prochlorococcus* strains to changing iron availability. *ISME J* **5**: 1580–1594.

Thrash JC, Temperton B, Swan BK, Landry ZC, Woyke T, DeLong EF, *et al.* (2014). Single-cell enabled comparative genomics of a deep ocean SAR11 bathytype. *ISME J* **8**: 1440–51.

Tichý M, Lupínková L, Sicora C, Vass I, Kuviková S, Prášil O, *et al.* (2003). *Synechocystis* 6803 mutants expressing distinct forms of the Photosystem II D1 protein from *Synechococcus* 7942: Relationship between the psbA coding region and sensitivity to visible and UV-B radiation. *Biochim Biophys Acta - Bioenerg* **1605**: 55–66.

Ting CS, Hsieh C, Sundararaman S, Mannella C, Marko M. (2007). Cryo-electron tomography reveals the comparative three-dimensional architecture of *Prochlorococcus*, a globally important marine cyanobacterium. *J Bacteriol* **189**: 4485–4493.

Toledo G, Palenik B. (1997). *Synechococcus* diversity in the California Current as seen by RNA polymerase (rpoC1) gene sequences of isolated strains. *Appl Environ Microbiol* **63**: 4298–4303.

Tolonen AC, Aach J, Lindell D, Johnson ZI, Rector T, Steen R, *et al.* (2006). Global gene expression of *Prochlorococcus* ecotypes in response to changes in nitrogen availability. *Mol Syst Biol* **2**: 1–15.

Turesson G. (1922). The species and the variety as ecological units. *Hereditas* **3**: 100–113.

Tyystjärvi E. (2008). Photoinhibition of Photosystem II and photodamage of the oxygen evolving manganese cluster. *Coord Chem Rev* **252**: 361–376.

U

Urbach E, Scanlan DJ, Distel DL, Waterbury JB, Chisholm SW. (1998). Rapid diversification of marine picophytoplankton with dissimilar light- harvesting structures inferred from sequences of *Prochlorococcus* and *Synechococcus* (cyanobacteria). *J Mol Evol* **46**: 188–201.

van Valen L. (1973). A new evolutionary law. *Evol Theory* **1**: 1–30.

V

Valladares A, Montesinos ML, Herrero A, Flores E. (2002). An ABC-type, high-affinity urea permease identified in cyanobacteria. *Mol Microbiol* **43**: 703–715.

Varet H, Brillet-Guéguen L, Coppée JY, Dillies MA. (2016). SARTools: A DESeq2- and edgeR-based R pipeline for comprehensive differential analysis of RNA-Seq data. *PLoS One* **11**: 1–8.

de Vargas C, Audic S, Henry N, Decelle J, Mahe F, Logares R, *et al.* (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**. e-pub ahead of print, doi: 10.1126/science.1261605.

Varkey D, Mazard S, Ostrowski M, Tetu SG, Haynes P, Paulsen IT. (2016). Effects of low temperature on tropical and temperate isolates of marine *Synechococcus*. *ISME J* **10**: 1252–1263.

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

Vaulot D, LeBot N, Marie D, Fukai E. (1996). Effect of phosphorus on the *Synechococcus* cell cycle in surface mediterranean waters during summer. *Appl Environ Microbiol* **62**: 2527–2533.

Vega-palas MA, Madueno F, Herrero A, Flores E. (1990). Identification and Cloning of a Regulatory Gene for Nitrogen Assimilation in the Cyanobacterium *Synechococcus* sp . Strain PCC 7942. *J Bacteriol* **172**: 643–647.

W

Waterbury JB, Watson SW, Guillard RRL, Brand LE. (1979). Widespread occurance of a unicellular, marine, planktonic, cyanobacterium. *Nature* **277**: 293–294.

Waterbury JB, Watson SW, Valois FW, Franks DG. (1986). Biological and ecological characterization of the marine unicellular cyanobacterium *Synechococcus*. *Can Bull Fish Aquat Sci* **214**: 71–120.

Waterbury JB, Willey JM, Franks DG, Valois FW, Watson SW. (1985). A Cyanobacterium Capable of Swimming Motility. *Science* **230**: 74–76.

Webb EA, Moffett JW, Waterbury JB. (2001). Iron Stress in Open-Ocean Cyanobacteria (*Synechococcus*, Trichodesmium, and Crocosphaera spp.): Identification of the IdiA Protein. *Appl Environ Microbiol* **67**: 5444–5452.

Welch RA, Burland V, Plunkett G, Redford P, Roesch P, Rasko D, *et al.* (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci U S A* **99**: 17020–4.

West NJ, Lebaron P, Strutton PG, Suzuki MT. (2011). A novel clade of *Prochlorococcus* found in high nutrient low chlorophyll waters in the South and Equatorial Pacific Ocean. *ISME J* **5**: 933–944.

West NJ, Scanlan DJ. (1999). Niche-partitioning of *Prochlorococcus* populations in a stratified water column in the eastern North Atlantic Ocean. *Appl Environ Microbiol* **65**: 2585–2591.

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

Willey JM, Waterbury JB. (1989). Chemotaxis toward Nitrogenous Compounds by Swimming Strains of Marine *Synechococcus* spp. *Appl Environ Microbiol* **55**: 1888–1894.

De Wit R, Bouvier T. (2006). 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? *Environ Microbiol* **8**: 755–758.

Worden AZ, Follows MJ, Giovannoni SJ, Wilken S, Zimmerman AE, Keeling PJ. (2015). Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science* **347**: 1257594.1-10.

Wyman M. (1992). An in vivo method for the estimation of phycoerythrin concentrations in marine cyanobacteria (*Synechococcus* spp.). *Limnol Oceanogr* **37**: 1300–1306.

Wyman M, Bird C. (2007). Lack of Control of Nitrite Assimilation by Ammonium in an Oceanic Picocyanobacterium, *Synechococcus* sp. Strain WH8103. *Appl Environ Microbiol* **73**: 3028–3033.

Wyman M, Gregory RPF, Carr N. (1985). Novel Role for Phycoerythrin in a Marine Cyanobacterium, *Synechococcus* Strain DC2. *Science* **230**: 818–920.

x

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

y

Yelton AP, Acinas SG, Sunagawa S, Bork P, Pedrós-Alió C, Chisholm SW. (2016). Global genetic capacity for mixotrophy in marine picocyanobacteria. *ISME J* 1–12.

Yooseph S, Nealson KH, Rusch DB, McCrow JP, Dupont CL, Kim M, *et al.* (2010). Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature* **468**: 60–6.

z

Zehr JP, Weitz JS, Joint I. (2017). How microbes survive in the open ocean. *Science (80-)* **357**: 646–647.

Zhang B, Horvath S. (2005). A General Framework for Weighted Gene Co-Expression Network Analysis. *Stat Appl*

Genet Mol Biol **4**. e-pub ahead of print, doi: 10.2202/1544-6115.1128.

Zhao Y, Jia X, Yang J, Ling Y, Zhang Z, Yu J, *et al.* (2014). PanGP: A tool for quickly analyzing bacterial pan-genome profile. *Bioinformatics* **30**: 1297–1299.

Zinser ER, Johnson ZI, Coe A, Karaca E, Veneziano D, Chisholm SW. (2007). Influence of light and temperature on *Prochlorococcus* ecotype distributions in the Atlantic Ocean. *Limnol Oceanogr* **52**: 2205–2220.

Zinser ER, Lindell D, Johnson ZI, Futschik ME, Steglich C, Coleman ML, *et al.* (2009). Choreography of the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph, *Prochlorococcus*. *PLoS One* **4**: e5135.

Zubkov M V, Fuchs BM, Tarran GA, Burkill PH, Amann R. (2003). High Rate of Uptake of Organic Nitrogen Compounds by *Prochlorococcus* Cyanobacteria as a Key to Their Dominance in Oligotrophic Oceanic Waters. *Appl Environ Microbiol* **69**: 1299–1304.

Zubkov M V, Mary I, Woodward EMS, Warwick PE, Fuchs BM, Scanlan DJ, *et al.* (2007). Microbial control of phosphate in the nutrient-depleted North Atlantic subtropical gyre. *Environ Microbiol* **9**: 2079–2089.

Zwirglmaier K, Heywood JL, Chamberlain K, Woodward EMS, Zubkov M V, Scanlan DJ. (2007). Basin-scale distribution patterns of picocyanobacterial lineages in the Atlantic Ocean. *Environ Microbiol* **9**: 1278–1290.

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

ANNEXES

ANNEXE 1 : Light color acclimation: a key process in the global ocean distribution of *Synechococcus* cyanobacteria.

Grébert, T., Doré, H., Partensky, F., Farrant, G.K., Emmanuel, B., Picheral, M., et al. Light color acclimation: a key process in the global ocean distribution of *Synechococcus* cyanobacteria. *Proc. Natl. Acad. Sci. in revision*.

Light coloracclimation: a key process in the global ocean distribution of *Synechococcus* cyanobacteria

Théophile Grébert[a], Hugo Doré[a], Frédéric Partensky[a], Gregory K. Farrant[a,1], Emmanuel Boss[b], Marc Picheral[c], Lionel Guidi[c], Stéphane Pesant[d,e], Patrick Wincker[f], Silvia G. Acinas[g], David M. Kehoe[h] and Laurence Garczarek[a,2]

[a]Sorbonne Universités-Université Paris 06 & Centre National de la Recherche Scientifique (CNRS), UMR 7144, Marine Phototrophic Prokaryotes Team, Station Biologique, CS 90074, 29688 Roscoff cedex, France; [b]Maine In-situ Sound and Color Lab, University of Maine, Orono, ME 04469; [c]Sorbonne Universités-Université Paris 06 & CNRS, UMR 7093, Observatoire océanologique, 06230 Villefranche-sur-mer, France; [d]PANGAEA, Data Publisher for Earth and Environmental Science, University of Bremen, Bremen, Germany; [e]MARUM, Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany; [f]Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), Institut de Génomique, Genoscope, 91057 Evry, France; [g]Department of Marine Biology and Oceanography, Institute of Marine Sciences (ICM), Consejo Superior de Investigaciones Científicas (CSIC), Barcelona ES-08003, Spain; [h]Department of Biology, Indiana University, Bloomington, IN 47405 [1]Present address: Food Safety, Environment, and Genetics, Matís Ltd., 113 Reykjavík, Iceland. [2]To whom correspondence should be addressed. Email: laurence.garczarek@sb-roscoff.fr.

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Marine *Synechococcus* cyanobacteria are major contributors to global oceanic primary production and exhibit a unique diversity of photosynthetic pigments, allowing them to exploit a wide range of light niches. However, the relationship between pigment content and niche partitioning has remained largely undetermined due to the lack of a single-genetic marker resolving all pigment types (PT). Here, we developed and employed a novel and robust method based on three distinct marker genes to estimate the relative abundance of all known *Synechococcus* PTs from metagenomes. Analysis of the *Tara* Oceans dataset allowed us, for the first time, to reveal the global distribution of *Synechococcus* PTs and to define their environmental niches. Green-light specialists (PT 3a) dominated in warm, green equatorial waters, whereas blue-light specialists (PT 3c) were particularly abundant in oligotrophic areas. Type IV chromatic acclimaters (CA4-A/B), which are able to dynamically modify their light absorption properties to maximally absorb green or blue light, were unexpectedly the most abundant PT in our dataset and predominated at depth and high latitudes. We also identified local populations in which CA4 might be inactive due to the lack of specific CA4 genes, notably in warm high nutrient low chlorophyll areas. Major ecotypes within clades I-IV and CRD1 were preferentially associated with a particular PT, while others exhibited a wide range of PTs. Altogether, this study provides unprecedented insights into the ecology of *Synechococcus* PTs and highlights the complex interactions between vertical phylogeny, pigmentation and environmental parameters that shape *Synechococcus* community structure and evolution.

metagenomics | synechococcus | pigment type | best paper

Introduction

Marine *Synechococcus* is the second most abundant phytoplanktonic organism in the world's oceans and constitutes a major contributor to global primary production and carbon cycling (1, 2). This genus displays a wide genetic diversity and several studies have shown that among the ~20 clades defined based on various genetic markers, five (I-IV and CRD1) predominate *in situ* and can be broadly associated with distinct sets of physico-chemical parameters (3–5). In a recent study, we further defined Ecologically Significant Taxonomic Units (ESTUs), i.e. organisms belonging to the same clade and co-occurring in the field, and highlighted that the three main parameters affecting the *in situ* distribution of these ESTU were temperature and availability of iron and phosphorus (6). Yet, marine *Synechococcus* also display a wide pigment diversity, suggesting that light could also influence their ecological distribution, both qualitatively and quantitatively (7, 8).

This diversity comes from differences in the composition of in their main light-harvesting antennae, called phycobilisomes (PBS, 7–9). These water-soluble macromolecular complexes consist of a central core anchoring six or eight radiating rods made of several distinct phycobiliproteins, i.e. proteins to which specific enzymes (phycobilin lyases) covalently attach chromophores called phycobilins (7, 10). Although the PBS core is conserved in all marine *Synechococcus*, rods have a very variable composition, and three main pigment types (PT) are usually distinguished (Fig. S1, 7, 11). In PT 1, PBS rods are solely made of phycocyanin (PC, encoded by the *cpcBA* operon) and bear the red-light absorbing chromophore phycocyanobilin (PCB; A_{max} = 620 nm) as the sole chromophore. In PT 2, rods are made of PC and phycoerythrin I (PE-I, encoded by *cpeBA*) and attach both PCB and the green-light (GL) absorbing phycoerythrobilin (PEB; A_{max} = 550 nm). All other marine *Synechococcus* belong to PT 3 and have rods made of PC, PE-I and PE-II (encoded by *mpeBA*) that bind PCB, PEB and the blue-light (BL) absorbing phycourobilin (PUB; A_{max} = 495 nm; Fig. S1). Several subtypes can be defined within PT3, based on the fluorescence excitation ratio at 495 nm and 545 nm with emission at 585 nm (hereafter $Ex_{495:545}$, Fig. S1), a proxy for the PUB:PEB ratio. This ratio is low ($Ex_{495:545}$ <0.6) in subtype 3a (GL specialists), intermediate in subtype 3b (0.6 ≤ $Ex_{495:545}$ <1.6) and high ($Ex_{495:545}$ ≥ 1.6) in subtype 3c (BL specialists; 7, 11). Additionally, some strains of subtype 3d are

Significance

Understanding the functional diversity of specific microbial groups on a global scale is critical yet poorly developed. By combining the considerable knowledge accumulated through recent years on the molecular bases of photosynthetic pigment diversity in marine *Synechococcus*, a major phytoplanktonic organism, with the wealth of metagenomic data provided by the *Tara* Oceans expedition, we have been able to reliably quantify all known pigment types along the transect and provide the first global distribution map. Unexpectedly, cells able to dynamically change their pigment content to match the ambient light color were ubiquitous and predominated in many environments. Altogether, our results enlighten the role of adaptation to light quality on niche partitioning in a key primary producer.

Reserved for Publication Footnotes

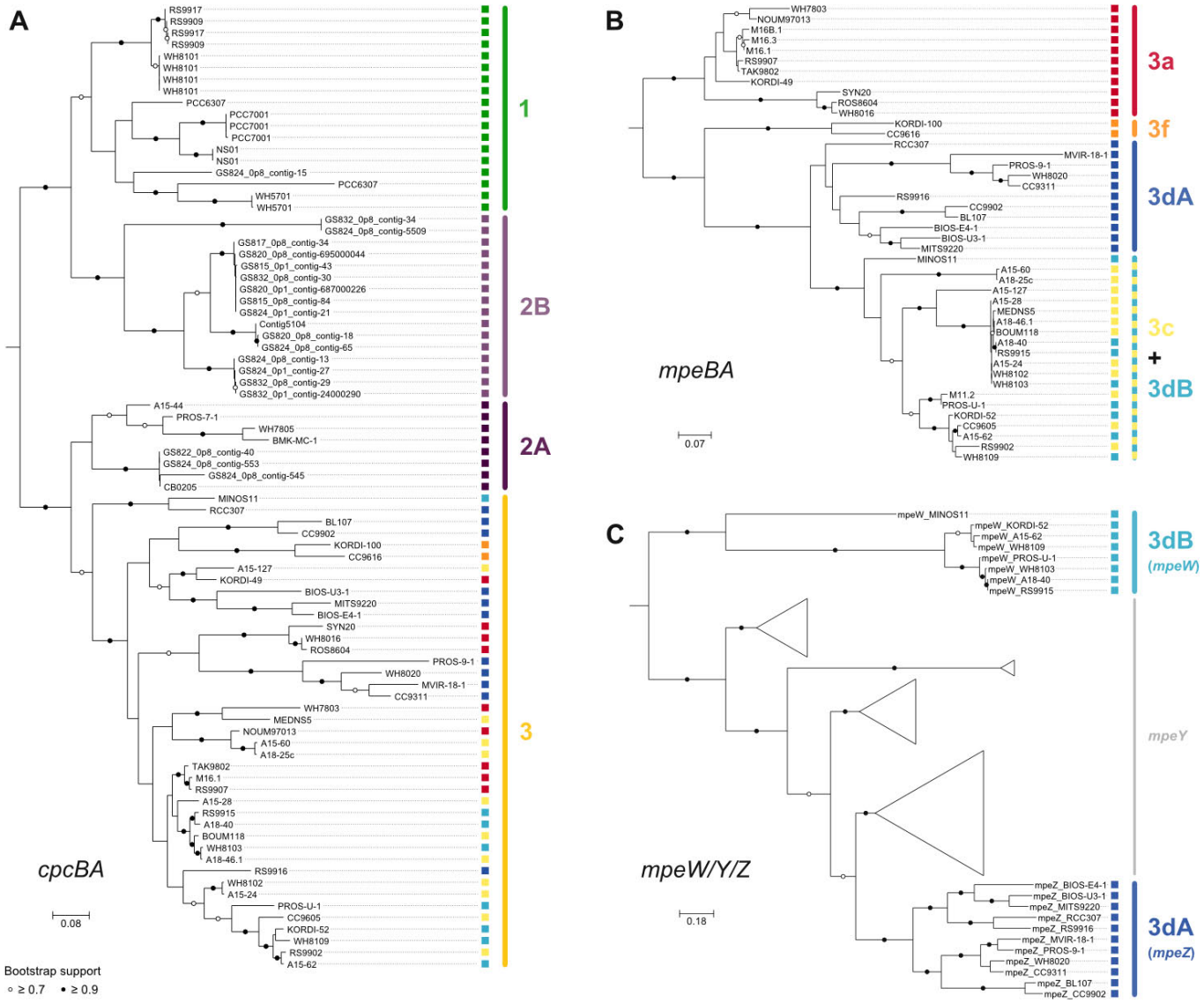


Fig. 1. : Maximum likelihood phylogenetic trees of (*A*) *cpcBA* operon, (*B*) *mpeBA* operon and (*C*) the *mpeW/Y/Z* gene family. The *cpcBA* tree includes both strains with characterized pigment type (PT) and environmental sequences (prefixed with GS) assembled from metagenomes of the Baltic Sea (38). Circles at nodes indicate bootstrap support (black: > 90 %; white: > 70 %). Note that for PT 2B clade, only environmental sequences are available. PT associated to each sequence is indicated as a colored square. The scale bar represents the number of substitutions per nucleotide position.

able to change their PUB:PEB ratio depending on ambient light color, a process called type IV chromatic acclimation (hereafter CA4), allowing them to maximally absorb BL or GL (11–14). Comparative genomic analyses further showed that genes involved in the synthesis and regulation of PBS rods are gathered into a dedicated genomic region, the content and organization of which correspond to the different PTs (7). Similarly, chromatic acclimation has been correlated with the presence of a small specific genomic island (CA4 genomic island) that exists in two distinct configurations (CA4-A and CA4-B, 11). Both contain two regulators (*fciA* and *fciB*) and a phycobilin lyase (*mpeZ* in CA4-A or *mpeW* in CA4-B), thus defining two distinct CA4 genotypes: 3dA and 3dB (11, 14, 15). Finally, some strains possess a complete or partial CA4 genomic island but are not able to perform CA4, displaying a fixed $Ex_{495:545}$ corresponding to 3a, 3b or 3c phenotypes (11).

As there is no correspondence between pigmentation and core genome phylogeny (7, 16, 17), deciphering the relative abundance and niche partitioning of *Synechococcus* PTs in the environment requires specific approaches. In the past 30 years, studies have been based either on i) proxies of the PUB:PEB ratio as assessed by flow cytometry (18–20), fluorescence excitation spectra (21–27), epifluorescence microscopy (28), or ii) phylogenetic analyses of *cpcBA* or *cpeBA* (17, 29–34). However, analyses based on optical properties could only describe the distribution of high- and low-PUB populations without being able to differentiate GL (3a) or BL (3c) specialists from CA4 cells (3d) acclimated to GL or BL, while genetic analysis solely based either on *cpcBA* or *cpeBA* could not differentiate all PTs. For instance, only two studies have reported CA4 populations *in situ* either in the western English Channel (17) or in sub-polar waters of the western Pacific Ocean (29) but none of them were able to differentiate CA4-B from high PUB (i.e. 3c) populations. As a consequence, the global relative abundance of CA4 as well as the link between genetic and pigment diversity has remained largely unclear.

Here, we analyzed 109 metagenomic samples collected from all major oceanic basins during the 2.5-yr *Tara* Oceans (2009-

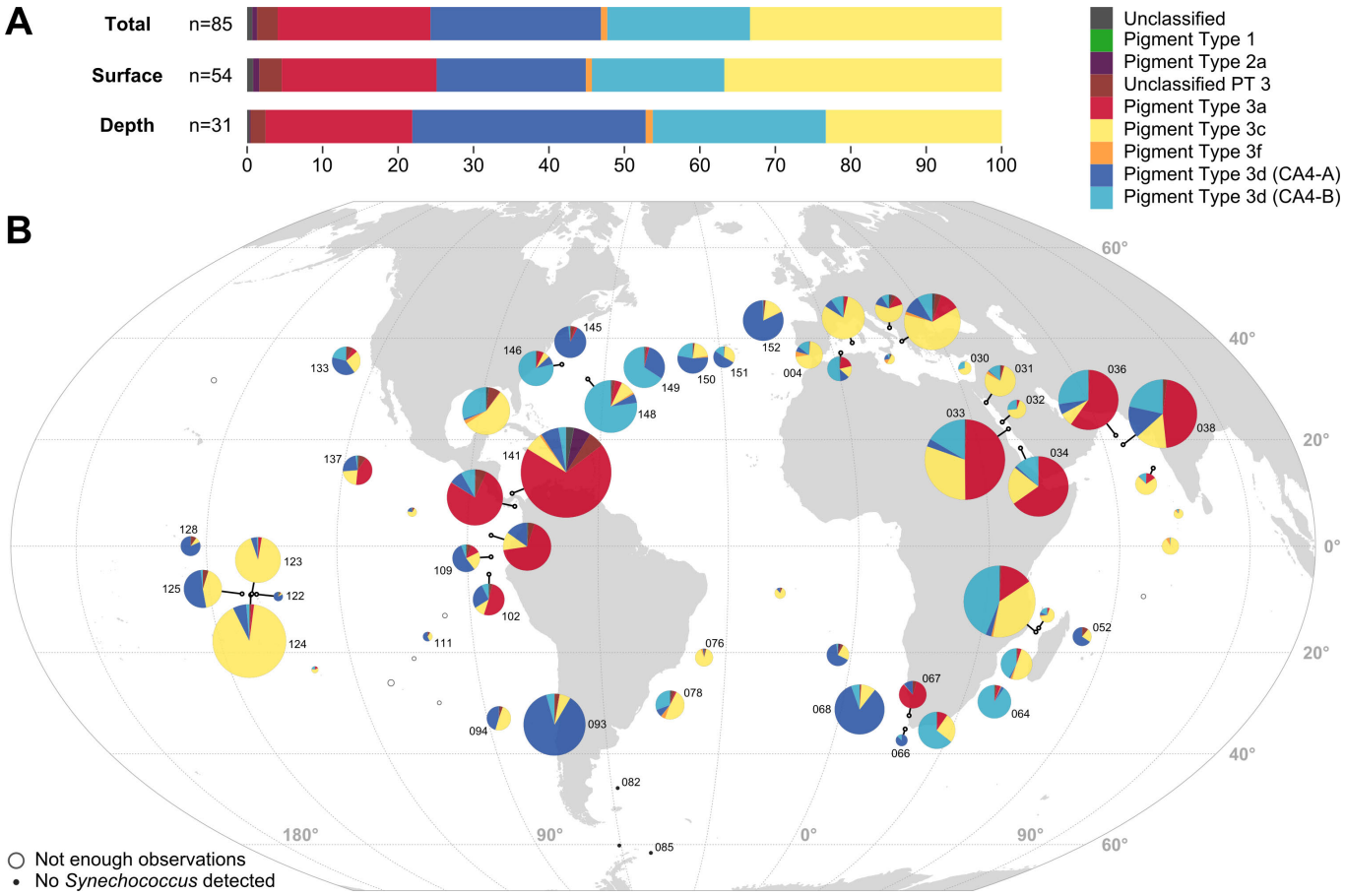


Fig. 2. Distribution of *Synechococcus* pigment types (PT). (*A*) Relative abundance of each PT in the whole dataset (Total), in surface and at the DCM (Deep Chlorophyll Maximum). (*B*) Map showing the global distribution of all *Synechococcus* PTs in surface waters along the *Tara* Oceans transect. Diameters of pies are proportional to the number of *cpcBA* reads normalized by the sequencing effort. Stations with less than 30 *cpcBA* or *mpeBA* reads are indicated by open circles and those with no *cpcBA* reads by black dots. Numbers next to pies correspond to *Tara* Oceans stations.

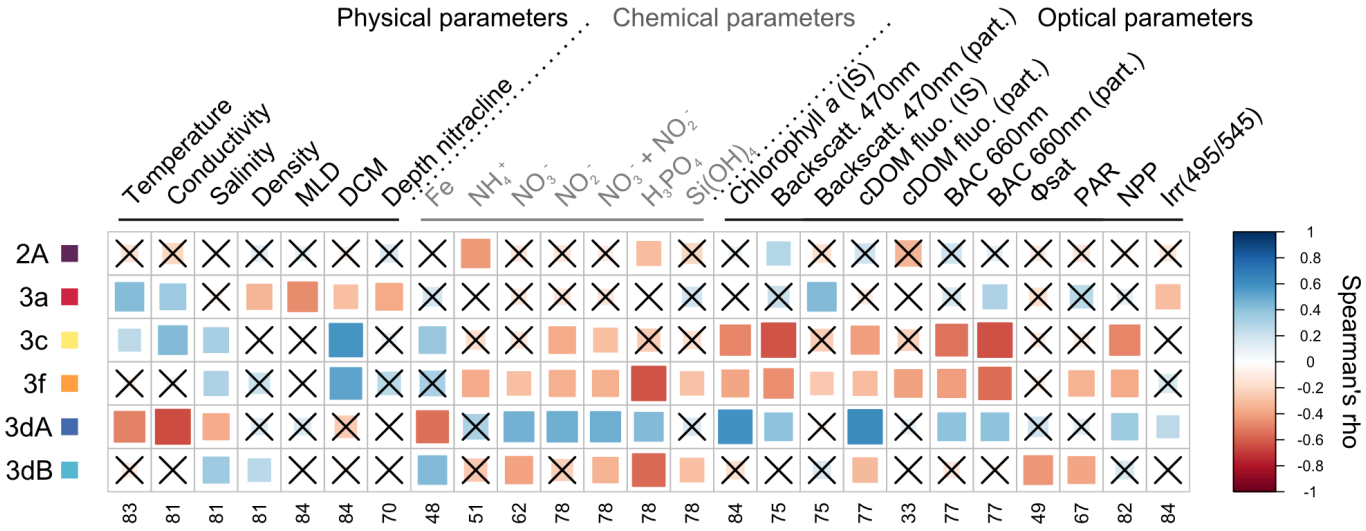


Fig. 3. Correlation analysis between *Synechococcus* pigment types (PT) and environmental parameters measured along the *Tara* Oceans transect for all sampled depths. The scale shows the degree of correlation (blue) or anti-correlation (red) between two variables. Non-significant correlations (adjusted *P* value > 0.05) are indicated by crosses. Number of observations for each environmental parameter is indicated at the bottom. Abbreviations: MLD, mixed layer depth; DCM, deep chlorophyll maximum; IS, *in situ*; Backscatt., backscattering; cDOM fluo, colored dissolved organic matter fluorescence; BAC, beam attenuation coefficient; Φsat, satellite-based non-photochemical quenching (NPQ)-corrected quantum yield of fluorescence (proxy for iron limitation); PAR, photosynthetically active radiation; NPP, net primary production; Irr(495/545), ratio of downwelling irradiance at 495 nm and 545 nm.

2011) expedition (35) using an original bioinformatic pipeline combining a $_{mi}$Tag approach (6, 36) to recruit single reads from multiple PBS gene markers, and placement of these reads in reference trees to assign them to a given PT. This pipeline has

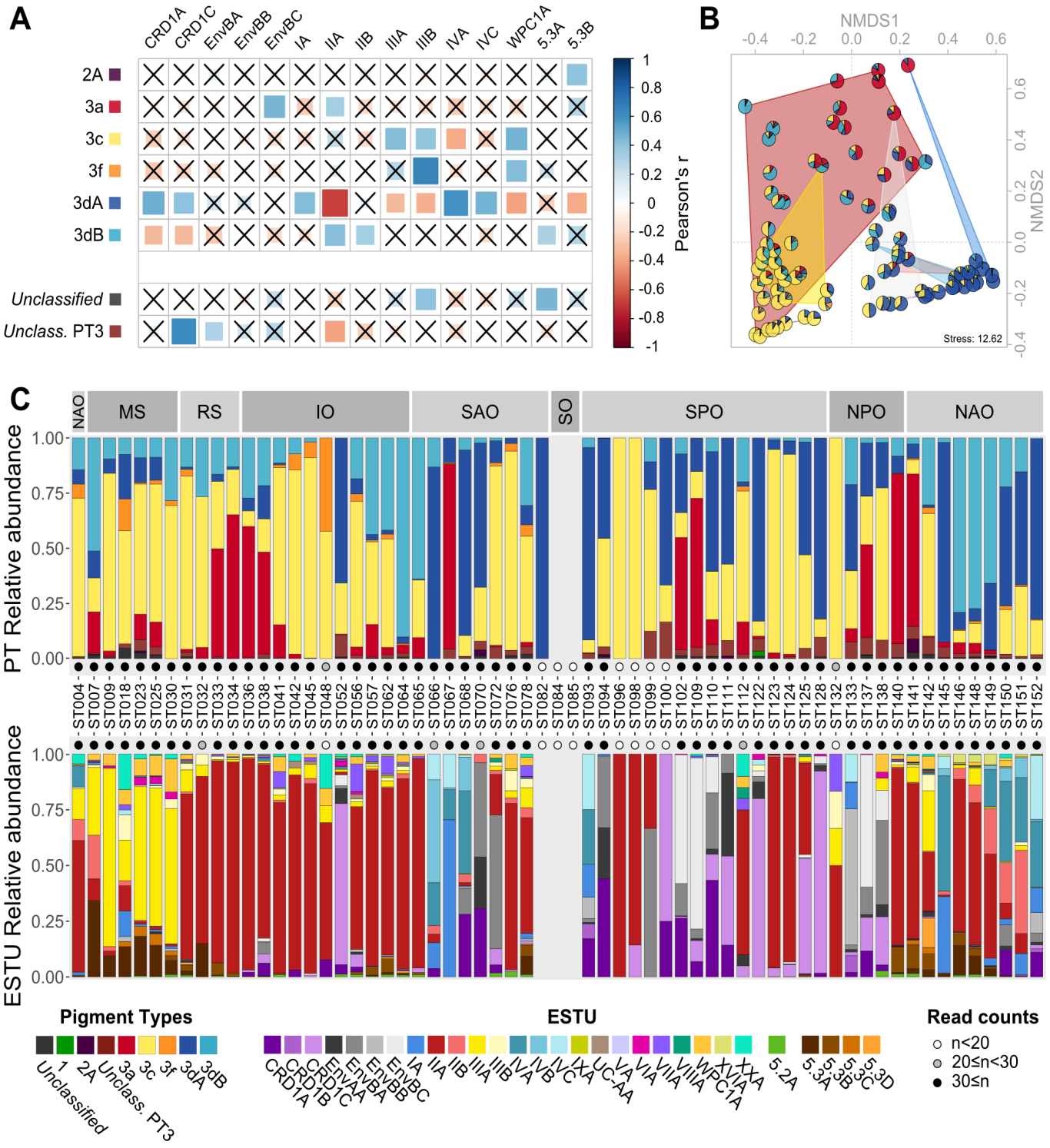


Fig. 4. Relationship between *Synechococcus* pigment types (PT) and Ecologically Significant Taxonomic Units (ESTUs, as defined in 6). (*A*) Correlation analysis between *Synechococcus* PTs and the most abundant ESTUs (>1% relative abundance) for all sampled depths (the complete dataset is shown in Fig. S5). Non-significant correlations (adjusted *P* value > 0.05) are indicated by crosses. Surface of station TARA_067, identified as an outlier (see Fig. S7), was removed for this analysis. (*B*) NMDS analysis of stations according to Bray–Curtis distance between PT assemblages. Samples that belong to the same ESTU assemblage have been contoured with a background color according to the color code used in (6), namely: red, assemblage 1 dominated by ESTU IIA; yellow, assemblage 2 dominated by ESTU IIIA; dark blue, assemblage 4 dominated by ESTUs IA and IVA/B; pink, assemblage 5 co- dominated by ESTUs IIB and IVA/B; grey, assemblage 6 co-dominated by ESTUs CRD1C and EnvBC; light blue, assemblage 8 co-dominated by ESTUs IVA/B, EnvBB and CRD1A/B. (*C*) PT and ESTU relative abundance at each surface station along the *Tara* Oceans transect. Oceanic provinces are indicated in the top gray panels. NAO, North Atlantic Ocean; MS, Mediterranean Sea; RS, Red Sea; IO, Indian Ocean; SAO, South Atlantic Ocean; SO, Southern Ocean; SPO, South Pacific Ocean; NPO, North Pacific Ocean.

allowed the first description of the worldwide distribution of all known *Synechococcus* PTs, as well as of their realized environmental niches (37). This study provides an unprecedented global

view of how a major photosynthetic organism adapts to natural light color and/or intensity gradients in the ocean.

Results

A novel, robust approach for estimating pigment types abundance from metagenomes

We developed a multi-marker approach combining phylogenetic information retrieved from three different genes or operons (*cpcBA*, *mpeBA* and *mpeW*, Fig. 1) to overcome the issue of fully resolving the whole range of PTs and subtypes. While *cpcBA* discriminated PT 1, 2 and 3 (Fig. 1A), only the *mpeBA* operon, a PT3 specific marker, was able to distinguish the different PT 3 subtypes (Fig. 1B), though as for *cpeBA* it could not differentiate the 3dB type (CA4-B) from BL specialists (i.e. PT 3c; 11, 29). The *mpeW* marker was thus selected to specifically target PT 3dB and, by subtraction, enumerate PT 3c (Fig. 1C). Interestingly, members of PT2 were split into two clear-cut clusters using the *cpcBA* marker, 2A and 2B (Fig. 1A), the latter corresponding to a purely environmental PT identified from assembled metagenomes of the Baltic Sea (38). Strains KORDI-100 and CC9616 also clustered apart from other strains in the *mpeBA* phylogeny, suggesting that they have a divergent evolutionary history from other PT 3 members (Fig. 1B). This was supported by the diverged gene content and order of their PBS rod genomic region, leading us to separate these strains from other PT and refer to them as PT 3f, even though they have a similar phenotype as PT 3c ($Ex_{495:545}$ ratio ≥ 1.6; 30). To investigate the phylogenetic resolution of small fragments of these three markers, 150 bp simulated reads (*Tara* Oceans cleaned/merged reads mean length: 164bp, median: 168bp, sd: 20bp) were generated from all sequences in our reference databases and assigned to a PT using our bioinformatic pipeline. Inferred and known PTs were then compared. The percentage of simulated reads assigned to the correct PT was between 93.2% and 97.0% for all three markers, with less than 2.1-5.6% of reads that could not be classified and an error-rate below 2%, showing that all three markers display a sufficient resolution to reliably assign the different PTs (Fig. S2B, D and F).

To ensure that the different markers could be quantitatively compared in a real dataset, we examined the correlations between the different markers estimates of PT abundances in the 109 metagenomes analyzed in this study. Total *cpcBA* counts were highly correlated (R^2=0.994, n=109, Fig. S3A) with total *Synechococcus* counts obtained with the *petB* gene, which was previously used to study the phylogeography of marine picocyanobacteria (6), and the correlation slope was not significantly different from 1 (slope: 1.040; Wilcoxon's paired difference test p-value=0.356). *cpcBA* is thus as good as *petB* at capturing the total population of *Synechococcus* reads. Moreover, counts of *cpcBA* reads assigned to PT 3 and total *mpeBA* counts (specific for PT 3) were also strongly correlated (R^2=0.996, n=109, Fig. S3B), and not skewed from 1 (slope of 0.991, Wilcoxon's p-value=0.607), indicating that *mpeBA* and *cpcBA* counts can be directly compared. Although no redundant information for PT 3dB is available with the three selected markers, another marker targeting 3dB (*fciAB*) was tested and produced results similar to *mpeW* (Fig. S3C). These results demonstrate that our multi-marker approach can be used to reliably and quantitatively infer the different *Synechococcus* PTs from short metagenomic reads, with PT 1, 2A, 2B abundances being assessed by *cpcBA* normalized counts, PT 3a, 3f and 3dA by *mpeBA* normalized counts, PT 3dB by *mpeW* normalized counts and PT 3c by the difference between *mpeBA* normalized counts for 3c + 3dB and *mpeW* normalized counts. We thus used this approach on the *Tara* Oceans metagenomes, generated from 109 samples collected at 65 stations located in the major oceanic basins (Fig. 2).

CA4 populations are widespread and predominate at depth and high latitude

The latitudinal distribution of *Synechococcus* inferred from *cpcBA* counts is globally consistent with previous studies (2, 6, 39), with *Synechococcus* being present in most oceanic waters, but quasi absent (< 20 *cpcBA* counts) below 60° S (Southern Ocean stations TARA_082 to TARA_085; Fig. 2B). Overall, the number of *cpcBA* recruited reads per station was between 0 and 8,151 (n=63, median: 449, mean: 924, sd: 1478) for surface and 0 and 3,200 (n=46, median:170, mean: 446, sd: 664) for deep chlorophyll maximum (DCM) samples, respectively. Stations with less than 30 *cpcBA* reads were excluded from further analysis.

As expected from *Tara* cruise sampling being mostly performed in oceanic waters, PT 1 and 2, both known to be mostly abundant in coastal waters (Fig. 2A-B; 29, 38, 40, 41; see also refs in Table S1), were almost absent from this dataset (total of 15 and 513 *cpcBA* reads, respectively). While PT 2A was mostly found at the surface at one station off Panama (TARA_141, 417 out of 6,637 reads at this station, Fig. 2B), PT 2B was virtually absent (total of 3 *cpcBA* reads) from our dataset and might thus be confined to the Baltic Sea (38). This low abundance of PT 1 and 2 precluded the correlation analysis between their distribution and physico-chemical parameters. PT 3 was by far the most abundant along the *Tara* Oceans transect, accounting for 99.1 ± 1.4% (mean ± sd) of *cpcBA* reads at stations with ≥30 *cpcBA* read counts. Interestingly, several PT 3 subtypes often co-occurred at a given station.

PT 3a (green light specialists) totaled 20.3% of read counts, with similar abundance between surface (20.5%) and DCM (19.4%) samples, and was particularly abundant in intertropical oceanic borders and regional seas, including the Red Sea, the Arabian Sea and the Panama/Gulf of Mexico area (Fig. 2B). Correlation analyses show that this PT is consistently associated with high temperatures but also with greenish (low blue to green downwelling irradiance ratio: $Irr_{495:545}$), particle-rich waters (high particle backscattering at 470 nm and beam attenuation coefficient at 660 nm; Fig. 3). Still, in contrast with previous studies that reported the distribution of low-PUB populations (18, 21, 23–26), this PT does not seem to be restricted to coastal waters, explaining the absence of correlation with chlorophyll concentration and colored dissolved organic matter (cDOM).

Blue light specialists (PT 3c) appear to be globally widespread, with the exception of high latitude North Atlantic waters, and accounted for 33.4% of reads, with a higher relative abundance at the surface (36.8%) than at the DCM (23.3%, Fig. 2A). This PT is dominant in transparent, oligotrophic, iron-replete areas such as the Mediterranean Sea or South Atlantic and Indian Ocean gyres (Fig. 2B and 4C). In the South Pacific, PT 3c was also found to be predominant in the Marquesas Islands area (TARA_123 and 124), where the coast proximity induced a local iron enrichment (6). Consistently, PT 3c was found to be positively associated with iron concentration, high temperature and deep DCM and anti-correlated with chlorophyll fluorescence, nitrogen concentrations, net primary production (NPP) as well as other related optical parameters, such as backscattering at 470 nm and beam attenuation coefficient at 660 nm (Fig. 3). Despite its rarity, PT 3f seems to thrive in a similar environment, with the highest relative abundances in the Indian Ocean and Mediterranean Sea (Fig. 2B and 4C). Its occurrence in the latter area might explain its strong anti-correlation with phosphorus availability.

The most striking result of this study was the widespread occurrence of both 3dA and 3dB CA4 types, which represented 22.6% and 18.9% of reads respectively, and could locally account for up to 95% of the *Synechococcus* population (Fig. 2, 4C and S4). In contrast to blue and green light specialists, both CA4 types were proportionally less abundant at the surface (19.8% and 17.5%, for 3dA and 3dB, respectively) than at depth (30.9% and 22.9%). Interestingly, PT 3dA and 3dB generally displayed

complementary distributions along the *Tara* Oceans transect (Fig. 2B). PT 3dA was predominant at high latitude in the northern hemisphere as well as in other vertically mixed waters such as in the Chilean upwelling (TARA_093) or in the Agulhas current (TARA_066 and 68; Fig 2B). Accordingly, PT 3dA distribution seems to be driven by low temperature, high nutrient and highly productive waters (high NPP, chlorophyll *a* and optical parameters), a combination of physico-chemical parameters almost opposite to those observed for blue light specialists (PT 3c; Fig. 3). In contrast, PT 3dB shares a number of characteristics with PT 3c, including the anti-correlation with nitrogen concentration and association with iron availability (as indicated by both a positive correlation with [Fe] and negative correlation with the iron limitation proxy ϕsat; Fig. 3), consistent with their widespread occurrence in iron replete oceanic areas. Also noteworthy, PT 3dB was one of the sole PT (with 3f) to be associated with low photosynthetically available radiation (PAR).

Niche partitioning of *Synechococcus* populations rely on a subtle combination of ESTU and PT niches

We previously showed that temperature, iron and phosphorus availability constituted major factors influencing the diversification and niche partitioning of *Synechococcus* ESTUs (i.e., genetically related subgroups within clades that co-occur in the field; 6). Yet, as PT does not follow the vertical phylogeny, its impact on niche partitioning could not be assessed in the previous study (7). In order to decipher the respective roles of genetic and pigment diversity in *Synechococcus* community structure, we examined the relationships between ESTUs and PTs *in situ* abundances through correlation and NMDS analyses (Fig. 4A-B) and compared their respective distributions (Figs. 4C and S4).

Interestingly, all PTs are either preferentially associated with or excluded from a subset of ESTUs. PT 2A is found at low abundance at a few stations along the *Tara* Oceans transect and, when present, it is seemingly associated with the rare ESTU 5.3B (Fig. 4A), an unusual PT/genotype combination so far only observed in metagenomes from freshwater reservoirs (42). PT 3a is associated with ESTUs EnvBC (occurring in low iron areas) and IIA, the major ESTU in the global ocean (Fig. 4A), a result consistent with NMDS analysis, which shows that PT 3a is found in assemblages dominated by these two ESTUs (indicated by red and grey backgrounds in Fig. 4B), as well as with independent observations on cultured strains (Table S1). PT 3c is associated with ESTU IIIA (the dominant ESTU in P-depleted areas), as observed on many isolates (Table S1), and is also linked, like the rare PT 3f, with ESTUs IIIB and WPC1A, both present at lower abundance in P-poor waters (Fig. 4A). PT 3f is also associated with the newly described and low-abundance ESTU XXA (previously EnvC; Fig. S5; 4, 6). Both PT 3f and ESTU XXA were rare in our dataset but systematically co-occurred, in agreement with the fact that the only culture representative of the latter clade belongs to PT 3f (Table S1).

PT 3dA appears to be associated with all ESTUs from clades CRD1 (specific of iron-depleted areas) as well as with those representative of coastal and cold waters (IA, IVA, IVC), but is strikingly anti-correlated with most other major ESTUs (IIA, IIIA and –B, WPC1A and 5.3B; Fig. 4A). This pattern is opposite to PT 3dB that is preferentially found with ESTU IIA, IIB and 5.3A, but not in CRD1A or –C (Fig. 4A). Thus, it seems that the two types of CA4 are found in distinct and complementary sets of clades. Interestingly, our analysis might suggest the occurrence of additional PTs not isolated so far, since a number of reads (0.7% and 2.7% for *cpcBA* and *mpeBA,* respectively, Fig. 2A) could not be assigned to any known PTs. For instance, while most CRD1C seem preferentially associated with PT 3dA, a fraction of the population could only be assigned at the PT 3 level (Fig. 4A). Similarly, a number of reads could not be assigned to any known PT in stations rich in ESTU 5.3A and XXA, although one cannot exclude that this observation might be due to a low number of representative strains, and thus PT reference sequences, for these ESTUs.

The preferred association of PTs with specific ESTUs is also well illustrated by some concomitant shifts of PTs and ESTU assemblages. For instance, in the wintertime North Atlantic Ocean, the shift from 3dB-dominated stations on the western side (TARA_142 and TARA_146-149) to 3dA-dominated stations near European coasts (TARA_150 to 152) and North of Gulf stream (TARA_145) is probably related to the shift in ESTU assemblages occurring along this transect, with ESTU IIA being gradually replaced by ESTU IVA (Fig. 4C; see also 6). Similarly, the takeover of CRD1C by IIA in the Marquesas Island area (TARA_123 to 125), which is iron-enriched with regard to surrounding high nutrient low chlorophyll (HNLC) waters (TARA_122 and 128), perfectly matched the corresponding replacement of PT 3dA by 3c. However, in several other cases, PT shifts were not associated with a concomitant ESTU shift or vice versa. One of the most striking examples of these dissociations is the transect from the Mediterranean Sea to the Indian Ocean, where the entry in the northern Red Sea through the Suez Canal triggered a sharp shift from a IIIA to a IIA-dominated community (TARA_030 and 031) which was not accompanied by any obvious change in PT. Conversely, a sharp rise in the relative abundance of PT 3a was observed in the southern Red Sea/northeastern Indian Ocean (TARA_033 to 038) without changes in the large dominance of ESTU IIA. Altogether, this strongly suggests that a subtle combination of ESTU and PTs respective niche occupancy is responsible for the observed niche partitioning of *Synechococcus* populations.

Deficient chromatic acclimaters are dominant in HNLC areas

Although our results clearly indicate that CA4 cells represent a large proportion of the *Synechococcus* community in a wide range of ecological niches, this must be somewhat tempered by the fact that, in culture, about 30% of the strains possessing a CA4-A or B genomic island are not able to chromatically acclimate (Table S1; 11). Some of these natural mutants have an incomplete CA4 genomic island (Fig. S6K). For example, strains WH8016 (ESTU IA) and KORDI-49 (WPC1A) both lack the CA4-A specific lyase-isomerase MpeZ, an enzyme shown to bind a PUB molecule on PE-II (14), and display a green light specialist phenotype (PT 3a; $Ex_{495:545}$~ 0.4) whatever the ambient light color (11). However, since they possess PT 3a *mpeBA* alleles, reads from field WH8016- or KORDI-49-like cells are adequately counted as PT 3a (Fig. S6K). Another CA4-deficient strain, BIOS-E4-1 (ESTU CRD1C), possesses *mpeZ* and a 3dA *mpeBA* allele but lacks the CA4 regulators FciA and FciB as well as the putative lyase MpeY and exhibits a fixed blue light specialist phenotype (PT 3c; $Ex_{495:545}$~ 1.7; Fig. S6K; 11, 15). Thus, reads from such natural *Synechococcus* CA4-incapable mutants in the field are counted as 3dA using the *mpeBA* marker. Lastly, the strain MVIR-18-1 possesses a complete CA4-A island and a 3dA *mpeBA* allele but lacks *mpeU*, a gene necessary for blue light acclimation (Fig. S6K; 43). While MVIR-18-1 displays a fixed green light phenotype, reads from such *Synechococcus* are also erroneously counted as 3dA.

To assess the significance of these genotypes in the field, we compared the normalized read counts obtained for 3dA with *mpeBA*, *fciAB*, *mpeZ, mpeU* and *mpeY* (Fig. S6A-J). Overall this analysis revealed a high consistency between these different markers ($0.860<R^2<0.986$), indicating that most *mpeZ*-containing populations also contained 3dA alleles for *fciAB*, *mpeY*, *mpeU* and *mpeBA* and are therefore likely able to perform CA4. However, a number of stations, all located in HNLC areas (TARA_094, 111 and 122 to 128 in the Pacific Ocean and TARA_052 located northwest of Madagascar, Fig. 2B), displayed

more than 10-fold higher *mpeBA, mpeU* and *mpeZ* counts than *fciAB* and *mpeY* counts (Fig. S6A, B, E, F, H, I). This indicates that a large proportion or even the whole population (TARA_122 and 124) of 3dA in these HNLC areas is probably lacking the FciA/B regulators and MpeY like strain BIOS-E4-1 (Fig. S6K), and is thus likely similarly stuck in the blue light specialist phenotype (PT 3c; 11). Conversely, station TARA_067 exhibited consistently more than twice the *fciAB* and *mpeZ* counts compared to *mpeBA*, *mpeY* or *mpeU* (Fig. S6B-E, G, H) and was a clear outlier when comparing pigment type and clade composition (Fig. S7). This suggests that the proportion of PT 3dA might have been underestimated at this station, as a significant proportion of this population probably corresponds to PT 3a genotypes that have acquired a CA4-A island by lateral gene transfer, as is likely the case for strains WH8016 and KORDI-49. Finally, no station exhibited markedly lower *mpeU* counts compared to all other genes, indicating that the strain MVIR-18-1 genotype is probably rare in the oceans.

It must be noted that two out of the six sequenced CA4-B strains (WH8103 and WH8109) also have a deficient CA4 phenotype and display a constant, intermediate $Ex_{495:545}$ ratio (0.7 and 1, respectively), despite any obvious PBS- or CA4-related gene deletion (11). Accordingly, the plot of 3dB normalized read counts obtained with *mpeW* vs. *fciAB* shows no clear outlier (Fig. S3C).

Discussion

Marine *Synechococcus* display a large pigment diversity, with different PTs preferentially harvesting distinct regions of the light spectrum. Previous studies based on optical properties or on a single genetic marker could not differentiate all PTs (17, 29–31), and thus could not assess the realized environmental niches (*sensu* Pearman *et al.*, 37) of each PT and the role of light quality on global *Synechococcus* distribution. Here, we showed that a *mi*Tag approach combining three genetic markers can be used to reliably predict all major PTs or subtypes. Applied to the extensive *Tara* Oceans dataset, this original approach, which avoids PCR amplification and cloning biases, allowed us to describe for the first time the distribution of the different *Synechococcus* PTs at a global scale and to refine our understanding of their ecology.

PT 3 was found to be largely dominant over PT 1 and 2 along the *Tara* Oceans transect, and biogeography and correlation analyses with environmental parameters provided several novel and important insights concerning niche partitioning of PT 3 subtypes. Green (PT 3a) and blue (PT 3c) light specialists were both shown to dominate in warm areas but display clearly distinct niches, with 3a dominating in *Synechococcus*-rich stations located on oceanic borders, while 3c predominated in purely oceanic areas where the global abundance of *Synechococcus* is low. These results are in agreement with the prevailing view of an increase in the PUB:PEB ratio from green onshore mesotrophic waters to blue offshore oligotrophic waters (18, 19, 21–26, 28, 29, 44, 45). Similarly, we showed here that PT 3dB, which could not be distinguished from PT 3c in previous studies (17, 29–31), prevails in more coastal and/or mixed temperate waters than do 3c populations. The realized environmental niche of the other type of CA4 (PT 3dA) is the best defined of all PTs as it is clearly associated with nutrient-rich waters and with the coldest stations of our dataset, occurring at high latitude, at depth or in vertically mixed waters (e.g., TARA_068, 093 and 133). This result is consistent with a recent study demonstrating the dominance of 3dA in sub-Arctic waters of the Northwest Pacific Ocean (29), suggesting that the prevalence of 3dA at high latitude can be generalized. Altogether, while little was previously known about the abundance and distribution of CA4 populations in the field, here we show that they are ubiquitous, dominate in a wide range of niches, are present both in coastal and oceanic mixed waters, and overall are the most abundant *Synechococcus* PT.

The relationship between ESTUs and PTs shows that some ESTUs are preferentially associated with only one PT, while others present a much larger pigment diversity. ESTU IIA, the most abundant and ubiquitous ESTU in the field (5, 6) displays the widest PT diversity (Fig. 4B), a finding confirmed by clade II isolates spanning the largest diversity of pigment content, with representative strains of PT 2, 3a, 3c and 3dB within this clade (Table S1; see also 7, 11, 46–48). This suggests that this ESTU can colonize all light color niches, an ability which might be partially responsible for its global ecological success. Our current results do not support the previously observed correlation between clade III and PT 3a (29) since the two ESTUs from this clade (IIIA and B) were associated with PT 3c and/or 3f. This discrepancy could be due either to different methods used in these studies or to the occurrence of distinct clade III populations in coastal areas of the northwestern Pacific Ocean and along the *Tara* Oceans transect. However, the pigment phenotype of strains isolated to date is more consistent with our findings (Table S1; 16, 41). In contrast to most other PTs, the association between PT 3dA and ESTUs was found to be nearly exclusive in the field, as ESTUs from clades I, IV, CRD1 and EnvA were not associated with any other PT, and reciprocally PT 3dA is only associated with these clades (Fig. 4A). An interesting exception to this general rule was observed in the Benguela upwelling (TARA_ 067), where the dominant ESTU IA population possesses both a 3a *mpeBA* allele and *fciA*/*B* and *mpeZ* genes (Fig. S6K and S7), suggesting that cells, which were initially green light specialists (PT 3a), might have inherited a complete CA4-A island through lateral gene transfer at this station. Interestingly, among the seven clade I strains sequenced to date, three possess a 3a *mpeBA* allele, among which WH8016 also has a CA4-A island. However, the CA4-A island in this strain is only partial as it lacks *mpeZ*, and therefore does not give strain WH8016 the ability to perform CA4 (11). It is thus difficult to conclude whether the lateral transfer of this island, likely a rare event since it was only observed in populations of the Benguela upwelling, has conferred these populations the ability to perform CA4.

Another striking result of this study was the unsuspected importance of populations that have likely lost the ability to chromatically acclimate, specifically in warm HNLC areas, which cover wide expanses of the South Pacific and correspond to the poorest areas of the World Ocean (49). Interestingly, populations living in these ultra-oligotrophic environments have a different genetic basis for their consistently elevated PUB phenotype than do typical blue light specialists (i.e., PT 3c), since they have lost the CA4 regulators *fciA*/*B* and accumulated mutations in *mpeY,* a yet uncharacterized member of the phycobilin lyase family, as observed in strain BIOS-E4-1 (Fig. S6K; 11). This finding, consistent with the previous observation that the south Pacific is dominated by "high PUB" *Synechococcus* (22), is further supported by the recent sequencing of three isolates from the Equatorial Pacific, strains MITS9504, MITS9509 (both CRD1C) and MITS9508 (CRD1A; 50), all of which contain, like BIOS-E4-1, a 3dA *mpeBA* allele, a CA4-A island lacking *fciA*/*B* and a partial (MITS9508) or highly degenerated (2 other MIT strains) *mpeY* gene sequence (Fig. S6K). These natural CA4-A mutants may have adapted to blue, ultra-oligotrophic waters by inactivating a likely energetically costly acclimation mechanism (positive selection) or might be a consequence of the lower selection efficiency associated to the probably reduced effective population size of *Synechococcus* found in such an extreme environment (genetic drift). If, as we hypothesize, all *Synechococcus* cells counted as 3dA at these stations are CA4-deficient, these natural mutants would represent about 15% of the total 3dA population. In contrast, CRD1-A populations of the eastern border of the Pacific

Ocean (TARA_102, 109-110, 137) are likely true CA4 populations as they possess all CA4 genes (Fig. S6K).

In conclusion, our study provided unprecedented insights into the distribution, ecology and adaptive value of all known *Synechococcus* PTs. Surprisingly, the sum of 3dA and 3dB constitutes about 40% of the total *Synechococcus* counts in the *Tara* Oceans dataset, making chromatic acclimaters (PT 3d) the most globally abundant PT, even when taking into account potential CA4-deficient natural mutants. In addition, this PT made up 95% of the *Synechococcus* population at high latitudes and was present in every one of the five major ESTUs in the field (I, II, III, IV and CRD1). This suggests that chromatic acclimation likely confers a strong adaptive advantage compared to strains with a fixed pigmentation, particularly in vertically mixed environments and at depth at stations with a stratified water column. The previously unsuspected abundance of CA4 populations could partially explain previous contradicting observations that the PUB:PEB ratio either increases with depth (18, 21, 26) or remains constant throughout the water column (22, 27, 28, 51).

The occurrence of natural CA4 mutants and evidence for lateral transfer of the CA4 genomic island suggest that not only temperature and nutrient availability but also light quality co-exert a selective pressure on *Synechococcus* evolution, and that changes in pigment diversity occurs in response to changes in light niches by acquisition or loss of specific PBS synthesis and/or regulation genes, as previously observed for phosphorus and nitrogen transport genes in *Prochlorococcus* (52–54). Still, the complex interactions between PT, vertical phylogeny and environmental parameters remain unclear and more work is needed to refine our understanding of the balance between the forces shaping community composition. Is it temperature and/or nutrient availability that primarily controls community clade composition, and after that the within-clade populations displaying the pigmentation that fits best the light condition are selected, or does light quality select for a given PT and then other environmental parameters for the most adapted clades? Part of the answer lies in the observation that in some environments and more specifically at the boundaries of *Synechococcus* environmental niche(s), where the harshest conditions are encountered, both pigment and clade diversities are drastically reduced. This could indicate either a genetic bottleneck and founder effect or that extreme selective pressure only allows the fittest organism to survive. However, this concomitant reduction of genetic and pigment diversity rather tends to support a co-selection by light quality and other environmental parameters. On the contrary, the diverse PTs occurring within some clades, as well as the co-occurrence of different PTs at most stations compared to more clear-cut clade shifts (e.g., in the Red Sea/Indian Ocean) might indicate that light quality is not the strongest selective force or that light changes are too transient to allow the dominance and fixation of a particular PT in a population. Future experimental work exploring the fitness effects of ESTU members and pigmentation under different controlled environmental conditions (including temperature, nutrients and light) might help to clarify their respective effects on the diversification of this ecological important photosynthetic organism.

Materials and Methods

Metagenomic samples

This study focused on 109 metagenomic samples corresponding to 65 stations from the worldwide oceans collected during the 2.5-year *Tara* Oceans circumnavigation (2009-2011). Sample processing and features are the same as described in (6). Sequencing depths ranged from 16 $\times 10^6$ to 258 $\times 10^6$ reads per sample after quality control and paired-reads merging, and corresponding fragments lengths averaged 164 ± 20 bp (median: 168 bp).

Databases: reference sequences and outgroup sequences

We created a reference database comprising the full-length gene/operon nucleotide sequence and the characterized pigment type of the cultivated isolate for each marker used (*cpcBA*, *mpeBA* and *mpeW*). Sequences were retrieved from public genomes and previous publications. The *cpcBA* database comprised 83 sequences (64 unique), of which 18 (12) PT1, 5 (5) PT2A, 19 (6) PT2B and 39 (39) PT3; for *mpeBA*, 11 PT3a, 2 PT3f, 11 3dA and 17 3dB (41 sequences, all unique); for *mpeW*, 5 unique sequences. For each marker, a reference alignment was generated with MAFFT L-INS-i v6.953b (55), and a reference phylogenetic tree was inferred with PhyML v. 20120412 (GTR+I+G, 10 random starting trees, best of SPR and NNI moves, 500 bootstraps; 56) and drawn using the ETE Toolkit (57).

A database of outgroups was also built. For *cpcBA* and *mpeBA*, it comprised 491 full-length sequences of paralogous sequences coming from marine *Synechococcus* or *Prochlorococcus* (e.g. *cpeBA*) as well as orthologous sequences from other marine and freshwater organisms retrieved from public databases (NCBI nt). For *mpeW*, this outgroup database was solely made of 116 full-length sequences of paralogs (*mpeZ*, *mpeY* and *cpeY*) from marine *Synechococcus* or *Prochlorococcus*, as no homolog could be identified in public databases.

Read assignation and estimation of PT abundance

Reads were preselected using BLAST+ (58) with relaxed parameters (blastn, maximum E-value of 1e-5, minimum percent identity 60%, minimum 75% of read length aligned), using reference sequences as subjects; the selection was then refined by a second BLAST+ round against databases of outgroups: reads with a best-hit to outgroup sequences were excluded from downstream analysis. Selected reads were then aligned to the marker reference alignment with MAFFT v.7.299b (--addfragments --adjustdirectionaccurately) and placed in the marker reference phylogenetic tree with *pplacer* (59). For each read, *pplacer* returns a list of possible positions (referred to as placements) at which it can be placed in the tree and their associated "likelihood weight ratio" (LWR, proxy for the probability of the placement; see *pplacer* publication and documentation for more details). Reads were then assigned to a pigment type using a custom classifier written in Python. Briefly, internal nodes of the reference tree were assigned a pigment type based on the pigmentation of descending nodes (PT of child reference sequences if the same for all of them, "unclassified" otherwise). For each read, placements were assigned to their nearest ascending or descending node based on their relative position on the edge, and the lowest common ancestor (LCA) of the set of nodes for which the cumulated LWR was greater than 0.95 (LCA of possible placements at 95% probability) was then computed. Finally, the read was assigned to the pigment type of this LCA. Different combinations of read assignment parameters (LCA at 90%, 95% or 100%; assignation of placements to the ascending, descending or nearest node) were also assessed, and resulted either in higher rates of unassigned reads or unacceptable error rates (Fig. S2).

Read counts were normalized by adjusted marker length: for each marker and each sequence file, counts were normalized by (L - ℓ + 1), with L the length of the marker gene (*cpcBA* mean length: 1053.7bp; *mpeBA* mean length: 1054.6bp; *mpeW* mean length: 1193.3bp) and ℓ the mean length of reads in the sequence file. Finally, the abundance of PT1, 2A and 2B was defined as the normalized *cpcBA* read counts of these PT, the abundance of PT3a, 3f and 3dA as the normalized *mpeBA* read counts of these PT, 3dB as the normalized *mpeW* count and 3c as the difference between the normalized *mpeBA* (3c + 3dB) read count and the PT3dB count assessed with *mpeW*. The abundance of unclassified sequences was also taken into account.

Detailed *petB* counts for clade and ESTU abundances were obtained from (6).

Read assignment simulations

For each marker, simulated reads were generated from one reference sequence at a time using a sliding window of 100, 125 or 150bp (TARA mean read length: 164.2bp; median 169bp) and steps of 5 bp. Simulated reads were then assigned to a pigment type with the aforementioned bioinformatic pipeline, using all reference sequences except the one used to simulate reads. Inferred pigment types of simulated fragments were then compared to known pigment types of originating reference sequence.

Statistical analyses

Environmental parameters were obtained from (6). Hierarchical clustering and NMDS analyses of stations were performed using R (60) packages cluster v1.14.4 (61) and MASS v7.3–29 (62), respectively. PT contingency tables were filtered by considering only stations with more than 30 *cpcBA* reads and 30 *mpeBA* reads, and only PT appearing in at least 2 stations and with more than 150 reads in the whole dataset. Contingency tables were normalized using Hellinger transformation that gives lower rates to rare PT. The Bray–Curtis distance was then used for ordination (isoMDS function; maxit, 100; k, 2). Correlations were performed with R package Hmisc_3.17-4 with Benjamini & Hochberg multiple comparison adjusted p-value (63).

Acknowledgements

We warmly thank Annick Bricaud for fruitful discussions on biooptics. This work was supported by the French "Agence Nationale de la Recherche" Programs SAMOSA (ANR-13-ADAP-0010) and France Génomique (ANR-10-INBS-09), the French Government "Investissements d'Avenir" programs OCEANOMICS (ANR-11-BTBR-0008), the European Union's Seventh Framework Programs FP7 MicroB3 (grant agreement 287589) and MaCuMBA (grant agreement 311975) and the Spanish Ministry of Science and Innovation grant MicroOcean PANGENOMICS (GL2011-26848/BOS). We also thank the support

and commitment of the *Tara* Oceans coordinators and consortium, Agnès b. and E. Bourgois, the Veolia Environment Foundation, Région Bretagne, Lorient Agglomeration, World Courier, Illumina, the EDF Foundation, FRB, the Prince Albert II de Monaco Foundation, the *Tara* schooner and its captains and crew. *Tara* Oceans would not exist without continuous support from 23 institutes (http://oceans.taraexpeditions.org). We would like to thank the ABIMS platform for providing computing infrastructure.

1. Guidi L, et al. (2016) Plankton networks driving carbon export in the oligotrophic ocean. *Nature* 532(7600):465–470.
2. Flombaum P, et al. (2013) Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc Natl Acad Sci* 110(24):9824–9829.
3. Zwirglmaier K, et al. (2008) Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ Microbiol* 10(1):147–161.
4. Mazard S, Ostrowski M, Partensky F, Scanlan DJ (2012) Multi-locus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*. *Environ Microbiol* 14(2):372–386.
5. Sohm JA, et al. (2016) Co-occurring *Synechococcus* ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. *ISME J* 10(2):333–345.
6. Farrant GK, et al. (2016) Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proc Natl Acad Sci* 113(24):E3365–E3374.
7. Six C, et al. (2007) Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.: a comparative genomics study. *Genome Biol* 8(12):R259.
8. Alberte RS, Wood AM, Kursar TA, Guillard RRL (1984) Novel Phycoerythrins in Marine *Synechococcus* spp. *Plant Physiol* 75(3):732–739.
9. Ong LJ, Glazer AN (1991) Phycoerythrins of marine unicellular cyanobacteria. I. Bilin types and locations and energy transfer pathways in *Synechococcus* spp. phycoerythrins. *J Biol Chem* 266(15):9515–9527.
10. Sidler WA (1994) Phycobilisome and Phycobiliprotein Structures. *The Molecular Biology of Cyanobacteria*, Advances in Photosynthesis. (Springer, Dordrecht), pp 139–216.
11. Humily F, et al. (2013) A Gene Island with Two Possible Configurations Is Involved in Chromatic Acclimation in Marine *Synechococcus*. *PLoS ONE* 8(12):e84459.
12. Palenik B (2001) Chromatic adaptation in marine *Synechococcus* strains. *Appl Environ Microbiol* 67(2):991–994.
13. Everroad C, et al. (2006) Biochemical Bases of Type IV Chromatic Adaptation in Marine *Synechococcus* spp. *J Bacteriol* 188(9):3345–3356.
14. Shukla A, et al. (2012) Phycoerythrin-specific bilin lyase–isomerase controls blue-green chromatic acclimation in marine *Synechococcus*. *Proc Natl Acad Sci* 109(49):20136–20141.
15. Sanfilippo JE, et al. (2016) Self-regulating genomic island encoding tandem regulators confers chromatic acclimation to marine *Synechococcus*. *Proc Natl Acad Sci* 113(21):6077–6082.
16. Toledo G, Palenik B, Brahamsha B (1999) Swimming marine *Synechococcus* strains with widely different photosynthetic pigment ratios form a monophyletic group. *Appl Environ Microbiol* 65(12):5247–5251.
17. Humily F, et al. (2014) Development of a targeted metagenomic approach to study a genomic region involved in light harvesting in marine *Synechococcus*. *FEMS Microbiol Ecol* 88(2):231–249.
18. Olson RJ, Chisholm SW, Zettler ER, Armbrust EV (1990) Pigments, size, and distributions of *Synechococcus* in the North Atlantic and Pacific Oceans. *Limnol Oceanogr* 35(1):45–58.
19. Sherry ND, Michelle Wood A (2001) Phycoerythrin-containing picocyanobacteria in the Arabian Sea in February 1995: diel patterns, spatial variability, and growth rates. *Deep Sea Res Part II* 48(6–7):1263–1283.
20. Jiang T, et al. (2016) Temporal and spatial variations of abundance of phycocyanin- and phycoerythrin-rich *Synechococcus* in Pearl River Estuary and adjacent coastal area. *J Ocean Univ China* 15(5):897–904.
21. Lantoine F, Neveux J (1997) Spatial and seasonal variations in abundance and spectral characteristics of phycoerythrins in the tropical northeastern Atlantic Ocean. *Deep Sea Res Part I* 44(2):223–246.
22. Neveux J, Lantoine F, Vaulot D, Marie D, Blanchot J (1999) Phycoerythrins in the southern tropical and equatorial Pacific Ocean: Evidence for new cyanobacterial types. *J Geophys Res Oceans* 104(C2):3311–3321.
23. Wood AM, Phinney DA, Yentsch CS (1998) Water column transparency and the distribution of spectrally distinct forms of phycoerythrin- containing organisms. *Mar Ecol Prog Ser* 162:25–31.
24. Campbell L, et al. (1998) Response of microbial community structure to environmental forcing in the Arabian Sea. *Deep Sea Res Part II Top Stud Oceanogr* 45(10):2301–2325.
25. Hoge FE, Wright CW, Kana TM, Swift RN, Yungel JK (1998) Spatial variability of oceanic phycoerythrin spectral types derived from airborne laser-induced fluorescence emissions. *Appl Opt* 37(21):4744–4749.
26. Wood AM, Lipsen M, Coble P (1999) Fluorescence-based characterization of phycoerythrin-containing cyanobacterial communities in the Arabian Sea during the Northeast and early Southwest Monsoon (1994–1995). *Deep Sea Res Part II Top Stud Oceanogr* 46(8):1769–1790.
27. Yona D, Park MO, Oh SJ, Shin WC (2015) Distribution of *Synechococcus* and its phycoerythrin pigment in relation to environmental factors in the East Sea, Korea. *Ocean Sci J* 49(4):367–382.
28. Campbell L, Iturriaga R (1988) Identification of *Synechococcus* spp. in the Sargasso Sea by immunofluorescence and fluorescence excitation spectroscopy performed on individual cells. *Limnol Oceanogr* 33(5):1196–1201.
29. Xia X, et al. (2017) Phylogeography and pigment type diversity of *Synechococcus* cyanobacteria in surface waters of the northwestern pacific ocean. *Environ Microbiol* 19(1):142–158.
30. Xia X, Liu H, Choi D, Noh JH (2017) Variation of *Synechococcus* Pigment Genetic Diversity Along Two Turbidity Gradients in the China Seas. *Microb Ecol*:1–12.
31. Xia X, Guo W, Tan S, Liu H (2017) *Synechococcus* Assemblages across the Salinity Gradient in a Salt Wedge Estuary. *Front Microbiol*
32. Liu H, Jing H, Wong THC, Chen B (2014) Co-occurrence of phycocyanin- and phycoerythrin-rich *Synechococcus* in subtropical estuarine and coastal waters of Hong Kong: PE-rich and PC-rich *Synechococcus* in subtropical coastal waters. *Environ Microbiol Rep* 6(1):90–99.
33. Chung C-C, Gong G-C, Huang C-Y, Lin J-Y, Lin Y-C (2015) Changes in the *Synechococcus* Assemblage Composition at the Surface of the East China Sea Due to Flooding of the Changjiang River. *Microb Ecol* 70(3):677–688.
34. Haverkamp T, et al. (2008) Diversity and phylogeny of Baltic Sea picocyanobacteria inferred from their ITS and phycobiliprotein operons. *Environ Microbiol* 10(1):174–188.
35. Sunagawa S, et al. (2015) Structure and function of the global ocean microbiome. *Science* 348(6237):1261359.
36. Logares R, et al. (2014) Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environ Microbiol* 16(9):2659–2671.
37. Pearman PB, Guisan A, Broennimann O, Randin CF (2008) Niche dynamics in space and time. *Trends Ecol Evol* 23(3):149–158.
38. Larsson J, et al. (2014) Picocyanobacteria containing a novel pigment gene cluster dominate the brackish water Baltic Sea. *ISME J* 8(9):1892–1903.
39. Paulsen ML, et al. (2016) *Synechococcus* in the Atlantic Gateway to the Arctic Ocean. *Front Mar Sci.*
40. Haverkamp THA, et al. (2008) Colorful microdiversity of *Synechococcus* strains (picocyanobacteria) isolated from the Baltic Sea. *ISME J* 3(4):397–408.
41. Hunter-Cevera KR, Post AF, Peacock EE, Sosik HM (2015) Diversity of *Synechococcus* at the Martha's Vineyard Coastal Observatory: Insights from Culture Isolations, Clone Libraries, and Flow Cytometry. *Microb Ecol* 71(2):276–289.
42. Cabello-Yeves PJ, et al. (2017) Novel *Synechococcus* Genomes Reconstructed from Freshwater Reservoirs. *Front Microbiol*.
43. Mahmoud RM, et al. (2017) Adaptation to Blue Light in Marine *Synechococcus* Requires MpeU, an Enzyme with Similarity to Phycoerythrobilin Lyase Isomerases. *Front Microbiol*.
44. Choi DH, Noh JH (2009) Phylogenetic diversity of *Synechococcus* strains isolated from the East China Sea and the East Sea. *FEMS Microbiol Ecol* 69(3):439–448.
45. Veldhuis MJW, Kraay GW (1993) Cell abundance and fluorescence of picoplankton in relation to growth irradiance and nitrogen availability in the Red Sea. *Neth. J. Sea Res.* 31(2):135–145.
46. Ahlgren NA, Rocap G (2006) Culture Isolation and Culture-Independent Clone Libraries Reveal New Marine *Synechococcus* Ecotypes with Distinctive Light and N Physiologies. *Appl Environ Microbiol* 72(11):7193–7204.
47. Bemal S, Anil AC (2016) Genetic and ecophysiological traits of *Synechococcus* strains isolated from coastal and open ocean waters of the Arabian Sea. *FEMS Microbiol Ecol* 92(11).
48. Everroad RC, Wood AM (2012) Phycoerythrin evolution and diversification of spectral phenotype in marine *Synechococcus* and related picocyanobacteria. *Mol Phylogenet Evol* 64(3):381–392.
49. Morel A, et al. (2007) Optical properties of the "clearest" natural waters. *Limnol Oceanogr* 52(1):217–229.
50. Cubillos-Ruiz A, Berta-Thompson JW, Becker JW, van der Donk WA, Chisholm SW (2017) Evolutionary radiation of lanthipeptides in marine cyanobacteria. *Proc Natl Acad Sci U S A* 114(27):E5424–E5433.
51. Katano T, Nakano S (2006) Growth rates of *Synechococcus* types with different phycoerythrin composition estimated by dual-laser flow cytometry in relationship to the light environment in the Uwa Sea. *J Sea Res* 55(3):182–190.
52. Martiny AC, Huang Y, Li W (2009) Occurrence of phosphate acquisition genes in *Prochlorococcus* cells from different ocean regions. *Environ Microbiol* 11(6):1340–1347.
53. Martiny AC, Coleman ML, Chisholm SW (2006) Phosphate acquisition genes in *Prochlorococcus* ecotypes: evidence for genome-wide adaptation. *Proc Natl Acad Sci U S A* 103(33):12552–12557.
54. Martiny AC, Kathuria S, Berube PM (2009) Widespread metabolic potential for nitrite and nitrate assimilation among *Prochlorococcus* ecotypes. *Proc Natl Acad Sci U S A* 106(26):10787–10792.
55. Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* 30(4):772–780.
56. Guindon S, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59(3):307–321.
57. Huerta-Cepas J, Serra F, Bork P (2016) ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol Biol Evol* 33(6):1635–1638.
58. Camacho C, et al. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.
59. Matsen FA, Kodner RB, Armbrust EV (2010) pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* 11:538.
60. R Core Team (2014) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria) Available at: http://www.R-project.org/.
61. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K (2017) *cluster: Cluster Analysis Basics and Extensions*.
62. Venables WN, Ripley BD (2002) *Modern Applied Statistics with S* (Springer, New York). Fourth Available at: http://www.stats.ox.ac.uk/pub/MASS4.
63. Harrell FE (2016) *Hmisc: Harrell Miscellaneous* Available at: http://CRAN.R-project.org/package=Hmisc.

Please review all the figures in this paginated PDF and check if the figure size is appropriate to allow reading of the text in the figure.

If readability needs to be improved then resize the figure again in 'Figure sizing' interface of Article Sizing Tool.

Modeling of the blue to green irradiance ratio (Ir495/Ir545) at *Tara* Oceans stations.

We used the clear sky surface irradiance model of Frouin and McPherson in Fortran and translated to Matlab by Werdell (see Frouin et al., 1989 and Tanre et al., 1979 for the analytical formula used) using the date, latitude and longitude of each station, assuming sunny sky and at noon.

The spectral light distribution averaged over the mixed layer was computed from:

$$< Ir(\lambda) >= \frac{\int_0^{MLD} E(\lambda,0^-)e^{-k(\lambda,chl)z}dz}{MLD} = \frac{I(\lambda,0^-)}{MLD\ k(\lambda,chl)}\left\{1 - e^{-k(chl)MLD}\right\}$$

where:
- *chl* denotes the average chlorophyll value in the mixed layer. [*chl*] was based on a fluorometer that was calibrated against HPLC data and corrected for non-photochemical quenching,
- MLD is the mixed layer depth that was computed based on a temperature threshold criterion
- $k(\lambda, chl)$ is the diffuse attenuation coefficient at wavelength λ (495 or 545 using a 10 nm bandwidth). This parameter was computed using Morel and Maritorena (2001)'s equation:

$$k(\lambda, chl) = k_w(\lambda) + \chi(\lambda)[chl]^{e(\lambda)}$$

k_w, χ and e are provided in Table 2 of Morel and Maritorena (2001) and have the following values for the wavelengths of interest:

Wavelength [nm]	$k_w(\lambda)$ [m^{-1}]	$\chi(\lambda)$	$e(\lambda)$
495	0.01885	0.06907	0.68947
545	0.05212	0.04253	0.65591

If the sampling depth was below the MLD, the irradiance was computed as follows:

$Ir(\lambda, sampling\ depth) = (\lambda, 0^-)e^{-k(\lambda,chl)sampling\ depth}$.

The ratio was then computed as *Ir*(*495*)*/Ir*(*545*).

References:

Frouin, R., D. W. Ligner, and C. Gautier, 1989: A Simple analytical formula to compute clear sky total and photosynthetically available solar irradiance CC at the ocean surface. J. Geophys. Res., 94, 9731-9742.

Morel, A. and S. Maritorena, 2001: Bio-optical properties of oceanic waters: A reappraisal. J. Geophys. Res., 106, 7163–7180.

Tanre, D., M. Herman, P.-Y. Deschamps, and A. De Leffe, 1979: Atmospheric modeling for Space measurements of ground reflectances, including bi-directional properties. Appl. Optics, 18, 21,3587-21,3597.

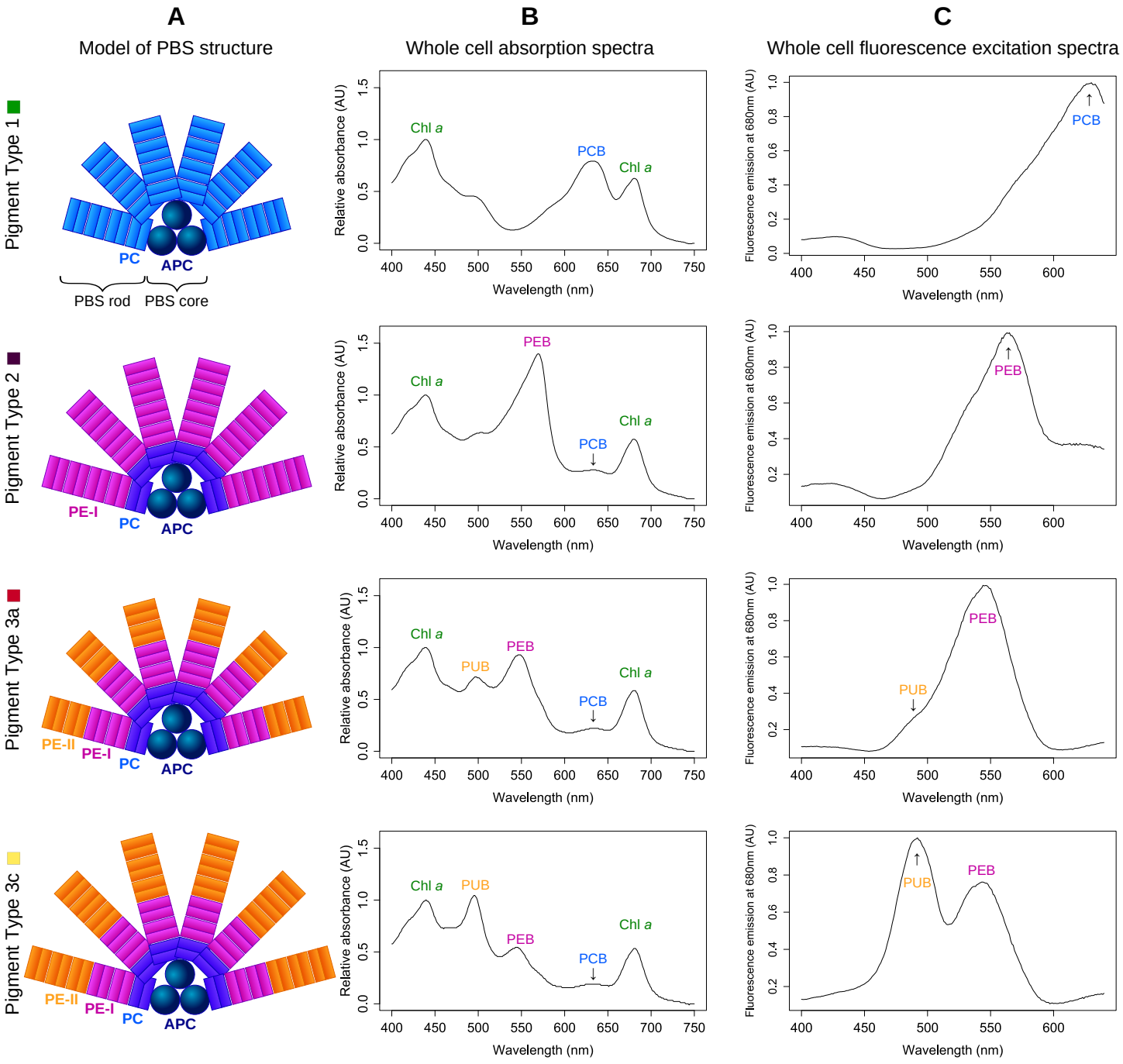


Figure S1: Biochemical composition and biooptical properties of phycobilisomes (PBS) of the main Synechococcus pigment types (PT). (*A*) Models of PBS structure, highlighting the conserved core and variable rods of increasing complexity from PT1 to PT3 (Redrawn after Six *et al.*, 2007). (*B*) Whole cell absorption spectra of the different PTs (Reproduced after Six *et al.*, 2007). Chromophores responsible of each absorption peaks are indicated. (*C*) Whole cell fluorescence excitation spectra with emission at 680 nm. Note that for chromatic acclimaters (PT 3d), the PBS structure is similar to other PT 3 but that the excitation ratio at 495 nm and 545 nm ($Ex_{495:545}$) varies from 0.6 in green light to 1.6 in blue light (not shown).

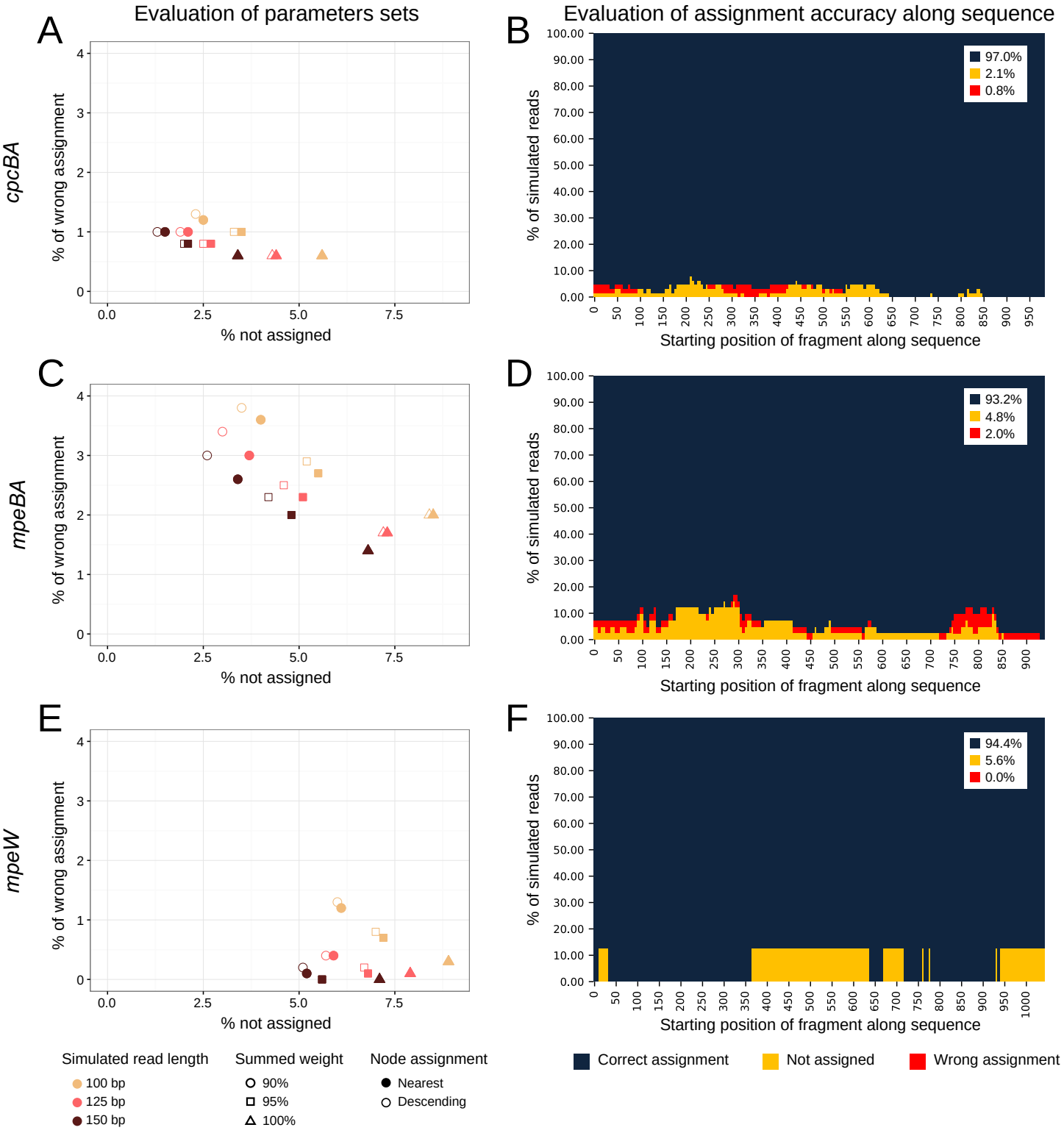


Figure S2: Evaluation of the assignment pipeline and the resolution power of the different markers used in this study. Simulated reads were generated from the reference dataset and assigned using a custom-designed pipeline (see materials and methods). (*A*, *C*, *E*) Evaluation of different sets of parameters tested for read assignment for the different markers: *cpcBA* (*A*), *mpeBA* (*C*) and *mpeW* (*E*). 100 (yellow), 125 (pink) and 150 bp (dark red) long reads were simulated. For each read, pplacer returns a list of possible positions in the tree, each associated with a likelihood weight. From these placements, we considered only those that reached a summed likelihood weight of either 90% (circle), 95% (square) or 100% (triangle). The assignment was then performed based on the phenotype of either the nearest node (solid symbol) in the tree or the descending (child) node (empty symbol). (*B*, *D*, *F*) Evaluation the resolution power along *cpcBA* (*B*), *mpeBA* (*D*) or *mpeW* (*F*) for 150 bp simulated reads assigned using the parameters selected for *Tara* Oceans metagenomic read assignment (i.e., nearest node assignment and summed weight of 95%). Note that *Tara* Oceans reads had a mean length of 164 bp.

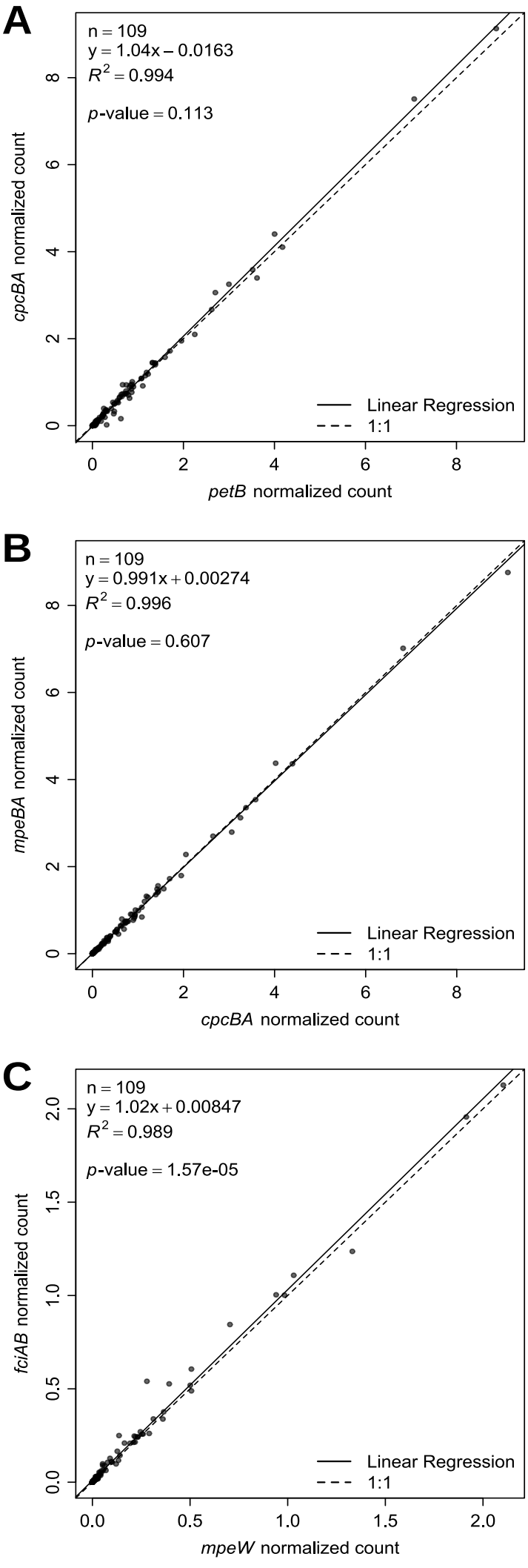


Figure S3: Correlations between the number of reads recruited using the main markers used in this study. (*A*) Correlation between *petB* (vertical phylogeny) and *cpcBA* counts used to discriminate pigment types (PT) 1, 2 and 3. (*B*) Correlation between PT 3 counts using *cpcBA* and total *mpeBA* counts. Note that *mpeBA* is a PT3 specific marker and is used to discriminate PTs 3a, 3dA, 3f and 3c +3dB. (*C*) Correlation between PT 3dB counts using *fciAB*, a PT 3dB- and 3dA-specific marker and total *mpeW* counts, a PT 3dB-specific marker.

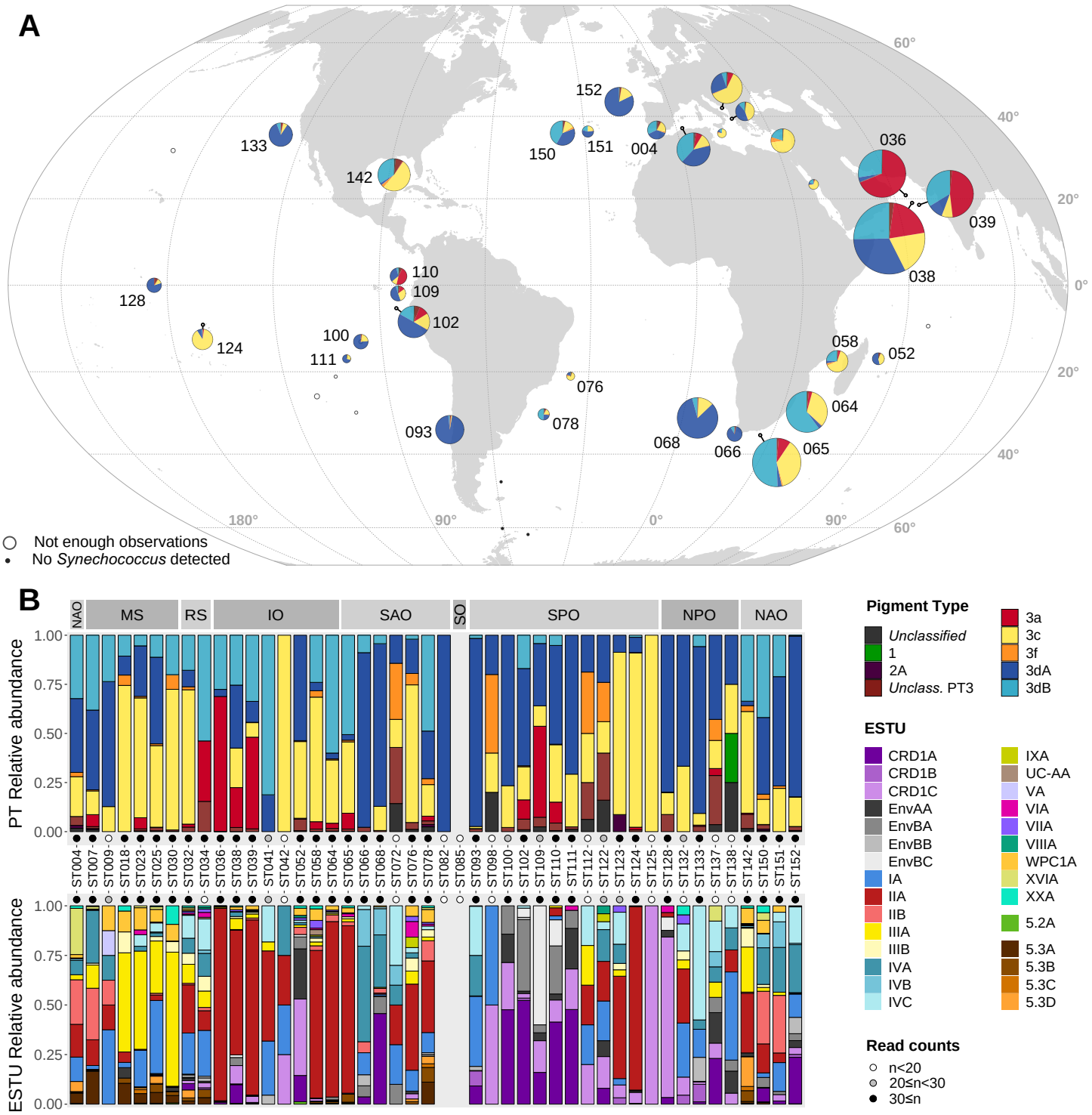


Figure S4: Distribution of Synechococcus pigment types (PTs) at depth (Deep Chlorophyll Maximum). (*A*) Map showing the global distribution of all *Synechococcus* PTs at depth along the *Tara* Oceans transect. Diameters of pies are proportional to the number of *cpcBA* reads normalized by the sequencing effort. Stations with less than 30 *cpcBA* or *mpeBA* reads are indicated by open circles and those with no *cpcBA* reads by black dots. Numbers next to pies correspond to *Tara* Oceans stations. (*B*) PTs and ESTU relative abundance at depth for sampling station along the *Tara* Oceans transect. Oceanic provinces are indicated in the top gray panels. NAO, North Atlantic Ocean; MS, Mediterranean Sea; RS, Red Sea; IO, Indian Ocean; SAO, South Atlantic Ocean; SO, Southern Ocean; SPO, South Pacific Ocean; NPO, North Pacific Ocean.

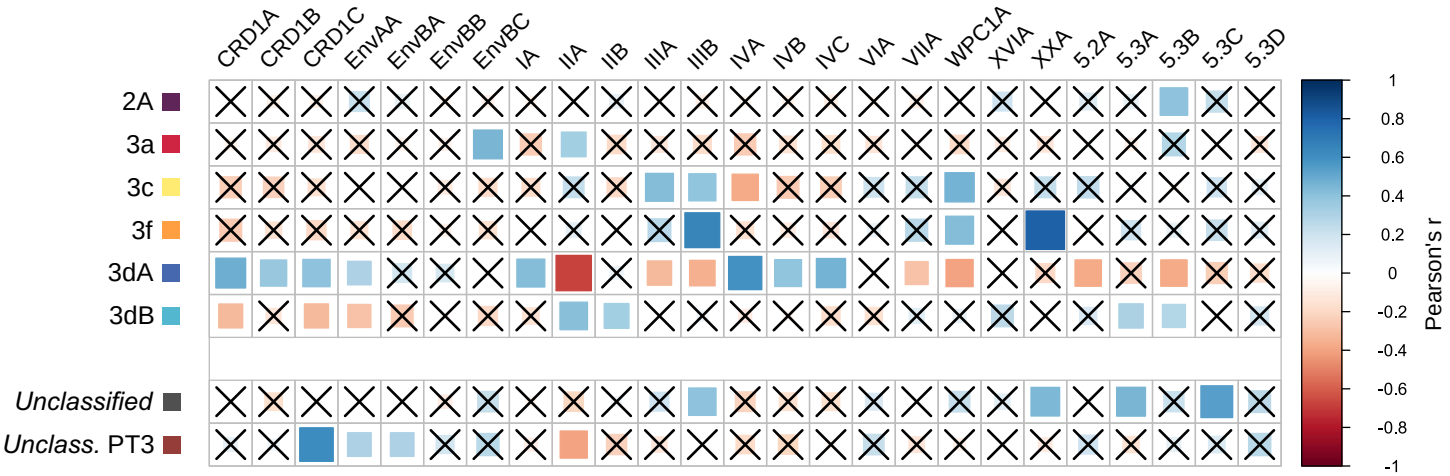


Figure S5: Same as Fig. *4A* but for all ESTUs. Unclass., unclassified.

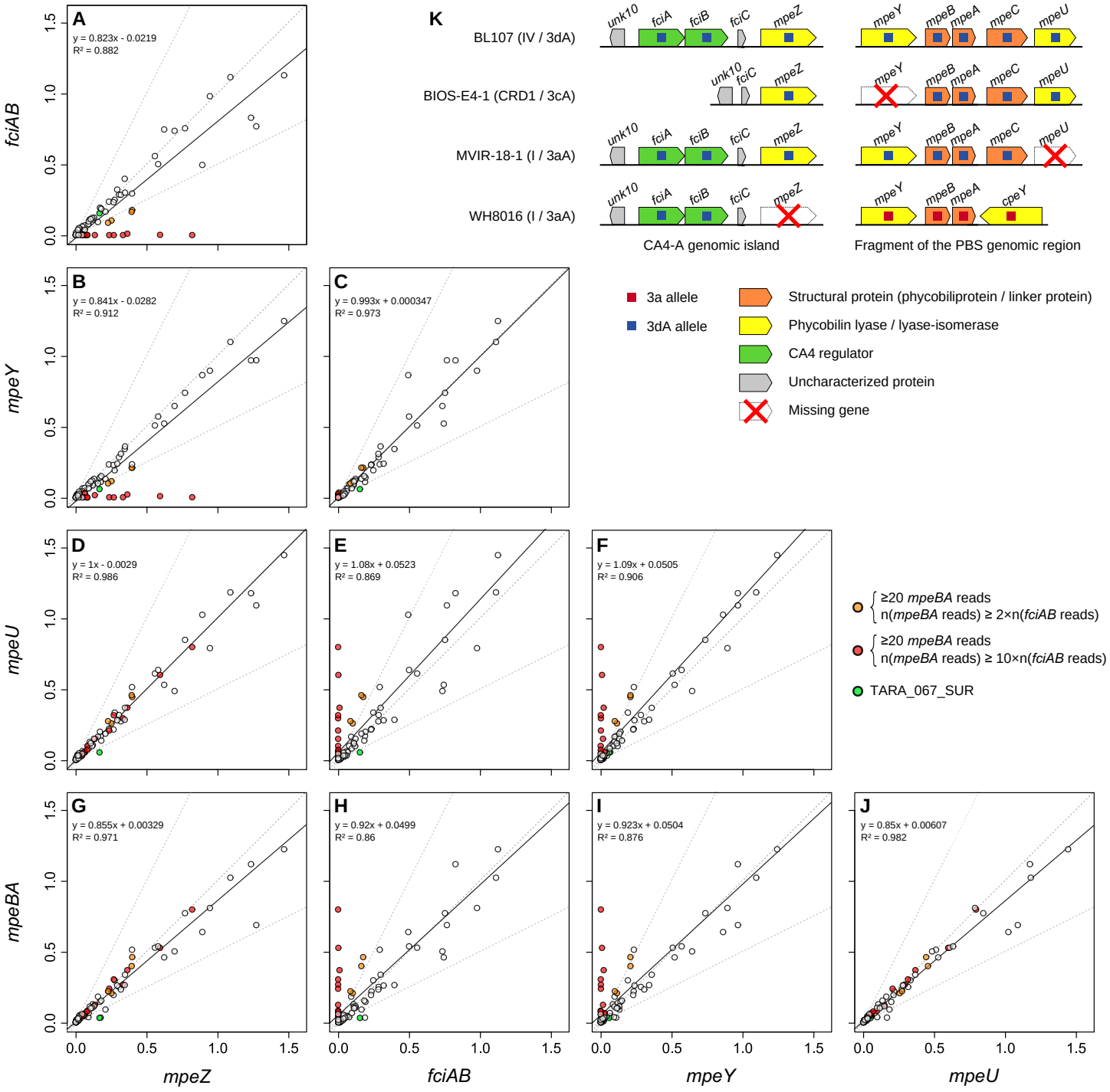


Figure S6: Focus on pigment type (PT) 3dA natural mutants, exhibiting an altered gene content with regard to typical PT 3dA. (*A-J*) Correlation between the number of reads assigned as PT 3dA using different markers (all present in single gene copy in typical 3dA). Each circle corresponds to a *Tara* Oceans station and depth. Orange circles: stations with at least 20 *mpeBA* reads assigned to PT 3dA and at least twice more 3dA counted with *mpeBA* than with fciAB, corresponding to the surface sample of stations TARA_070, TARA_110 and TARA_137 and the DCM of stations TARA_038, TARA_058 and TARA_110. Red circles: same but with more than 10-fold 3dA counted with *mpeBA* than *fciAB*, corresponding to the surface sample of stations TARA_052, TARA_094, TARA_111 and TARA_122 to TARA_128, and DCM of stations TARA_052, TARA_100, TARA_111 and TARA_128. Green circle: surface of station TARA_067. (*K*) CA4-A genomic island and fragment of the phycobilisome (PBS) genomic region for a typical, CA4-able 3dA strain (strain BL107), and 3 CA4-deficient strains, which are stuck either in blue light phenotype (similar to strain BIOS-E4-1), or green light phenotype (as strains MVIR-18-1 and WH8016). Note that KORDI-49 and WH8016 strains have identical PBS gene complement and genomic arrangement. The complete PBS genomic region of the BL107 strain can be found in Six *et al.*, 2007. Note that for readability, surface of station TARA_093 has been omitted since it has the highest normalized counts (2.7-3.2) for all markers and exhibited a good agreement between markers (ratio close to 1:1).

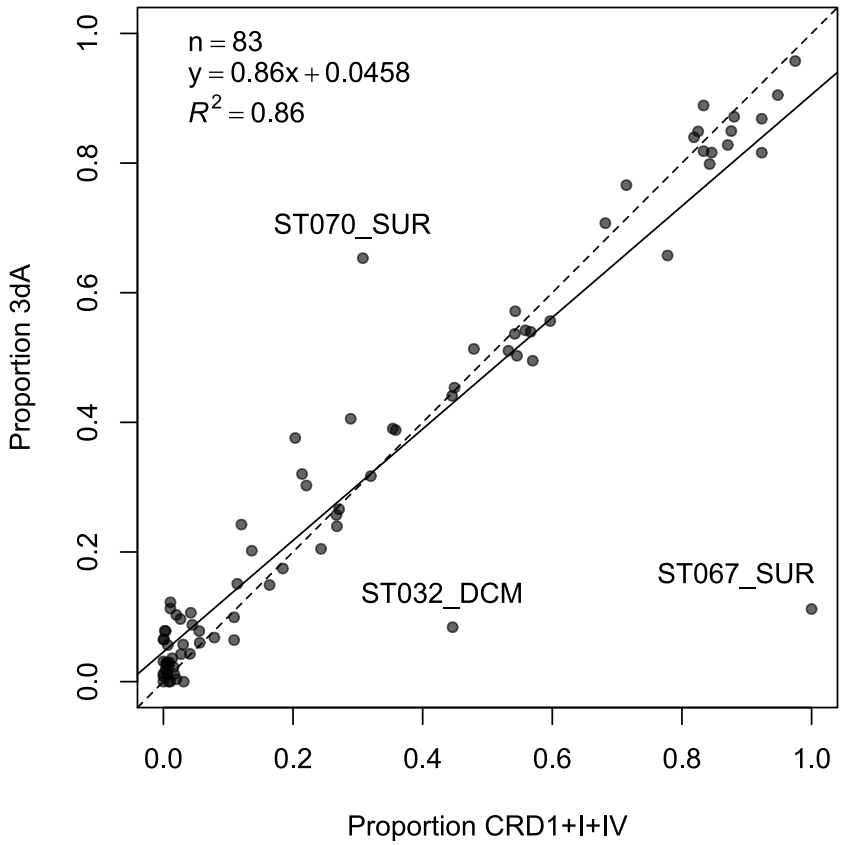


Figure S7: Correlation between the proportion of clades I, IV and CRD1, as assessed with *petB*, and the proportion of pigment type 3dA, as assessed with *mpeBA*, at each station.

ANNEXE 2 : Community-level response to natural perturbations of iron in open ocean planktonic ecosystems.

Caputi L., Carradec Q., Eveillard D., Kirilovsky A., Pelletier E., Pierella Karlusich J. J., Rocha Jimenez Vieira F., Villar E., Chaffron S., Malviya S., Scalco E., Acinas S. G., Alberti A., Aury J-M., Benoiston A-S., Bertrand A., Biard T., Bittner L., Boccara M., Brum J. R, Brunet C., Busseni G., Carratalà A., Claustre H., Coelho L. P., Colin S., D'Aniello S., Da Silva C., Del Core M., **Doré H.**, Gasparini S., Kokoszka F., Jamet J-L., Lejeusne C., Lepoivre C., Lescot M., Lima-Mendez G., Lombard F., Lukeš J., Maillet N., Madoui M-A., Martinez E., Mazzocchi M. G., Néou M. B., Paz-Yepes J., Poulain J., Ramondenc S., Romagnan J-B., Roux S., Salvagio Manta D., Sanges R., Speich S., Sprovieri M., Sunagawa S., Taillander V., Tanaka A., Tirichine L., Trottier C., Uitz J., Veluchamy A., Veselá J., Vincent F., Yau S., Kandels-Lewis S., Searson S., Dimier C., Picheral M., *Tara* Oceans Coordinators, Bork P., Boss E., de Vargas C., Follows M. J., Grimsley N., Guidi L., Hingamp P., Karsenti E., Sordino P., Stemmann L., Sullivan M. B., Tagliabue A., Zingone A., Garczarek L., d'Ortenzio F., Testor P., Not F., Ribera d'Alcalà M., Wincker P., Bowler C., Iudicone D. (*in prep.*). **Community-level response to natural perturbations of iron in open ocean planktonic ecosystems.**

Community-level response to natural perturbations of iron in open ocean planktonic ecosystems

Luigi Caputi[†1], Quentin Carradec[†2,3,4], Damien Eveillard[†5], Amos Kirilovsky[†6,7], Eric Pelletier[†2,3,4], Juan J. Pierella Karlusich[†6], Fabio Rocha Jimenez Vieira[†6], Emilie Villar[†6,8], Samuel Chaffron[5], Shruti Malviya[6,9], Eleonora Scalco[1], Silvia G. Acinas[10], Adriana Alberti[2], Jean-Marc Aury[2], Anne-Sophie Benoiston[6,11], Alexis Bertrand[2], Tristan Biard[12,13], Lucie Bittner[6,8,11], Martine Boccara[6], Jennifer R. Brum[14], Christophe Brunet[1], Greta Busseni[1], Anna Carratalà[15], Hervé Claustre[13], Luis Pedro Coelho[16], Sébastien Colin[6,8], Salvatore D'Aniello[1], Corinne Da Silva[2], Marianna Del Core[17], Hugo Doré[8], Stéphane Gasparini[13], Florian Kokoszka[6,18], Jean-Louis Jamet[19], Christophe Lejeusne[1,8], Cyrille Lepoivre[20], Magali Lescot[21], Gipsi Lima-Mendez[22,23,24], Fabien Lombard[13], Julius Lukeš[25], Nicolas Maillet[1,26], Elodie Martinez[27], MariaGrazia Mazzocchi[1], Mario B. Néou[2,3,4], Javier Paz-Yepes[6], Julie Poulain[2], Simon Ramondenc[13], Jean-Baptiste Romagnan[13], Simon Roux[14], Daniela Salvagio Manta[17], Remo Sanges[1], Sabrina Speich[18], Mario Sprovieri[17], Shinichi Sunagawa[16,28], Alessandro Tagliabue[29], Vincent Taillander[13], Atsuko Tanaka[6], Leila Tirichine[6], Camille Trottier[30], Julia Uitz[13], Alaguraj Veluchamy[6,31], Jana Veselý[25], Flora Vincent[6], Sheree Yau[32], Stefanie Kandels-Lewis[16,33], Sarah Searson[13], Céline Dimier[6,8], Marc Picheral[13], *Tara Oceans Coordinators,* Emmanuel Boss[34], Colomban de Vargas[8], Mick Follows[35], Nigel Grimsley[32], Lionel Guidi[13,36], Pascal Hingamp[20], Eric Karsenti[6,33], Paolo Sordino[1], Lars Stemmann[13], Matthew B. Sullivan[14], Adriana Zingone[1], Laurence Garczarek[8], Fabrizio d'Ortenzio[13], Pierre Testor[37], Fabrice Not[8], Maurizio Ribera d'Alcalà[*1], Patrick Wincker[*2,3,4], Chris Bowler[*6], Daniele Iudicone[*1]

Tara Oceans Coordinators Silvia G. Acinas[10], Peer Bork[38,39,40], Emmanuel Boss[34], Chris Bowler[*6], Colomban de Vargas[8], Mick Follows[35], Gabriel Gorsky[13], Nigel Grimsley[32], Pascal Hingamp[20], Daniele Iudicone[*1], Olivier Jaillon[2,3], Stefanie Kandels-Lewis[16,33], Lee Karp-Boss[34], Eric Karsenti[6,33], Uros Krzic[41], Fabrice Not[8], Hiroyuki Ogata[42], Stéphane Pesant[43,44], Jeroen Raes[22], Emmanuel G. Reynaud[45], Christian Sardet[13], Mike Sieracki[46,47], Sabrina Speich[18], Lars Stemmann[13], Matthew B. Sullivan[14], Shinichi Sunagawa[16,29], Didier Velayoudon[48], Jean Weissenbach[2,3,4], Patrick Wincker[*2,3,4]

[†] : These authors contributed equally to this work. [*] : Corresponding authors

Author Information

[1]: Stazione Zoologica Anton Dohrn, Villa Comunale, 80121, Naples, Italy. [2]: CEA - Institut de Génomique, Genoscope, 2 rue Gaston Crémieux, 91057 Evry France. [3]: CNRS UMR 8030, Evry. [4]: Université d'Evry Val d'Essonne, Evry, France. [5]: Laboratoire des Sciences du Numérique de Nantes (LS2N) – CNRS, Université de Nantes, École Centrale de Nantes, IMT Atlantique, 2 rue de la Houssinière, 44322 Nantes, France. [6]: Ecole Normale Supérieure, PSL Research University, Institut de Biologie de l'Ecole Normale Supérieure (IBENS), CNRS UMR 8197, INSERM U1024, 46 rue d'Ulm, F-75005 Paris, France. [7]: INSERM, UMRS1138, Laboratory of Integrative Cancer Immunology, Centre de Recherche des Cordeliers, 15 Rue de l'Ecole de Medecine, 75006 Paris, France. [8]: CNRS, UMR 7144, Station Biologique de Roscoff, Place Georges Teissier, 29680

Roscoff, France. [9]: Simons Centre for the Study of Living Machines, National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore 560065, India. [10]: Department of Marine Biology and Oceanography, Institute of Marine Sciences (ICM)-CSIC, Pg. Marítim de la Barceloneta 37-49, Barcelona E0800, Spain. [11]: Sorbonne Universite´s, UPMC Univ Paris 06, Univ Antilles Guyane, Univ Nice Sophia Antipolis, CNRS, Evolution Paris Seine - Institut de Biologie Paris Seine (EPS - IBPS), 75005 Paris, France. [12]: Sorbonne Universités, UPMC Université Paris 06, CNRS, Laboratoire Adaptation et Diversité en Milieu Marin UMR7144, Station Biologique de Roscoff, 29688 Roscoff, France. [13]: Sorbonne Universités, UPMC Université Paris 06, CNRS, Laboratoire d'Océanographie de Villefranche (LOV) UMR7093, Observatoire Océanologique, 06230 Villefranche-sur-Mer, France. [14:] Departments of Microbiology and Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus OH 43210, USA. [15]: Laboratory of Environmental Chemistry, School of Architecture, Civil and Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. [16]: Structural and Computational Biology Unit, European Molecular Biology Laboratory, 69117 Heidelberg, Germany. [17]: Institute for Coastal Marine Environment (IAMC) - CNR of Capo Granitola, Via del Mare 3, Torretta Granitola (TP), 91021, Italy. [18]: LMD Laboratoire de météorologie dynamique. Ecole normale supérieure de Paris, PSL Research University, 24 rue Lhomond, 75231 Paris Cedex 05, France. [19]: Université de Toulon, Laboratoire PROTEE-EBMA E.A. 3819, BP 20132, 83957 La Garde Cedex, France. [20]: Aix Marseille Univ, Université de Toulon, CNRS, IRD, MIO, Marseille, France. [21]: Information Génomique et Structurale, UMR7256, CNRS, Aix-Marseille Université, Institut de Microbiologie de la Méditerranée (FR3479), Parc Scientifique de Luminy, Marseille, France.[22]: Department of Microbiology and Immunology, Rega Institute, KU Leuven, Herestraat 49, 3000 Leuven, Belgium. [23]: Center for the Biology of Disease, VIB, Herestraat 49, 3000 Leuven, Belgium. [24]: Department of Applied Biological Sciences, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium. [25]: Institute of Parasitology, Biology Centre CAS, Branišovská 1160/31, 370 05 České Budějovice, Czech Republic. [26]: Institut Pasteur - Bioinformatics and Biostatistics Hub - C3BI, USR 3756 IP CNRS - Paris, France. [27]: IRD - BP 529 - 98713 Papeete – TAHITI. [28]: Institute of Microbiology, Department of Biology, ETH Zurich, Vladimir-Prelog-Weg 4, 8093 Zurich, Switzerland. [29]: Department of Earth Ocean and Ecological Sciences, School of Environmental Sciences, University of Liverpool, Liverpool, UK. [30]: IFREMER, Physiology and Biotechnology of Algae Laboratory, rue de l'Ile d'Yeu, 44311, Nantes, France. [31]: Biological and Environmental Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia. [32]: Sorbonne Universités, UPMC Univ Paris 06, CNRS, Biologie Intégrative des Organismes Marins (BIOM, UMR 7232), Observatoire Océanologique, Banyuls sur Mer, France. [33]: Directors' Research European Molecular Biology Laboratory Meyerhofstr. 1, 69117 Heidelberg, Germany. [34]: School of Marine Sciences, University of Maine, Orono, Maine, USA. Dept of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [36]: Department of Oceanography, University of Hawaii, Honolulu, Hawaii 96822, USA. [37]: Sorbonne Universités (UPMC, Univ Paris 06)-CNRS-IRD-MNHN, Laboratoire LOCEAN, Paris, France. [38]: Structural and Computational Biology, European Molecular Biology Laboratory, Meyerhofstr. 1, 69117 Heidelberg, Germany. [39]: Max Delbrück Centre for Molecular Medicine, 13125 Berlin, Germany. [40]: Department of Bioinformatics, Biocenter,University of Würzburg, 97074 Würzburg, Germany. [41]: Cell Biology and Biophysics, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. [42]: Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto, 611-001,

Japan. [43]: MARUM, Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany. [44]: PANGAEA, Data Publisher for Earth and Environmental Science, University of Bremen, Bremen, Germany. [45]: Earth Institute, University College Dublin, Dublin, Ireland. [46] National Science Foundation, Arlington, VA 22230, USA. [47]: Bigelow Laboratory for Ocean Sciences East Boothbay, ME, USA. [48]: DVIP Consulting, Sèvres, France.

Quantifying ocean plankton impacts on nutrient and energy cycling, and predicting their response to change is challenged by community complexity and processes that span from genes to ecosystems. Here we use environmental and metagenomic data gathered during the Tara Oceans expedition[1-4] to identify coherent assemblages of taxa co-varying with iron, a key yet often limiting micronutrient[5-7] in the global ocean. These iron associated assemblages (IAAs) contained not only eukaryotic phytoplankton such as diatoms that are well known to respond to iron[8], but also a range of consumers, parasites and viruses, inferring the importance of biotic interactions in structuring community composition as a result of increased primary production in response to supply of nutrients. To assess the relevance of the global IAAs in a dynamic environment, we followed community composition and gene expression changes during a localized perturbation downstream of the Marquesas archipelago in the equatorial Pacific Ocean, where frequent plankton blooms are believed to be the result of iron injection[9]. Our analysis revealed coherent changes wherein the globally identified IAAs and taxa within them were also responsive locally. Responses were sustained in part by changes in the expression of iron-related genes throughout the community but also by changes in the relevance of individual genotypes such as of cyanobacteria and diatoms, suggesting that community plasticity may have an important impact on large scale equilibria in marine ecosystems in addition to the responses of specialized species. By integrating global data with a localized response we provide a framework for understanding the resilience of plankton ecosystems in a changing environment.

Marine plankton play critical roles in pelagic oceanic ecosystems, with their photosynthetic component (eukaryotic phytoplankton and cyanobacteria) accounting for approximately 50% of Earth's Net Primary Production (NPP), that fuels marine food webs and sequesters organic carbon to the ocean interior[10,11]. Transient accumulations in plankton biomass (blooms) often arise after the injection of a previously limiting nutrient such as iron, a critical trace metal limiting productivity in High Nutrient Low Chlorophyll (HNLC) regions of the open ocean[12,13]. While study of such phenomena is extensive, knowledge of organismal responses is generally restricted to individual plankton groups[14,15] or localities[16,17] because of the challenges of sampling complex communities in nature and the lack of reference gene and organism catalogues that are needed to interpret community data. Here we use a combination of global and local studies enabled by data from the Tara Oceans project to attempt to overcome these limitations.

A systematic and comprehensive method to determine iron concentrations in the global ocean has not yet been developed[18]. We therefore used the outputs of two different models that provide estimates at different resolutions – the ECCO2-DARWIN derived values represent simulated five year averages[19] whereas PISCES2 generates iron estimates based on actual concentrations measured by the GeoTraces consortium[20] (see Methods and Extended Data Table 1).

To test the reliability of the model outputs, genes encoding the iron-starvation induced protein (ISIP) family from diatoms (*Bacillariophyta*) were selected as molecular indicators of iron status because they have been shown previously to display robust transcriptional responses to iron limitation[21-25]. From analysis of the patterns of *ISIP* gene abundance and expression as a function of simulated iron concentrations at globally distributed sampling sites we found highly congruent patterns (Fig. 1, Supplementary Table 1), providing a strong indication that the iron concentration estimates from the two models were indeed reliable. In addition to diatoms, we carried out a global analysis of *ISIP* gene abundance and expression in the other main groups of photosynthetic eukaryotes (*Chlorophyta*, *Haptophyceae*, *Pelagophyceae*, *Dinophyceae*), as well as other genes encoding iron-sensitive proteins (ferritin, ferredoxin, flavodoxin, cytochrome c_6, plastocyanin, fructose bisphosphate aldolase, and proteorhodopsin), which further confirmed the reliability of the iron estimates from the two models (Supplementary Information S1 and S2).

To delineate the systematic co-occurrence of planktonic biota as a function of iron we then applied weighted gene correlation network analysis (WGCNA)[1,26]. This approach deciphers dense sub-communities (modules) of organisms within a global co-occurrence network, and because of the high levels of co-variation of individual taxa they can be inferred to interact ecologically (see Supplementary Information SI1). Five eukaryotic modules significantly associated with iron were identified, that we denoted iron associated assemblages (IAAs; Black, DarkRed, Turquoise, White and Yellow in Fig. 2a and Extended Data Fig. 1), containing between 43 (DarkRed) and 592 (Turquoise) different Operational Taxonomic Units[27] (OTUs -Extended Data Table 2a and Supplementary Information SI3). For each taxon in each IAA, we computed its correlation with iron and its centrality to the subnetwork (see Methods and Supplementary Information SI3). In the five subnetworks, the most central taxa showing high positive correlations with iron were heterotrophs having relatively low prey selectivity, mixotrophs or fungi. Only in the Yellow subnetwork a group of autotrophic protists (*Mamiellophyceae*) showed both high centrality and high positive correlations to iron (Extended Data Table 2a). Unlike the other IAAs, the DarkRed module was strongly anti-correlated with iron (ECCO2-DARWIN-estimates), in particular mostly because of the diatom, dinoflagellate and foraminifera taxa within it (Extended Data Fig. 2 and Table 2a).

To further pinpoint the most relevant taxa for each IAA we used a partial least square (PLS) analysis to compute variable importance in projection (VIP) scores (Fig. 2a and Extended Data Table 2a). Each IAA contained between 9 (DarkRed) and 44 (Turquoise) lineages with VIP>1 that predicted as much as 73% of the variability of iron in the ocean (Supplementary Information SI3). Of the photosynthetic groups, dinoflagellate taxa appeared particularly relevant in the Black and Turquoise modules (>50% of the total autotrophs), diatoms in the DarkRed module (57% of the autotrophs), whereas haptophytes were only relevant in the Yellow module (35%). On average, autotrophs represented 27% of the total OTUs in the five modules. In all modules heterotrophs (heterotrophic protists and metazoans) were the most relevant trophic class, ranging from 54% (Yellow module) to 63% of the total OTUs (Black module). Marine fungi were almost undetectableuin the Black module (3%) and ranged from 15% to 18% in the others.

By projecting each sampling station onto the IAA module 'eigenlineage' (see Supplementary Information SI3), we observed that they displayed different contributions at global scale (Fig. 2b and Supplementary Information SI3). For example, the Yellow and Black modules invert their eigenvalue in the open oceans compared to the very iron-rich Mediterranean basin while the

DarkRed module inverts its eigenvalue in the Marquesas Islands as well as around the Galapagos Islands, i.e., where the equatorial upwelling of nutrient-rich, iron-depleted waters is maximal[28].

WGCNA analysis was also performed on prokaryotic and viral communities, as well as on prokaryotic genes from the Ocean Microbial Reference Gene Catalogue[29,30]. We could identify several gene modules significantly associated with iron (up to r=0.54, P=3x10-9; Supplementary Information SI3.3) but failed to reveal significant IAAs for prokaryotic taxa (at best r=0.19, P=9x10-2). This result is congruent with the proposed decoupling of taxonomy and function in bacterial assemblages[31], but may be because of lack of resolution of the taxonomic marker (16S rDNA miTags). Genes with high VIP scores included some assigned to iron and siderophore metabolism from *Prochlorococcus*, *Alteromonas*, *Blastococcus* and *Marinobacter* (Supplementary Information SI3,3). Analysis of viral communities identified four IAAs (up to r=0.63, P=10-8) containing a majority of viruses with unknown host ranges (Supplementary Information SI3.2), perhaps indicative that they are eukaryotic viruses, for which very little is known[32].

To explore the relevance of the globally defined IAAs to an ecosystem experiencing a dynamic natural perturbation related to iron bioavailability[9], we focused on the Marquesas archipelago (Supplementary Information SI4). This area was selected because of the perennial plankton blooms that are visible from space. Although iron concentrations have never been measured around the archipelago, these blooms are believed to be the result of iron injection coupled with altered circulation due to an island mass effect[9]. Our analyses revealed a highly turbulent environment, with mixing up to 100 m and strong lateral shearing downstream of the islands, which generated an area of recirculation in the wake of the main island (sampling Station TARA_123) and the formation of small eddies (sampling Station TARA_124) where blooms were occurring (Fig. 3). Because the chlorophyll-enriched patch extended over 400 km to the west of the islands, we performed an additional sampling station in the bloom area corresponding to about 2/3 weeks downstream from the islands (Station TARA_125). These sites were compared to an HNLC site upstream of the islands (TARA_122) where chlorophyll levels were lower than those observed inside the bloom (Fig. 3, Extended Data Table 3, and Supplementary Information SI4).

Consistent with chlorophyll and NPP estimates, the biomass of primary producers was around 50% higher at the downstream stations than at HNLC Station TARA_122, with noticeable increases in diatoms, haptophytes, pelagophytes, and *Synechococcus* (Figs. 4a-d and Extended Data Fig. 3). The higher productivity fueled increases in zooplankton standing stock at these stations, in particular copepods, chaetognaths, and appendicularians (Fig. 4b,e and Extended Data Fig. 3), although carbon export to depth was only increased substantially at Station TARA_123 (Fig. 4f,g), which correlates most strongly with the increase in larger sized phytoplankton at this station, as has often been noted in earlier studies[33]. Although WGCNA analysis of 16S sequences did not reveal any prokaryotic IAAs, taxonomic analysis of cyanobacteria using the more resolutive petB marker revealed changes in terms of composition, abundance and diversity in the wake of the islands (Extended Data Fig. 3 and Supplementary Information SI3). For example, we observed an almost complete shift of *Synechococcus* community composition from clade CRD1 at TARA_122 to clade II at TARA_123 and TARA_124, while relative abundance of Prochlorococcus HLIII and IV, previously shown to dominate in iron-depleted waters[34,35], was significantly reduced (Fig. 4c and Supplementary Information SI3).

Eukaryotic phytoplankton diversity increased at TARA_123 and TARA_124 (Extended Data Fig. 3d and Supplementary Information SI3), contrasting with the idea that overall diversity is reduced in blooms[36]. At these stations the increased numbers of diatoms was due principally to *Thalassiosira* and *Pseudo-nitzschia*, representatives of which are well known to respond to fluctuations in iron[13,22](Supplementary Information SI3). Increases in haptophyte and pelagophyte abundance were principally due to *Phaeocystis* and *Pelagomonas*, respectively. By contrast, the community at Station TARA_122 was more characteristic of an oligotrophic environment, with an abundance of Rhizaria[37], *Planktoniella* diatoms[38], *Chrysochromulina* haptophytes[39], and *Pelagococcus* pelagophytes[40].

Comparison of metagenomic and metatranscriptomic data further allowed us to differentiate between qualitative shifts in genotypes adapted to specific conditions and changes in transcriptional outputs in cyanobacteria (Extended Data Fig. 4), eukaryotic phytoplankton (Fig. 5a, Extended Data Figs. 5 and 6), and metazoans (Extended Data Fig. 7) using data from marine prokaryote and micro-eukaryote transcriptomes and metatranscriptomes[3,29,41] (see Supplementary Information SI4). Most importantly, study of gene switches proposed to be responsive to ambient iron concentrations such as ferredoxin/flavodoxin, plastocyanin/cytochrome c_6, and FBAI/FBAII[8,22,42-46] revealed patterns generally consistent with increased bioavailability at stations downstream of the islands with respect to HNLC Station TARA_122 both in *Synechococcus* (Extended Data Fig. 4 and Supplementary Information SI4.1) and in the major groups of eukaryotic phytoplankton (Fig. 5a, Extended Data Figs. 5 and 6, and Supplementary Information SI4). The expression patterns of proteorhodopsin and ferritin genes displayed the same trends in diatoms, as did genes encoding ISIP[8,21-25] (Extended Data Figs. 5 and 6).

Importantly, analysis of the global IAAs revealed that their behaviour is dynamic at the four Marquesas sampling sites (Fig. 2c) and that the most prominently responsive eukaryotic taxa were present in them with high scores (with the exception of *Pelagomonas*) (Extended Data Table 2a). The DarkRed modules show an opposite patterns compared to others modules, and dramatically changes its eigenvalue score from negative to positiv at the onset of the bloom(Fig. 2c).

The diatom genera displaying the strongest transcriptional responses at the Marquesas archipelago stations, the genus *Pseudo-nitzschia* show relatively high VIP score and in the DarkRed module, where it is negatively correlated to iron (Fig. 2a, Extended Data Fig. 6 and Extended Data Table 2a). The other strong-responding diatom genus, *Thalassiosira,* is found in the Turquoise and in the Black modules. Other diatom-rich modules not associated with iron contain predominantly diatoms from other genera that are not responsive in the Marquesas stations and are not negatively correlated with iron, as in the DarkRed module (Extended Data Fig. 2 and Table 2a, Supplementary Information SI1-3). Clearly then, while diatoms as a group have evolved a remarkable capacity to occupy a wide range of niches in the contemporary ocean, some taxa are likely to be specialists in specific habitats, as inferred from previous more focused studies[8].

Beyond the remarkable consistency between the global and local analyses, we further noted that *Thalassiosira* and *Pseudo-nitzschia* appear to employ different mechanisms to respond to iron. Specifically, small ferritin-containing *Thalassiosira* cells expressing cytochrome c_6 genes increase in abundance at Station TARA_123, replacing larger *Thalassiosirales* genetically adapted to low

iron at Station TARA_122 by their almost exclusive expression of plastocyanin with respect to cytochrome c_6 (Fig. 5b-c and Extended Data Fig. 6c). On the other hand, *Pseudo-nitzschia* cells respond with a strong differential expression of plastocyanin, indicative of a more flexible acclimation response. Similarly, in pelagophytes and haptophytes gene switches for iron-requiring enzymes such as ferredoxin and FBAII appear to be regulated at the transcriptional level (Fig. 5). We therefore conclude that the delineation of co-responsive sub-communities at global level can provide a valuable framework for identifying key lineages whose adaptive capacities can be compared and contrasted in specific contexts. Our analyses also illustrate that while individual taxa may specialize in the mechanisms they employ to respond to inputs of iron, there is significant redundancy in the response of each IAA as a whole, which may be important to ensure ecosystem resilience in a highly unpredictable environment.

In summary, we show that the island mass effect in the wake of the Marquesas islands impacts organisms from the base to higher trophic levels via both adaptive processes leading to the selection of preferred genotypes at the community level and acclimation to fine-tune metabolic functioning via transcriptional responses. Our local observations of the most affected organisms are consistent with IAAs identified in the global ocean, suggesting that large scale equilibria are in fact dynamic and responsive to small scale perturbations. Although nutrient limitation is considered as a bottom-up phenomenon impacting principally primary producers, our results highlight the additional significance of organisms at higher trophic levels that influence top-down processes and cause a reorganization of the different 'modules' within the community. This reorganization reflects ontogenetic responses essentially driven by modulation of gene expression patterns[3] and phylogenetic responses driven by the turnover of organisms, processes that are likely to occur over different evolutionary timescales and thus be affected differently by oceanic dispersal[4]. Numerical simulations of ocean processes that aim to capture the fluxes of key elements, as well as other forcings due to local and large scale phenomena[47], are currently based on just a handful of plankton functional types[48]. Our results infer, on the other hand, the need to incorporate the response of entire plankton assemblages to accurately encompass the different levels of biological organization operating from genes through organisms to communities. The IAAs and other modules described herein may provide such a framework, and assist our capacity to predict the responses of planktonic ecosystems to natural and human-induced perturbations.

References

1. Guidi, L. *et al.* Plankton networks driving carbon export in the oligotrophic ocean. *Nature* **532,** 465–470 (2015).

2. Bork, P. *et al.* Tara Oceans studies plankton at planetary scale. *Science* **348,** 873 (2015).

3. Carradec, Q., Pelletier, E., et al. A global ocean atlas of eukaryotic genes. *Nature*, in review (2017).

4. Richter, D. J. et al. Global plankton biogeography is shaped via ocean circulation dynamics. *Nature*, in review (2017).

5. Butler, A. Acquisition and utilization of transition metal ions by marine organisms. *Science* **281,** 207–210 (1998).

6. de Baar, H. J. W. *et al.* Synthesis of iron fertilization experiments: From the iron age in the age of enlightenment. *J. Geophys. Res. C Ocean.* **110,** 1–24 (2005).

7. Boyd, P. W. *et al.* Mesoscale iron enrichment experiments 1993-2005: synthesis and future directions. *Science* **315,** 612–617 (2007).

8. Marchetti, a. *et al.* Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proc. Natl. Acad. Sci.* **109,** E317–E325 (2012).

9. Martinez, E. & Maamaatuaiahutapu, K. Island mass effect in the Marquesas Islands: Time variation. *Geophys. Res. Lett.* **31,** 1–4 (2004).

10. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. G. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science* **281,** 237–240 (1998).

11.Behrenfeld, M. J. & Boss, E. S. Resurrecting the Ecological Underpinnings of Ocean Plankton Blooms. *Ann. Rev. Mar. Sci.* **6,** 167–194 (2014).

12. Behrenfeld, M. J. *et al.* Controls on tropical Pacific Ocean productivity revealed through nutrient stress diagnostics. *Nature* **442,** 1025–1028 (2006).

13. Coale, K. H., Fitzwater, S. E., Gordon, R. M., Johnson, K. S. & Barber, R. T. Control of community growth and export production by upwelled iron in the equatorial Pacific Ocean Kenneth. *Nature* **379,** 1994–1997 (1996).

14. Bertrand, E. M. *et al.* Phytoplankton-bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 9938–43 (2015).

15. Gong, W. *et al.* Molecular insights into a dinoflagellate bloom. *ISME J.* 1–14 (2016). doi:10.1038/ismej.2016.129

16. Smetacek, V. *et al.* Deep carbon export from a Southern Ocean iron-fertilized diatom bloom. *Nature* **487,** 313–319 (2012).

17.Clayton, S., Dutkiewicz, S., Jahn, O. & Follows, M. J. Dispersal, eddies, and the diversity of marine phytoplankton. *Limnol. Oceanogr. Fluids Environ.* **3,** 182–197 (2013).

18. Tagliabue, A. *et al.* The integral role of iron in ocean biogeochemistry. *Nature* **543,** 51–59 (2017).

19. Menemenlis, B. D., Campin, J., Heimbach, P., Hill, C. & Lee, T. ECCO2 : High Resolution Global Ocean and Sea Ice Data Synthesis. *Mercat. Ocean Q. Newsl.* **31,** 13–21 (2008).

20. Aumont, O., Tagliabue, A., Bopp, L. & Gehlen, M. PISCES-v2 : an ocean biogeochemical model for carbon and. *Geosci. Model Dev.* **8,** 2465–2513 (2015).

21. Allen, A. E. *et al.* Whole-cell response of the pennate diatom Phaeodactylum tricornutum to iron starvation. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 10438–10443 (2008).

22. Lommer, M. *et al.* Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biol.* **13,** R66 (2012).

23. Chappell, P. D. *et al.* Genetic indicators of iron limitation in wild populations of Thalassiosira oceanica from the northeast Pacific Ocean. *Isme J* **9,** 592–602 (2015).

24. Morrissey, J. *et al.* A novel protein, ubiquitous in marine phytoplankton, concentrates iron at the cell surface and facilitates uptake. *Curr. Biol.* **25,** 364–371 (2015).

25. Marchetti, A. *et al.* Development of a molecular-based index for assessing iron status in bloom-forming pennate diatoms. *J. Phycol.* 1–13 (2017). doi:10.1111/jpy.12539

26. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9,** 559 (2008).

27. de Vargas, C., Audic, S., Henry, N., Decelle, J. & Mahé, F. Eukaryotic plankton diversity in the sunlit ocean. *Science* **348,** 1–12 (2015).

28. Blain, S., Bonnet, S. & Guieu, C. Dissolved iron distribution in the tropical and sub tropical South Eastern Pacific. *Biogeosciences* **5,** 269–280 (2008).

29.Sunagawa, S. *et al.* Structure and function of the global ocean microbiome. *Science* **348,** 1261359 (2015).

30. Roux, S. *et al.* Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* **537,** 689–693 (2016).

31. Louca, S., Parfrey, L. W. & Doebeli, M. Decoupling function and taxonomy in the global ocean microbiome. *Science* **353,** 1272–1277 (2016).

32. Chow, C.-E. T. & Suttle, C. A. Biogeography of Viruses in the Sea. *Annu. Rev. Virol.* **2,** 41–66 (2015).

33. Sarmiento, J. L., Gruber, N., Brzezinski, M. A. & Dunne, J. P. High-latitude controls of thermocline nutrients and low latitude biological productivity. *Nature* **427,** 56–60 (2004).

34. Rusch, D. B., Martiny, A. C., Dupont, C. L., Halpern, A. L. & Venter, J. C. Characterization of Prochlorococcus clades from iron-depleted oceanic regions. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 16184–16189 (2010).

35. West, N. J., Lebaron, P., Strutton, P. G. & Suzuki, M. T. A novel clade of Prochlorococcus found in high nutrient low chlorophyll waters in the South and Equatorial Pacific Ocean. *ISME J.* **5,** 933–944 (2011).

36. Vallina, S. M. *et al.* Global relationship between phytoplankton diversity and productivity in the ocean. *Nat. Commun.* **5,** 4299 (2014).

37.Biard, T. *et al.* In situ imaging reveals the biomass of giant protists in the global ocean. *Nature* **532,** 504–507 (2016).

38. Malviya, S. *et al.* Insights into global diatom distribution and diversity in the world's ocean. *Proc. Natl. Acad. Sci.* **113,** E1516–E1525 (2016).

39.Stibor, H. & Sommer, U. Mixotrophy of a photosynthetic flagellate viewed from an optimal foraging perspective. *Protist* **154,** 91–98 (2003).

40. Guillou, L., Moon-Van Der Staay, S. Y., Claustre, H., Partensky, F. & Vaulot, D. Diversity and abundance of Bolidophiceae (Heterokonta) in two oceanic regions. *Appl. Environ. Microbiol.* **65,** 4528–4536 (1999).

41. Keeling, P. J. *et al.* The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. *PLoS Biol.* **12,** (2014).

42. Allen, A. E. *et al.* Evolution and functional diversification of fructose bisphosphate aldolase genes in photosynthetic marine diatoms. *Mol. Biol. Evol.* **29,** 367–379 (2012).

43. Peers, G. & Price, N. M. N. Copper-containing plastocyanin used for electron transport by an oceanic diatom. *Nature* **441,** 341–4 (2006).

44. Pierella Karlusich, J. J., Ceccoli, R. D., Graña, M., Romero, H. & Carrillo, N. Environmental selection pressures related to iron utilization are involved in the loss of the flavodoxin gene from the plant genome. *Genome Biol. Evol.* **7,** 750–767 (2015).

45 Mackey, K. R. M. *et al.* Divergent responses of Atlantic coastal and oceanic Synechococcus to iron limitation. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 9944–9 (2015).

46. Thompson, A. W., Huang, K., Saito, M. A. & Chisholm, S. W. Transcriptome response of high- and low-light-adapted Prochlorococcus strains to changing iron availability. *ISME J.* **5,** 1580–94 (2011).

47. IPCC Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.). IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. 151 (IPCC, Geneva, Switzerland, 2014).

48. Le Quere, C. *et al.* Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models\rdoi:10.1111/j.1365-2486.2005.1004.x. *Glob. Chang. Biol.* **11,** 2016–2040 (2005).

Figure 1: Abundance and expression of diatom *ISIP* genes with respect to iron concentration estimates.

Figure shows abundance (metagenome) and expression (metatranscriptome) of *ISIP* genes in surface samples from all size fractions (0.8-5, 5-20, 20-180, 180-2000 µm). (a) Global distribution of relative abundance (left) and expression (right) of *ISIP* genes. The circle colours represent iron concentration estimates at each *Tara* Oceans sampling site according to the PISCES2 model (Extended Data Table 1. The relative levels of *ISIP* genes are represented by the circle area. (b,c) Correlation of the relative abundance (b) and expression (c) of *ISIP* genes with respect to iron concentration estimates. Violin plots show the results when stations were partitioned into low iron (< 0.615 µmol/m^3; 22 stations; light grey) and high iron (> 0.615 µmol/m^3; 46 stations; dark grey), and the scatter plots display correlations normalized between 0 and 1. (d) Correlations between relative abundance and expression of *ISIP* genes, expressed as a ratio of their total levels. Scatter plots indicate Pearson correlation coefficients (pcc) and p-values in blue, and x's represent missing values. The colors in the scatter plots were used just to clarify the visualization of overlapping points.

Figure 2: Iron-associated assemblages (IAAs) of eukaryotic plankton organisms in the global ocean and in the Marquesas Islands stations.
a) Description of modules associated to iron. Relative abundances and co-occurrences of eukaryotic lineages were used to decipher modules. Five modules can predict iron with high accuracy: Black (DARWIN: R=0.37, P=$6x10^{-4}$; PISCES: R=0.38, P=$3x10^{-4}$), DarkRed (DARWIN: R=-0.43, P=$5x10^{-5}$; PISCES: R=0.19, P=0.08), Turquoise (DARWIN: R=0.46, P= $1x10^{-5}$; PISCES: R=0.42, P=$9x10^{-5}$), and Yellow (DARWIN: R=0.19, P=$^{0.09}$;PISCES: R=0.56, P=$5x10^{-8}$). To explain the role of each lineage within each IAA, lineages are associated to their score of centrality, which depicts the relative importance of each lineage to the module subnetwork, the correlation of their relative abundance with iron concentrations, and their VIP score. Representative major lineages within each module are emphasized by circles and named. b) Contribution of Tara Oceans stations to the global variance of IAAs. For each selected IAA, we represent the projection of stations on the first principal component, which highlights stations that explain most of the variance (see Supplementary Information Section SI3). c) Contributions of IAAs to the Marquesas archipelago stations. With the exception of the DarkRed module, all module show negative eigenlineage scores. The DarkRed module show positive eigenlineage scores in all the Marquesas archipelago stations, which are however significantly decreasing within the bloom stations.

Figure 3: The Marquesas study site, showing sampling sites and surface chlorophyll concentrations.

Map of surface chlorophyll in the Marquesas area. Drifter and Provbio trajectories are indicated as well as the location of the Tara Oceans sampling stations, with a zoom on Stations TARA_123 and TARA_124. The pre-bloom Station TARA_122 was located 27 km upstream of the islands. These waters were characterized by low chlorophyll concentrations (integrated total chlorophyll A (0 - 118 meters depth) $[Chl\text{-}A]_{int}$: 16.6 mg.m^{-2}; Figure 4a), but high concentrations of nutrients (NO_2: 0.12 mmol·m^{-3}, PO_4: 0.57 mmol·m^{-3}, NO_2NO_3: 5.5 mmol·m^{-3} , Si: 2.2 mmol·m^{-3}; Supplementary

Information Section SI4) characteristic of an HNLC region 11,12. Station TARA_123 is coastal, 8 km downstream of Nuku Hiva island and with a seabed depth of 1903 m and $[Chl\text{-}A]_{int}$ of 33.6 $mg \cdot m^{-2}$. Nutrients were as elevated as in the pre-bloom HNLC area, and with a particular increase of NO2 around 150m depth (1.47 $mmol \cdot m^{-3}$). Station TARA_124 is away from the coast, 43 km from Nuku Hiva, in even deeper water (2414 m bottom depth), and in an eddy also characterized by high chlorophyll content with respect to Station TARA_122 (28.5$mg \cdot m^{-2}$). Currents transported the chlorophyll patch from the coast, but also induced the eddy that sustained primary productivity by water mixing (see Supplementary Information Section SI4). Station TARA_125 is located 300 km downstream of the islands and was also characterized by high chlorophyll concentrations (27.6 $mg \cdot m^{-2}$).

Figure 4: **Main contributors to chlorophyll and carbon export.**

a) Relative contribution of different autotrophic lineages to the total chlorophyll concentration in the euphotic zone (0 - 120 meters), derived from photosynthetic pigment analysis and expressed as percent of the total measured Chlorophyll. b) Depth-integrated biomass ($mg \cdot C \cdot m^{-2}$) of autotrophs and mesozooplankton (> 300 µm) in the euphotic zone (0 – 120 m). Autotroph biomass was calculated from HPLC data using the Phyto C:Chla ratio from[25] except for *Synechococcus*, *Prochlorococcus* and pico-eukaryotes, for which the flow cytometry data were used. Zooplankton biomass was inferred from Zooscan-calculated biovolumes (see methods for further details). c) Relative abundance of *Prochlorococcus* and *Synechococcus* cyanobacteria genera expressed as percent of the total *Prochlorococcus* plus *Synechococcus* abundance estimated from flow cytometry data.Genetic markers (petB) showed exactly the same trends (see Supplementary Information SI3). d) Relative abundance (%) of ribotypes (18S-V9 Tags, see[27]) assigned to autotrophic eukaryotes lineages at the sea surface (5 meters depth). Abundances were computed for the two size fractions containing the majority of autotrophic lineages, namely 0.8 - 5 µm and 5 – 20 µm size fractions. e) Relative abundance (%) of ribotypes (18S-V9 Tags, see20) assigned to metazoan lineages lineages at the sea surface (5 meters depth). Abundances were computed for the two size fractions containing the majority of metazoan, namely the 20 – 180 µm and 180 – 2000 µm size fractions. f) Relative contribution (%) of metazoan (x axis) and of large protists (y axis) to the total carbon export (z axis - $mg \cdot m^{-2} \cdot d^{-1}$) as calculated from the Underwater Vertical Profiler[37] (UVP). Data show that Metazoa lineages played a major role in the significative increase of the carbon export detected in station Tara_123. g) Profile of carbon flux in $mg \cdot m^{-2} \cdot d^{-1}$ along the water column in stations Tara_122-Tara_125 estimated from particle size distribution and abundance and measured by UVP. A significant change in particle size spectra was observed at TARA_123, indicating strong vertical export of organic matter, whereas the carbon flux to depth at TARA_124 was similar to what was found at TARA_122 (30 $mg \cdot m^{-2} \cdot d^{-1}$).

Figure 5: Normalized mRNA abundances at Station TARA_123 with respect to Station TARA_122 for the five major groups of eukaryotic phytoplankton.

a) Distribution of fold-changes in Pfam abundance in metatranscriptomic data between the oligotrophic Station TARA_122 and the blooming Station TARA_123 in the five main photosynthetic taxonomic groups. Dinophyceae and Bacillariophyta were analyzed in the 5 – 20 µm size fractions, whereas Pelagophyceae, Haptophyceae and Chlorophyta data are derived from the 0.8 – 5µm filter. Pfam abundances were normalized in percentage within each taxonomic group and Pfam with abundances lower than 0.05% were excluded from the analysis. Pfams with interesting

expression patterns are indicated (LHC = PF00504; Ferredoxin=PF00111; FBA II = PF01116; FBA I = PF00274; Plastocyanin = PF00127, PF13473; Ribosomal_Prot = Sum of all ribosomal Pfam domains). b) Relative abundances and mRNA levels (disk sizes in RKPM) of genes potentially responsive to iron in metagenome and metatranscriptome datasets. Values were normalized by total number of genes within each grouping (*Pseudo-nitzschia* and *Thalassiosira*). For clarity we focused only on changes in 5-20 micron size fractions from Stations TARA_122 and 123. Colors indicate the contribution of each station in the expression levels. c) Relative ratios between pairs of genes whose presence in the genome or transcriptional activity has been reported previously to be potentially responsive to iron bioavailability. For clarity, ferritin levels have been multiplied by a factor of 10 to be comparable with ISIP levels, and only 5-20 μm size fractions from Stations TARA_122 and 123 are compared.

Extended Data Tables

Extended Data Table 1: Iron concentrations: PISCES v.2 estimates, ECCO2-DARWIN estimates and Observed data.

In bold: Tara sampling stations and depths used for the WGCNA analyses. Observed data: Iron concentrations measured at sampling station which are the closest to the Tara stations. Only stations in a range of 2 degrees are considered reliable in order to make a comparison with models-derived data.

Extended Data Table 2: Characteristics of lineages associated with modules.

a) Iron Associated Assemblages: Lineage, ECCO-2-DARWIN VIP scoress and correlation to iron, PISCES2 VIP scoress and correlation to iron, centrality. For selected groups, VIP scores and Iron corelation median values are given. **b**) DarkGrey and Red modules: ECCO-2-DARWIN VIP scoress and correlation to iron, PISCES2 VIP scoress and correlation to iron, centrality.

Extended Data Table 3: Vertical profiles (0 - 118 meters depth) of taxonomically relevant photosynthetic pigments.

Extended Data Figures.

Extended Data Figure 1: Heatmap showing correlation between eukaryotic modules associated to environmental parameters, as revealed by weighted gene correlation network analysis (WGCNA).

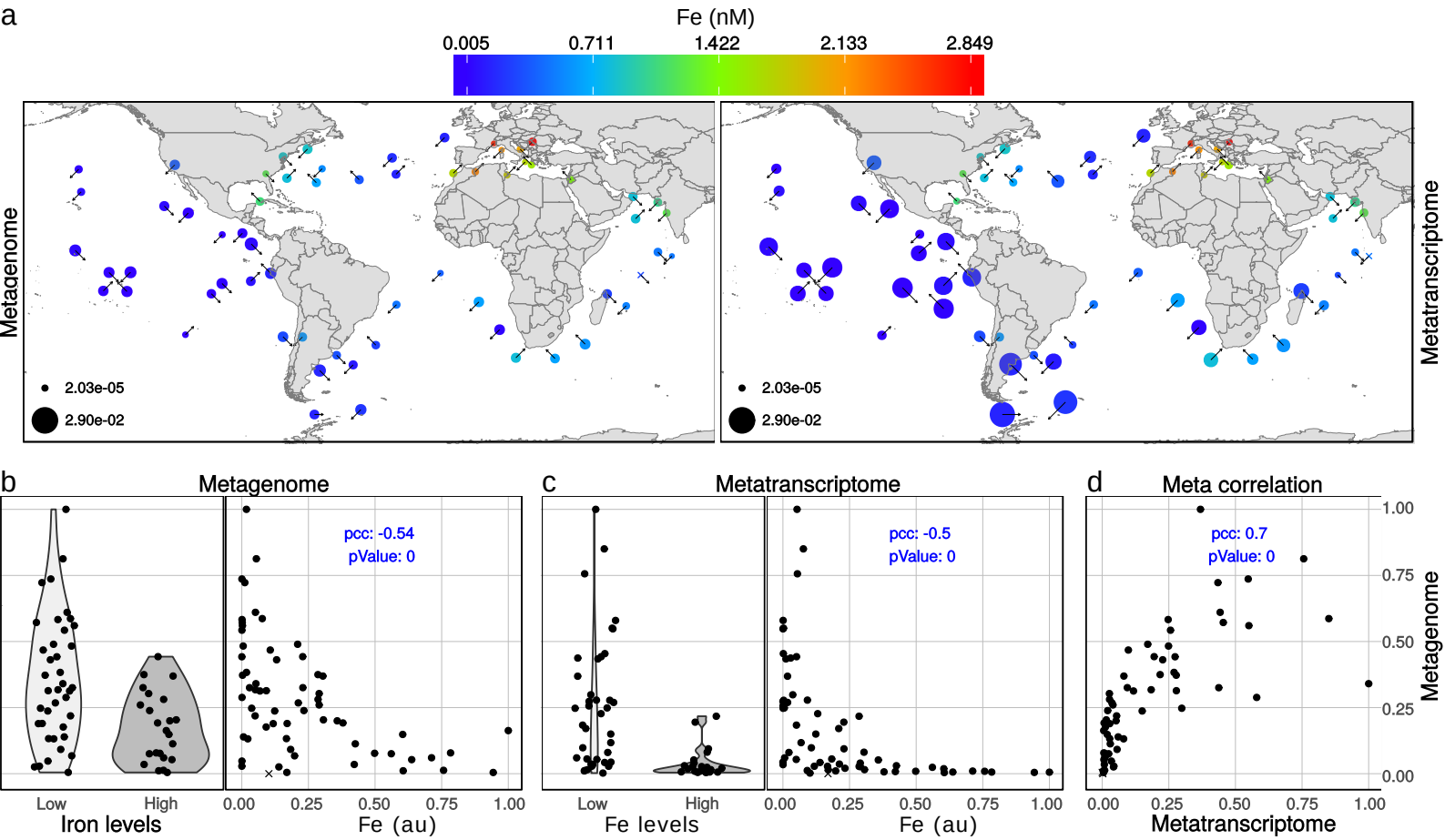

a
Fe (nM)
0.005
0.711
1.422
2.133
2.849
Metagenome
Metatranscriptome
2.03e-05
2.90e-02
2.03e-05
2.90e-02
b
Metagenome
pcc: -0.54
pValue: 0
Low
High
Iron levels
0.00
0.25
0.50
0.75
1.00
Fe (au)
c
Metatranscriptome
pcc: -0.5
pValue: 0
Low
High
Fe levels
0.00
0.25
0.50
0.75
1.00
Fe (au)
d
Meta correlation
pcc: 0.7
pValue: 0
1.00
0.75
0.50
0.25
0.00
Metagenome
0.00
0.25
0.50
0.75
1.00
Metatranscriptome

Figure 1

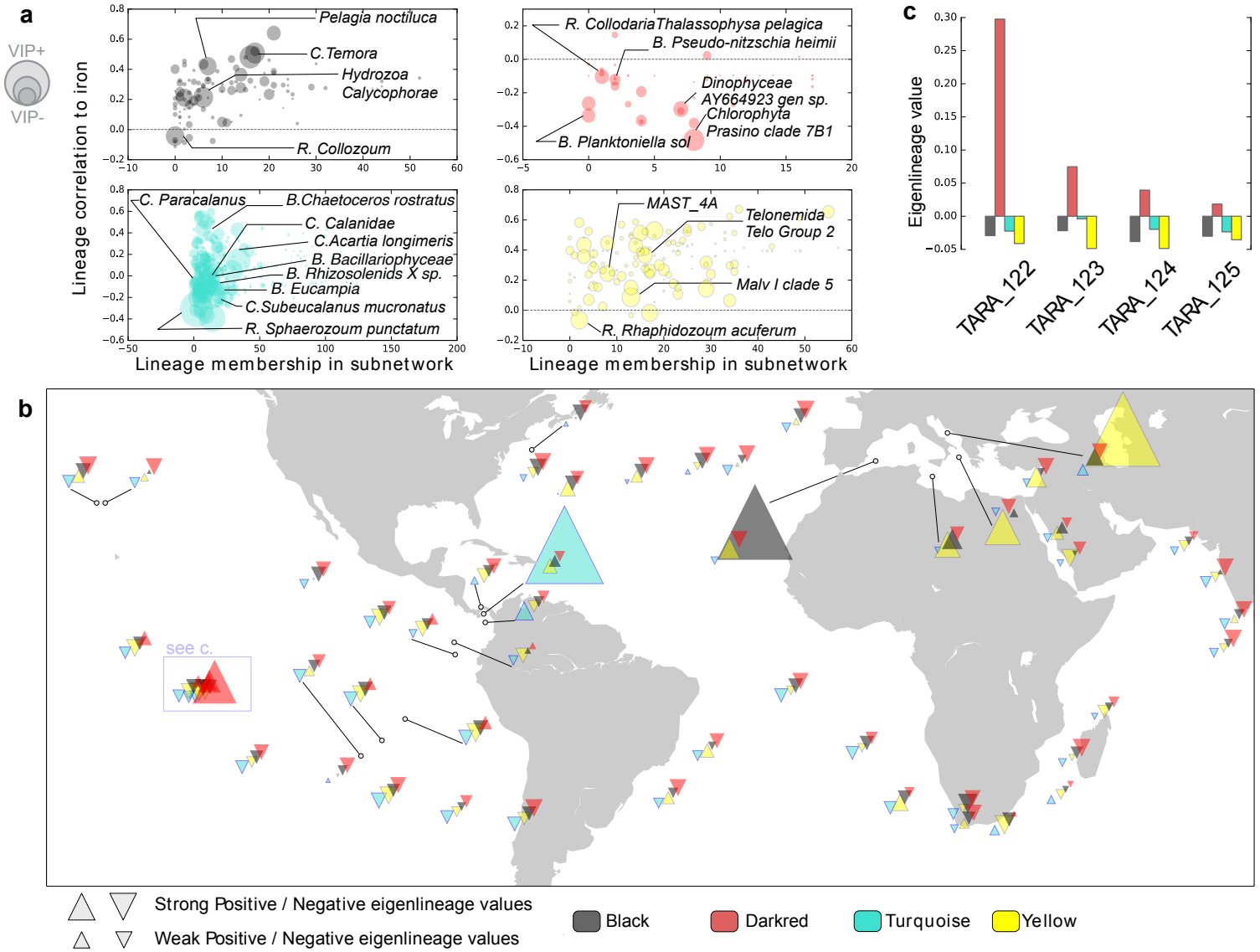

a
VIP+
VIP-
Lineage correlation to iron
Pelagia noctiluca
C.Temora
Hydrozoa
Calycophorae
R. Collozoum
C. Paracalanus
B.Chaetoceros rostratus
C. Calanidae
C.Acartia longimeris
B. Bacillariophyceae
B. Rhizosolenids X sp.
B. Eucampia
C.Subeucalanus mucronatus
R. Sphaerozoum punctatum
Lineage membership in subnetwork
R. CollodariaThalassophysa pelagica
B. Pseudo-nitzschia heimii
Dinophyceae
AY664923 gen sp.
Chlorophyta
Prasino clade 7B1
B. Planktoniella sol
MAST_4A
Telonemida
Telo Group 2
Malv I clade 5
R. Rhaphidozoum acuferum
Lineage membership in subnetwork
c
Eigenlineage value
TARA_122
TARA_123
TARA_124
TARA_125
b
see c.
Strong Positive / Negative eigenlineage values
Weak Positive / Negative eigenlineage values
Black
Darkred
Turquoise
Yellow

Figure 2

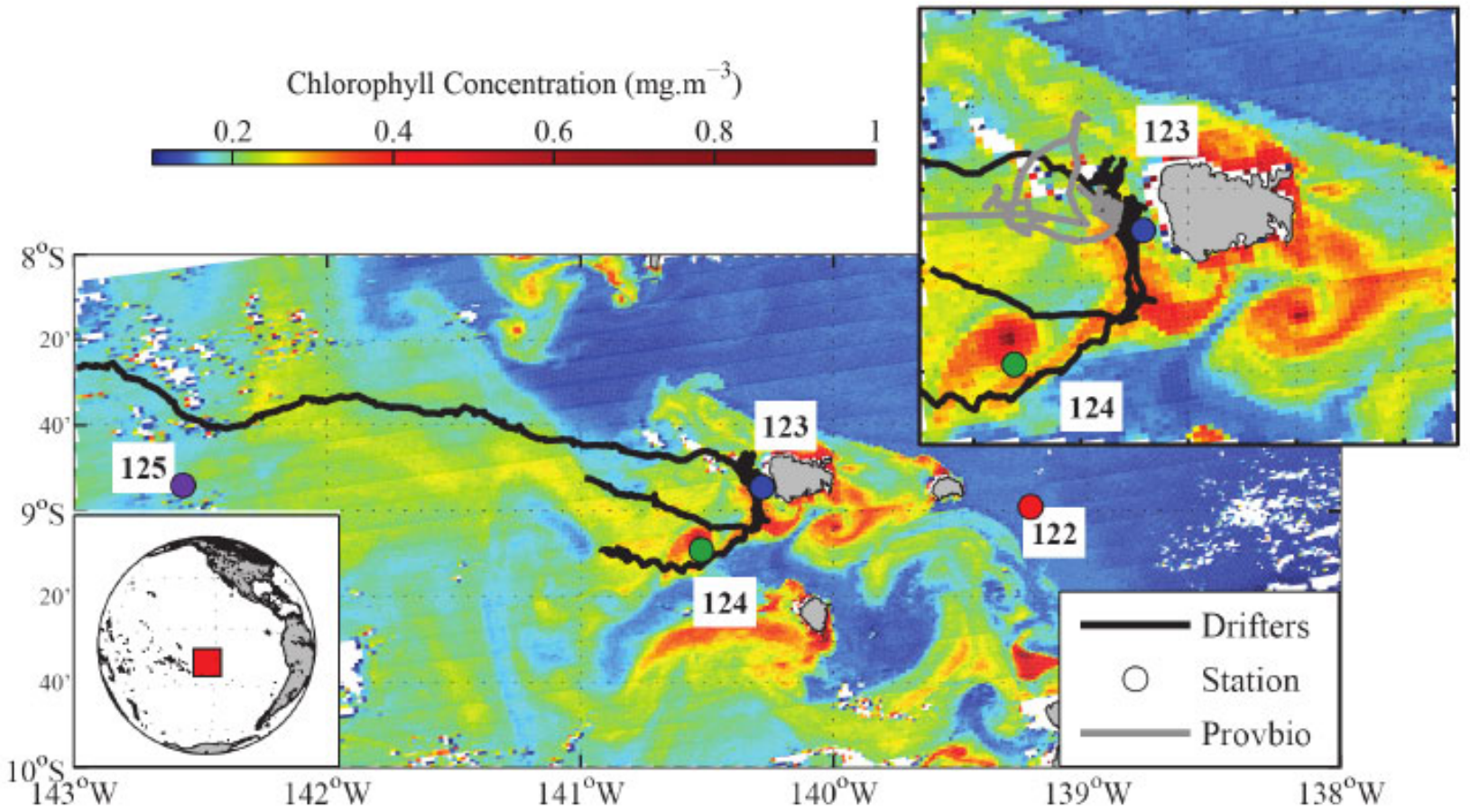

Chlorophyll Concentration (mg.m^{-3})
0.2
0.4
0.6
0.8
1
8°S
20'
40'
9°S
20'
40'
10°S
143°W
142°W
141°W
140°W
139°W
138°W
123
124
125
122
Drifters
Station
Provbio

Figure 3

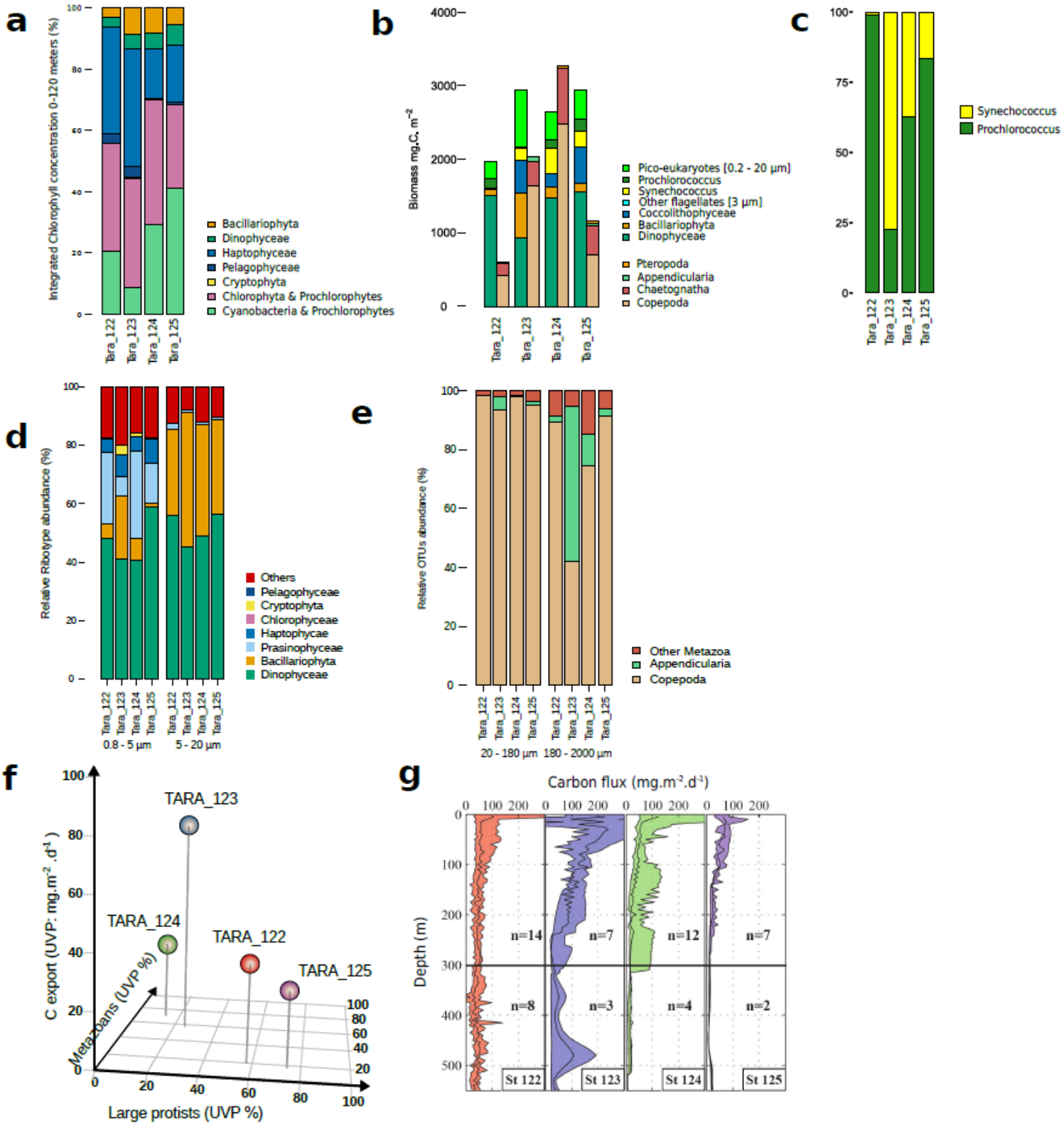

a
Integrated Chlorophyll concentration 0-120 meters (%)
Bacillariophyta
Dinophyceae
Haptophyceae
Pelagophyceae
Cryptophyta
Chlorophyta & Prochlorophytes
Cyanobacteria & Prochlorophytes
Tara_122
Tara_123
Tara_124
Tara_125
b
Biomass mg.C. m-2
Pico-eukaryotes [0.2 - 20 µm]
Prochlorococcus
Synechococcus
Other flagellates [3 µm]
Coccolithophyceae
Bacillariophyta
Dinophyceae
Pteropoda
Appendicularia
Chaetognatha
Copepoda
c
Synechococcus
Prochlorococcus
d
Relative Ribotype abundance (%)
Others
Pelagophyceae
Cryptophyta
Chlorophyceae
Haptophycae
Prasinophyceae
Bacillariophyta
Dinophyceae
0.8 - 5 µm
5 - 20 µm
e
Relative OTUs abundance (%)
Other Metazoa
Appendicularia
Copepoda
20 - 180 µm
180 - 2000 µm
f
C export (UVP: mg.m-2 .d-1)
TARA_123
TARA_124
TARA_122
TARA_125
Metazoans (UVP %)
Large protists (UVP %)
g
Carbon flux (mg.m-2.d-1)
Depth (m)
n=14
n=7
n=12
n=7
n=8
n=3
n=4
n=2
St 122
St 123
St 124
St 125

Figure 4

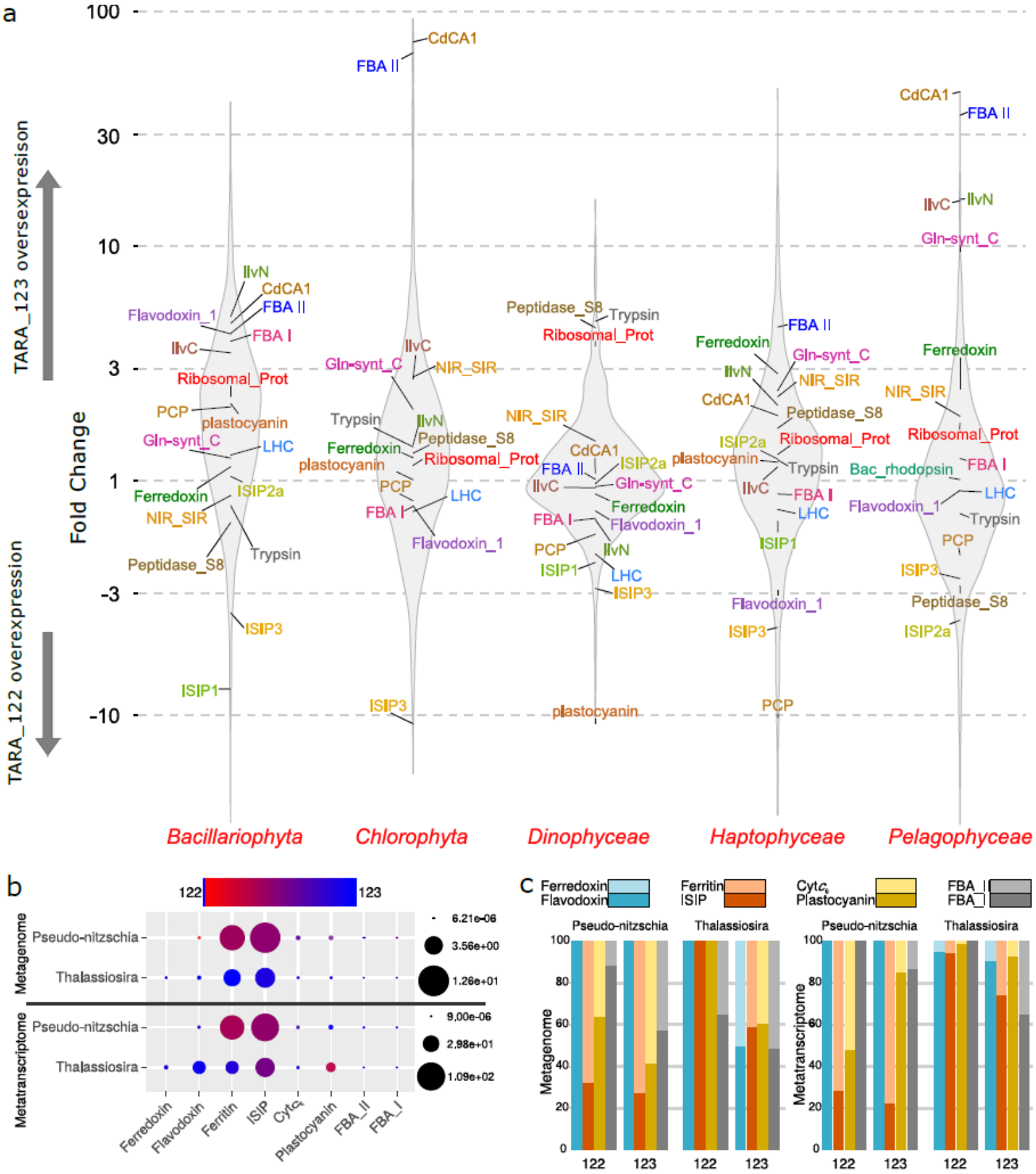
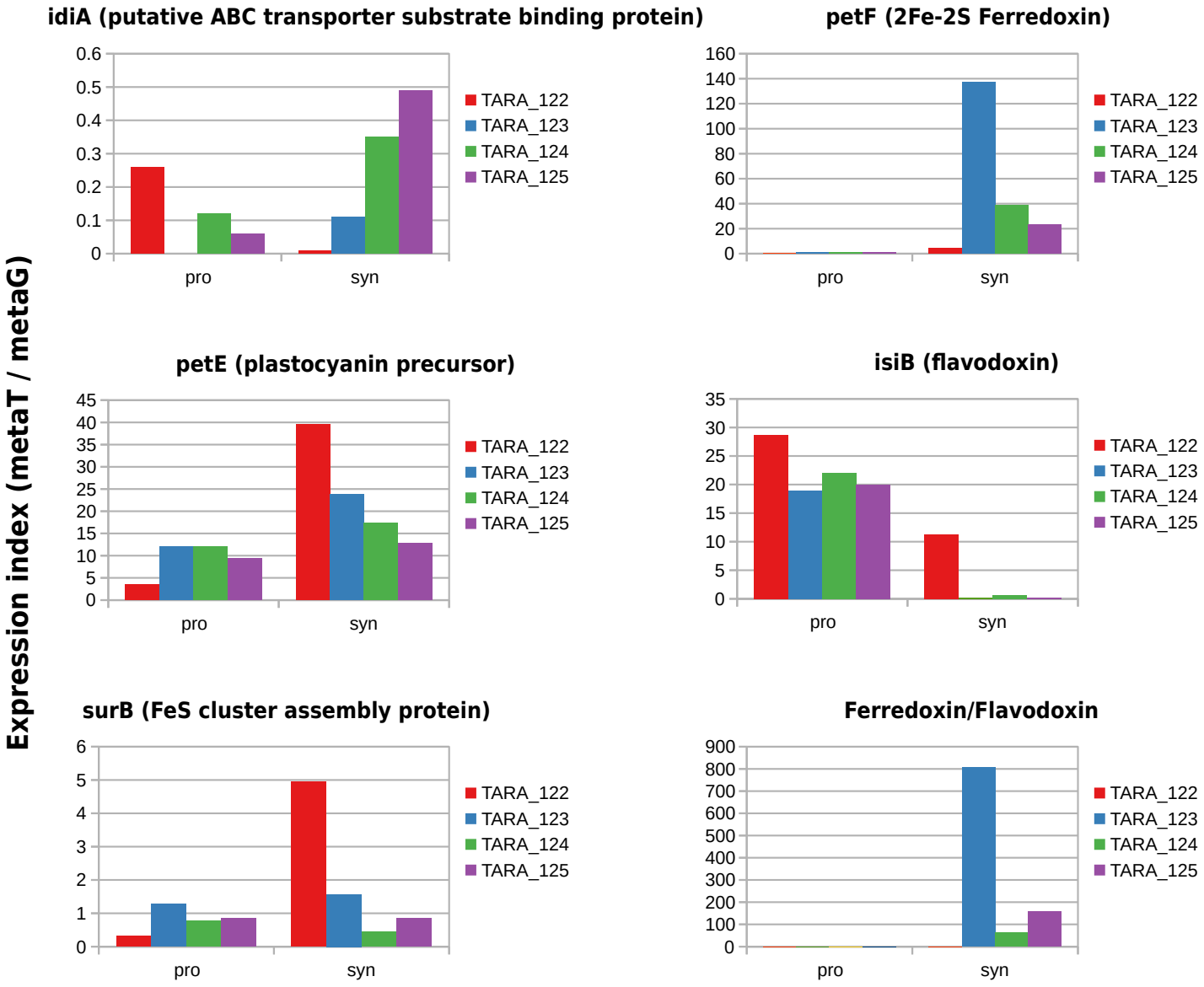
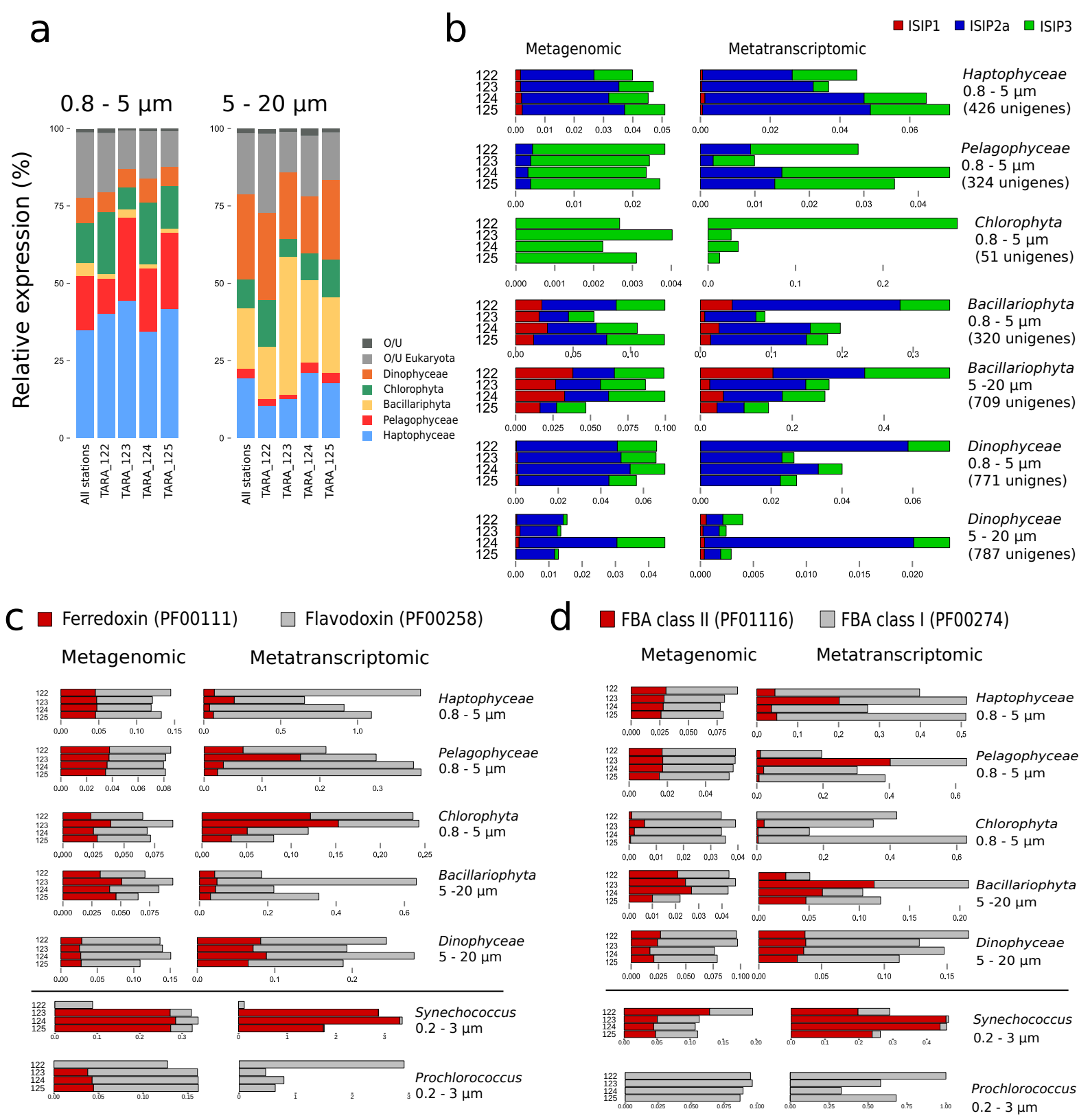

a
Fold Change
TARA_123 oversexpresison
TARA_122 overexpression
100
30
10
3
1
-3
-10
Bacillariophyta
Chlorophyta
Dinophyceae
Haptophyceae
Pelagophyceae
b
122
123
Metagenome
Metatranscriptome
Pseudo-nitzschia
Thalassiosira
6.21e-06
3.56e+00
1.26e+01
9.00e-06
2.98e+01
1.09e+02
Ferredoxin
Flavodoxin
Ferritin
ISIP
Cytc6
Plastocyanin
FBA_II
FBA_I
C
Ferredoxin
Flavodoxin
Ferritin
ISIP
Cytc6
Plastocyanin
FBA_I
FBA_II
Pseudo-nitzschia
Thalassiosira
Metagenome
Metatranscriptome
100
80
60
40
20
0
122
123

Figure 5



Extended Data Figure 4: Differential expression patterns of cyanobacterial iron-related genes among Tara Oceans stations in the Marquesas archipelago. Differential expression patterns (transcription values normalized over genomic occurrence) of several iron-related genes from cyanobacteria (pro : Prochlorococcus; syn : Synechococcus) are plotted relative to the levels observed in Station TARA_122 (index 100). The flavodoxin/ferredoxin ratio is also plotted (petB/isiB). For more information, see Supplementary Information SI4.1.



Extended Data Figure 5: Specific genes show differential abundances and expression patterns with group-specific responses.

a) Contribution of main taxonomic groups to the expression of Light-Harvesting Chlorophyll-binding proteins (LHC). The breakdown of the relative expression (in percentage) of the LHC Pfam domain (PF00504) in the different taxonomic groups is indicated by a colour code in each Marquesas station (TARA_122 to TARA_125). The average mRNA levels in all Tara Oceans stations is also shown (left column). The two smallest size fractions (0.8 - 5 µm and 5 - 20 µm) are shown. O/U = Other or Unassigned.

b) Relative genomic abundance (left panel) and mRNA levels (right panel) of the three ISIP genes in five photosynthetic eukaryote lineages in the four Marquesas Islands stations. Taxonomic groups were analyzed in the 0.8 - 5 µm size filter, except for Bacillariophyta and Dinophyceae which were also studied in the 5- 20 µm filter. Horizontal scales represent the relative abundance or expression of genes as percentages in the metatranscriptome or metagenome datasets for each taxonomic group.

c) and d) Relative genomic abundance (left panels) and mRNA levels (right panels) of Ferredoxin and Flavodoxin genes (c) and Fructose bisphosphate aldolase class II and class I (d). Seven photosynthetic taxonomic groups are represented; five eukaryote lineages studied in the 0.8 - 5 µm size filter (Haptophyceae, Pelagophyceae and Chlorophyta) and in the 5 - 20 µm size filter for Bacillariophyta and Dinophyceae. The two cyanobacteria genera (Synechococcus and Prochlorococcus) were analyzed from 0.2 - 3 µm size filters. Horizontal scales represent the relative abundance or expression of genes in percentage of the metagenome or metatranscriptome for each taxonomic group.For more information, see Supplementary Information SI4.2

ADAPTATION A LA NICHE ECOLOGIQUE CHEZ DEUX REPRESENTANTS MAJEURS DU PHYTOPLANCTON MARIN, *SYNECHOCOCCUS* ET *PROCHLOROCOCCUS* : DES GENES A L'ECOSYSTEME

Les picocyanobactéries marines *Prochlorococcus* et *Synechococcus* sont les organismes photosynthétiques les plus abondants sur la planète et sont présentes dans tous les océans à l'exception de l'océan Austral, un succès écologique sans doute rendu possible par leur grande diversité génétique. Au cours de cette thèse, j'ai cherché à mieux comprendre les liens entre diversité génétique et adaptation à la niche écologique chez ces deux genres. Tout d'abord, l'étude de la répartition des populations de picocyanobactéries à l'échelle mondiale à l'aide d'un marqueur taxonomique très résolutif m'a permis de définir des unités taxonomiques écologiquement significatives pour les taxons dominants, d'améliorer la délimitation de leurs niches écologiques et d'identifier les principaux facteurs abiotiques influençant leur répartition *in situ*. La deuxième partie de ce travail de thèse a visé à identifier les bases génétiques de l'adaptation des picocyanobactéries marines à des niches écologiques distinctes. D'un point de vue évolutif, l'étude comparative de 81 génomes de ces organismes a révélé le rôle combiné des gains et pertes de gènes et des substitutions d'acides aminés dans la diversification de ces deux genres. L'analyse de la répartition de l'ensemble des gènes connus de picocyanobactéries marines dans l'océan mondial a également permis de montrer que chaque communauté, adaptée à des conditions environnementales données, possède un répertoire de gènes distinct. Enfin, le dernier volet de cette thèse a consisté en la caractérisation physiologique d'une souche modèle de *Synechococcus* et de 4 souches représentatives des écotypes dominant *in situ* soumises à des stress lumineux et thermiques afin de mieux comprendre les bases moléculaires de l'adaptation et la variabilité écotypique de la réponse au stress chez *Synechococcus*, notamment par une approche de transcriptomique à grande échelle. Les résultats obtenus au cours de cette thèse ont donc permis non seulement d'améliorer notre connaissance des niches écologiques occupées par les picocyanobactéries marines et du niveau taxonomique à considérer pour différencier les écotypes au sein de ces deux genres, mais également de mieux comprendre les mécanismes leur ayant permis de s'adapter à ces niches variées et de coloniser presque tous les océans.

NICHE ADAPTATION IN TWO MAJOR MEMBERS OF MARINE PHYTOPLANKTON, *SYNECHOCOCCUS* AND *PROCHLOROCOCCUS*: FROM GENES TO ECOSYSTEM

The marine picocyanobacteria *Synechococcus* and *Prochlorococcus* are the two most abundant photosynthetic organisms on the planet and are present in all oceans, except the Antarctic Ocean, an ecological success likely related to their wide genetic diversity. During this PhD thesis, I explored the links between genetic diversity and niche adaptation in these two genera. First, the analysis of the distribution of picocyanobacterial populations at the global scale using a high-resolution taxonomic marker, allowed me to define ecologically significant taxonomic units for all major taxa, to improve the delineation of their ecological niches and to identify the main abiotic factors influencing their in situ distribution. The second part of this work aimed at identifying the genetic bases of adaptation of marine picocyanobacteria to distinct niches. From an evolutionary point of view, the comparative analysis of 81 genomes of these organisms revealed the combined role of gene gains and losses and of substitutions in protein sequences in the diversification of both genera. Analysis of the distribution of all known picocyanobacterial genes in the global ocean also allowed to show that each community, adapted to specific environmental conditions, possesses a distinct gene repertoire. Finally, the last part of this work has consisted in the physiological characterization of one model *Synechococcus* strain and four other strains representative of the predominating ecotypes in the field, which were submitted to light and thermal stresses in order to better understand the molecular bases of niche adaptation and the ecotypic variability of the stress response in *Synechococcus*, notably through an extensive transcriptomic study. Altogether, results obtained during this PhD have allowed not only to improve our knowledge of the ecological niches occupied by marine picocyanobacteria and of the taxonomical level to consider to differentiate ecotypes within both genera, but also to better understand the mechanisms that allowed them to adapt to these various niches and to colonize almost all the oceans.