HAL
open science

Étude des bases moléculaires de l'adaptation et de l'acclimatation aux changements environnementaux chez les picocyanobactéries marines, des organismes clés du phytoplancton

Ulysse Guyet

To cite this version:

Ulysse Guyet. Étude des bases moléculaires de l'adaptation et de l'acclimatation aux changements environnementaux chez les picocyanobactéries marines, des organismes clés du phytoplancton. Bio-informatique [q-bio.QM]. Sorbonne Université, 2020. Français. NNT : 2020SORUS185 . tel-03709569

HAL Id: tel-03709569

https://theses.hal.science/tel-03709569

Submitted on 30 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE DE DOCTORAT DE SORBONNE UNIVERSITE

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

Présentée par

Ulysse GUYET

En vue de l'obtention du grade de

DOCTEUR de SORBONNE UNIVERSITE

Étude des bases moléculaires de l'adaptation et de l'acclimatation aux changements environnementaux chez les picocyanobactéries marines, des organismes clés du phytoplancton

Soutenue le 29 juin 2020, devant le jury composé de :

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

Dr. PELLETIER Eric .....Rapporteur
CEA-Genoscope, Evry

Dr. JEAN Géraldine .....Examinatrice
Université de Nantes

Dr. MORALES Julia .....Examinatrice
CNRS, Roscoff

Dr. EVEILLARD Damien.....Co-encadrant de thèse
Université de Nantes

Dr. GARCZAREK Laurence.....Directrice de thèse
CNRS, Roscoff

Station Biologique de Roscoff
UMR 7144 Adaptation et Diversité en Milieu Marin
Equipe ECOMAP (Ecology of Marine Plankton)

RESUMÉ

Les deux picocyanobactéries marines *Prochlorococcus* et *Synechococcus* sont les organismes photosynthétiques les plus abondants de la planète, leur vaste distribution étant sans doute liée à leur importante diversité génomique. La première partie de mon travail a consisté à étudier les bases génétiques de l'adaptation de ces organismes à des niches écologiques distinctes et a révélé, en comparant 81 génomes de ces deux genres, le rôle des gains/pertes de gènes et des variations des séquences nucléotidiques dans la diversification de ces organismes. Une deuxième partie a consisté en l'analyse des réponses physiologiques d'une souche modèle de *Synechococcus* (WH7803) et de quatre autres souches représentatives des écotypes dominants *in situ* (clades I à IV) à divers stress environnementaux, afin d'identifier les gènes impliqués dans les réponses spécifiques ou communes à ces différents stress et/ou écotypes. Enfin, la dernière partie de ma thèse a visé à identifier la distribution de l'ensemble des gènes de picocyanobactéries dans l'océan mondial et à la relier aux paramètres environnementaux et à la distribution des écotypes, ce qui a permis de mettre en évidence les répertoires de gènes spécifiques de niches et/ou d'écotypes. L'intégration de ces résultats m'a permis de mieux comprendre les mécanismes d'adaptation et d'acclimatation, qui ont permis aux picocyanobactéries de coloniser la quasi-totalité des niches écologiques éclairées des océans.

ABSTRACT

The two marine picocyanobacteria Prochlorococcus and Synechococcus are the most abundant photosynthetic organisms on the planet, their wide distribution being probably linked to their large genomic diversity. The first part of my work has consisted in studying the genetic bases of the adaptation of these organisms to distinct ecological niches and revealed, by comparing 81 genomes of these two genera, the role of gene gains/losses and variations in nucleotide sequences in the diversification of these organisms. A second part has consisted in the analysis of the physiological responses of a Synechococcus model strain (WH7803) and of four other strains, representative of the dominant ecotypes in the field (clades I to IV), to various environmental factors in order to identify the genes that are involved in specific or common responses to these different stresses and/or ecotypes. Finally, the last part of my thesis aimed at identifying the distribution of all picocyanobacterial genes in the world ocean and to link it to environmental parameters and to the distribution of ecotypes, which made it possible to highlight the gene repertoires specific of niches and/or ecotypes. The integration of these results led me to get a better understanding of the adaptation and acclimation mechanisms, which allowed marine picocyanobacteria to colonize virtually all lit ecological niches of the oceans.

Remerciements

Tout d'abord, j'aimerais remercier les membres du jury d'avoir accepté de lire et juger ce travail. Je remercie donc les rapporteurs, José-Manuel Garcia Fernandez et Éric Pelletier ainsi que les examinatrices, Géraldine Jean et Julia Morales. Merci également aux membres de mon comité de suivi de thèse pour leur intérêt et leurs suggestions tout au long de cette thèse : Alexis Dufresne, François Thomas, Erwan Corre et Christophe Destombe.

Merci à toi, Laurence, de m'avoir fait découvrir les picocyanobactéries et de m'avoir fait confiance tout au long de cette thèse. C'est un plaisir de travailler avec toi, j'ai appris beaucoup durant cette période et je t'en suis infiniment redevable. Je pense aussi à Fred, avec qui tu formes une bonne équipe. Merci Fred pour tes remarques pertinentes, ton regard critique et ton trait d'humour. Merci à vous deux pour votre travail et votre persévérance qui ont permis à cette thèse de voir le jour !

Merci Damien pour m'avoir co-encadré durant cette thèse, tu as toujours des idées originales pour mettre en valeurs les résultats. Et merci à Julie pour sa participation dans les travaux présentés dans cette thèse.

Je souhaite aussi remercier l'ensemble des personnes impliqués dans les études présentées dans cette thèse. Je pense tout d'abord à la part colossale de travail qu'ont représentées les manip SAMOSA, c'est pourquoi je remercie Momo, An, Justine, Théophile, Hugo, Solène, Christophe, Fred et Laurence qui ont permis d'obtenir un si beau jeu de données et de tels résultats. Je remercie les membres de la plateforme ABIMS, en particulier Mark pour le développement et l'amélioration constante de Cyanorak et pour toutes les demandes de nouveaux exports. Je pense aussi à Gildas et Jean-Michel, pour tous les outils installés et pour m'avoir parfois laisser coloniser les nœuds du cluster. J'aimerais aussi remercier Hugo et Greg pour leur contribution bioinformatique.

La station biologique de Roscoff est un lieu privilégié pour travailler et ce grâce à l'ensemble des personnes qui veillent à ce que la vie à la station soit plus douce. Je remercie particulièrement Nathalie et Léna pour la gestion, ainsi que Céline qui est incroyablement douée pour nous extirper des méandres administratifs les plus complexes.

Merci à l'ensemble de l'équipe ECOMAP qui est comme une grande famille et merci pour ces pauses cafés qui nous redonne le sourire, à Flo, Momo, Fabienne, Domi, Léna, Mathilde, Pris, Estelle, Sarah, Louison, Martin, Peche, (je ne peux pas citer tout le monde !) mais aussi aux anciens du groupes : Pym, Emilie, Laure, Charles, Margot, Théophile, Hugo, Laura, Solène, Adriana, Justine …

Je pense aussi aux fidèles du Gulf Stream, bravo à la team « no panais » qui un jour vaincra ! Et merci Flo pour les graines, j'aurais maintenant toujours du panais à portée de main !

Roscoff c'est aussi les mercredis du Ty Pierre et pleins de bons moments, pour tout ça je remercie Mathilde, Camille, Léna, Damien, Florian, Erwan, Bertille, Alicia, Martin, Jade, Morgane, Adrien, Ewen, Kévin, Yacine, Guillaume, Louison, , Victor, Antoine, Yasmine, Marion, Lisa, Laure, Jérémy, Mariana, Peche, Hugo, Laura, Théophile, Justine, Alexis, Marine, Eloïse, qui m'ont tous permis de garder des souvenirs mémorables. Merci à Camille pour les délires incroyables et de partager la passion de la « grande musique », le duc n'a plus aucun secret pour nous ! Léna, merci pour avoir supporté toutes mes « visions » de Newtnewt. Bien sûr j'ai forcément une pensée pour Mathilde, merci pour ta bonne humeur contagieuse, tes histoires géniales et toutes les pauses café, « celle-ci c'est une bonne personne » ! Merci aussi à Louison et Bertille et tous les habitués de cette fameuse pause café du milieu d'aprem !

Enfin je ne peux pas oublier mes collègues musiciens : Kévin, Yacine, Ewen, Guillaume, Eléonore, merci pour votre passion et d'avoir la même folie pour la musique, que de bonnes sessions !

Je souhaite remercier l'ensemble de ma promo du master Bioinfo et particulièrement Dimitri, Jennifer, Victor, Anne-Sophie, Clara, Xavier, Kenzo et Leïla. On en aura vécu des histoires mémorables et j'espère qu'on aura l'occasion d'en rigoler pour encore longtemps. Du point de vue professionnel, vous êtes de très bons scientifiques et je pense notre réseau nous sera toujours aussi bénéfique.

J'ai une pensée pour les manceaux : Julien, Clémentine et Mathieu, les quatre Fantastiques ! C'est sûrement en partie grâce à vous que je me suis lancé dans la thèse, je vous en remercie ! En particulier Mathieu, que de souvenirs à Nantes avec cette coloc' ! Brad, tu es mon gars sûr, toujours là quand il faut !

Enfin, merci à mes parents pour votre soutien et votre dévouement sans faille, je vous dois énormément.

Et Merci à toi Axelle, celle qui est à mes côtés depuis si longtemps et qui a toujours su me supporter. Merci de partager ma vie.

Table des matières

INTRODUCTION 1

I. Les picocyanobactéries, des membres clés du phytoplancton 3

I.1 Le phytoplancton 3

I.1.1 Généralités 3

I.1.2 Rôle dans la pompe biologique 3

I.2 Les picocyanobactéries marines 5

I.2.1 Découverte 5

I.2.2 L'appareil photosynthétique 6

I.2.3 Importance écologique des picocyanobactéries et distribution 11

II. Diversité génétique et biogéographie des picocyanobactéries marines 13

II.1 Diversité génétique des picocyanobactéries marines 13

II.2 Répartition biogéographique et microdiversité des picocyanobactéries marines 15

III. Réponses des cyanobactéries aux variations des conditions environnementales 20

III.1 Bases génétiques de l'adaptation chez les picocyanobactéries marines 20

III.1.1 Caractéristiques générales des génomes de picocyanobactéries 20

III.1.2 Génome commun, accessoire et pan-génome 21

III.1.3 Rôles des transferts horizontaux et îlots génomiques 23

III.1.4 Rôle des substitutions nucléotidiques 26

III.2 Réponses physiologiques aux variations des conditions environnementales 27

III.2.1 Réponse de l'appareil photosynthétique et mécanismes de photoprotection 27

III.2.2 Les mécanismes d'ajustement de la fluidité membranaire 35

III.2.3 Effet de la carence en nutriments 35

Objectifs de thèse 39

CHAPITRE I 3

Contexte de l'étude.....43

I.1 Cyanorak v2.1, a scalable information system dedicated to the visualization and expert curation of picocyanobacteria genomes.....45

I.2 Evolutionary mechanisms of long-term genome diversification associated with niche partitioning in marine picocyanobacteria 72

CHAPITRE II 139

Contexte de l'étude.....141

Synergic effects of temperature and irradiance on the physiology of the marine Synechococcus strain WH7803 143

CHAPITRE III 201

Contexte de l'étude.....203

III.1 Global distribution of *Synechococcus* and *Prochlorococcus* gene repertoires reveals adaptive strategies of picocyanobacterial communities 205

III.2 Distribution globale des gènes de désaturases, impliqués dans la thermoregulation de membranes, chez les picocyanobactéries marines 246

III.3 Physiologie comparative et acclimatation aux facteurs environnementaux des principaux écotypes de *Synechococcus*.....250

CONCLUSIONS ET PERSPECTIVES 262

I. Mécanismes évolutifs des génomes de picocyanobactéries et rôle dans l'adaptation à la niche écologique 264

II. Caractéristiques générales et spécificités de la réponse transcriptomique de *Synechococcus* aux variations des paramètres environnementaux.....269

III. Importance du maintien et de l'amélioration des bases de données génomiques de référence dans un contexte de croissance exponentielle des données omiques marines 272

REFERENCES 279

ANNEXES..... 297

Liste des figures

Figure 1 - Cycle du carbone océanique 4

Figure 2 - Réactions de la photosynthèse 7

Figure 3 - Antennes collectrices de lumière de *Synechococcus* et *Prochlorococcus* 9

Figure 4 - Composition du phycobilisome et propriétés d'absorption des différents types pigmentaires.. 10

Figure 5 - Distribution dans l'océan mondial de *Prochlorococcus* et *Synechococcus* 11

Figure 6 - Arbres phylogénétiques des picocyanobactéries marines 14

Figure 7 - Relation entre la phylogénie *Prochlorococcus* et l'adaptation aux paramètres environnementaux 16

Figure 8 - Biogéographie des ESTU de Prochlorococcus à partir des métagénomes de surface de *Tara* Oceans et leur relation avec les paramètres physico-chimiques 17

Figure 9 - Biogéographie des ESTUs de *Synechococcus* à partir des métagénomes de surface de *Tara* Oceans et leur relation avec les paramètres physico-chimiques 19

Figure 10 -Pan-génome de *Prochlorococcus* 22

Figure 11 Distributions des genes de nitrate reductase (*narB*) et nitrite reductase (*nirB*) parmi 329 génomes de *Prochlorococcus* et *Synechococcus* 24

Figure 12 - Contenu en gènes et synténie de la région codant pour la biosynthèse et la régulation des bras des phycobilisomes chez *Synechococcus* 25

Figure 13 - Sites des dommages induits par la lumière dans le PSII 28

Figure 14 - Cycle de réparation de la protéine D1 du photosystème II (PSII) 29

Figure 15 - Représentation du flux d'électron d'une molécule d'eau jusqu'au $NADP^+$ et des voies alternatives de transport d'électron chez les cyanobactéries 34

Figure 16 - Localisation des stations de l'expédition *Tara* Océans utilisées dans cette étude (a), et abondances relatives des gènes de désaturases lipidiques Δ9 *desC3* (b) et *desC4* (c) et de désaturases lipidiques Δ12 *desA2* (d) et *desA3* (e), en fonction de la température de surface aux différentes stations 250

Figure 17 - Variation du rendement quantique du PSII (FV/FM) en réponse au stress dû aux fortes lumières (HL, 250 µE.m-2.s-1) et aux basses températures (LT, 13°C) pour différentes souches de Synechococcus acclimatées en faible lumière (20 µE.m-2.s-1) et à 22°C 254

Figure 18 - Taux de réparation cumulatif du PSII pendant les 6 heures de l'expérience pour les cultures contrôle (Ct, panneau de gauche) ou soumises à un stress lumineux (HL, panneau de droite) 255

Figure 19 - Variations du rapport d'émission de fluorescence de phycoérythrine (PE) à phycocyanine (PC) chez différentes souches de *Synechococcus* acclimatées à faible lumière (LL, 25 µE.m^{-2}.s^{-1}) 256

Figure 20 - Clustering des gènes par profils d'expression (log2FC) 258

Figure 21- Description du workflow d'intégration de nouveaux génomes..... 275

Figure 22 - Représentation de la complétude des génomes en fonction de leur nombre total de protéines communes..... 276

Figure 23 - Représentation du nombre de génomes par clades pour *Prochlorococcus et Synechococcus* 277

Figure 24 - Analyse des pangénomes de *Prochlorococcus* (vert) et *Synechococcus* (orange) calculés à partir des génomes nouvellement téléchargés et des génomes de Cyanorak 278

Liste des tableaux

Tableau 1 - Criblage du génome pour les gènes présumés de désaturase dans 53 génomes marins de *Synechococcus* et *Cyanobium*..... 248

Tableau 2 - Caractéristiques des souches de *Synechococcus* utilisées dans cette étude 251

Tableau 3 – Description du plan expérimental..... 252

INTRODUCTION

I. Les picocyanobactéries, des membres clés du phytoplancton

I.1 Le phytoplancton

I.1.1 Généralités

L'habitat marin constitue le plus large écosystème sur Terre, couvrant plus de 70% de la surface terrestre et atteignant une profondeur de 3,7 km en moyenne. La grande majorité des organismes l'occupant font partie du plancton (du grec : πλαγκτός ou planktos, à la dérive), c'est à dire l'ensemble des organismes pélagiques se déplaçant passivement au gré des courants marins. Ce groupe est composé d'organismes présentant une grande diversité de taille, de morphologie, de taxonomie, de régime trophique (phototrophe, hétérotrophe, mixotrophe) et mode de vie (libre, symbiose). Les espèces composant le plancton couvrent cinq groupes d'organismes (les virus, les bactéries, les archées, les protistes et les métazoaires) et ont été divisées historiquement suivant un critère de taille (allant de pico à méga) mais aussi suivant leur mode de vie (Breitbart, 2012; Sieburth et al., 1978; Suttle, 2007). Le phytoplancton regroupe l'ensemble des organismes photosynthétiques microscopiques (procaryotes ou eucaryotes) vivant dans la zone euphotique, qui correspond à la couche éclairée des océans et qui peut atteindre jusqu'à 250 mètres de profondeur.

I.1.2 Rôle dans la pompe biologique

Les organismes photosynthétiques ne représentent que 0,2% de la biomasse globale mais ont néanmoins un rôle majeur dans les cycles biogéochimiques globaux et au sein des écosystèmes marins puisqu'ils sont responsables d'environ 46% de la production primaire nette globale (Field et al., 1998). L'activité photosynthétique permet à ces organismes de fixer le CO_2 atmosphérique afin de générer de la matière organique qui constitue la base d'un réseau trophique complexe. Environ la moitié de cette matière organique est utilisée par les niveaux trophiques supérieurs lorsque le phytoplancton est consommé par les organismes hétérotrophes ou mixotrophes qui nourrissent, à leur tour, les métazoaires (Figure 1). Cependant le reste de la matière organique est excrétée durant la croissance et à la mort des organismes phytoplanctoniques sous forme dissoute (DOM) ou particulaire (POM). Ces substrats riches en énergie sont consommés par des bactéries hétérotrophes ou mixotrophes, libérant au passage des nutriments sous forme inorganique qui seront à nouveau disponibles pour le phytoplancton. La prédation de ces bactéries est aussi une source d'énergie pour les niveaux trophiques supérieurs. L'ensemble de ces phénomènes est appelé « boucle microbienne » (Azam et al., 1983; Sherr and Sherr, 2002). Enfin, Une fraction de la matière organique va couler vers les fonds marins

et sera séquestrée dans les sédiments océaniques pendant plusieurs milliers d'années, c'est la pompe biologique à carbone (Volk and Hoffert, 1985).

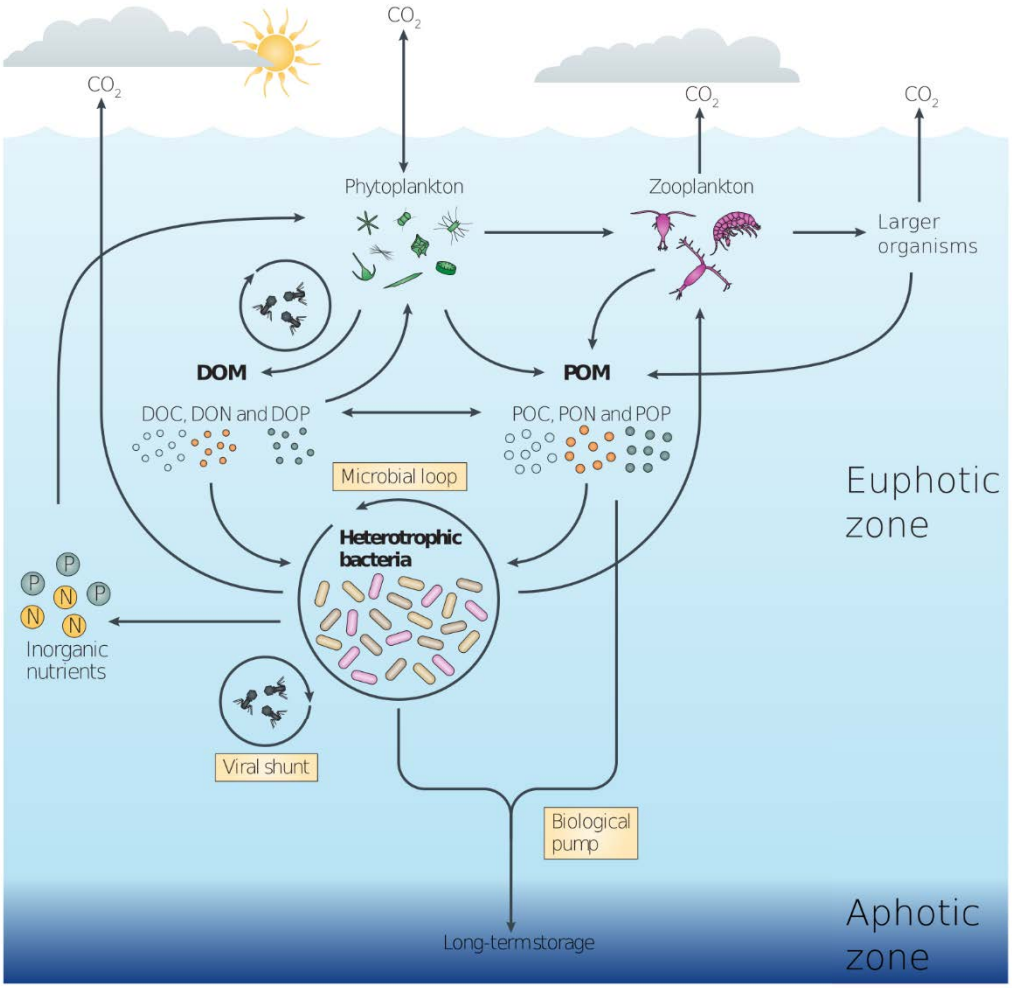


Figure 1 - Cycle du carbone océanique. DOM : matière organique dissoute (incluant DOC, DON et DOP, pour Carbone, Azote et Phosphate organiques dissouts, respectivement) et POM : matière organique particulaire (incluant POC, PON et POP, pour Carbone, Azote et Phosphate organiques particulaires, respectivement ; d'après Buchan et al., 2014.

I.2 Les picocyanobactéries marines

I.2.1 Découverte

Les organismes du picoplancton ne sont étudiés que depuis la fin des années des années 1970 grâce à l'avènement de nouvelles techniques et technologies améliorant le prélèvement d'échantillons et leur analyse. L'introduction de la microscopie à épifluorescence en biologie marine a permis la découverte de cyanobactéries planctoniques sphériques et de petite taille (entre 0,8 et 2 µm) et fluoresçant dans l'orange, dû à la présence de phycoérythrine parmi les pigments composants leurs antennes photosynthétiques. Ces organismes sont retrouvés jusqu'à des profondeurs allant jusqu'à 400 mètres et à des concentrations pouvant atteindre 10^5 cellules par millilitre parmi des échantillons prélevés dans diverses zones océaniques (Waterbury et al., 1979). Ils ont pu être classés dans le genre *Synechococcus* sur la base de caractères physiologiques et morphologiques, préalablement décrits chez des souches d'eau douce appartenant à ce genre (Waterbury et al., 1979). En 1988, Sallie Chisholm met en évidence durant des campagnes océaniques des populations de cellules beaucoup plus petites (0,5 à 0,8 µm de diamètre) que *Synechococcus* grâce à la cytométrie en flux. Ces populations émettent de la fluorescence rouge, ce qui indique qu'elles contiennent de la chlorophylle et n'ont pas les pigments accessoires présents chez les membres du genre *Synechococcus* qui les font fluorescer dans l'orange (Chisholm et al., 1988). Cette cyanobactérie très proche de la lignée de *Synechococcus* sera plus tard nommée *Prochlorococcus* (Chisholm et al., 1992).

Les cellules des cyanobactéries ont deux membranes : une membrane externe, qui forme la paroi cellulaire, et une membrane interne, la membrane cytoplasmique, qui sépare le cytoplasme du périplasme. Les deux membranes sont séparées par une couche de peptidoglycanes. Les cyanobactéries sont donc similaires aux bactéries de type Gram négatif, mais avec une couche de peptidoglycane membranaire plus épaisse, proche de celle que l'on trouve chez les bactéries de type Gram positif (Hoiczyk and Hansel, 2000). Les réactions lumineuses de la photosynthèse oxygénique ont lieu dans les thylakoïdes qui forment un réseau de membranes intracellulaires. Chez *Synechococcus*, ces membranes sont disposées en cercles concentriques très espacés alors que chez *Prochlorococcus*, elles sont plus compactes. Cette différence de disposition est dû au fait que l'antenne photosynthétique de *Synechococcus* est constituée est un macro-complexe (phycobilisome) alors que *Prochlorococcus* récolte la lumière grâce à des protéines intra-membranaires, nécessitant beaucoup moins d'espacement des membranes thylakoïdales.

I.2.2 L'appareil photosynthétique

I.2.2.1 Photosynthèse oxygénique

La photosynthèse permet aux organismes phototrophes d'utiliser l'énergie lumineuse afin de fixer le carbone inorganique utilisé dans la synthèse de matière organique. Chez les cyanobactéries, la photosynthèse est dite oxygénique, au cours de laquelle le donneur d'électron est une molécule d'eau. La première phase de ce processus, dite phase claire, rassemble les réactions dépendantes de la lumière qui ont toutes lieu au sein des thylakoïdes, et implique 4 complexes macromoléculaires : les photosystèmes (PS)I et II, le cytochrome b_6/f et l'ATP synthase (Pessarakli, 2016). L'énergie lumineuse, d'abord absorbée par l'antenne collectrice, est ensuite transférée sous forme d'exciton à une paire spéciale de molécules de chlorophylle (Chl) *a* (P680) dans le centre réactionnel du PSII, composé d'un dimère de protéines D1/D2. Ces molécules de Chl *a* excitées (P680*) vont passer à l'état oxydé ($P680^+$) en dissipant l'énergie d'excitation par le transfert d'un électron à une molécule de phéophytine ($Pheo_{D1}$). Cet électron sera transféré successivement à deux plastoquinones (Q_A puis Q_B). Après un second turnover photochimique, la plastoquinone Q_B, totalement réduite en plastoquinol, capture deux protons du cytoplasme (PQH_2). Les molécules de Chl *a* oxydées ($P680^+$) vont récupérer un électron du complexe d'évolution de l'oxygène résultant de la dissociation de l'eau. L'oxydation successive de ce complexe permet l'oxydation de deux molécules d'eau et de relarguer une molécule de dioxygène et quatre protons dans le lumen du thylakoïde. Le cytochrome b_6/f joue un rôle clé dans la médiation du transfert d'électrons entre les deux photosystèmes et dans l'augmentation du nombre de protons pompés à travers la membrane vers le lumen. En effet, les réactions au sein du cytochrome b_6/f permettent le transfert d'électrons entre la plastoquinol réduite (PQH_2) provenant du PSII et le transporteur d'électrons (cytochrome c_6). Au cours de cette étape, deux protons passent du stroma vers le lumen du thylakoïde pour chaque électron transféré du PSII au PSI. La génération de ce gradient de protons servira plus tard à fournir de l'énergie à l'ATP synthase afin de phosphoryler l'ADP en ATP. Le PSI va ensuite catalyser le transfert d'un électron induit par la lumière entre le donneur d'électrons (cytochrome c_6) dans le lumen du thylakoïde et une molécule acceptrice (ferrédoxine) dans le cytoplasme. Le PSI va collecter la lumière via des pigments (molécules de Chl *a* et β-carotène) et transférer cette énergie vers une paire spéciale de molécules de Chl *a* (P700). Les molécules de Chl *a* excitées (P700*) vont ainsi émettre des électrons qui seront transférés successivement par des transporteurs vers le dernier accepteur $NADP^+$, par l'intermédiaire de la flavoprotéine soluble ferrédoxine-$NADP^+$ réductase (FNR), ce qui conduira finalement à la production de NADPH (Grotjohann and Fromme, 2005; Pessarakli, 2016).

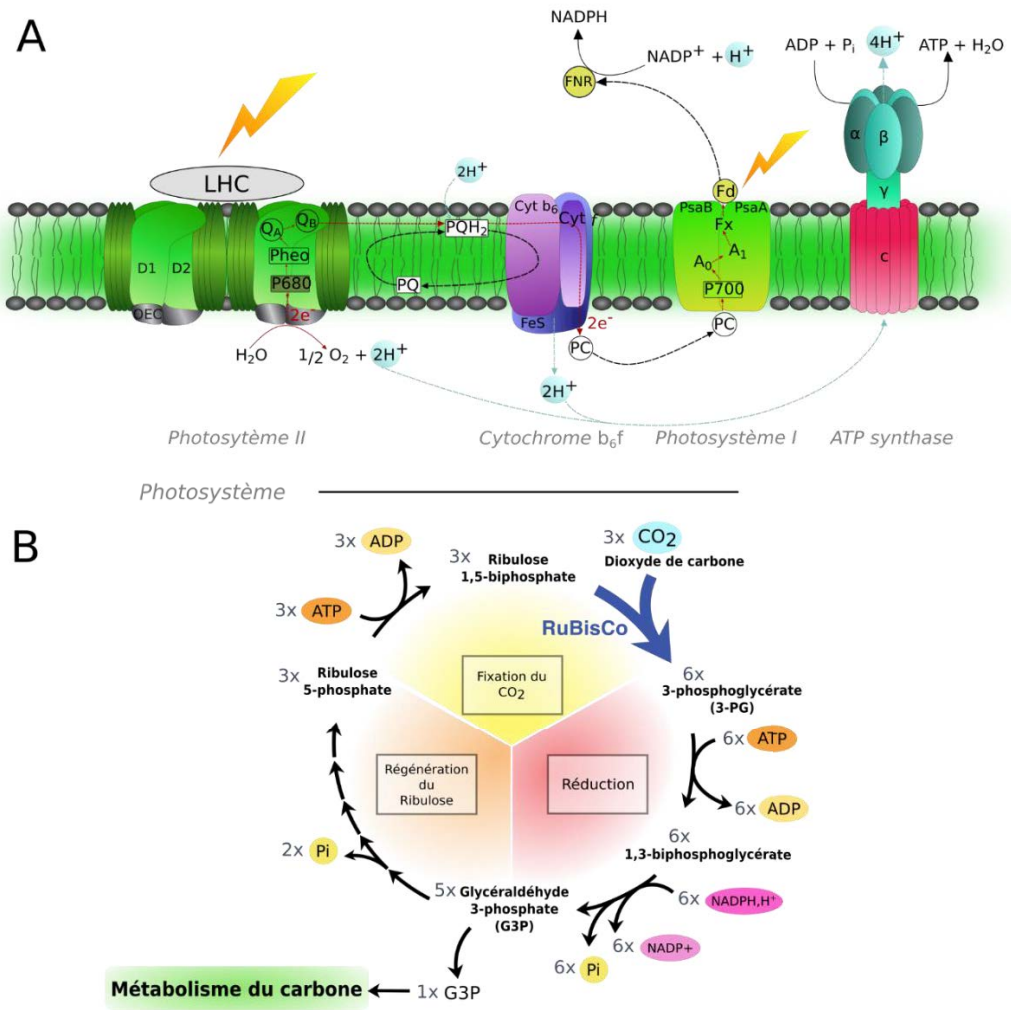


Figure 2 - Réactions de la photosynthèse. A. Complexes impliqués dans la phase claire de la photosynthèse. L'énergie lumineuse captée par l'antenne collectrice (LHC : Light Harvesting Complex) est transférée au photosystème II et déclenche une chaîne de transfert d'électrons responsable ensuite d'un gradient de protons transmembranaire, de la dissociation de l'eau et de la production de NADPH. Le gradient de protons est utilisé par l'ATP synthase pour générer de l'ATP. Schéma réalisé par Justine Pittera (Pittera, 2015). B. Phase sombre de la photosynthèse ou cycle de Benson-Calvin. Au cours de cette phase, le CO_2 est fixé en utilisant l'ATP et le NADPH produit durant la phase claire (PQ : plastoquinone, PC : plastocyanine).

Durant la deuxième phase de la photosynthèse, ou phase sombre, l'ATP et le pouvoir réducteur du NADPH précédemment produits sont utilisés dans le cycle de Benson-Calvin pour fixer le CO_2 afin de synthétiser des carbohydrates. Ce cycle est composé de trois grandes phases. La première, la carboxylation, catalysée par l'enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO), ajoute un groupement carboxyl à une molécule de ribulose-1,5-bisphosphate (RuBP), puis clive ce produit pour former deux molécules de 3-phosphoglycerate (PGA). Vient ensuite l'étape de réduction, au cours de laquelle la PGA est réduite en glyceraldehyde-3-phosphate (GA3P) par l'action consécutive de la phosphoglycerate kinase (PGK) et de la GA3P dehydrogenase (GAPDH) en utilisant l'ATP et le NADPH générés plus

tôt. Cette GA3P est isomérisée avant la dernière étape du cycle, au cours de laquelle elle sera utilisée pour regénérer l'accepteur primaire de CO_2, la RuBP, ou servir de substrat à la synthèse de carbohydrates qui sont les précurseurs de la plupart des molécules organiques (ADN, ARN, protéines, etc…).

I.2.2.2 Diversité pigmentaire

Synechococcus est pourvu, à l'instar de la plupart des cyanobactéries, d'une antenne collectrice de lumière, le phycobilisome (Figure 3A). Ce complexe multiprotéique est ancré dans la membrane thylakoïdale par un cœur d'allophycocyanine (APC) entouré de six bras, composés de phycobiliprotéines (PBP) reliées entre elles par des protéines de liaisons (linkers). Il existe quatre PBPs différentes chez *Synechococcus*, chacune capable de lier des chromophores (phycobilines) différents et donc d'absorber des longueurs d'ondes spécifiques (Figure 4 ; Sidler, 1994). La composition en PBP des bras des phycobilisomes a permis de définir trois principaux types pigmentaires : les bras des souches de type 1 sont composés uniquement de phycocyanine (PC), le type 2 possède des bras composés de PC et de phycoérythrine I (PEI) et le type 3 a des bras constitués de PC, de PEI et de phycoérythrine II (PEII). Ces compositions distinctes ont une incidence sur les chromophores fixés par chaque type pigmentaire : le type 1 ne possède que de la phycocyanobiline (PCB) et absorbe préférentiellement la lumière rouge ; le type 2, qui contient à la fois de la PCB et de la phycoérythrobiline (PEB), absorbe la lumière verte alors que chez le type 3, on retrouve de la PCB, de la PEB et de la phycourobiline (PUB), cette dernière absorbant la lumière bleue. Pour chacun de ces types pigmentaires, le cœur d'allophycocyanine du phycobilisome lie de la PCB (Six et al., 2007a). Le type pigmentaire 3 a pu être divisé en sous-types sur la base du rapport des chromophores PUB et PEB. Ce rapport est faible chez les souches de type 3a qui absorbent préférentiellement la lumière verte, intermédiaire chez le type 3b et haut pour le type 3c qui absorbent tous deux préférentiellement de la lumière bleue (Humily et al., 2013; Six et al., 2007a). Les souches du type 3d sont capables de modifier leur rapport PUB:PEB en fonction de la couleur de la lumière, ainsi leur phénotype sera de type 3a sous une lumière verte et de type 3c sous une lumière bleue (Everroad et al., 2006; Humily et al., 2013; Palenik, 2001). A noter qu'il existe deux types d'acclimatateurs chromatiques (3dA et 3dB) qui sont phénotypiquement équivalent mais diffèrent d'un point de vue génétique par le contenu en gènes et la localisation dans les génomes d'un ilot génomique spécifique de l'acclimatation chromatique (Humily et al., 2013).

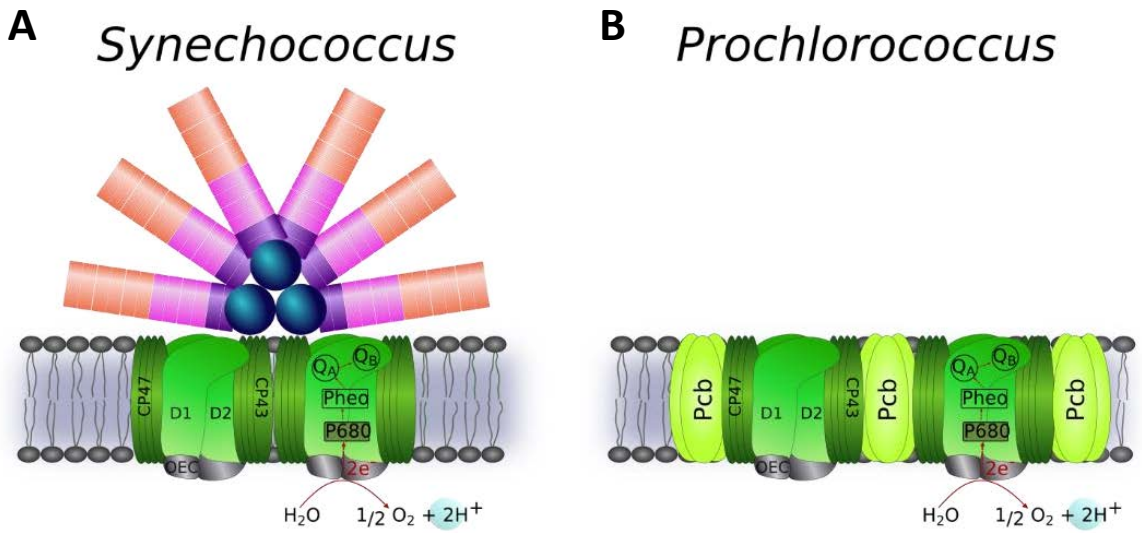


Figure 3 - Antennes collectrices de lumière de *Synechococcus* et *Prochlorococcus*. A. *Synechococcus* collecte la lumière grâce à un complexe multiprotéique, le phycobilisome. B. *Prochlorococcus* possède une antenne composée de protéines intra-membranaires (Pcb) liant des molécules de DV-Chl *a* et de DV-Chl *b*. D'après Pittera, 2015.

Les études de la distribution *in situ* des types pigmentaires, utilisant le rapport PUB:PEB mesuré par des techniques optiques, telles que la spectrofluorimétrie ou la cytométrie en flux (Lantoine and Neveux, 1997; Olson and Chisholm, 1990), ainsi que l'analyse de marqueurs moléculaires, tels que *cpcBA* ou *cpeBA,* codant respectivement pour la PC et la PEI (Humily et al., 2014; Xia et al., 2017, 2018), ont mis en évidence que le type 1 domine dans les eaux de surface à faible teneur en sel, le type 2 est retrouvé dans les eaux côtières ou dans des zones intermédiaires entre les milieux saumâtres et océaniques, alors que le type 3 est retrouvé dans les eaux mésotrophes, où la lumière est plutôt verte, ou les eaux oligotrophes, où la lumière bleue pénètre plus en profondeur. Néanmoins, ces études basées sur la fluorescence ou un seul marqueur moléculaire ne permettent pas de différencier les sous-types pigmentaires de *Synechococcus* au sein du type 3. Récemment, la combinaison de trois gènes marqueurs, permettant de discriminer l'ensemble des types et sous-types pigmentaires connus, a permis de révéler que le type 3a domine dans les régions chaudes (Mer Rouge, golfe du Mexique, nord de l'océan Indien), le 3c est retrouvé en Méditerranée et dans les zones enrichies en fer de l'océan Pacifique, alors que le type 3d est très répandu et domine à hautes latitudes. Les types 3dA et 3dB présentent des distributions complémentaires : le type 3dA est le plus abondant à hautes latitudes alors que le type 3dB domine en profondeur dans les eaux chaudes. Les acclimateurs chromatiques représentent à eux seuls 40% de la population totale de *Synechococcus* (Grébert et al., 2018).

Prochlorococcus ne possède quant à lui pas de phycobilisome mais capte l'énergie lumineuse grâce à des protéines intramembranaires collectrices de lumière (Pcb), liant des molécules dérivées de la chlorophylle, les divinyl-Chl *a* et *b* (Figure 3B). Sa diversité pigmentaire est beaucoup moins importante que celle de *Synechococcus*, mais deux principaux écotypes ont été mis en évidence : l'un en colonisant les eaux superficielles dont le rapport Chl *b*/*a* est faible (appelé HL pour « high light ») et l'autre, plus en profondeur, dont le rapport Chl *b*/*a* est plus fort (appelé LL pour « low light ») et qui absorbe mieux la lumière bleue qui prédomine en profondeur que l'écotype de surface (Moore et al., 1995; Partensky et al., 1993). Ces deux écotypes diffèrent également par leur optima lumineux de croissance (Moore et al., 1998).

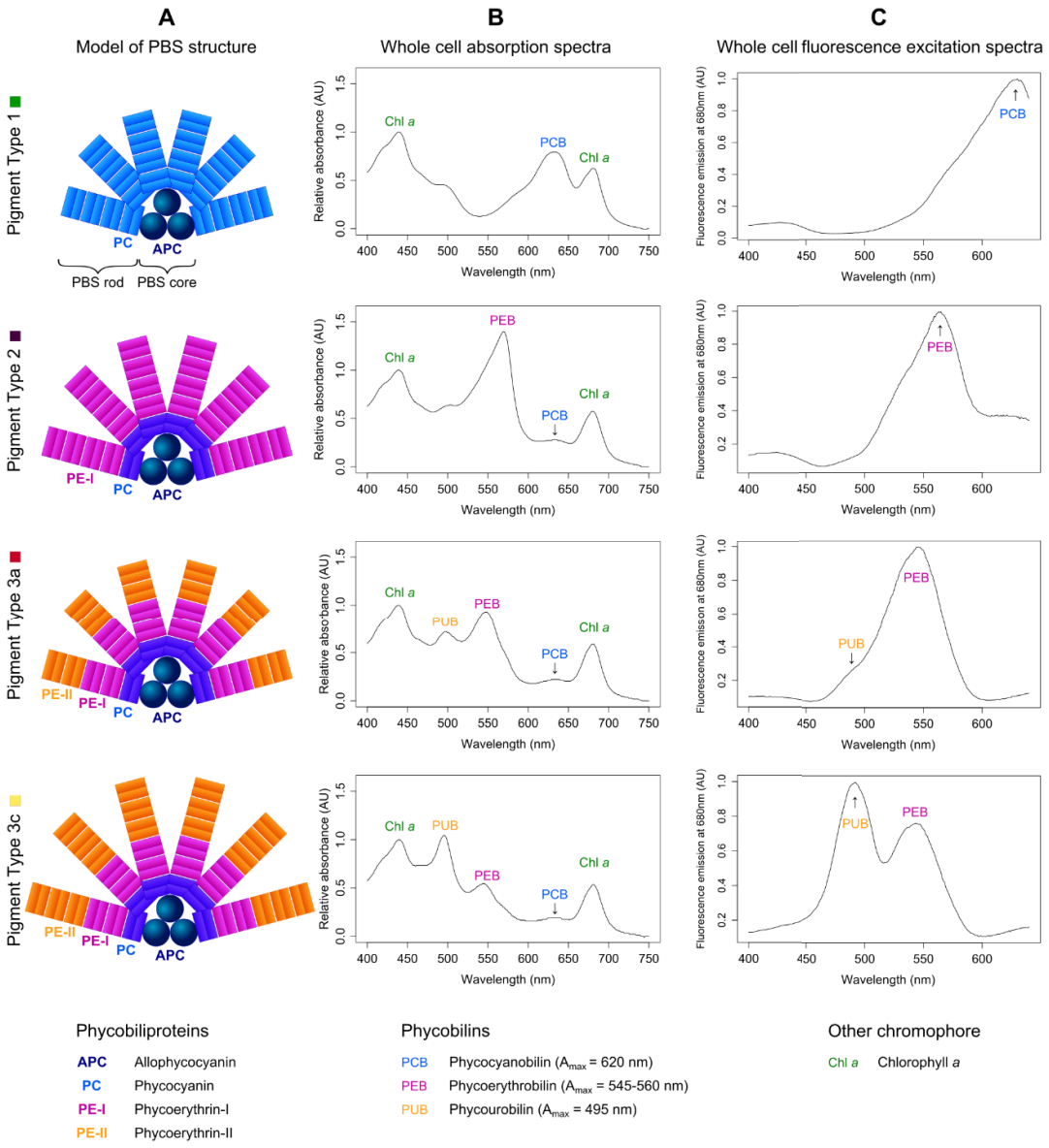


Figure 4 - Composition du phycobilisome et propriétés d'absorption des différents types pigmentaires. A. Structure des phycobilisomes. B. Spectres d'absorption des différents types pigmentaires. C. Spectres d'excitation à 680 nm. Figure extraite de Grébert et al., 2018.

I.2.3 Importance écologique des picocyanobactéries et distribution

Les picocyanobactéries marines sont les organismes photosynthétiques les plus abondants de l'océan global et représentent 10% des organismes phytoplanctoniques des 200 premiers mètres de la colonne d'eau, avec une abondance globale estimée de $2{,}9 \times 10^{27}$ et $7{,}0 \times 10^{26}$ cellules pour *Prochlorococcus* et *Synechococcus,* respectivement (Flombaum et al., 2013; Whitman et al., 1998).

Le genre *Synechococcus* possède une très large distribution géographique (Figure 5B) puisqu'on le retrouve dans la majeure partie des régions océaniques à l'exception des régions polaires. Il est particulièrement abondant dans les eaux riches en nutriments tels que les zones côtières, les eaux mélangées ou les *upwellings* avec des concentrations pouvant atteindre entre 1,5 et 3,7 x 10^6 cellules/mL (Partensky et al., 1999a; Saito et al., 2005). On le retrouve à de plus faibles concentrations (1-3 × 10^3) dans les zones très oligotrophes tels que les gyres océaniques (Partensky et al., 1999a; Zwirglmaier et al., 2007).

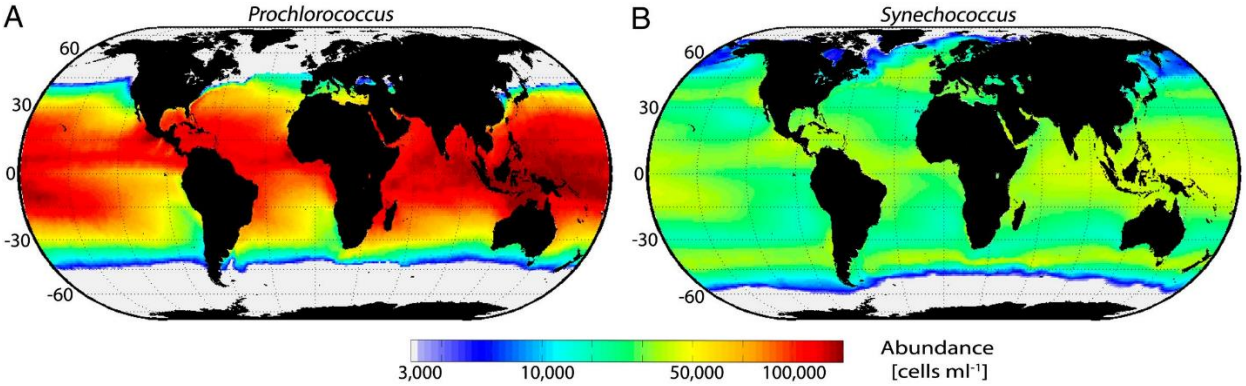


Figure 5 - Distribution dans l'océan mondial de *Prochlorococcus* et *Synechococcus*. Figure issue de (Flombaum et al., 2013).

L'aire de distribution de *Prochlorococcus* est plus restreinte (Figure 5A), puisqu'il est retrouvé principalement entre 45°N et 40°S (Johnson et al., 2006; Olson et al., 1990; Partensky et al., 1999a; Zwirglmaier et al., 2007), dans des eaux tempérées et chaudes (supérieures à 10°C), stratifiées et oligotrophes (Campbell and Vaulot, 1993; Flombaum et al., 2013; Johnson et al., 2006; Partensky et al., 1999b). De plus, *Prochlorococcus* est numériquement plus abondant que *Synechococcus* dans la plupart des régions où les deux genres coexistent, en particulier dans les zones oligotrophes. *Synechococcus* est plus abondant en surface qu'en profondeur et n'est pas retrouvé en dessous de 150 m au contraire de *Prochlorococcus* qui peut croître jusqu'à 200-250 m, à la limite de la zone euphotique (Partensky et al., 1999a). De par leur abondance dans les océans, *Prochlorococcus* et *Synechococcus* seraient responsables de 8,5% (4 Gt de carbone fixé)

et 16,7% (8 Gt de carbone fixé) de la production primaire net océanique, respectivement (Flombaum et al., 2013).

Ces organismes, largement répandus dans les zones océaniques, sont donc capables de coloniser une très large gamme de niches écologiques aux caractéristiques bien distinctes, ce qui laisse penser que ces genres possèdent de remarquables capacités d'adaptation et d'acclimatation aux variations des conditions environnementales, ce qui se traduit par une très grande diversité génétique et fonctionnelle.

II. Diversité génétique et biogéographie des picocyanobactéries marines

II.1 Diversité génétique des picocyanobactéries marines

Synechococcus est un genre polyphylétique contenant des souches marines et d'eau douce. Les *Synechococcus* marins ont d'abord été classés en trois clusters (MC-A, MC-B et MC-C) sur la base de leur composition pigmentaire, leur tolérance vis-à-vis de la salinité et leur GC% (Waterbury and Rippka, 1989). Plus tard, les clusters MC-A et B ont été rassemblés dans un même groupe : le cluster 5 qui inclut les *Synechococcus* marins et les *Cyanobium* (Herdman et al., 2001). Ce groupe a été subdivisé en 3 sous-clusters (SC): le SC 5.1, qui domine les régions océaniques, semble être le plus diversifié et regroupe la majorité des souches décrites ; le SC 5.2 regroupe des membres halotolérants colonisant les zones estuariennes (Cai et al., 2010; Chen et al., 2006) ou des mers peu salées comme la Baltique (Haverkamp et al., 2008); enfin le SC 5.3, encore peu connu, comporte des membres marins, limités à quelques zones spécifiques de l'océan (Dufresne et al., 2008; Farrant et al., 2016) mais également des membres lacustres qui n'ont été que récemment découverts (Cabello-Yeves et al., 2017). Alors qu'une dizaine de clades ont été définis au sein du SC 5.1 sur la base de l'ARNr 16S (Fuller et al., 2003; Moore et al., 1998), l'utilisation de nouveaux gènes marqueurs tels que la région intergénique entre les ARNr 16S et 23S (ITS ; Ahlgren and Rocap, 2006; Rocap et al., 2002), le gène *rpoC1*, codant pour la sous-unité β de l'ARN polymerase (Ferris and Palenik, 1998; Mühling et al., 2006; Palenik, 1994; Toledo and Palenik, 1997), *ntcA*, le régulateur global du métabolisme de l'azote (Penno et al., 2006) et *petB,* codant pour le cytochrome b_6 (Mazard et al., 2012a), ont permis d'augmenter la résolution taxonomique au sein de ce groupe. Par exemple, le gène *petB* permet de distinguer plus de 40 sous-clades au sein des clades du SC 5.1 (Mazard et al., 2012a).

Prochlorococcus constitue une branche monophylétique au sein de la radiation des *Synechococcus* marins chez laquelle on distingue une branche monophylétique correspondant au phototype HL alors que le phototype LL constituent un groupe polyphylétique à la base des HL (Figure 6 ; Moore et al., 1998). Par la suite, le phototype de surface a été divisé en six clades distincts (HLI-VI ; Biller et al., 2015; Huang et al., 2012; Malmstrom et al., 2013; West and Scanlan, 1999) et celui de profondeur en 5 clades (LLI, LLII/III, LLIV, LLV et LLVI; Biller et al., 2015; Huang et al., 2012; Johnson et al., 2006; Lavin et al., 2010; Malmstrom et al., 2013; Martiny et al., 2009a).

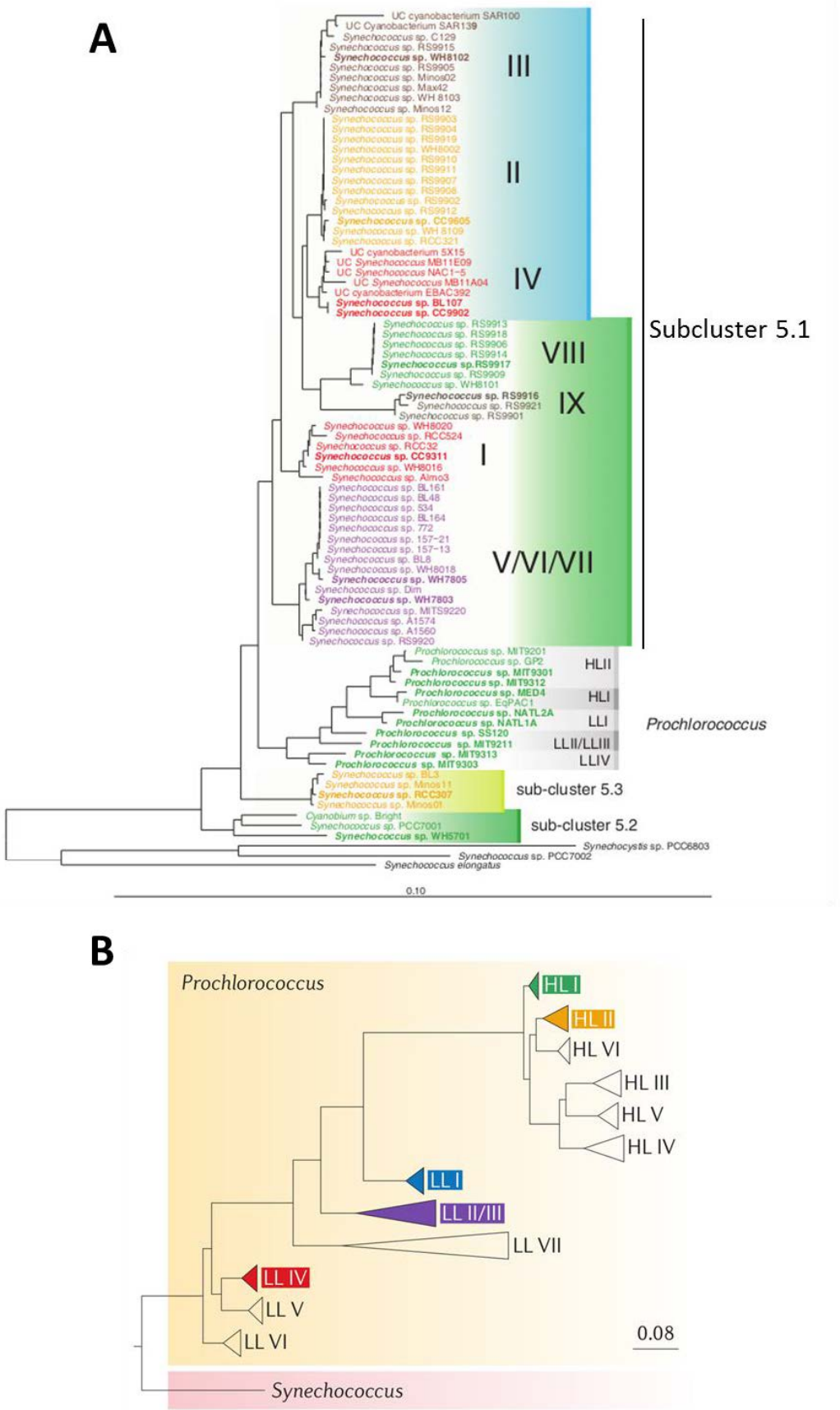


Figure 6 - Arbres phylogénétiques des picocyanobactéries marines. A. Arbre construit en Neighbor-Joining sur la base de l'ARNr 16S, issu de Scanlan, 2012. B. Arbre phylogénétique des clades de *Prochlorococcus* construit en utilisant le marqueur ITS, issu de Biller et al., 2015**.**

II.2 Répartition biogéographique et microdiversité des picocyanobactéries marines

En plus d'une distribution verticale bien spécifique des deux phototypes de *Prochlorococcus,* reflétant les différences de préférenda lumineux de croissance déterminées sur les souches en culture (Moore and Chisholm, 1999; Moore et al., 1998), les études biogéographiques ont permis de mettre en évidence des différences de distribution spatiale des clades de *Prochlorococcus*. Le clade HLI est particulièrement abondant à hautes latitudes dans les eaux tempérées froides, riches et mélangées alors que le clade HLII domine à basses latitudes dans les eaux chaudes et stratifiées (Chandler et al., 2016; Farrant et al., 2016; Johnson et al., 2006; Zinser et al., 2007; Zwirglmaier et al., 2007, 2008). En accord avec cette distribution *in situ*, des différences d'optima de température et de limites de croissance ont pu être observées en laboratoire entre des souches représentatives de ces deux clades (Johnson et al., 2006). Les clades HLIII et HLIV quant à eux dominent, tout comme le clade HLII, dans les eaux chaudes mais sont restreints aux zones dites HNLC ('high-nutrient, low-chlorophyll' ; Behrenfeld and Kolber, 1999; Fung et al., 2000; Martin et al., 1994), zones riches en nutriments mais où la concentration en fer est très faible, ce qui a pour effet de limiter la croissance et le développement des organismes phytoplanctoniques (Farrant et al., 2016; Huang et al., 2012; Kent et al., 2016; Malmstrom et al., 2013; Rusch et al., 2010; West et al., 2011). Les clades HLV et VI n'ont été retrouvés dans l'environnement qu'à faible concentration, ce qui fait que l'on connait encore mal leur distribution, si ce n'est que le clade HLV cohabite avec les clades HLIII-IV dans les régions HNLC (West et al., 2011), alors que le clade HLVI semble être présent plus en profondeur par rapport aux autres populations HL (Huang et al., 2012).

La distribution géographique des clades définis au sein du phototype LL est moins bien définie que celle des clades HL. Le clade LLI est retrouvé à une profondeur intermédiaire par rapport aux clades HL et aux autres clades LL dans les eaux tropicales stratifiées, alors qu'on le retrouve jusqu'en surface dans les eaux mélangées à plus hautes latitudes (Johnson et al., 2006; Zinser et al., 2006). Au contraire du clade LLI, les clades LLII/III et LLIV sont limités à la base de la zone euphotique, en accord avec leur sensibilité à l'exposition aux fortes intensités lumineuses (Malmstrom et al., 2010; Zinser et al., 2006). Enfin, les clades LLV et VI ont été détectés uniquement dans les zones dites à minimum d'oxygène (ou OMZ pour 'oxygen minimum zone'), retrouvées principalement dans le Pacifique tropical Est et la mer d'Arabie (Lavin et al., 2010).

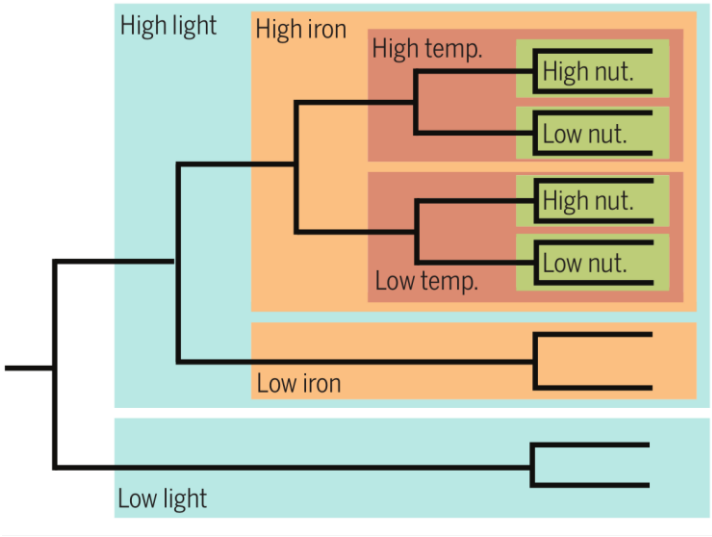


Figure 7 - Relation entre la phylogénie *Prochlorococcus* et l'adaptation aux paramètres environnementaux. Les traits impliqués dans l'adaptation à la lumière, la carence en fer, la température, et la concentration en macronutriments sont liés à la phylogénie du genre d'une manière hiérarchique. Figure issue de Martiny et al., 2015.

À l'échelle des clades, il semble donc que la distribution des membres du genre *Prochlorococcus* soit principalement influencée par quelques paramètres physico-chimiques : l'intensité lumineuse, la température et la disponibilité en fer (Figure 7). Plus récemment, une microdiversité intra-clade a également été mise en évidence en examinant les patrons de distribution de sous-groupes phylogénétiques au sein des différents clades de *Prochlorococcus* (Larkin et al., 2016; Martiny et al., 2009a). Dans ce contexte, Farrant et al. (2016) ont délimité des ESTUs (Ecologically Significant Taxonomic Units), c'est-à-dire un ensemble d'unités taxonomiques opérationnelles (OTU) appartenant au même clade et présentant des aires de distribution géographique semblables. Le clustering des stations de *Tara* Oceans en fonction de leur composition en ESTUs et la corrélation de l'abondance relative de ces ESTUs avec les paramètres abiotiques ont permis d'identifier trois assemblages majeurs d'ESTUs dans les eaux de surface et de mettre en évidence les caractéristiques des niches occupées par ces différents assemblages (Figure 8). Ainsi, l'assemblage 1 est caractéristiques des eaux tempérées et est dominé par l'ESTU HLIA, l'assemblage 2 retrouvé dans les zones chaudes et pauvres en azote est dominé par l'ESTU HLIIA, et enfin l'assemblage 3, caractéristique des zones HNLC, est co-dominé par les ESTUs HLIIIA et HLIVA (Farrant et al., 2016). En dehors de ces ESTUs dominantes, d'autres présentes en relativement faible abondance se sont révélées intéressantes car leur écologie diffère de celle du clade. En effet, alors que l'ESTU HLIA, qui domine le clade HLI, colonise les eaux tempérées froides, l'ESTU HLIC semble avoir une tolérance plus importante vis-à-vis de la température puisqu'elle est également présente dans les eaux chaudes dominées par l'ESTU HLIIA. De même, l'abondance relative maximale de l'ESTU HLIIC a été observée au maximum profond de

chlorophylle (DCM ; Deep Chlorophyll Maximum) dans l'Océan Indien équatorial, ce qui suggère que cette ESTU pourrait être adaptée à des profondeurs intermédiaires, tout comme l'ESTU dominante du clade LLI (LLIA). Enfin, au sein de ce dernier clade, une seconde ESTU (LLIB) présente la particularité de dominer le clade LLI dans les eaux de surface des zones HNLC de l'Océan Indien et de l'Océan Pacifique, ce qui montre que l'adaptation à la carence en fer n'est pas spécifique des clades HLIII et HLIV de *Prochlorococcus* (Farrant et al., 2016).

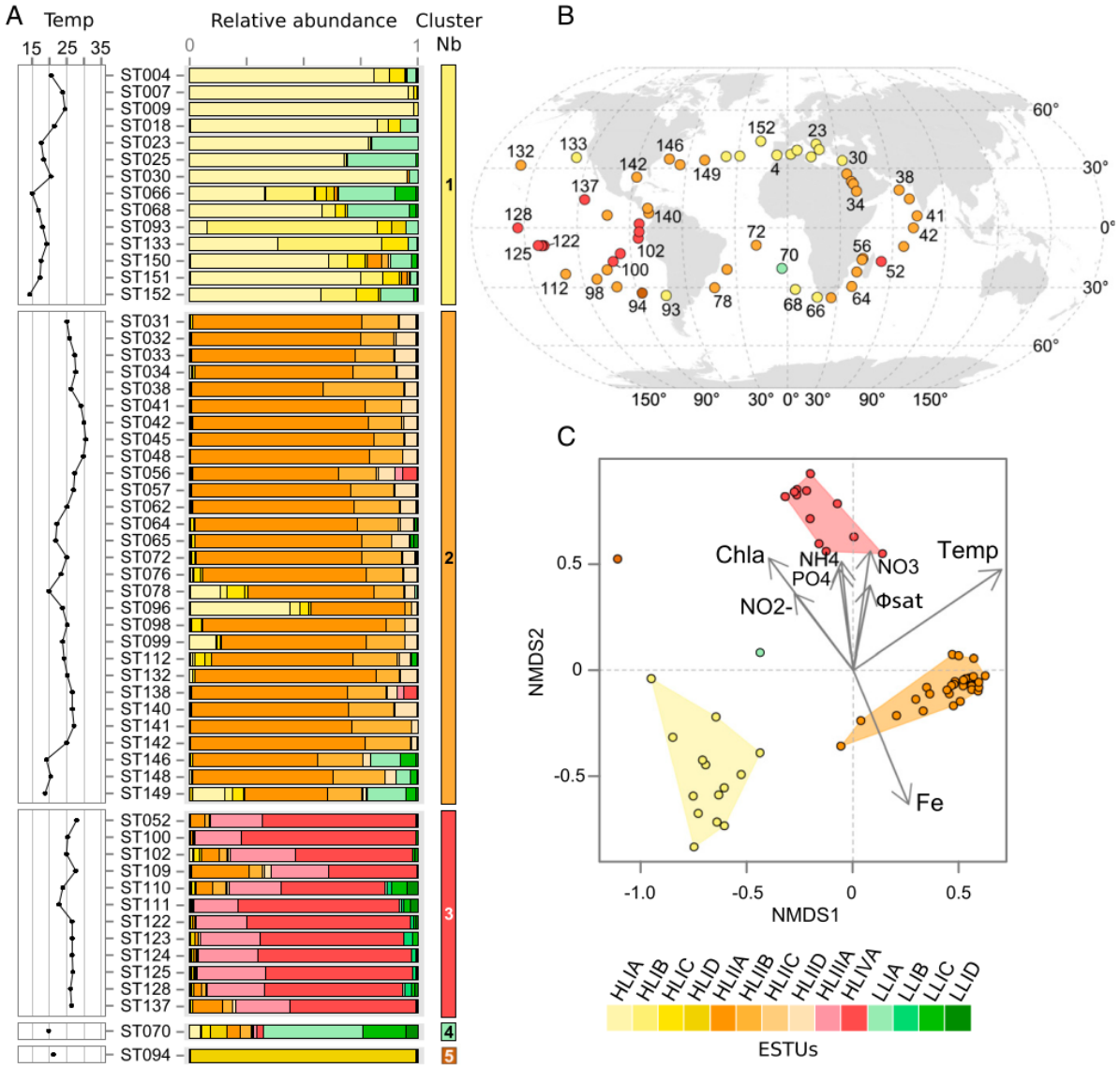


Figure 8 - Biogéographie des ESTU de Prochlorococcus à partir des métagénomes de surface de *Tara* Oceans et leur relation avec les paramètres physico-chimiques. A. Abondances relatives des ESTU de *Prochlorococcus* à chaque station regroupée en cluster B. Distribution des assemblages d'ESTU de *Prochlorococcus* colorées par cluster C. NMDS des paramètres physico-chimiques. Les stations ont été colorée en fonction de leur cluster d'assemblages. NMDS stress value: 0.0985. Figure extraite de Farrant et al., 2016.

Chez *Synechococcus*, des études ont également montré une répartition spécifique des différents clades. Au sein du SC 5.1, les clades I et IV co-dominent dans les zones côtières et océaniques à hautes latitudes, caractérisées par des eaux relativement froides et riches en sels nutritifs (Farrant et al., 2016; Sohm et al., 2015; Tai and Palenik, 2009; Zwirglmaier et al., 2007, 2008). Le clade II domine dans les eaux oligotrophes et mésotrophes tropicales et subtropicales (Farrant et al., 2016; Huang et al., 2012; Sohm et al., 2015; Zwirglmaier et al., 2007, 2008). Le clade III est retrouvé, en cooccurrence avec le SC 5.3, dans les eaux oligotrophes, pauvres en phosphate, telles que la mer Méditerranée et le golfe du Mexique (Farrant et al., 2016; Mella-Flores et al., 2011). Enfin, le clade CRD1 est le clade dominant dans des régions pauvres en fer de l'océan, et notamment une grande partie du Pacifique sud (Farrant et al., 2016; Sohm et al., 2015). A part ces cinq clades dominants, la distribution des autres clades de *Synechococcus* reste peu ou mal connue, à l'exception du clade VIII que l'on retrouve essentiellement dans les estuaires et les eaux euryhalines (Hunter-Cevera et al., 2016; Xia et al., 2017).

La délimitation d'ESTU au sein de ce genre a permis, à l'instar de *Prochlorococcus*, de mettre en évidence une importante microdiversité au sein de certains clades, jusqu'alors considérés comme des écotypes. En effet, deux ESTUs ont été définis au sein du clade II : alors que l'ESTU IIA, qui est l'ESTU dominant du clade II, est retrouvé comme celui-ci dans les eaux chaudes oligotrophes ou mésotrophes, l'ESTU IIB colonise spécifiquement les eaux froides et mélangées. Dans le clade CRD1, trois ESTUs ont été définis, tous présents dans les zones carencées en fer mais colonisant des niches thermiques distinctes: CRD1B est spécifique des eaux froides, CRD1C des eaux chaudes tandis que les CRD1A sont présents dans les deux types de niches, suggérant qu'il seraient capables de s'adapter à une plus large gamme de température (Farrant et al., 2016). Le clustering des stations, basé sur leur composition en ESTUs, a mis en évidence 8 assemblages distincts d'ESTUs le long du transect de *Tara* Oceans (Figure 9A). Ces assemblages ne semblent pas spécifiques d'une région océanique donnée, même si on peut noter que le cluster 2, dominé par l'ESTU IIIA qui coexiste avec les ESTU 5.3A et WPC1A, est principalement retrouvé en mer Méditerranée et anti-corrélé avec la concentration en phosphate. L'assemblage 1, dominé par l'ESTU IIA est retrouvé dans les eaux chaudes, non limitées en fer, mésotrophes ou oligotrophes (Figure 9). Les assemblages 4, 5 et 8, dominés par les ESTUs IA, IVA et IVB, sont retrouvés dans les eaux océaniques mélangées, les régions côtières et les eaux tempérées froides à hautes latitude, ce qui correspond aux paramètres de niche précédemment déterminés pour l'ensemble des clades I et IV. Globalement, le partitionnement des niches occupées par les ESTUs de *Synechococcus* les plus abondants dans l'environnement

semble contrôlé par trois facteurs principaux : la température, la disponibilité en fer et en phosphate (Farrant et al., 2016; Sohm et al., 2015; Zwirglmaier et al., 2008).

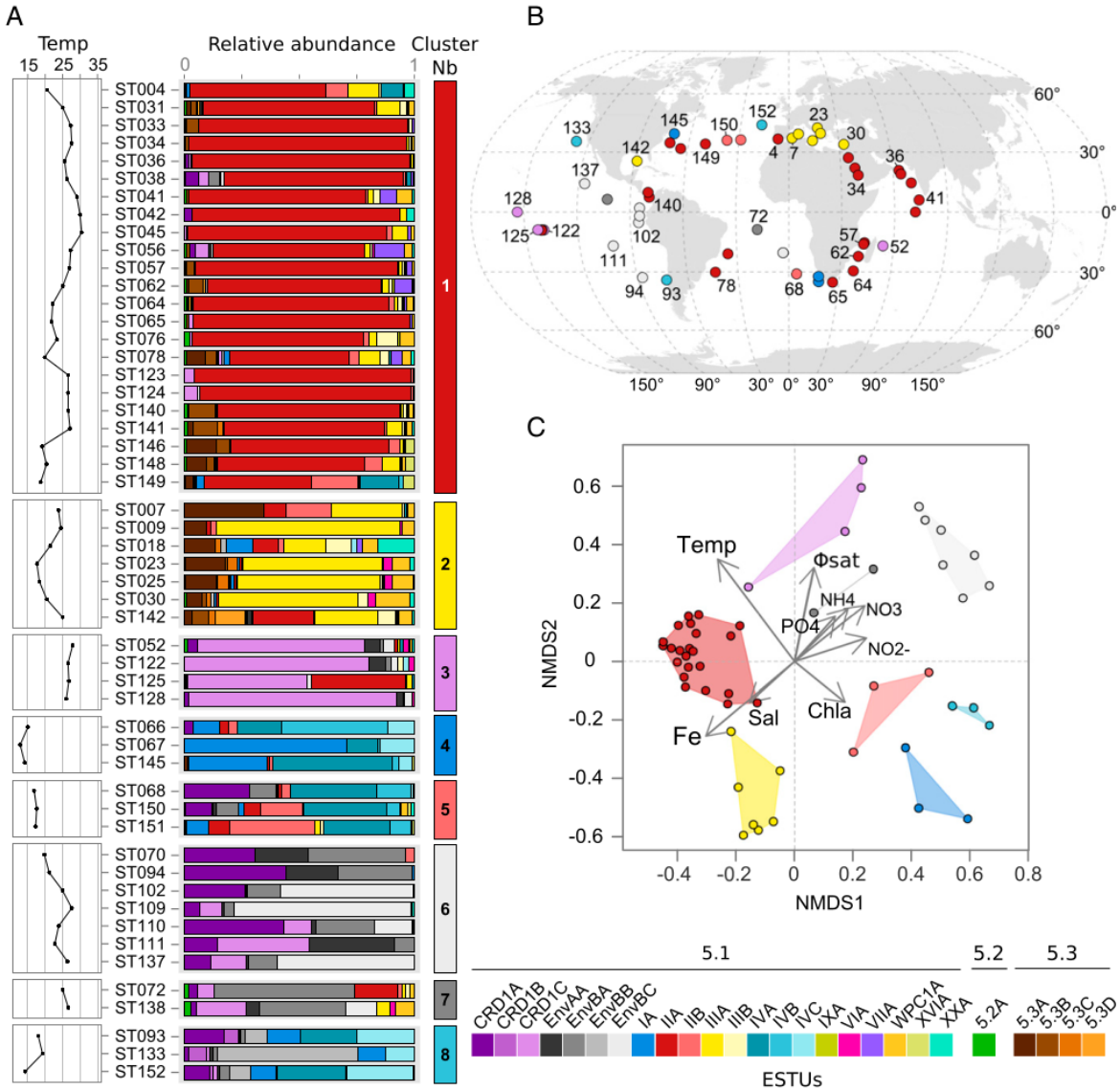


Figure 9 - Biogéographie des ESTUs de *Synechococcus* à partir des métagénomes de surface de *Tara* Oceans et leur relation avec les paramètres physico-chimiques. A. Abondances relatives des ESTUs de *Synechococcus* à chaque station, regroupées en assemblages (clusters). B. Distribution des assemblages d'ESTUs de *Synechococcus* le long du transect *Tara* Oceans. C. Analyse NMDS des paramètres physico-chimiques. Les stations ont été colorées en fonction de leur appartenance à un assemblage. NMDS stress value: 0.1369. Figure extraite de Farrant et al., 2016.

III. Réponses des cyanobactéries aux variations des conditions environnementales

III.1 Bases génétiques de l'adaptation chez les picocyanobactéries marines

III.1.1 Caractéristiques générales des génomes de picocyanobactéries

Du fait de leur petite taille et de leur intérêt écologique, le séquençage des quatre premiers génomes de picocyanobactéries a démarré dès les années 2000 (Dufresne et al., 2003; Palenik et al., 2003; Rocap et al., 2003) et leur nombre n'a cessé de croitre depuis, avec en particulier trois études comparatives majeures : l'une centrée sur *Prochlorococcus* (Kettler et al., 2007) et les deux autres sur *Synechococcus* (Dufresne et al., 2008; Scanlan et al., 2009). Ces études ont permis de montrer que les génomes de picocyanobactéries marines sont particulièrement petits par rapport aux génomes de cyanobactéries d'eau douce ou à ceux des bactéries en général : 1,62 à 2,7 Mb pour *Prochlorococcus* et 2,11 à 3,34 Mb pour *Synechococcus* ; Kettler et al., 2007; Scanlan et al., 2009; Shih et al., 2013). La petite taille de la plupart des génomes de *Prochlorococcus* (< 2 Mb, à l'exception des membres du clade LLIV) serait dû à un phénomène de réduction du génome qui se serait déroulé au sein de ce genre, de façon concomitante à la diminution du contenu en GC%. Ces phénomènes, qui ont été attribués à la perte de plusieurs gènes codant pour des enzymes de réparation des transversions de G:C en A:T (Dufresne et al., 2005; Partensky and Garczarek, 2010; Rocap et al., 2003), pourraient constituer un avantage adaptatif dans les zones oligotrophes où les nutriments sont rares. En effet, ils permettraient pour les cellules de réduire la quantité de phosphate et d'azote utilisés, notamment pour la synthèse de l'ADN. De plus, un génome réduit permet aux cellules d'être plus petites, ce qui aurait pour conséquences i) de diminuer l'effet de diffusion lumineuse et d'auto-ombrage (Morel et al., 1993), qui constitue un avantage pour les clades LL à la base de la zone euphotique en améliorant l'apport lumineux, mais également ii) d'augmenter leur rapport surface sur volume, ce qui favoriserait l'acquisition des nutriments, une caractéristique particulièrement avantageuse pour les clades HL qui dominent dans les eaux pauvres de surface (Chisholm, 1992; Dufresne et al., 2005). Les génomes des clades HLI et HLII sont généralement plus réduits que ceux des clades LL. Les souches du clade LLIV, situés à la base de la radiation des *Prochlorococcus*, possèdent quant à eux les plus gros génomes du genre *Prochlorococcus*, et sont phylogénétiquement les membres les plus proches du genre *Synechococcus* (Biller et al., 2015).

Une seconde hypothèse complémentaire avec la première, permettant d'expliquer cette réduction génomique, est basée sur la théorie de coévolution dite de la « Reine Noire », qui pose comme postulat que certaines fonctions biologiques sont redondantes au sein de l'ensemble des

organismes d'une communauté, d'une niche ou d'un environnement donné. Ces fonctions sont considérées comme coûteuses à l'échelle de la communauté et leur élimination constitue donc un avantage sélectif. Néanmoins, ces fonctions doivent être maintenues par d'autres membres de la communauté microbienne et fournir, par relargage dans le milieu, les produits de ces fonctions dites de « bien commun » (Ma et al., 2018; Morris et al., 2008, 2011, 2012). Par exemple, bien que tous les génomes de *Prochlorococcus* soient dépourvus de gène codant pour la catalase, il a été démontré que les cellules peuvent assimiler celle de bactéries hétérotrophes coexistant dans le milieu et utiliser leurs propriétés anti-oxydantes (Morris et al., 2011). Cette théorie pourrait donc également expliquer la perte de certains gènes des génomes de *Prochlorococcus*.

III.1.2 Génome commun, accessoire et pan-génome

La comparaison des génomes de picocyanobactéries disponibles a également permis de montrer que ces génomes ne comportent que peu de paralogues, le nombre de gènes variant de 1,716 à 3,022 chez *Prochlorococcus* et de 2,358 à 3,129 chez *Synechococcus* (Scanlan et al., 2009). Les gènes communs, constituant 52 à 67 % du génome total, permettent d'assurer les fonctions de biosynthèse nécessaires à leur mode de vie autotrophe (Dufresne et al., 2008; Kettler et al., 2007; Scanlan et al., 2009), alors que les gènes accessoires, c'est-à-dire retrouvés dans un sous-ensemble de génomes, sont les plus susceptibles de jouer un rôle dans l'adaptation de ces organismes à diverses niches écologiques. Alors que le génome commun de ces organismes commence à être bien connu (environ 1000 gènes pour *Prochlorococcus* et 1500 pour *Synechococcus* ; Biller et al., 2015; Dufresne et al., 2008), la taille de leur pan-génome, c'est-à-dire de la somme des gènes communs et accessoires pour un genre donné, est considérable et beaucoup plus difficile à appréhender puisque chaque nouveau génome séquencé apporte en moyenne 160 nouveaux gènes dans le cas de *Prochlorococcus* (Figure 10 ; Biller et al., 2015). Une projection théorique basée sur 13 génomes de *Prochlorococcus* a prédit que la population mondiale de ce genre contiendrait 57,792 gènes (Baumdicker et al., 2012).

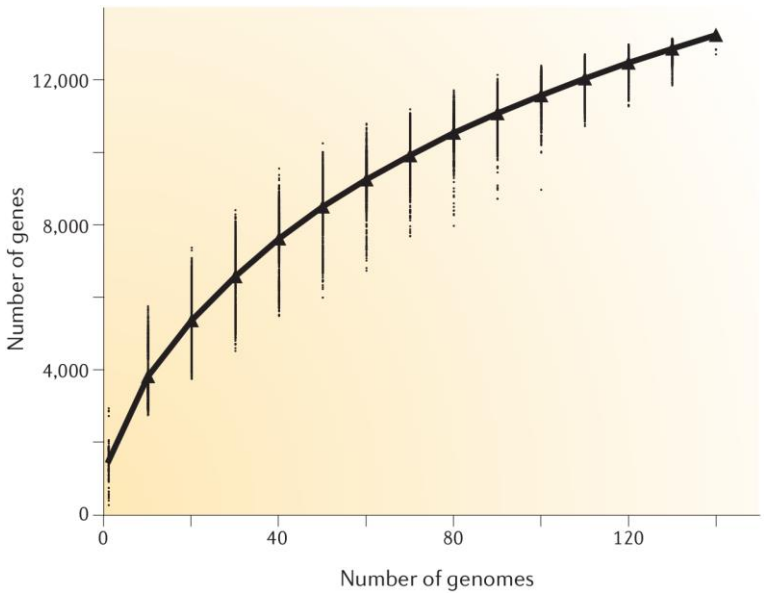


Figure 10 - Pan-génome de *Prochlorococcus*. La figure représente le nombre de nouveaux gènes découvert à chaque nouveau génome séquencé de *Prochlorococcus*. Le pan-génome est loin de la saturation, c'est-à-dire que l'ajout de chaque nouveau génome permet de mettre en évidence de nombreux gènes au sein de ce genre. Figure issue de Biller et al., 2015.

La comparaison des 11 premiers génomes de *Synechococcus* et des 3 premiers génomes de *Prochlorococcus* a également permis de mettre en évidence les gènes spécifiques des deux genres de picocyanobactéries. Ainsi, 70 familles de gènes se sont avérées spécifiques de *Synechococcus*, c'est-à-dire qu'elles auraient été perdues par l'ancêtre commun de tous les *Prochlorococcus* (Dufresne et al., 2008). Parmi ceux-ci, 23 codent pour des protéines impliquées dans la photosynthèse, tels que les gènes impliqués dans la biosynthèse d'allophycocyanine et de phycocyanine, les gènes *psbA* codant pour l'isoforme D1:2 de la protéine D1 conférant une meilleure résistance du PSII à la photoinhibition (Garczarek et al., 2008) ou encore des gènes impliqués dans le transport et l'assimilation du carbone (Dufresne et al., 2008). La perte du gène *kaiA*, qui code pour un des trois composants de l'horloge circadienne, ce qui explique que *Prochlorococcus* ait une horloge minimaliste puisqu'elle semble être remise à 0 chaque matin (Holtzendorff et al., 2008). Un certain nombre de transporteurs ABC de solutés compatibles, des composés participant au maintien de l'osmolarité des cellules, sont également absents des génomes de *Prochlorococcus,* ce qui pourrait induire une plus faible tolérance aux fortes salinités par rapport à *Synechococcus* (Hagemann et al., 1997; Scanlan et al., 2009). On peut également noter que des pertes sélectives survenues après la différentiation du clade LLIV, puisque les gènes *psbU* et *psbV*, codant pour deux protéines impliquées dans la stabilisation du cluster manganèse (ou OEC; Roose et al., 2007), sont présents dans ce clade basal et ont été perdus par toutes les lignées de *Prochlorococcus* à génomes réduits. A l'inverse, seuls 10 gènes se sont avérés spécifiques de *Prochlorococcus* (Partensky and Garczarek, 2010). Parmi ceux-ci on retrouve

également des gènes photosynthétiques, comme les gènes *pcb*, codant pour les protéines d'antenne associées au PSII, le gène *PcCao* codant pour une Chl *b* synthase permettant la synthèse de Chl b_2 et un homologue de *crtL-e*, impliqué dans la synthèse d'α-carotène (Satoh and Tanaka, 2006; Stickforth et al., 2003).

III.1.3 Rôles des transferts horizontaux et îlots génomiques

A l'instar de l'ensemble des procaryotes, le contenu en gènes accessoires et uniques est très variable d'une souche à l'autre au sein de ces genres, indiquant que de nombreux évènements de perte et gain de gènes ont eu lieu au cours de l'évolution de ces génomes (Coleman et al., 2006; Kettler et al., 2007). Ces gènes accessoires ou uniques sont fréquemment retrouvés dans des régions hypervariables du génome, appelées des îlots génomiques, qui sont entourées de régions conservées contenant majoritairement des gènes communs à l'ensemble du genre (Dufresne et al., 2008; Kettler et al., 2007). Ces ilots génomiques sont enrichis en gènes issus de transfert latéraux, qui chez *Synechococcus* et *Prochlorococcus* qui a priori ne possèdent pas de plasmides, auraient lieu grâce aux phages qui peuvent servir de vecteurs (Dufresne et al., 2008; Lindell et al., 2004) mais également potentiellement grâce à des vésicules extracellulaires qui pourraient permettre l'échange de composants cellulaires entre cellules et notamment de l'ADN (Biller et al., 2014). Le rôle précis joué par ces îlots génomiques dans l'adaptation à la niche écologique reste difficile à déterminer, puisque la fonction de la plupart des gènes contenus dans ces régions est le plus souvent inconnue (Scanlan et al., 2009).

Chez *Prochlorococcus*, les premiers ilots ont été identifiés en comparant les deux premiers génomes de souches HL (MED4 et MIT9312; Coleman et al., 2006), ce qui a permis d'identifier cinq ilots majeurs au niveau desquels le recrutement des lectures environnementales de l'expédition Global Ocean Survey (GOS; Rusch et al., 2007) ne retournait presqu'aucun 'hits' au contraire des régions génomiques conservées adjacentes. Parmi ces cinq ilots, l'un est enrichi en gènes impliqués dans le transport, l'assimilation et la régulation du métabolisme du phosphate et un second en gènes impliqués dans l'assimilation des nitrites et des nitrates (Martiny et al., 2006, 2009b). Contrairement aux gènes impliqués dans l'adaptation à la lumière, le contenu en gènes de ces ilots génomiques est directement relié au lieu d'isolement des souches (ou aux sites de prélèvement pour les métagénomes) et non pas à la phylogénie verticale des cellules puisque des souches très proches phylogénétiquement peuvent avoir des contenus en gènes différents. Par exemple, l'abondance relative des gènes impliqués dans l'assimilation des nitrates et des nitrites s'est révélée plus importante dans la mer des Caraïbes et dans l'océan Indien que dans la mer des Sargasses et l'océan Pacifique oriental, ce qui semble lié à la disponibilité en azote dans

ces différentes régions (Martiny et al., 2009c) (Martiny et al., 2009). Une étude plus récente a de plus montré que la répartition des gènes impliqués dans l'assimilation des nitrates au sein des clades semble largement régie par la perte stochastique de gènes et la recombinaison homologue entre cellules étroitement apparentées (Figure 11 ; Berube et al., 2019).

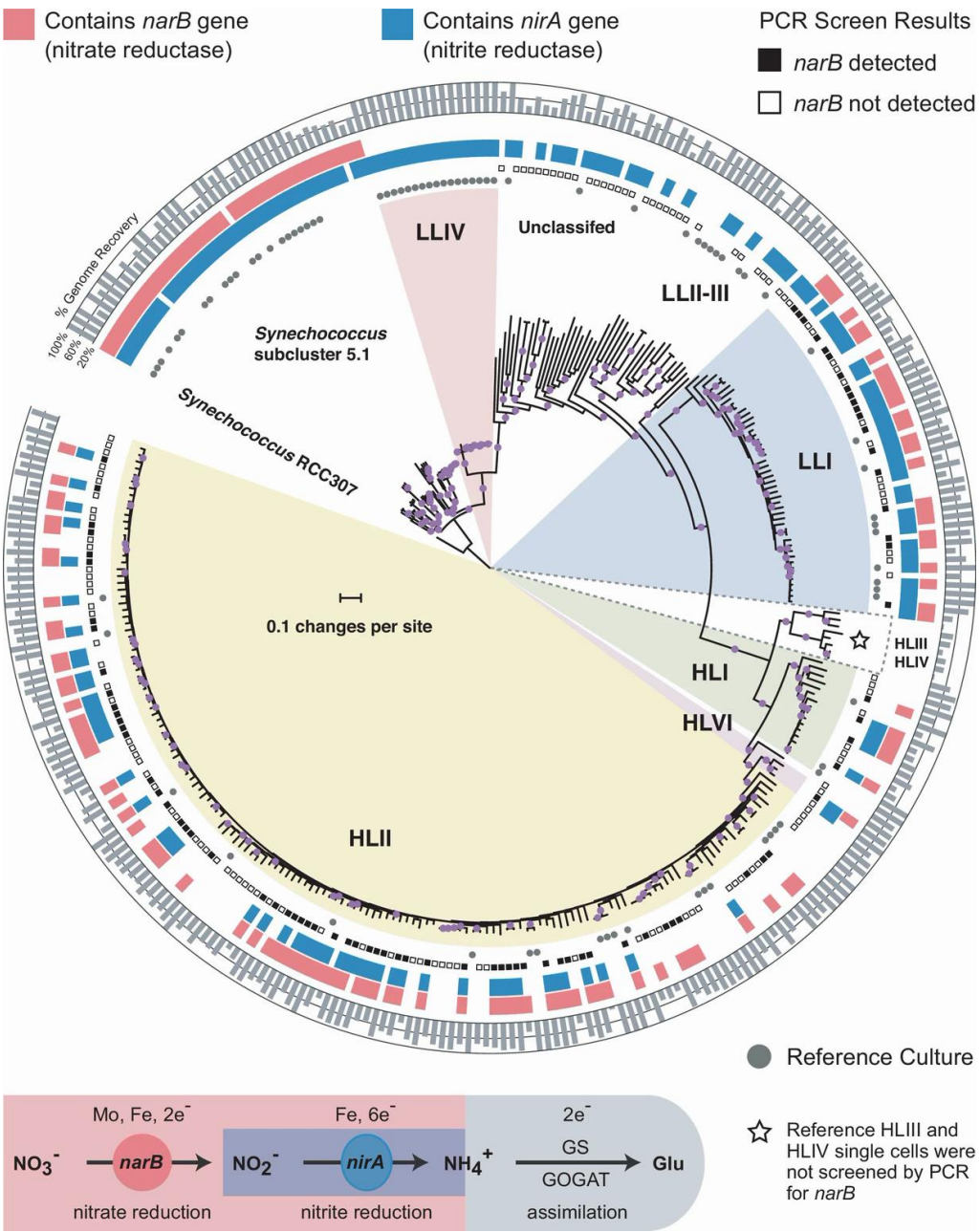


Figure 11 Distributions des genes de nitrate reductase (*narB*) et nitrite reductase (*nirB*) parmi 329 génomes de *Prochlorococcus* et *Synechococcus*. Les génomes pouvant utiliser le nitrate (présence du gène *narB*) et les génomes possédant *nirB* sont indiqués avec une barre rouge ou bleue, respectivement. Les barplots gris indique la complétion de chaque génome. Les génomes provenant de souches de références sont indiqués par des cercles gris et le résultat des screens de PCR pour *narB* sont indiqués par des carrés pleins ou vides, pour la présence ou l'absence de ce gène respectivement. Les cercles violets sur les branches de l'arbre indiquent que les génomes associés clusterisent ensemble dans au moins 75 % des 250 arbres réplicats réalisés en bootstrap. Figure issue de Berube et al., 2019.

Chez *Synechococcus*, les ilots ont essentiellement été identifiés en utilisant la déviation de la fréquence tétranucléotidique par rapport à la moyenne du génome (Dufresne et al., 2008). Ces ilots sont en moyenne plus nombreux que chez *Prochlorococcus* et contiennent notamment des gènes impliqués dans le transport de nutriments et la modification de la surface des cellules et surtout de très nombreux gènes de fonctions inconnues (Dufresne et al., 2008; Palenik et al., 2006). Ces îlots présentent des niveaux de conservation entre souches très variables, ce qui suggère que ces ilots ont été acquis plus ou moins récemment au cours de l'évolution de ces génomes (Dufresne et al., 2008). L'un de ces ilots, regroupant les gènes codant pour des protéines impliquées dans la biosynthèse et la régulation des bras des phycobilisomes, a la particularité d'être conservé chez toutes les souches de ce genre. Cependant, sa taille et son contenu en gènes sont très variables entre les souches, et sont directement reliés au type pigmentaire auquel la souche appartient, indépendamment de sa phylogénie (Figure 12; cf. I.2.2.2 Diversité pigmentaire. Dans ce contexte, on peut noter que les souches du type pigmentaire 3d possèdent un ilot supplémentaire, l'îlot CA4, qui est responsable de leur capacité à varier leur contenu pigmentaire en fonction de la couleur de la lumière ambiante. On retrouve cet îlot sous deux formes : CA4-A qui est toujours en dehors de la région phycobilisome et CA4-B qui est directement inséré au milieu de cette région (Humily et al., 2013).

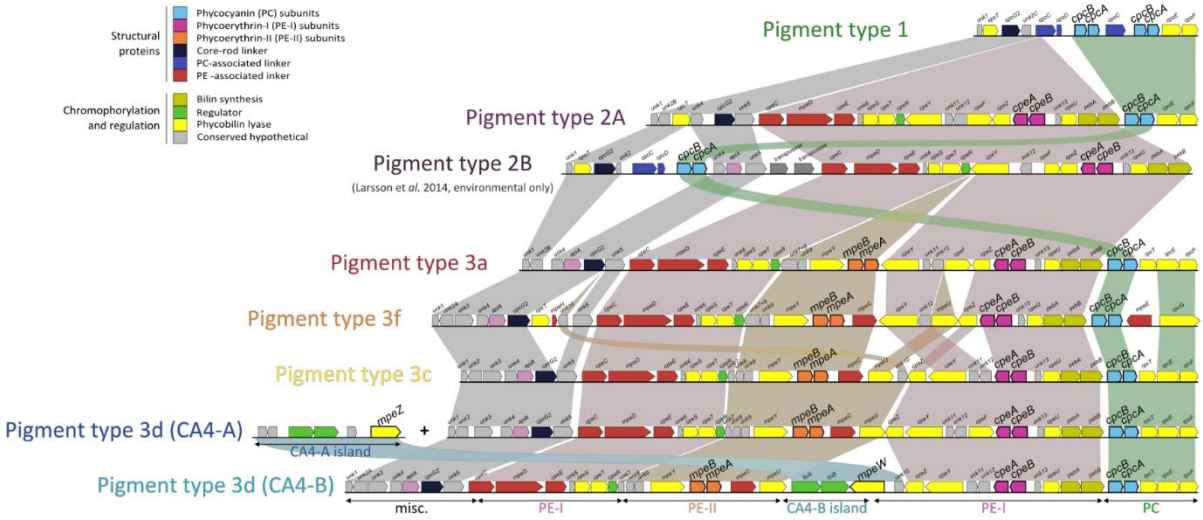


Figure 12 - Contenu en gènes et synténie de la région codant pour la biosynthèse et la régulation des bras des phycobilisomes chez *Synechococcus*. Les rubans de couleurs permettent de repérer la synténie de bloc de gènes partagés entre les différentes versions de cet îlot. En partant du type pigmentaire 1 jusqu'aux 3, on remarque que cette région est devenue plus complexe par l'acquisition de blocs de gènes. De plus, une région supplémentaire est retrouvée chez les deux types pigmentaires 3d, la forme CA4-B, insérée dans la région PBS et la forme CA4-A, distante de la région PBS sur le chromosome (Grébert et al, en préparation).

III.1.4 Rôle des substitutions nucléotidiques

L'acquisition de nouveaux gènes et donc de nouvelles fonctions ne peut pas expliquer à elle seule l'adaptation à certaines niches environnementales. En effet, on trouve parfois des organismes très proches d'un point de vue phylogénétique et occupant des niches distinctes mais dont le pool de gènes est très similaire, et l'un des meilleurs exemples concerne les clades de *Prochlorococcus* colonisant les zones tempérées (HLI) et chaudes (HLII) de l'Océan (Coleman et al, 2006 ; Martiny et al, 2006 ; Kettler et al, 2007 ; Larkin et Martiny, 2017). Ces observations posent la question du rôle des variations des séquences nucléotidiques dans la diversification des génomes et dans l'adaptation à des environnements spécifiques (McDonald and Currie, 2017; Thrash et al., 2014). Le modèle "Maestro" postule que certains traits, correspondant à des préférenda d'un paramètre donné, vont évoluer par un ajustement progressif qui n'est pas induit par des gains de gènes mais par des altérations de séquences de gènes préexistants dans le génome ou par la duplication de certains gènes (Larkin and Martiny, 2017). Dans ce contexte, plusieurs études ont montré que les genres *Prochlorococcus* et *Synechococcus* présentent une diversité génétique particulièrement importante. En séquençant le génome de cellules issues d'un même échantillon d'eau de mer, Kashtan et al. (2014) ont par exemple montré que les populations d'un clade donné de *Prochlorococcus* (p.ex. HLII) sont constituées de centaines de sous-populations avec des 'ossatures génomiques' (genomic backbones) distinctes, chaque ossature étant constitué d'une majorité des gènes communs et d'un petit set de gènes accessoires. De plus, les cellules au sein de sous-populations possédant la même ossature différaient en moyenne de 19,000 positions nucléotidiques alors que cette diversité atteignait 77,000 positions entre des cellules appartenant à des sous-populations possédant des ossatures différentes.

III.2 Réponses physiologiques aux variations des conditions environnementales

III.2.1 Réponse de l'appareil photosynthétique et mécanismes de photoprotection

La réponse de l'appareil photosynthétique aux variations de lumière, qui constitue la principale source d'énergie de ces organismes, joue un rôle prépondérant dans la réponse au stress en général chez tous les organismes photosynthétiques. De plus, l'intensité lumineuse est le seul paramètre à avoir des fluctuations pouvant atteindre des fréquences très élevées (de l'échelle de l'heure à celle de la milliseconde) et une amplitude aussi importante (de l'obscurité totale durant la nuit à un ensoleillement maximal au midi solaire ; Schubert and Forster, 1997). Les changements brusques de ce paramètre et en particulier une augmentation très importante de l'intensité lumineuse peuvent induire une diminution de l'efficacité photosynthétique (photoinhibition) et donc avoir un impact sur la croissance et la survie de ces organismes.

III.2.1.1 Photoinhibition et génération d'espèces réactives de l'oxygène

Le centre réactionnel du PSII (RCII) est le site majeur de photoinhibition puisqu'il est le composant de l'appareil photosynthétique le plus sensible à la lumière (Vass, 2012). Un excès d'énergie absorbée par le PSII aura pour conséquence une surréduction du pool de plastoquinones (PQ). Le manque de molécules de PQ réductibles laisse le site de liaison de la plastoquinone Q_B inoccupé, ce qui conduit à la stabilisation d'une semiquinone (Q_A^-), phénomène bloquant ou limitant le transport d'électrons (Vass, 2012). Cela va faciliter la formation d'une molécule de chlorophylle (P680) à l'état de triplet excité, qui interagit ensuite avec une molécule d'oxygène pour générer un radical $^1O_2$, qui constitue l'espèce réactive de l'oxygène (ROS pour Reactive Oxygen Species) la plus nocive pour les cellules photosynthétiques (Triantaphylidès et al., 2008). Des ROS peuvent être générées à partir de la réduction du dioxygène par des transporteurs d'électrons réduits du côté accepteur du PSII (Pospíšil, 2009). Classiquement, on parle de photoinhibition, lorsque Les ROS affectent directement les composants du PSII et en particulier la protéine D1, qu'ils vont endommager par leur pouvoir oxydant, conduisant à l'inactivation du transport des électrons dans le PSII (Krieger-Liszkay et al., 2008). La réactivation est dépendante de la synthèse d'une nouvelle protéine D1, constituant l'une des deux protéines majeures du RCII. Cependant, la photoinhibition peut également se dérouler du côté donneur du PSII par la dégradation du complexe manganèse (Mn_4CaO_5), impliqué dans l'étape de dissociation des molécules d'eau. Ce phénomène engendre un arrêt de l'apport d'électrons dans le RCII qui

induit une accumulation de radicaux oxydés dans ce dernier (Ohnishi et al., 2005), qui vont provoquer la dégradation des protéines D1 et D2 (Sass et al., 1997; Takahashi and Badger, 2011). Dans les deux cas de figure, les ROS accumulés dans le PSII lors d'une exposition à un excès de lumière sont responsables de l'endommagement du RCII. Le PSI est moins sensible que le PSII au stress lumineux mais peut toutefois être endommagé lorsque le flux d'électrons provenant du PSII dépasse la capacité des accepteurs d'électrons du PSI, produisant ainsi des ROS capables de détruire les protéines du RCI. Ces dommages sont particulièrement dangereux pour les cellules puisque pratiquement permanents en raison de la réparation très lente des protéines du PSI (Sonoike, 2011).

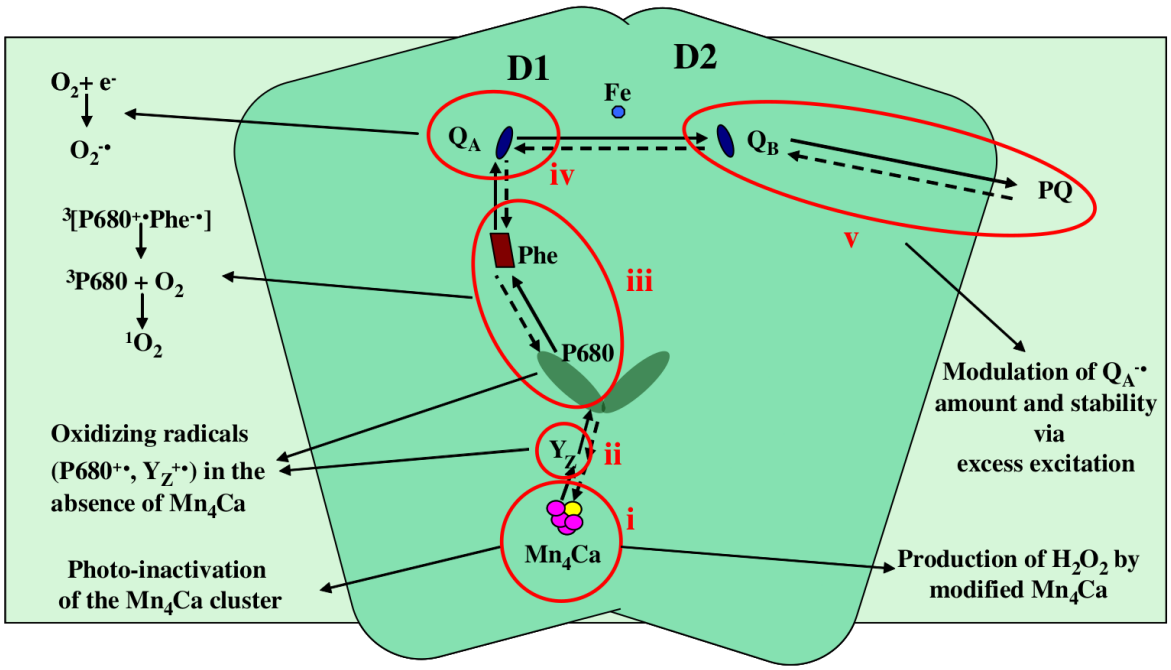


Figure 13 - Sites des dommages induits par la lumière dans le PSII. Figure extraite de Vass, 2012.

III.2.1.2 Les mécanismes de photoprotection

Les cyanobactéries, comme tous les autres organismes réalisant la photosynthèse oxygénique, ont développé une série de mécanismes de protection permettant de limiter les dommages de l'appareil photosynthétique, en particulier au niveau du PSII.

III.2.1.2.1 Cycle de réparation de la protéine D1

Le cycle de réparation de la protéine D1 constitue un élément essentiel dans la réponse au stress des organismes photosynthétiques en général. En conditions normales et stables de croissance, les taux de dégradation et de néosynthèse des protéines D1 s'équilibrent, permettant de maintenir une activité photosynthétique constante, et ce 'turnover' est d'autant plus rapide que l'intensité lumineuse d'acclimatation est importante. En revanche, en conditions d'excès de lumière ou tout autre stress environnemental entraînant une diminution du métabolisme, la vitesse de dégradation peut excéder la vitesse de néosynthèse de la protéine D1, et l'appareil photosynthétique est alors photoinhibé, entraînant une diminution de la croissance pouvant mener à la mort cellulaire (Six et al., 2007b). D'un point de vue moléculaire, la comparaison des génomes et l'analyse transcriptomique des gènes impliqués a également permis de mieux comprendre la régulation du PSII. Il existe en fait jusqu'à 6 gènes *psbA* par souche de *Synechococcus* codant pour la protéine D1, avec toujours une copie unique codant une isoforme D1:1, qui présente des performances photosynthétiques optimales et est exprimée lorsque l'intensité lumineuse est faible, et 2 à 5 copies par souches codant pour l'isoforme D1:2, moins performante mais aussi moins sensible à la photoinhibition et qui est exprimée en condition de forte intensité lumineuse (Figure 14 ; Garczarek et al., 2008; Sander et al., 2010; Sugiura et al., 2014). En ce qui concerne les souches du genre *Prochlorococcus*, elles possèdent de 1 à 3 copies de *psbA* codant toutes pour l'isoforme D1:1 (Garczarek et al., 2008). Ainsi, le fait que *Synechococcus* possède ces deux isoformes pourrait en partie expliquer que les souches de *Prochlorococcus* semblent plus sensibles à la photoinhibition que celles de *Synechococcus* (Garczarek et al., 2008; Six et al., 2007b).

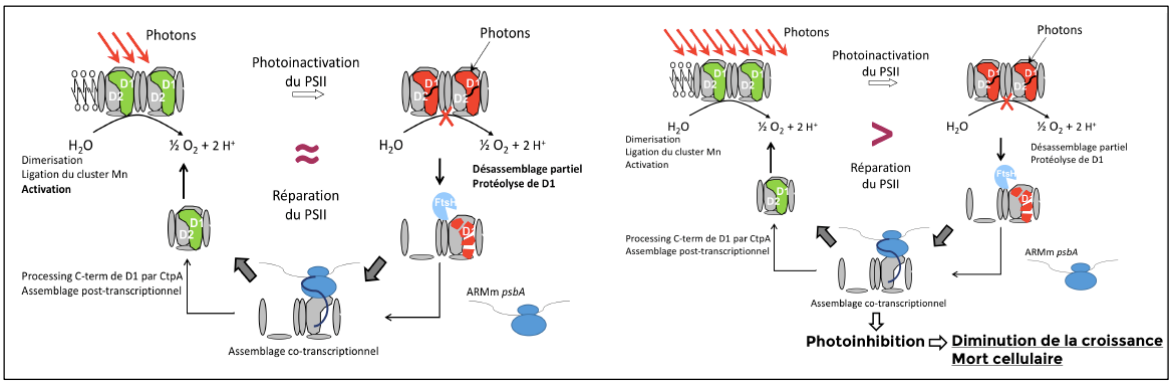


Figure 14 - Cycle de réparation de la protéine D1 du photosystème II (PSII) d'après Garczarek, 2018.

Le suivi des différents paramètres impliqués dans ce cycle de réparation a également permis de mieux comprendre la réponse à divers stress environnementaux chez ces organismes et l'influence de 'l'histoire lumineuse' des cultures sur l'intensité de cette réponse. L'étude de la réponse de *Synechococcus* sp. WH7803 au stress dû aux fortes lumières (HL pour « high light ») et aux radiations UV a permis de montrer qu'une acclimatation préalable des souches en HL leur permet de subir une réduction moindre de leur activité photosynthétique que les souches acclimatées en LL (« low light » pour basse lumière ; Garczarek et al., 2008). Par ailleurs, les cultures HL possèdent i) un pool beaucoup plus important de transcrits codant l'isoforme D1:2, ce qui permettrait au PSII d'avoir une meilleure résistance à la photoinactivation et également ii) un taux de réparation de la protéine D1 beaucoup plus important que les cultures LL. Les cellules acclimatées aux HL semblent donc bien mieux préparées à subir un stress additionnel, peut-être du fait d'une meilleure gestion du stress oxydant par rapport aux cellules acclimatées en LL (Garczarek et al., 2008). Cependant, étonnamment l'induction d'un stress oxydant directement induit par l'ajout de peroxyde d'hydrogène ou de méthylviologène dans le milieu de culture a montré que les cellules préalablement acclimatées en HL sont au contraire plus sensibles au stress oxydant que les cellules acclimatées en LL (Blot et al., 2011). Les données photophysiologiques recueillies au cours de cette étude indiquent que l'effet synergique des ROS et de l'intensité lumineuse sur la photoinhibition des cellules serait dû à la fois à un plus fort taux de dégradation de la protéine D1 et à une inhibition de la synthèse *de novo* de cette protéine par les ROS, alors que les cellules acclimatées à HL nécessitent une réparation beaucoup plus active de la protéine D1 pour maintenir leur activité photosynthétique.

Enfin, on observe également qu'en condition de cycle jour/nuit, l'activité photosynthétique des cellules décroit proportionnellement à l'intensité lumineuse pour atteindre un minimum au midi solaire mais que cette décroissance est beaucoup plus forte chez *Prochlorococcus* que chez *Synechococcus* (Mella-Flores et al., 2012). Cette observation pourrait être liée au fait que le taux de réparation de la protéine D1 au midi solaire est beaucoup plus élevé chez *Synechococcus* que chez *Prochlorococcus*. De plus, les analyses transcriptomiques ont également révélé des différences importantes entre ces organismes dans les profils d'expression journalier de plusieurs gènes, impliqués notamment dans la biosynthèse et/ou la réparation des dommages aux photosystèmes, des antennes collectrices, de la fixation du CO2 ou encore des mécanismes de protection contre les UV et les stress oxydant, traduisant des différences importantes dans la régulation de ces processus par la lumière. Ces résultats suggèrent qu'alors que *Synechococcus* a développé des moyens efficaces pour gérer le stress dû à la lumière et au UV, *Prochlorococcus* semble survivre aux heures stressantes de la journée en mettant place un

set minimum de mécanismes de protection et en réduisant temporairement plusieurs processus métaboliques clés. Ces propriétés pourraient jouer un rôle bien plus important qu'on ne le pensait sur la distribution et la dynamique de ces deux picocyanobactéries *in situ* (Mella-Flores et al., 2012).

III.2.1.2.2 Effet des paramètres environnementaux sur les phycobilisomes

Plusieurs études ont révélé que *Synechococcus* était capable de moduler la taille des bras de ses PBS en réponse à des variations de l'intensité lumineuse, notamment en détachant les disques distaux de PEII afin réduire le flux de photons parvenant au PSII, et donc les dommages générés (Six et al., 2004, 2007c). Les PBS constituent également des cibles privilégiées des radiations UV, qui non seulement entrainent une déconnection des phycobiliprotéines au niveau de l'extrémité des bras des PBS, et plus spécifiquement au niveau de la jonction entre les phycoérythrines I et II, mais également une déconnection des PBS de la membrane thylacoïdale par protéolyse du linker reliant le cœur du PBS au photosystème (Lcm; Six et al., 2007c). Enfin, il a été démontré que ces complexes seraient également impliqués dans l'adaptation différentielle à la température des thermotypes de *Synechococcus* puisque les phycobiliprotéines présentent différents niveaux de thermostabilité moléculaire. En effet, la phycocyanine s'est avérée être la moins stable de toutes les phycobiliprotéines en réponse au stress thermique, une sensibilité potentiellement liée à deux substitutions dont la distribution varie selon les thermotypes (Pittera et al., 2017).

III.2.1.2.3 'Quenching' non photochimique de la fluorescence (NPQf)

Afin de réguler la quantité d'énergie lumineuse parvenant aux photosystèmes, un mécanisme consiste à dissiper l'excès d'énergie sous forme de chaleur au niveau de l'antenne du PSII avant qu'elle n'atteigne le RC et ainsi limiter la dégradation des protéines du RCII. Ce phénomène, qui est appelé NPQf (pour Non-Photochemical Quenching of fluorescence), se déroule sur une échelle de temps allant de quelques secondes à quelques minutes (Niyogi and Truong, 2013). Chez *Synechococcus*, ce processus se produit lors de l'activation par la lumière de la protéine OCP (orange carotenoid protein) qui lie un keto-caroténoïde. Lors de l'exposition des cellules aux fortes lumières, l'OCP va changer de conformation pour devenir active, puis se lier au cœur du PBS au niveau de l'APC (Kirilovsky and A. Kerfeld, 2013). Cette interaction va permettre la dissipation de l'énergie lumineuse par transfert de charge ou d'énergie de la biline excitée de l'APC vers l'OCP (Tian et al., 2011). Cette protéine est donc à la fois le senseur et le dissipateur de

l'excès d'énergie lumineuse, qui permet d'éviter que 80% de l'excitation du PBS n'atteigne le RC (Kirilovsky and A. Kerfeld, 2013; Tian et al., 2011). Lorsque la cellule ne subit plus de stress lumineux, l'OCP se détache du PBS et retourne à sa forme inactive grâce à l'action de la protéine FRP (fluorescence recovery protein ; Boulay et al., 2010). Malgré le rôle essentiel de ce mécanisme, seule une partie des souches de *Synechococcus* marins possèdent l'opéron constitué des gènes codant pour le keto-carotenoide, l'OCP et la FRP. De même, *Prochlorococcus* ne possèdent aucun de ces trois gènes, ce qui est en accord avec l'absence de PBS dans ces souches (Scanlan et al., 2009). Néanmoins, les souches de *Prochlorococcus* sont capables d'avoir une réponse de type NPQ, en particulier les souches HL qui sont plus susceptibles d'être soumises à des variations de l'intensité lumineuse, même si le mécanisme en jeu n'est pas encore connu (Bailey et al., 2005; Six et al., 2007b).

III.2.1.2.4 Autres mécanismes de photoprotection

D'autres mécanismes de photoprotection impliquent des protéines inductibles par la lumière ou d'autres stress : les high light-inducible protein (HLIP) et IsiA (iron-stress induced), cette dernière étant également impliquée dans la réponse à la carence en fer (Havaux et al., 2003). Les protéines HLIP, présentes aussi bien chez *Synechococcus* que chez *Prochlorococcus*, sont notamment impliquées dans le phénomène de dissipation non photochimique de l'énergie lumineuse absorbée en excès au niveau des antennes (Havaux et al., 2003). Elles peuvent également avoir la fonction de transporteur de pigment, en capturant la chlorophylle libérée lors du cycle de réparation du PSII afin d'éviter les potentiels effets phototoxiques des pigments libres (Xu et al., 2004). La protéine IsiA, uniquement présente chez certaines souches de *Synechococcus* marins, s'accumule chez les cyanobactéries lors de carence en fer ou d'exposition aux fortes lumières permettant à la fois de protéger le PSII des dommages liés à la formation de ROS et d'augmenter les capacités de capture de la lumière par le PSI (Havaux et al., 2005; Yeremenko et al., 2004). L'OCP et les HLIP sont aussi impliquées dans l'élimination des radicaux $^1O_2$ qui s'accumule dans le RCII en cas de photoinhibition (He et al., 2001; Kerfeld et al., 2017).

III.2.1.2.5 Transitions d'état

Le PSII et le PSI ont des propriétés d'absorption de la lumière distinctes et leur degré d'excitation peut être déséquilibré par des changements quantitatifs et/ou qualitatifs de lumière exposant la machinerie photosynthétique à des risques de photoinhibition. Afin de maintenir une efficacité photosynthétique maximale, l'énergie peut être redistribuée entre les photosystèmes à une échelle de temps de quelques secondes par le mécanisme de transition d'état (Allen, 2003;

Lemeille and Rochaix, 2010). Chez les cyanobactéries, cette transition d'état se caractérise par le déplacement des PBS à la surface des membranes thylakoïdales (Yang et al., 2009). En réponse à des déséquilibres d'excitation des photosystèmes, ce phénomène est activé en fonction de l'état d'oxydoréduction du pool de plastoquinones (Mullineaux, 2014). Ainsi, l'oxydation des plastoquinones induit l'état 1, tandis que leur réduction induit l'état 2. Les conditions de lumière favorisant la surexcitation du PSII induisent une transition vers l'état 2, dans lequel une grande partie de l'énergie est détournée du PSII au PSI, tandis que les conditions de lumière qui provoquent une excitation excessive du PSI induisent une transition vers l'état 1, dans lequel l'énergie est préférentiellement transférée au PSII (Allen, 2003). En changeant le niveau de transfert d'énergie vers les photosystèmes, les transitions d'état agissent pour contrôler l'équilibre du transport d'électrons linéaire (impliquant les deux PS et générant à la fois la force proton-motrice qui est génératrice d'ATP et un pouvoir réducteur) et du transport cyclique d'électrons du PSI (impliquant uniquement le PSI et générant la force proton-motrice sans production de pouvoir réducteur). L'état 2 permet donc d'accroitre l'activité du PSI par rapport au PSII, favorisant ainsi le transport cyclique d'électrons (Mullineaux, 2014).

III.2.1.2.6 Voies alternatives du transport des électrons

Afin d'évacuer les électrons en surplus dans la chaîne de transport des électrons et ainsi éviter la surréduction des transporteurs, il existe de multiples voies alternatives du transport des électrons, qui permettent aux cyanobactéries de gérer les variations rapides de l'intensité lumineuse. Ce rôle protecteur de l'appareil photosynthétique est géré chez les cyanobactéries d'eau douces par des flavoprotéines (Bersanini et al., 2014; Zhang et al., 2009, 2012). Chez *Synechocystis* sp. PCC 6803, il existe 4 gènes codant pour des flavoprotéines (*flv1-4*), fortement surexprimés lors de l'exposition à de fortes lumières ou en cas de limitation en carbone (Zhang et al., 2009). Flv1 et 3 protègent le PSI en jouant un rôle de puits à électrons grâce à leur activité oxydoréductase afin de récupérer les électrons du côté accepteur du PSI et les délivrer à une molécule d'O_2, qui est réduite en H_2O sans danger pour la cellule, un processus appelé la réaction de Mehler (Allahverdiyeva et al., 2011, 2013). FLv2 et 4 sont spécifiques des cyanobactéries et permettent de protéger le PSII de la photoinhibition. De façon similaire à Flv1 et 3, FLv2 et 4 jouent un rôle de puit à électrons mais au niveau du PSII, permettant d'éviter que le pool de plastoquinones soit surréduit et donc la production de ROS (Allahverdiyeva et al., 2011, 2013; Bersanini et al., 2014; Zhang et al., 2012). Le complexe COX (aa_3-type cytochrome *c* oxidase) peut quant à lui recevoir des électrons du cytochrome b_6/f qui a potentiellement également une fonction de puit d'électrons en cas de baisse d'activité du PSI (Lea-Smith et al., 2013). Chez

certaines souches de *Synechococcus* marins et la plupart des souches de *Prochlorococcus* HL et LLI, on retrouve la protéine PTOX (pour 'plastoquinol terminal oxidase') qui permet elle aussi l'accumulation de plastoquinones réduites. De façon intéressante, PTOX est composé de 2 atomes de fer alors que le cytochrome b_6/f et le PSI contiennent 6 et 12 atomes, respectivement. L'utilisation de PTOX lors de carence en fer peut être un avantage puisqu'il permet d'économiser du fer pour d'autres fonctions (Bailey et al., 2008). A noter que chez les picocyanobactéries marines, seul un couple de flavoprotéines, équivalent à Flv1 et 3 est présent et seules quelques souches de *Synechococcus* possèdent PTOX tandis que presque toutes les souches de *Prochlorococcus* possèdent une forme spécifique de cette protéine (Scanlan et al., 2009).

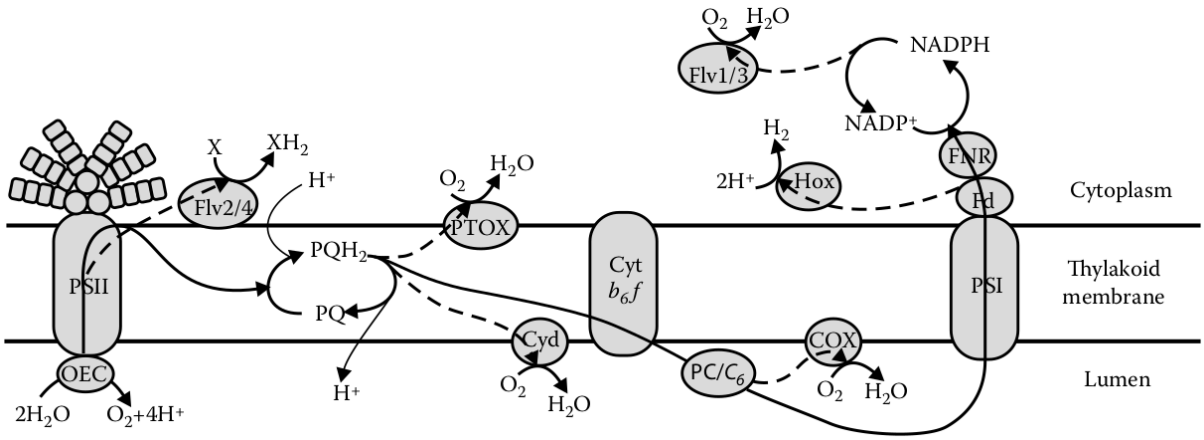


Figure 15 - Représentation du flux d'électron d'une molécule d'eau jusqu'au NADP⁺ et des voies alternatives de transport d'électron chez les cyanobactéries. Le trait plein correspond à la voie classique alors le trait en pointillé indique les voies alternatives jouant le rôle de puits à électrons. (C_6 : cytochrome c_6; COX : aa_3 -type cytochrome *c* oxidase; Cyd : cytochrome *bd*-quinol oxidase; Cyt b_6f : cytochrome b_6; Hox : [Ni-Fe]-hydrogenase; Fd : ferrédoxine; Flv1/3 et Flv2/4 : flavoprotéines; FNR : ferredoxin-NADP reductase; OEC : oxygen evolving complex; PC : plastocyanine; PQ : plastoquinone; PSI : photosystème I; PSII : photosystème II; PTOX : plastoquinol terminal oxidase). Figure de Pessarakli, 2016.

Les multiples stratégies mises au point pour préserver l'intégrité et le fonctionnement de la machinerie photosynthétique permettent aux cyanobactéries de survivre et de croître dans des conditions de lumière très variées en réagissant rapidement aux variations quantitatives et qualitatives du rayonnement incident dans les différentes niches environnementales occupées.

III.2.2 Les mécanismes d'ajustement de la fluidité membranaire

La température et le stress osmotique, entre autres, ont un effet direct sur la fluidité membranaire qui est elle-même influence la conformation et donc l'activité des protéines membranaires (Mikami and Murata, 2003) et notamment des photosystèmes I et II. C'est pourquoi la capacité des cellules à moduler la fluidité de ces membranes est indispensable au maintien de la 'fitness' des organismes photosynthétiques dans une niche thermique donnée (Mackey et al., 2013, Varkey et al., 2016). Dans ce contexte, une étude récente a montré qu'en réponse au froid *Synechococcus* sp. WH7803 était capable d'induire un raccourcissement des chaînes d'acides gras composant les lipides membranaires (conduisant ainsi à un amincissement de la membrane) et d'activer la désaturation des acides gras à des sites spécifiques des trois glycolipides présents chez cet organisme (Pittera et al., 2018). Des insaturations sont insérées dans les chaînes d'acides gras par des enzymes de la famille des désaturases, permettant de maintenir la fluidité de la membrane (Varkey et al., 2016, Xiaoyuan et al, 2008). La comparaison des génomes de *Synechococcus* disponibles a permis d'identifier quatre membres principaux de cette famille, codant deux désaturases Δ9 (*desC3* et *desC4*) et deux Δ12 (*desA2* et *desA3*) dont le pattern phylétique diffère entre les thermotypes chauds (clades II et III) d'une part et les thermotypes froids (clades I et IV) d'autre part, suggérant que les capacités de désaturation pourraient faire partie des stratégies physiologiques sous-jacentes à la spécialisation des lignées de *Synechococcus* à différentes niches thermiques (Pittera et al., 2018).

III.2.3 Effet de la carence en nutriments

Les limitations en nutriments sont les principaux facteurs limitant la croissance du phytoplancton dans le milieu naturel. En fonction des zones et de la saison, le phytoplancton va être soumis à différentes carences en particulier pour les nutriments suivants : l'azote (Hunt et al., 2013; Shilova et al., 2017), le fer (Mann and Chisholm, 2000; Shilova et al., 2017) et le phosphate (Fuller et al., 2005; Saito et al., 2014). Ces carences ont des effets particuliers sur la croissance et la physiologie des cellules et peuvent déclencher des mécanismes de réponse, permettant de gérer ces conditions peu favorables. Les picocyanobactéries étant des composants majeurs des communautés phytoplanctoniques colonisant les environnements oligotrophes, elles constituent de très bons modèles pour comprendre les mécanismes permettant de survivre à des carences en nutriments.

III.2.3.1 Carence en phosphate

Dans le cas d'une carence en phosphate, il a été montré que les picocyanobactéries sont capables de réduire leur quota en phosphate intracellulaire par 4, jusqu'à un point où la moitié du phosphate intracellulaire est retrouvé dans l'ADN, le reste du phosphate étant retrouvé dans d'autres composés cellulaire et dans l'ARN (Bertilsson et al., 2003; Fu et al., 2006). L'un des mécanismes mis en place pour limiter l'utilisation de phosphate consiste en le remplacement des phospholipides de la membrane par des lipides dépourvus de phosphate (Van Mooy et al., 2006, 2009). D'autres mécanismes d'adaptation moins connus semblent être mis en place en cas de faible concentration de phosphate dans le milieu, tel que suppression de voies métaboliques dépendantes du phosphate et la limitation de production d'ARN chez *Prochlorococcus* MED4 (Casey et al., 2016). En cas de limitation en phosphate dans le milieu des picocyanobactéries, celles-ci sont également capables d'accroitre leur capacité d'assimilation de ce macronutriment notamment en surexprimant des gènes impliqués dans i) le transport du phosphate (*pstABC*), ii) la captation du phosphate inorganique (la forme privilégiée par la cellule) dans l'espace périplasmique par la protéine PstS, codée par une famille de paralogues, iii) l'assimilation de phosphate organique par hydrolyse d'une macromolécule biologique, une réaction réalisée par les phosphatases et enfin iv) la motilité des cellules de *Synechococcus,* qui seraient capables de chimiotactisme par rapport au phosphate comme c'est le cas pour l'azote (Martiny et al., 2006; Mazard et al., 2012b; Moore et al., 2005; Pitt et al., 2010; Tetu et al., 2009; Willey and Waterbury, 1989).

Afin de mettre en place ces réponses liées au manque de phosphate, il faut qu'en amont l'organisme détecte cette carence. Chez *Synechococcus* et *Prochlorococcus*, c'est l'histidine kinase PhoR qui joue ce rôle et déclenche l'activation du régulateur transcriptionnel PhoB, qui à son tour induit la surexpression des gènes contenus dans le régulon Pho (*pstS*, *pstABC*, *PhoD*, etc ; Su et al., 2006, 2007; Tetu et al., 2009). Le complexe PhoR/PhoB, permettant d'induire le transport du phosphate, fonctionne en synergie avec le régulateur PrtA, qui joue un rôle dans l'activation des gènes de phosphatases (Ostrowski et al., 2010). Comme mentionné précédemment (cf. §III.1.3 Rôles des transferts horizontaux et îlots génomiques), la présence/absence de ces gènes est directement reliée aux sites d'isolement des souches, certains de ces gènes étant par exemple absents certains *Prochlorococcus* HL et les souches côtières de *Synechococcus*, colonisant des zones non limitées en phosphate (Moore et al., 2005; Palenik et al., 2006; Scanlan et al., 2009).

III.2.3.2 Carence en azote

La carence en azote est connue pour inhiber la croissance cellulaire et impacter l'efficacité de la photosynthèse (Lindell and Post, 2001; Lindell et al., 2002; Steglich et al., 2001; Tolonen et al., 2006). Cette carence peut notamment déclencher l'utilisation de sources alternatives d'azote. L'ammonium, directement assimilable par les cellules, est la source d'azote privilégiée par les picocyanobactéries alors que les nitrates et les nitrites nécessitent des étapes préalables de réduction plus couteuses en énergie. En absence d'ammonium et en présence de nitrates et/ou de nitrites, le gène *ntcA*, un régulateur transcriptionnel retrouvé chez toutes les picocyanobactéries, est exprimé et active l'expression des gènes impliqués dans le transport et l'assimilation de ces composés, son expression étant maximale en l'absence de source d'azote (Lindell and Padan, 1998; Lindell and Post, 2001; Lindell et al., 2002; Scanlan et al., 2009; Tolonen et al., 2006). Après incorporation des formes oxydées de l'azote par le transporteur transmembranaire NrtP, le nitrate est réduit par la nitrate reductase (NarB) en nitrite, lui-même réduit en ammonium par la nitrite reductase (NirA). En plus de ces sources inorganiques d'azote, les picocyanobactéries sont capables d'utiliser des sources organiques, telle que l'urée, incorporée via une perméase de type ABC (UrtA-D), puis dégradée par une uréase codée par les gènes *ureA-G* et libérant de l'ammonium (Collier et al., 1999; Scanlan et al., 2009; Valladares et al., 2002).

III.2.3.3 Carences en micronutriments

Les macronutriments ne sont pas les seuls à être potentiellement limitants pour la croissance de ces organismes, c'est aussi le cas des micronutriments, en particulier des métaux traces. En effet, les concentrations de fer, manganèse, zinc et cobalt ont un impact sur la croissance et la composition des espèces phytoplanctoniques (Boyd et al., 2000; Coale et al., 1996; Hassler et al., 2012; Moore et al., 2004; Saito et al., 2004; Sunda and Huntsman, 1995). Les micronutriments sont retrouvés en très faibles quantités dans l'environnement (pico- à nano-molaire) et sont pourtant indispensables au métabolisme de ces organismes. Les micronutriments sont en effet présents dans les cofacteurs et les enzymes du métabolisme cellulaire. Par exemple, l'appareil photosynthétique est composé de multiples cofacteurs contenant des métaux, notamment le fer dans les groupements fer-soufre (Fe-S) et les cytochromes, du manganèse dans le groupement manganèse du complexe d'évolution de l'oxygène (OEC) dans le PSII, du cuivre dans les plastocyanines, le zinc dans l'anhydrase carbonique, le cobalt dans la vitamine B12 et du magnésium dans chaque molécule de chlorophylle (Scanlan et al., 2009; Shcolnick and Keren, 2006). Les études décrivant les effets chez les cyanobactéries d'une carence ou d'un excès de micronutriments portent en majeure partie sur le fer.

Le fer est particulièrement essentiel pour les organismes photosynthétiques puisque l'on retrouve environ une vingtaine d'atomes de fer au sein de l'appareil photosynthétique (Bailey et al., 2008; Raven et al., 1999). En conditions de carence en ce nutriment, il a été montré que les gènes impliqués dans la biosynthèse des phycobilisomes sont sous-exprimés, réduisant par conséquent le nombre de ces complexes à la surface des thylacoïdes, et il en est de même pour les gènes responsables de la synthèse du PSI, qui à lui seul contient 12 atomes de fer (Bailey et al., 2008; Mackey et al., 2015; Morrissey and Bowler, 2012; Thompson et al., 2011). En revanche, ce stress induit la surexpression des gènes codant pour les transporteurs de fer (*idiA*, *futB*, *futC* ; Thompson et al., 2011). Afin de réduire le quota en fer des cellules, certaines enzymes aux fonctions équivalentes à des enzymes contenant du fer, utilisent un autre métal comme cofacteurs. C'est notamment le cas de flavodoxine (IsiB), utilisant le cuivre, qui remplace la ferrédoxine (PetF) mais également de la plastocyanine (PetE, contenant également du cuivre, qui remplace le cytochrome c_6 (PetJ) ou encore de la super oxyde dismutase (SOD) contenant du nickel qui remplace la Fe-SOD (Erdner and Anderson, 1999; Morrissey and Bowler, 2012; Rocap et al., 2003). Certaines souches vont également surexprimer la protéine IsiA, homologue du gène codant pour l'antenne photosynthétique de *Prochlorococcus*, qui en formant un anneau de protéines pigmentées autour du PSI, serait capable de compenser la diminution de complexes consommateurs en fer, tels que le PSI ou les PBS, en augmentant le rendement quantique du PSI (Bibby et al., 2001, 2003; Boekema et al., 2001; Thompson et al., 2011). Il a été suggéré que ces variations d'expression serait sous le contrôle d'un régulateur transcriptionnel, Fur, que l'on retrouve dans l'ensemble des génomes de picocyanobactéries (Scanlan et al., 2009).

Objectifs de thèse

Bien que la diversité génétique des deux genres de picocyanobactéries ainsi que leur biogéographie soient maintenant relativement bien décrites, les mécanismes impliqués dans l'adaptation et l'acclimatation à leurs niches respectives ainsi que les différences physiologiques entre écotypes restent encore mal compris. Ce manque de connaissance serait dû d'une part au nombre encore limité de génomes disponibles au début de ma thèse qui n'avaient pas permis de mettre en évidence les spécificités génomiques de chaque écotype, et d'autre part au fait que la plupart des études physiologiques n'ont pour le moment porté que sur quelques souches modèles, peu ou pas représentatives d'un point de vue écologique, et sur un nombre limité de facteurs environnementaux.

Le premier chapitre de cette thèse a visé à identifier les processus évolutifs qui sont l'origine de la diversification génétique des génomes de picocyanobactéries et à comprendre dans quelle mesure celle-ci a contribué au succès écologique de ces organismes et à leur adaptation à des niches spécifiques. J'ai pour cela utilisé des génomes encore non publiés de *Synechococcus* qui permettent enfin d'avoir accès à plusieurs génomes par clade, voire par sous-clade au sein de ce genre.

Bien que diverses études physiologiques aient étudié la réponse des picocyanobactéries à des changements des paramètres environnementaux, peu d'études ont pour le moment comparé les réponses physiologiques et transcriptionnelles de ces organismes à divers stress environnementaux et encore moins la synergie potentielle entre les différents facteurs environnementaux affectant la croissance et l'état physiologique de ces organismes. Le deuxième chapitre de cette thèse décrit une étude de la variabilité de la réponse physiologique de la souche modèle *Synechococcus* sp. WH7803 à différents stress lumineux et thermiques.

En dehors de quelques souches modèles de picocyanobactéries, la physiologie comparative de souches représentatives de clades/écotypes abondants *in situ* a encore été très peu abordée. Le dernier axe de cette thèse a visé à identifier les mécanismes d'acclimatation aux variations des conditions environnementales spécifiques ou communs aux quatre principaux écotypes du genre *Synechococcus* (clades I-IV). Enfin, l'analyse de données *in situ* à l'échelle globale à partir des données de l'expédition de *Tara* Oceans a permis de valider une partie des hypothèses formulées durant les différentes parties de cette thèse.

CHAPITRE I

Adaptation à la niche et
évolution des génomes de picocyanobactéries
marines

Contexte de l'étude

La première partie de ce chapitre consiste en une description du système d'information Cyanorak v2.1, regroupant actuellement 97 génomes de picocyanobactéries (dont 31 génomes de *Synechococcus* non publiés), un outil qui permet d'affiner manuellement l'annotation fonctionnelle des génomes et de faciliter les analyses de génomique comparative en termes de présence/absence de gènes dans l'ensemble de ces génomes. En effet dans ce système, les gènes prédits dans les différents génomes sont rassemblés en groupes de gènes orthologues (CLOG pour 'Clusters of Likely-Orthologous Genes'), tels que déterminés par Best-Hit réciproques. En résumé, un CLOG rassemble un groupe de gènes dont la séquence, et supposément la fonction, sont similaires chez les différentes souches. L'annotation réalisée à l'échelle du CLOG est automatiquement transférée à l'ensemble des séquences qui le compose et devient prioritaire sur l'annotation initiale du gène. Cyanorak v2.1 permet donc non seulement de réaliser une annotation simultanée et évolutive de l'ensemble des génomes de la base, mais également de générer des exports concernant diverses informations relatives à chaque souche ou à chaque CLOG dans différents formats. Ces exports ont notamment été utilisés dans l'étude de Doré et al., présentée dans la deuxième partie de ce chapitre, afin de déterminer aisément les gènes présents dans l'ensemble des souches, c-à-d le génome commun, ceux présents dans au moins une souche, c-à-d le génome accessoire, et enfin les gènes spécifiques de différents niveaux taxonomiques et/ou de groupes de souches occupant la même niche écologique. En effet, comme nous avons pu le voir dans l'introduction, diverses études récentes décrivant la distribution des picocyanobactéries marines ont permis de mettre en évidence une microdiversité importante chez ces organismes et l'existence au sein des principaux clades de populations génétiquement distinctes occupant des niches environnementales spécifiques. Par ailleurs, d'autres études ont démontré que ces deux genres présentent également une très large diversité génomique et certaines différences en termes de contenu en gènes ont pu être mis en évidence entre les différents groupes phylogénétiques, même si le nombre limité de génomes par clade n'avait jusqu'à présent pas permis d'identifier de façon fiable les gènes spécifiques de clades ou d'écotypes. Dans ce contexte, la deuxième partie de ce chapitre a contribué à répondre aux questions suivantes : i) quelles sont les bases génétiques de l'adaptation à différentes niches environnementales ? et ii) quels mécanismes évolutifs ont-ils permis d'aboutir à cette remarquable diversité génomique ?

Contribution

Les étapes d'extraction d'ADN, de séquençage, d'assemblage, d'annotation automatique des 32 nouveaux génomes présentés dans cette étude ainsi que les étapes de clustering en CLOGs et de développement de l'application Cyanorak v2.1 avaient été réalisées avant le début de ma thèse, notamment à travers une étroite collaboration entre l'équipe ECOMAP et la plateforme ABIMS. En ce qui me concerne, j'ai activement participé à l'amélioration du système d'information, notamment en listant les gènes nécessitant une curation de leur annotation fonctionnelle, en participant à la mise au point avec Mark Hoebeke d'un outil de curation des starts (*StartCorrector,* non finalisé à ce jour) et en travaillant sur les différents formats d'export de données. Par ailleurs, j'ai travaillé en collaboration avec Loraine Brillet-Guéguen à la génération d'une interface de visualisation des génomes (JBrowse), notamment en générant et formatant les données de génomique (gènes communs, accessoires et uniques, prédiction d'opérons, de gènes gagnés et/ou présents dans des ilots génomiques, etc.) et de transcriptomique intégrées à cet outil, incluant des données de RNAseq acquises dans le cadre de ma thèse et de microarrays acquises précédemment (Blot et al., 2011).

Concernant le deuxième manuscrit inclus dans ce chapitre et portant les mécanismes évolutifs de la diversification des génomes de picocyanobactéries marine et leur adaptation aux niches écologiques, tels qu'analysés sur les 81 génomes non-redondants de la base Cyanorak v2.1, une grande partie des analyses de génomique comparative avaient déjà été réalisées dans le cadre de la thèse d'Hugo Doré. Ma principale contribution a été de développer un workflow original permettant de détecter les îlots génomiques partagés par les différents génomes de picocyanobactéries analysés dans le cadre de cette étude.

I.1 Cyanorak v2.1, a scalable information system dedicated to the visualization and expert curation of picocyanobacteria genomes

Résumé de l'article

Cyanorak v2.1 (http://www.sb-roscoff.fr/cyanorak) est un système d'information dédié à l'annotation, la comparaison et la visualisation de génomes de *Prochlorococcus*, *Synechococcus* et *Cyanobium*, les microorganismes photosynthétiques les plus abondants sur Terre. La version publique actuelle de la base de données comprend 97 génomes, couvrant la plupart de la grande diversité génétique connue jusqu'à présent au sein de ces groupes. Les séquences ont été divisées en 25 834 clusters de groupes de gènes orthologues-like (CLOG). L'interface utilisateur donne accès aux informations détaillées sur les souches, y compris les caractéristiques génomiques, les numéros d'accession, les métadonnées ainsi qu'une carte interactive montrant leur site d'isolement. Un visualiseur de génome basé sur JBrowse est également disponible et permet aux utilisateurs de visualiser des informations supplémentaires liées à chaque génome, telles que les gènes acquis, les îlots génomiques, les opérons prédits ou les données transcriptomiques, lorsqu'elles sont disponibles. L'entrée principale dans la base de données se fait par une recherche rapide ou avancée d'un terme (nom du gène, produit, numéro EC, code PFAM, etc.), ce qui donne une liste de CLOGs et de gènes individuels. Chaque CLOG bénéficie d'une riche annotation fonctionnelle comprenant les COG, EggNOG, les numéros EC et K, les GO terms, les TIGR Roles, les Cyanorak Roles ainsi que les motifs TIGRFAM, PFAM, ProSite et InterPro. Un profil phylétique indique le génotype et le type pigmentaire de toutes les souches possédant l'orthologie sélectionnée. Ce système d'information comprend également plusieurs plugins utiles, notamment des outils de recherche BLAST, un contexte génomique comparatif, ainsi que diverses options d'exportation. En résumé, Cyanorak v2.1 constitue un outil inestimable et modulable de génomique comparative pour l'un des groupes de micro-organismes marins les plus pertinents sur le plan écologique.

Cyanorak v2.1, a scalable information system dedicated to the visualization and expert curation of picocyanobacteria genomes

Laurence Garczarek[1§*], Ulysse Guyet[1§], Hugo Doré[1], Gregory K. Farrant[1,2], Mark Hoebeke[2], Loraine Brillet-Guéguen[2,3], Antoine Bisch[1,2], Jukka Siltanen[2], Erwan Corre[2], Gildas Le Corguillé[2], Christophe Caron[2†], Morgane Ratin[1], Frances D. Pitt[4], Martin Ostrowski[4,5], Maël Conan[6], Anne Siegel[6], Karine Labadie[7], Jean-Marc Aury[7], Patrick Wincker[8], David J. Scanlan[4] and Frédéric Partensky[1]

[1]Sorbonne Université & CNRS, UMR 7144 'Adaptation & Diversity in the Marine Environment' (AD2M), Station Biologique de Roscoff (SBR), 29680 Roscoff, France

[2] CNRS, FR 2424, ABiMS Platform, Station Biologique de Roscoff (SBR), Roscoff, France

[3] Sorbonne Université, CNRS, UMR 8227, Integrative Biology of Marine Models (LBI2M), Station Biologique de Roscoff (SBR), Roscoff, France

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

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

[6] DYLISS (INRIA–IRISA)–INRIA, CNRS UMR 6074, Université de Rennes 1, 35042 Rennes, France

[7] Genoscope, Institut de biologie François-Jacob, Commissariat à l'Energie Atomique (CEA), Université Paris-Saclay, Evry, France

[8] Génomique Métabolique, Genoscope, Institut de biologie François Jacob, CEA, CNRS, Université d'Evry, Université Paris-Saclay, Evry, France

[§] These authors contributed equally to this study

[†] Deceased, May 5th, 2018

*Corresponding author. Email: laurence.garczarek@sb-roscoff.fr

IN PREPARATION

Abstract

Cyanorak v2.1 (http://www.sb-roscoff.fr/cyanorak) is an information system dedicated to the expert annotation, comparison and visualization of the genomes of *Prochlorococcus*, *Synechococcus* and *Cyanobium,* the most abundant photosynthetic microorganisms on Earth. The current public version of the database encompasses 97 genomes, covering most of the wide genetic diversity known so far within these groups. Sequences were split into 25,834 clusters of likely orthologous groups (CLOGs). The user interface gives access to the detailed information about strains, including genomic characteristics, accession numbers, metadata as well as an interactive map showing their isolation site. A genome viewer based on JBrowse is also available and allows users to visualize additional information related to each genome such as gained genes, genomic islands, predicted operons or transcriptomic data, when available. The main entry to the database is through either fast or advanced search for a term (gene name, product, EC number, PFAM code, etc.), resulting in a list of CLOGs and individual genes. Each CLOG benefits from a rich functional annotation including COG, EggNOG, EC and K numbers, GO terms, TIGR Roles, custom-designed Cyanorak roles as well as TIGRFAM, PFAM, ProSite and InterPro motifs. A phyletic profile indicates the genotype and pigment type of all strains possessing the selected ortholog. This information system also includes several useful plugins including BLAST search tools, comparative genomic context, as well as various export options. Altogether, Cyanorak v2.1 constitutes an invaluable, scalable tool for comparative genomics for some of the most ecologically relevant groups of marine microorganisms.

Introduction

The regular decrease in sequencing costs associated with the rapid development of Next Genome Sequencing (NGS) technologies has led to the multiplication of microbial genomes (1, 2), making possible extensive comparative genomics studies. Genomes are generally annotated using gene calling programs, such as e.g. RAST (3) or Prokka (4), which can provide fairly reliable annotations for the most conserved core genes involved in general metabolism (e.g., ribosomal proteins, Krebs or Calvin cycle enzymes, DNA replication, tRNA, etc.) or more specific but well characterized functions shared by many sequenced organisms (chlorophyll biosynthesis, nitrogen fixation, etc.). However, these automatic annotations are much less reliable for the least conserved accessory genes, such as those encoding enzymes responsible for cell wall biosynthesis that are often multi-domains, with highly variable domain composition, or those coding for species- or even strain-specific functions (e.g. biosynthesis of secondary metabolites, carotenoids, etc.). Thus, even though an initial step of automatic annotation is mandatory, functional annotation of predicted coding sequences (CDS) still require extensive expert human curation to be reliable, especially for non-model organisms. With the exponential increase of newly sequenced genomes, manually curating individual genomes is however a highly time-consuming and inefficient approach. A smart alternative is to curate several phylogenetically related genomes at a time, after gathering sequences into Clusters of Likely Orthologous Genes (CLOGs), i.e. genes that exhibit reciprocal best hits to one another and are hypothesized to have the same function in the different members of the dataset (5, 6). This strategy, notably used in the NCBI's 'prokaryotic genome annotation pipeline' (7) for annotating new genomes or re-annotating older genomes before inclusion in the RefSeq database, allows propagating rich, functional annotations made at CLOG level to all proteins composing the CLOG and makes it possible to unify and standardize these annotations across all sequenced strains.

Here we present Cyanorak v2.1, an information system based on CLOGs that is dedicated to the annotation and visualization of picocyanobacterial genomes. Initially created in the mid 2000's by A. Dufresne and co-authors to compare the first 14 genomes of marine and brackish picocyanobacteria (8), the Cyanorak database has tremendously increased since then and relies on a completely redesigned and tremendously enriched information system (v2.1) which, contrary to Cyanorak v1, is scalable. The current database encompasses 95 genomes and two metagenome-assembled genomes (MAGs), including 31 newly released *Synechococcus* and *Cyanobium* genomes (9), which have been closed using a custom-designed assembly and scaffolding pipeline (10). All strains whose genomes have been included in the Cyanorak v2.1 database belong to Cyanobacteria Subsection I, Cluster 5 *sensu* Herdman (11), a coccoid-shaped group that forms a deep monophyletic branch within this ancient phylum (12). The common ancestor of all Cluster 5 members is thought to have diverged from other cyanobacteria about 1 Gy ago, during the Mesoproterozoic period (13). Members of Cluster 5 are also called 'α−cyanobacteria', based on the occurrence in their cytoplasm of specific α-type carboxysomes, phylogenetically and structurally closer to that of thiobacilles than to the β-type carboxysomes found in all other cyanobacteria, so-called 'β−cyanobacteria' (14). Cluster 5 is itself split into four major groups, including the monophyletic, strictly marine *Prochlorococcus* lineage and three deeply branching groups, called sub-clusters (SC) 5.1 through 5.3 (8, 15, 16), recently suggested to be renamed *Ca.* Marinosynechococcus (SC 5.1), *Cyanobium* (SC 5.2) or *Ca.* Juxtasynechococcus (SC 5.3; 9). SC 5.1 is the most diversified of all these lineages, with about 10 phylogenetically distinct clades based on 16s rDNA phylogeny (16), all strictly marine except clade VIII that specifically gathers halotolerant strains. SC 5.2 also mostly encompasses halotolerant strains as well as one freshwater representative (*Cyanobium gracile* PCC 6307). While members of SC 5.3 were initially thought to be strictly marine (8, 16), freshwater members of this group were recently discovered in various lakes (17, 18). The current version of Cyanorak v2.1 encompasses representatives of most of the lineages (SC and clades) known to date in Cluster 5, with the exception of the newly described

freshwater members of SC 5.3 as well as members of the yet-uncultured SC 5.1 clades EnvA and EnvB (10, 19). Since all Cluster 5 members possess a similar morphology (spherical to rod shaped) and lifestyle (aquatic, non-diazotrophic oxyphototrophs; 11) and form a monophyletic branch within the Cyanobacteria phylum, we assume that members of most CLOGs defined within this genetically homogeneous group exhibit the same function, while this might not be true when considering more distant organisms, notably cyanobacteria exhibiting different lifestyles. Here we describe the construction of the cyanorak V2.1 database, the rich functional annotation available for each CLOG and the tools and plugins that were developed to explore the genomic diversity of this ecologically relevant group of organisms.

Material and Methods

Clustering of likely orthologous sequences and CLOG curation

Following the delineation of a first series of CLOGs based on the 14 first sequenced picocyanobacterial genomes (8), Cyanorak v1 CLOG numbers have been cited in a number of publications (see e.g., 20–28). In order to preserve at best these preexisting CLOG numbers after addition of 83 new genomes including new genomes either retrieved from Genbank or newly sequenced at our initiative (see Fig. 1 in Doré et al, Unpub, CHAPTER I.2), all genes from the 97 genomes were first clustered using all-against-all BLASTp+ comparison and the OrthoMCL clustering algorithm (29) with an e-value threshold of 10^{-5} and new CLOGs were then mapped to previously defined Cyanorak v1 CLOGs. New CLOGs containing all sequences from a v1 CLOG plus additional sequences from new genomes as well as manually curated CLOGs were assigned the previous v1 CLOG numbers. All other sequences were then tentatively assigned to preexisting CLOGs using HMMER (30) with an e-value threshold of 10^{-20} and remaining sequences were finally

clustered using OrthoMCL to define new CLOGs or left as singletons in individual CLOGs if not clustered.

Since Cyanorak v.1 contained only CDS, these steps also allowed us to generate CLOGs for rRNA, tRNA, tmRNAs using all-against-all BLASTn+ and the OrthoMCL algorithm using the same threshold as for CDS (30). After this semi-automatic clustering step, a large number of CLOGs (about 3,800) were further manually curated in order to i) check and complement the functional description of CLOGs, ii) verify that members of a given CLOG were truly orthologs, based on their phyletic pattern, alignments and phylogenetic trees. Paralogs were moved into different CLOGs when they grouped together into different branches as the *bona fide* orthologs. In order to refine assessment of the core genome (9), more than 1,750 genes missed by gene prediction softwares, either because they were too short (e.g. *petM*, *psbM*) or partially overlapping with other genes notably in the case of long 3'-3' overlaps (e.g. for *pyrB-ndbA or panB-hemN*), were also manually added to different genomes. Furthermore, many over-predictions (e.g., short CDS of unknown function totally overlapping long annotated CDS) were eliminated, and this even in genomes retrieved from Genbank. At last, starts of sequences obviously too short or too long were corrected, based on an alignment of all CLOG members and/or 5'-end extensions using TBLASTN searches.

Implementation of the Cyanorak v2.1 information system

The development of Cyanorak v2.1 was done in two steps. The first version (v2.0) of this information system included a history feature keeping track of every change and allowing to readily restore the database content at any previous date and/state. This version of the information system is still currently used for the manual curation of the database. However, the history was found to considerably slow down page loading times, especially when the number of genomes and MAG's in the database climbed to 97, and a completely new version of the Cyanorak information system, devoid

of the history feature (v2.1), was therefore recently developed and proved to be two to three time faster than v2.0. Several instances of the Cyanorak information system therefore co-exist on our server: i) the access-restricted, editable cyanorak V2.0 allowing expert curators to edit most fields of the 'CLOG' and 'gene' pages, and ii) the public, non-editable Cyanorak V2.1 versions, the latter being frozen in the state in which the Cyanorak database was at the time of the co-publication of the present paper and the comparative genomics study of the 81 non-redundant, high quality genomes of the database (9). This public version will be regularly updated in the future, with concomitant changes in version number, when new genomes, SAGs, MAGs and/or transcriptomes will be added to the Cyanorak database and described in the frame of forthcoming publications. An access-restricted, editable instance based on the v2.1 implementation is currently being developed and should replace in near future the current v2.0 instance for expert curation purposes.

Results

General characteristics of the database

Built from 97 picocyanobacterial genomes, including 43 *Prochlorococcus* and 54 *Synechococcus/Cyanobium*, which are representative of the wide genetic and pigment diversity existing within these genera, Cyanorak v2.1 encompasses 252,176 genes that were split into 25,834 CLOGs. A plot of the distribution of the number of sequences per CLOG (Supplementary Figure S1) expectedly shows that the most frequent category are CLOGs with one sequence, i.e. unique genes (15,283 CLOGs), and CLOGs with few (2-5) members. Although most of these CLOGs (e.g. 91% of unique genes) are annotated as "hypothetical" or "conserved hypothetical" proteins, a number of them display a more precise functional annotation, since they share some similarities to genes or domains of known function, with among the most abundant: glycosyltransferases, restriction-modification system proteins, integrases, transposases, methyltransferases, NAD-dependent epimerases/dehydratases and tetratricopeptide repeat family proteins. The next most abundant CLOG

category (611 CLOGs) is the one containing 97 sequences, which corresponds to the picocyanobacterial core genome *sensu stricto*. As expected, this number is significantly lower than the picocyanobacterial strict core gene set (911 genes) estimated by Doré et al. (9) using the 81 non-redundant, high-quality genomes of the Cyanorak database. Yet, given than some of the 97 genomes or MAG's, especially those not included in this 81-genome set, are incomplete and/or contain frameshifted genes (in this case, the two or more gene fragments resulting from a frameshift have been put in the same CLOG), many CLOGs contain a number of genes that is close but not exactly equal to 97. So, the picocyanobacterial core genome *sensu lato* is likely much larger than 611 genes, and we estimated it using a relaxed definition of core genes (CLOG is considered as core of a taxonomical group if it is present in ≥ 90% of the strains within this group) to be 893 genes. A small number of CLOGs contain a large number of sequences, i.e. between 105 and 337 sequences. These CLOGs most often contain paralogous sequences that cannot be split into different CLOGs based on phylogenetic analyses. These notably include the identical multi-copy genes coding for the photosystem II core proteins D1.2 and D2, porins, AbrB-like transcriptional regulators, the phosphate ABC transporter phosphate-binding protein PstS, etc.

Cyanorak pages and tools

The homepage of the Cyanorak v2.1 information system shortly describe the origin of the genomes used to build the CLOGs database, the history of its construction and the main reference that used it. The top banner available from all pages encompasses several clickable menus, including the 'organisms page' that lists the different genomes of the database and their characteristics, a 'search page' with different options to access to CLOG or gene pages of interest, a 'Blast search page', a

direct link to the JBrowse viewer of each genome, as well as several other menus providing useful information about the database (References, Links, About us).

Organisms page

The 'organisms page' consists of two tables, the first one listing *Prochlorococcus* genomes and the other one *Synechococcus* and *Cyanobium* genomes. They provide taxonomy, pigment type, sequencing center as well as various genomic characteristics (size, GC%, status, accession numbers, number of CDS, etc.) for each genome included in the database. Next to each strain name is a clickable 'J' logo that gives access to the JBrowse page of the corresponding genome (see below). A distribution map of all of the strains drawn with OpenStreetMap® (https://www.openstreetmap.org/) is also available in this section (Fig. 1), allowing the user to visualize only a subset of strains (e.g., *Prochlorococcus* only) as well as to display the different strains isolated in the same area as a cluster of strains or individually. In this section, each strain name in the table can be clicked to get more detailed information (e.g., isolation site, isolator, environment ontology ENVO code, etc.) and also allows the user to export gene and protein sequences in FASTA format as well as whole genomes in Genbank format. In each of these files, the annotation of every gene corresponds to that found on CLOG pages (see below), which was given priority over the initial gene annotation, even if the genome was retrieved from Genbank.

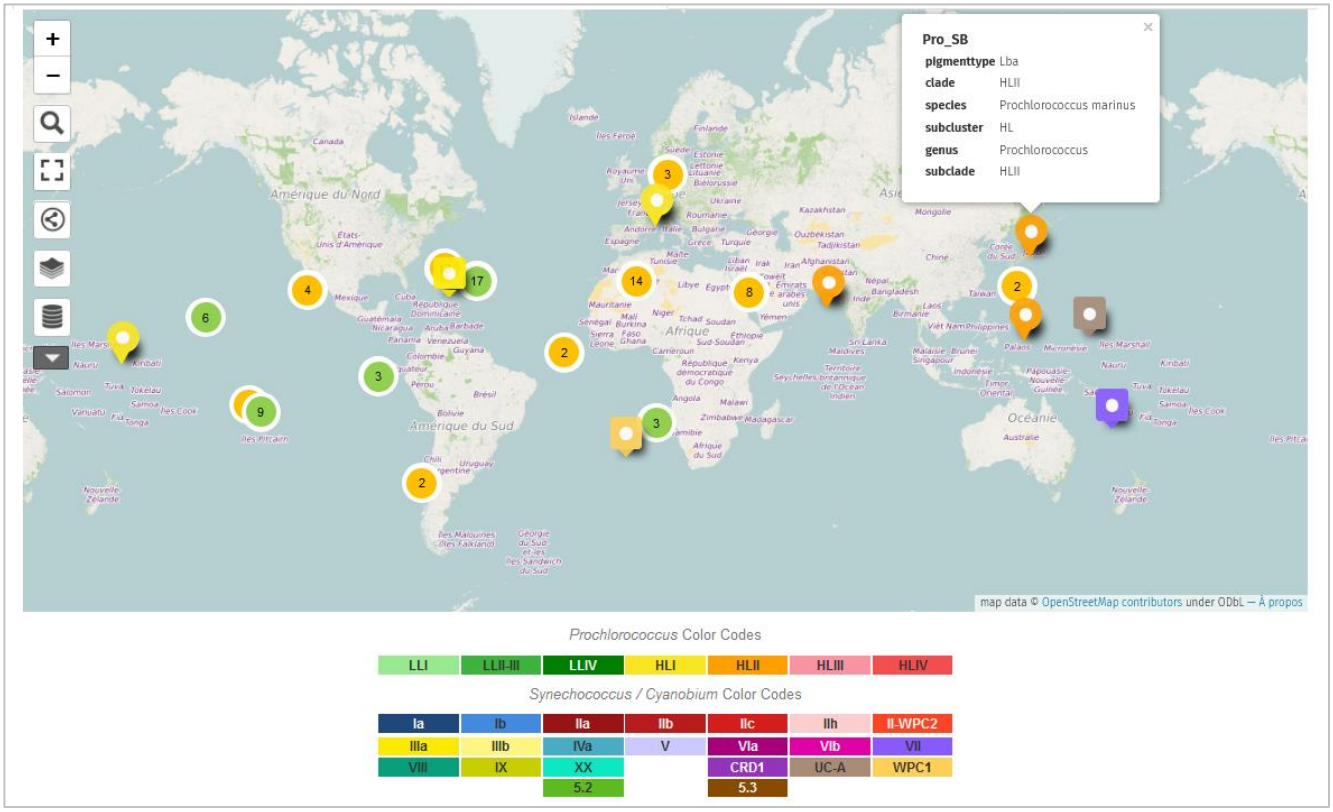


Figure 1: Map of the isolation sites of the different sequenced strains included in the Cyanorak v2.1 database. Green markers indicate *Prochlorococcus* strains and orange markers *Synechococcus* ones. Each marker can be expanded to reveal a 'call-out' that shows the strain name, isolation coordinates and depth, as shown for the North Atlantic Sea isolate *Prochlorococcus marinus* SB strain. The Search Data boxes shown on the left side of the map allows to search for specific strain(s), genotype, etc.

Genomic Search tools

The main entry to the Cyanorak v.2.1 database content is through the 'genomic search' menu, with three possible options. The first one is a 'quick search' of any term mainly through 'cluster number', 'gene names' and 'product descriptions' fields, a term that can be completed by stars before and/or after the term to extend the search. For instance, a search for dna* will provide a list a CDS clusters whose gene name annotation starts with "dna" (e.g. *dnaA*, *dnaB*, etc.) as well as a list of products whose description starts with DNA (e.g. DNA gyrase, DNA primase, etc.). Besides matching clusters, results also return under distinct tabs a list of matching CDS and (if any) RNAs that contain the searched term in their description. The 'advanced search' option allows the user to look into any field

documented in the Cluster, CDS or RNA pages, including functional categories (e.g., EC or K number, InterPro entries, GO terms, etc) and to select one, all or a specific set of strains. Finally, the 'phyletic cluster search' option is used to search for clusters, CDS and RNA that are shared by a selection of genomes and that can be either present or absent in other strains depending on the selected option. This search option is for instance most useful to identify genes specific of a particular strain combination such as all *Prochlorococcus*, a given clade or a given pigment type.

CLOG page

By clicking on a Cyanorak CLOG number (format: CK_XXXXXXXX) in the result list of any term search (see above), the user is sent to a cluster page providing a full description of the function and phyletic pattern of the corresponding CLOG. An indication in the upper right corner of the CLOG page specifies whether the cluster has or not been manually annotated, i.e. whether an expert has edited and validated its sequence content and functional annotation. The cluster page contains several fields, including from top to bottom: i) the corresponding gene identifier in the Cyanorak v1 database (if any), generally corresponding to the last digits of the Cyanorak v2.1 CLOG number, ii) a gene name and its synonyms (if any) as well as the product description, iii) functional categories including COG (5) and EggNOG (6) identifiers and their description, CyOG number (as reported in (31)), Enzyme Commission (EC) and K numbers with hyperlinks to their descriptions in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (www.kegg.jp), 'TIGR Roles' and custom-designed 'Cyanorak Roles', the latter being largely derived from the previous but providing more details on photosynthetic and other key cyanobacterial processes, as well as gene ontology (GO) terms and their description; iv) results of protein domain and motif searches, including TIGRFAMs, PFAMs, ProSite patterns and profiles, as well as InterPro entries, v) numbers of related clusters, i.e. possible paralogs, and vi) a phyletic pattern providing the distribution of the genes in the different genomes, classified by taxonomy (genus, clade and, for *Synechococcus* SC 5.1 strains only,

subclades, according to (19)), and indicating the pigment type of the corresponding strain (Fig. S2, (25, 32, 33)). The bottom of the page lists the different members of the CLOG, with their initial annotation, a useful information when the annotation was made either automatically or by other research groups.

On the top left of the 'cluster page' is a link to the 'genomic context', which displays the four genes upstream and downstream the selected gene in all members of the CLOG. Two possible representations of the context are accessible through a toggle button: genes are shown either in relative size or all at the same size (Fig. 2A-B). To ease comparisons, the central gene is always represented in forward direction whatever its original direction in the genome and the context is arranged accordingly. Each CLOG has a given (random) color and genome context can be regenerated around any gene of the current context by clicking on the name of the corresponding gene, while clicking on a CLOG number, in the relative size mode, opens the corresponding CLOG page.

Links at the bottom of the cluster page allows the user to export all sequences of the CLOG at a time as an amino acid or nucleotide FASTA file. The descriptor of each sequence in the export is standardized and provides the genus abbreviated to the three first letters (Cya|Syn|Pro); strain name; SC, clade or subclade depending on finest taxonomic level available for the strain as in Farrant et al., 2016; pigment type as in (25); Cyanorak ORF_ID; gene positions and strand in the genome; Cyanorak CLOG number and gene name if any (e.g. >Syn_A15-24_IIIa_3c|CK_Syn_A15-24_02629:2153016-2154431:1|CK_00000125|dnaB).

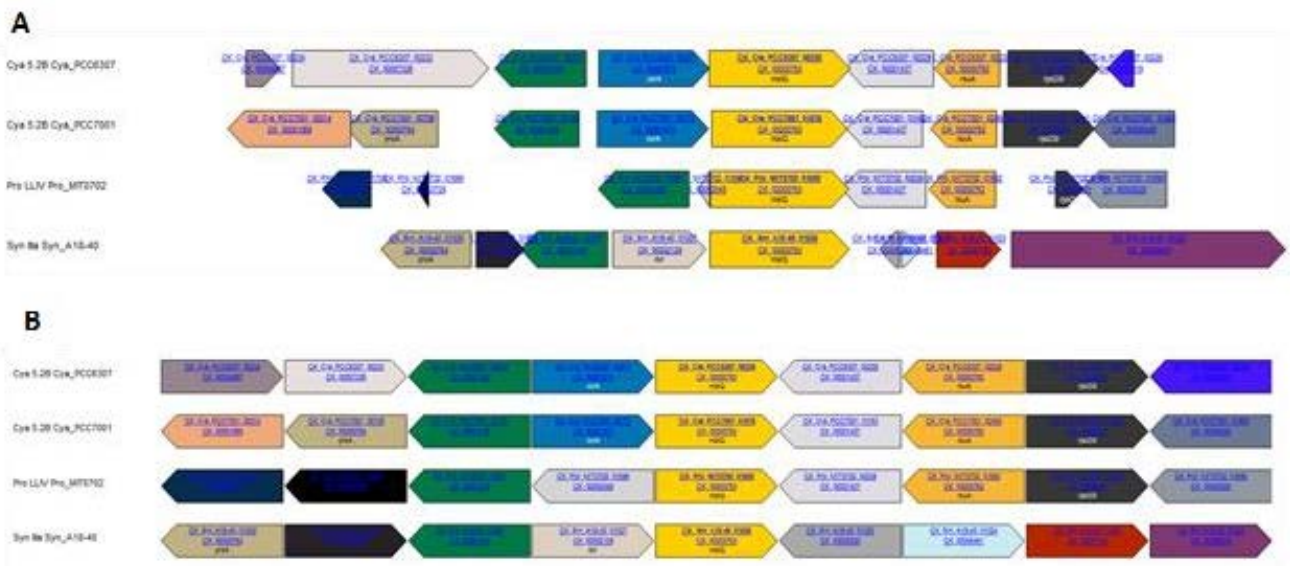


Figure 2: Genomic context of the malQ gene. A, genes represented in relative size. B, genes represented at the same size.

Gene page

By clicking on any gene in the cluster page or the relevant search result tab, the user accesses to the 'gene page', which includes most fields previously described for the cluster page. Specificities compared to the latter includes i) the source and location of the gene, namely the strain name, contig and gene location (position and strand) on the contig, generally consisting in the whole chromosome, ii) a series of identifiers in Cyanorak and, if relevant, in other databases (Genbank, RefSeq, etc.) and iii) the gene sequence in nucleotides and (for CDS) in amino acids. It must be stressed that this page contains the initial annotation of the gene (e.g. if the genome was retrieved from Genbank), and the latter often differs from the CLOG annotation, typically much more complete and extensive, especially if the cluster was manually curated. All genes included in Cyanorak (even when retrieved from public databases) were given in addition to their initial ORF_ID a unique Cyanorak ORF_ID with the standardized format CK_Genus_Strain_XXXXX (e.g. CK_Syn_PROS-U-1_00601) in order to homogenize the gene names and ease the identification of the genome source. Also noteworthy is that the genomes included in Cyanorak, even those that have been sequenced by other teams, have all been manually curated to some extent, including prediction of missing genes or removal of wrong

predictions, and thus differ from their counterpart in other public databases (Genbank, RefSeq, etc.) not only regarding their annotation (made at CLOG level) but also their gene content.

BLAST

An indispensable complement to the Cyanorak database is the possibility for users to search any sequence in all genomes or proteomes of the database using several blast options available from the Blast page. This includes a Blast page allowing to blast a nucleotide or protein sequence against a selection of up to 59 genomes ('Blast a selection') or all genomes ('Blast all') using a classical blast interface. A third option is to use a Galaxy interface ('Galaxy blast') allowing users to blast several sequences in batch using the cluster of the ABiMs platform. Result of the Blast returns the Cyanorak ORF_ID, the strain taxonomy (at the SC, clades and/or subclade level) and pigment type, the CLOG number as well as the CLOG gene name and product (e.g. CK_Syn_A15-24_00652_III_IIIa_3c CK_00001060!rpoC1!DNA-directed RNA polymerase complex, gamma subunit).

JBrowse page

An additional menu allows to list the data generated by our group and available through the JBrowse viewer (34) for each strain (core/accessory genes, gained genes, genomic islands, operon predictions, transcriptomes, etc.) that can be quickly accessed by clicking on the 'J' logo next to each strain name (Fig. 3). This environment allows the user to visualize the whole annotated genome and to zoom in and out on each genome and to see the local context and detailed annotation of any gene, as derived from the 'CLOG page' (see above). A 'select tracks' menu give access (when available) to additional data associated to each genome, which can be selected by project, publication, category (e.g., core/accessory/HGT, microarrays differential expression, RNAseq differential expression, RNAseq coverage) and, in the case of transcriptomic data, acclimation conditions (e.g. low or high light), stress conditions (e.g. low temperature, UV radiations, etc), stress duration, etc. To date the

largest set of complementary data is available for *Synechococcus* sp. WH7803 given that this strain has been the subject of several comparative genomics (Doré et al., submitted) and physiological studies (Blot et al., 2011; Guyet et al., submitted).

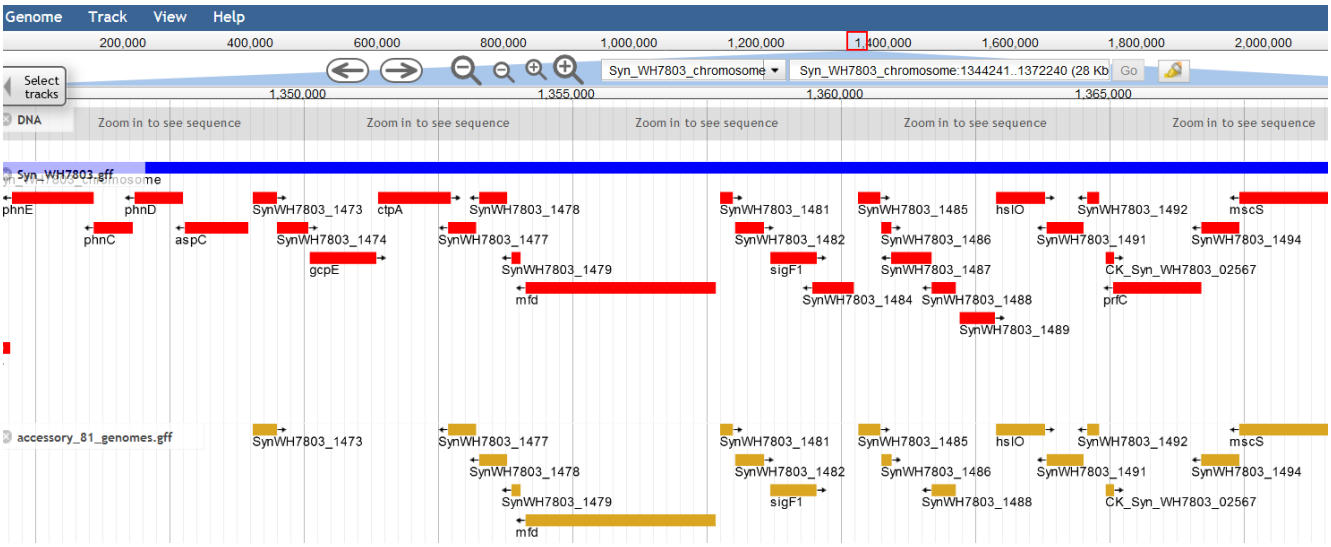


Figure 3: An example of genome visualization using the JBrowse plugin of Cyanorak v2.1.

Exports

Various exports are available from different pages of the Cyanorak v2.1 information system, including strains and genomes characteristics from the organisms page, annotated complete genomes from the individual genome pages, individual fasta protein and nucleotide sequences from gene pages and multifasta protein and nucleotide sequences from cluster pages. Furthermore for an internal use, the private version of Cyanorak automatically generates weekly exports on the local server including all previously mentioned exports available from the database but also a 'cluster attribute file', encompassing most information included in the cluster and gene pages for all Cyanorak CLOGs (see above), all genomes in multiple formats (fasta, gbk, gff), nucleotides and amino acid multifasta files per genome and per CLOG as well as the corresponding Blast databases. These exports, taking into

account the most recent modifications of the database, can be used to perform extensive comparative (meta)genomic/transcriptomic analyses using the tools available on the ABiMs calculation cluster.

Discussion

As the most ancient photosynthetic organisms, cyanobacteria had a key role in the oxidation of the primitive Earth atmosphere (35) but also in the primary endosymbiosis, an event that led to the advent of green and red algae and ultimately to all eukaryotic oxygenic phototrophs (36). Besides their relevance in evolutionary biology, cyanobacteria have also an interest in ecology, given their ubiquity and abundance in many ecosystems, including oceans and deserts (37) and the noxious impacts of their bloom-forming toxic representatives in freshwater environments (38). For all these reasons, but also thanks to their fairly small genomes sizes, ranging from 1.4 to 11.6 Mbp, these microorganisms have been the subject of many sequencing projects (8, 12, 39), which in turn triggered the generation of a number of dedicated genome databases. The most ancient of these databases is Cyanobase (http://genome.microbedb.jp/CyanoBase), initially created to provide access to the first sequenced cyanobacterial genome, the model freshwater *Synechocystis* sp. PCC 6803 (40, 41). This database has since then been extended to host many more recently sequenced freshwater and marine cyanobacterial genomes (376 entries, including 86 complete genomes in April 2019), but it is 'under maintenance' since summer 2019. Although this database provides many useful genomic information, this is not a CLOG database and is therefore not conceived to make extensive genomic comparisons. Also worth noting, CyanoClust (http://gclust.c.u-tokyo.ac.jp/CyanoClust/) is a database of homologous groups initially limited to cyanobacteria and plastids and which was more recently extended to a heterotrophic bacteria and Archaea (42). It provides lists of orthologs generated by the program Gclust, but the functional annotation is limited to the original product description associated with individual members of each CLOG. The database that was most similar to Cyanorak v2.1 but which is currently no more accessible, was the '*Prochlorococcus* portal' a.k.a. 'Proportal' (43), which

was also listing CLOGs from *Prochlorococcus* and marine *Synechococcus* with a strong focus on the former organism.

Compared to these databases, Cyanorak is particularly function-oriented and aims at providing rich and up-to-date functional annotations to CLOGs, with a preference for those derived from genes or proteins that were characterized in cyanobacteria. In contrast to most large CLOG databases currently available, such as COG (5) or EggNOG (6) that encompass very distantly related organisms, Cyanorak is focused on *Prochlorococcus* and *marine Synechococcus/Cyanobium*, i.e. a monophyletic and homogenous group of microorganisms sharing a similar morphology and life style, making more reliable the assumption that reciprocal best hits in different genomes truly correspond to orthologs. Thanks to this CLOG-based approach, the continuous expert curation efforts deployed since the mid-2000's have benefited to all genomes of the Cyanorak database, even those initially retrieved from Genbank. Furthermore, a number of genes that were missing in these often automatically annotated genomes have been added, while many overpredictions have been suppressed, so that both gene content and annotations differ between genomes in Cyanorak and their counterparts in large public databases.

Another important asset of Cyanorak is the variety of tools for exploring and exporting genomes from the database. For instance, one can search for CLOGs common or specific to a particular phylogenetic group of interest, an approach which can provide clues to identify genes coding for a specific function. Thus, searching Cyanorak for homologs of the MpeZ, an enzyme involved in type IV chromatic acclimation (CA4), i.e. a reversible pigmentation change occurring in some *Synechococcus* strains when shifted from blue to green light (44), has allowed us to identify a second type of chromatic acclimaters, so-called CA4-B, which possess MpeW, a MpeZ homolog. Both *mpeZ* and *mpeW* genes are located in a specific genomic island, but the gene content, organization and genomic context were found to differ between CA4-A and CA4-B islands (25). Another interesting example concerns the chlorophyll (Chl) biosynthesis pathway. It is well known that *Prochlorococcus*

lacks monovinyl-Chl *a*, which is replaced by divinyl-Chl *a,* even in reaction centers (45, 46). Comparing the genomic context of the core *malQ* gene (coding for a glucanotransferase) between all genomes of the Cyanorak database shows that in marine *Synechococcus*, this gene is always preceded by an enzyme that reduces divinyl-Chl *a* into mononyl-Chl *a*, but surprisingly there are two possible mutually exclusive reductase genes depending on strains, either *dvr* (47) in strains from clades I-IV, VII, CRD1, WPC1 XX and UC-A or *cvrA* (48) in all other *Synechococcus* lineages (Fig. S3). Dvr and CvrA possess the same enzymatic function but share no sequence identity, and thus are analogs. In *Prochlorococcus* genomes, neither *dvr* nor *cvrA* are found upstream *malQ*, and none of these genes is found elsewhere in the genome, explaining why these strains are all incapable to produce monovinyl-Chl *a*.

Cyanorak v2.1 is also a repository for a variety of transcriptomic data, the interpretation of which heavily relies on the quality of genome annotation. In Cyanorak, these data are connected to the genome database through a JBrowse interface, which also give access to genomic features such as predicted operons, gained genes or the core or accessory nature of genes, which can be used to further refine the interpretation of gene expression data (see e.g. (49)).

FUTURE DEVELOPMENTS

The current version of the database includes rRNA, tRNA and tmRNA, but still no small RNAs, so we plan to add such information in a forthcoming release, at least for the most conserved sRNAs. Among other planned improvements of the database, although many gene starts have been corrected manually, amino acid alignments readily made from exports from CLOG pages show that a large number of those starts are still seemingly too short or too long in a number of genomes and we will develop an application that allows to automatically correct likely wrong starts, at least when N-termini are not too variable. Also, new fields will be added on the CLOG page, including for instance orthologs of each CLOG in relevant biological models, such as the freshwater cyanobacteria

Synechocystis PCC 6803 and *Synechococcus* sp. PCC 7942, the heterotrophic bacteria *Escherichia coli* and *Bacillus subtilis* or the higher plant *Arabidopsis thaliana*.

Future versions of the Cyanorak database will include a number of transcriptomes for several *Synechococcus* strains recently generated by our team but that are still unpublished as well as genomes newly released from public databases. It must be stressed that Cyanorak does not have the vocation to host all of the rapidly-growing number of incomplete single amplified- and metagenome assembled- genomes (SAGs and MAGs), apart from a few representative uncultivated lineages (e.g., *Synechococcus* EnvA/B, *Prochlorococcus* HLIII-VI). Thus, a pipeline is currently being developed to easily transfer the rich functional annotation of the Cyanorak genomes to these new partial genomic sequences and, more generally, to any picocyanobacterial environmental reads retrieved from metagenomes and metatranscriptomes. A few previous studies, where annotations made in Cyanorak were used to analyze omic data, have notably allowed us to i) compare the nitrogen assimilation capacities of *Prochlorococcus* populations from inside and outside the Agulhas rings in the South Atlantic Ocean (50), ii) highlight differences between *Prochlorococcus* and *Synechococcus* populations in their adaption and acclimation responses to iron deficiency in the vicinity of the Marquesas island (51) and iii) demonstrate through the global oceanic distribution of desaturase genes the key role of these enzymes involved in the modulation of membrane fluidity for the colonization of different thermal niches by distinct *Synechococcus* lineages (28). We envision that Cyanorak will become a reference genome database for the taxonomic and functional annotation of not only newly released genomes and transcriptomes of marine picocyanobacteria, but also the ever-increasing meta-omes of these microorganisms, which given the natural abundance and ubiquity in the marine environment constitute a significant part of all reads retrieved from the upper lit layer of marine waters.

Acknowledgments

We warmly thank Michelle G. Giglio and the Institute of Genome Science (IGS) staff for annotating 32 *Synechococcus* and *Cyanobium* genomes using the IGS Manatee pipeline. We also thank Justine Pittera, Garance Monier and Théo Sciandra for participating in the curation of the Cyanorak v2.1 database, the Roscoff Culture Collection and Sophie Mazard for maintaining and isolating some of the *Synechococcus* strains used in this study as well as the ABiMS platform for providing computational support for this work. This work was supported by the French "Agence Nationale de la Recherche" Programs SAMOSA (ANR-13-ADAP-0010) and CINNAMON (ANR-17-CE2-0014-01), the Genoscope project METASYN, the Natural Environment Research Council grant NE/I00985X/1 and the European Union's Seventh Framework Programs FP7 MicroB3 (Grant 287589) and MaCuMBA (Grant 311975).

References

1. Nordberg,H., Cantor,M., Dusheyko,S., Hua,S., Poliakov,A., Shabalov,I., Smirnova,T., Grigoriev,I. V. and Dubchak,I. (2014) The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Res.*, **42**, D26–D31.
2. Vallenet,D., Engelen,S., Mornico,D., Cruveiller,S., Fleury,L., Lajus,A., Rouy,Z., Roche,D., Salvignol,G., Scarpelli,C., *et al.* (2009) MicroScope: A platform for microbial genome annotation and comparative genomics. *Database*, **2009**, bap021.
3. Aziz,R.K.R.K., Bartels,D., Best,A.A., DeJongh,M., Disz,T., Edwards,R.A.R.A., Formsma,K., Gerdes,S., Glass,E.M.E.M., Kubal,M., *et al.* (2008) The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics*, **9**, 75.
4. Seemann,T. (2014) Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, **30**, 2068–2069.
5. Galperin,M.Y., Makarova,K.S., Wolf,Y.I. and Koonin,E. V. (2015) Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.*, **43**, D261–D269.
6. Huerta-Cepas,J., Szklarczyk,D., Heller,D., Hernández-Plaza,A., Forslund,S.K., Cook,H., Mende,D.R., Letunic,I., Rattei,T., Jensen,L.J., *et al.* (2019) EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.*, **47**, D309–D314.
7. Tatusova,T., Dicuccio,M., Badretdin,A., Chetvernin,V., Nawrocki,E.P., Zaslavsky,L., Lomsadze,A., Pruitt,K.D., Borodovsky,M. and Ostell,J. (2016) NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.*, **44**, 6614–6624.
8. Dufresne,A., Ostrowski,M., Scanlan,D.J.J., Garczarek,L., Mazard,S., Palenik,B.P.P., Paulsen,I.T.T., de Marsac,N.T.T., Wincker,P., Dossat,C., *et al.* (2008) Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol*, **9**, R90.
9. Doré,H., Farrant,G.K., Guyet,U., Haguait,J., Humily,F., Ratin,M., Pitt,F.D., Ostrowski,M., Six,C., Brillet-Guéguen,L., *et al.* (2020) Evolutionary mechanisms of genome diversification in marine picocyanobacteria. *Genome Biol.*, **submitted**.
10. Farrant,G.K., Doré,H., Cornejo-Castillo,F.M., Partensky,F., Ratin,M., Ostrowski,M., Pitt,F.D., Wincker,P., Scanlan,D.J., Iudicone,D., *et al.* (2016) Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proc. Natl. Acad. Sci.*, **113**, E3365–E3374.
11. Herdman,M., Castenholz,R.W., Waterbury,J.B. and Rippka,R. (2001) Form-genus XIII. *Synechococcus*. In Boone,D.R., Castenholz,R.W. (eds), *Bergey's Manual of Systematic Bacteriology*. Springer-Verlag, New York, Vol. 1, pp. 508–512.
12. Shih,P.M., Wu,D., Latifi,A., Axen,S.D., Fewer,D.P., Talla,E., Calteau,A., Cai,F., Tandeau de Marsac,N., Rippka,R., *et al.* (2013) Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc. Natl. Acad. Sci. USA*, **110**, 1053–1058.
13. Sánchez-Baracaldo,P. (2015) Origin of marine planktonic cyanobacteria. *Sci. Rep.*, **5**, 17418.
14. Badger,M.R. and Price,G.D. (2003) CO2 concentrating mechanisms in cyanobacteria: Molecular components, their diversity and evolution. *J. Exp. Bot.*, **54**, 609–622.
15. Huang,S., Wilhelm,S.W., Harvey,H.R., Taylor,K., Jiao,N. and Chen,F. (2012) Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *ISME J.*, **6**, 285–297.
16. Scanlan,D.J., Ostrowski,M., Mazard,S., Dufresne,A., Garczarek,L., Hess,W.R., Post,A.F., Hagemann,M., Paulsen,I. and Partensky,F. (2009) Ecological genomics of marine picocyanobacteria. *Microbiol. Mol. Biol. Rev.*, **73**, 249–299.
17. Cabello-Yeves,P.J., Haro-Moreno,J.M., Martin-Cuadrado,A.B., Ghai,R., Picazo,A., Camacho,A. and Rodriguez-Valera,F. (2017) Novel *Synechococcus* genomes reconstructed from freshwater reservoirs. *Front. Microbiol.*, **8**, 1151.
18. Cabello-Yeves,P.J., Picazo,A., Camacho,A., Callieri,C., Rosselli,R., Roda-Garcia,J.J., Coutinho,F.H. and Rodriguez-Valera,F. (2018) Ecological and genomic features of two widespread freshwater picocyanobacteria. *Environ. Microbiol.*, **20**, 3757–3771.
19. 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.

20. Blot,N., Mella-Flores,D., Six,C., Le Corguillé,G., Boutte,C., Peyrat,A., Monnier,A., Ratin,M., Gourvil,P., Campbell,D. a, *et al.* (2011) Light history influences the response of the marine cyanobacterium *Synechococcus* sp. WH7803 to oxidative stress. *Plant Physiol.*, **156**, 1934–1954.
21. Garczarek,L., Dufresne,A., Blot,N., Cockshutt,A.M., Peyrat,A., Campbell,D.A., Joubin,L. and Six,C. (2008) Function and evolution of the *psbA* gene family in marine *Synechococcus*: *Synechococcus* sp. WH7803 as a case study. *ISME J.*, **2**, 937–953.
22. Kolowrat,C., Partensky,F., Mella-Flores,D., Le Corguillé,G., Boutte,C., Blot,N., Ratin,M., Ferréol,M., Lecomte,X., Gourvil,P., *et al.* (2010) Ultraviolet stress delays chromosome replication in light/dark synchronized cells of the marine cyanobacterium *Prochlorococcus marinus* PCC9511. *BMC Microbiol.*, **10**, 204.
23. Partensky,F. and Garczarek,L. (2010) *Prochlorococcus*: Advantages and limits of minimalism. *Ann. Rev. Mar. Sci.*, **2**, 305–331.
24. Mella-Flores,D., Six,C., Ratin,M., Partensky,F., Boutte,C., Le Corguillé,G., Marie,D., Blot,N., Gourvil,P., Kolowrat,C., *et al.* (2012) *Prochlorococcus* and *Synechococcus* have evolved different adaptive mechanisms to cope with light and UV stress. *Front. Microbiol.*, **3**, 285.
25. Humily,F., Partensky,F., Six,C., Farrant,G.K., Ratin,M., Marie,D. and Garczarek,L. (2013) A gene island with two possible configurations is involved in chromatic acclimation in marine *Synechococcus*. *PLoS One*, **8**, e84459.
26. Partensky,F., Flores,D.M., Six,C., Garczarek,L. and Czjzek,M. (2018) Comparison of photosynthetic performances of marine picocyanobacteria with different configurations of the oxygen- evolving complex. *Photosynth. Res.*, **0**, 0.
27. Pittera,J., Jouhet,J., Breton,S., Garczarek,L., Partensky,F., Maréchal,É., Nguyen,N.A., Doré,H., Ratin,M., Pitt,F.D., *et al.* (2018) Thermoacclimation and genome adaptation of the membrane lipidome in marine *Synechococcus*. *Environ. Microbiol.*, **20**, 612–631.
28. Breton,S., Jouhet,J., Guyet,U., Gros,V., Pittera,J., Demory,D., Partensky,F., Doré,H., Ratin,M., Maréchal,E., *et al.* (2020) Unveiling membrane thermoregulation strategies in marine picocyanobacteria. *New Phytol.*, **225**, 2396–2410.
29. Li,L., Stoeckert,C.J.J. and Roos,D.S. (2003) OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.*, **13**, 2178–2189.
30. Finn,R.D., Clements,J. and Eddy,S.R. (2011) HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.*, **39**, W29–W37.
31. Mulkidjanian,A.Y.Y., Koonin,E.V. V, Makarova,K.S.S., Mekhedov,S.L.L., Sorokin,A., Wolf,Y.I.I., Dufresne,A., Partensky,F., Burd,H., Kaznadzey,D., *et al.* (2006) The cyanobacterial genome core and the origin of photosynthesis. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 13126–13131.
32. Moore,L.R. and Chisholm,S.W. (1999) Photophysiology of the marine cyanobacterium *Prochlorococcus*: Ecotypic differences among cultured isolates. *Limnol. Oceanogr.*, **44**, 628–638.
33. Grébert,T., Doré,H., Partensky,F., Farrant,G.K., Boss,E.S., Picheral,M., Guidi,L., Pesant,S., Scanlan,D.J., Wincker,P., *et al.* (2018) Light color acclimation is a key process in the global ocean distribution of *Synechococcus* cyanobacteria. *Proc. Natl. Acad. Sci.*, **115**, E2010–E2019.
34. Buels,R., Yao,E., Diesh,C.M., Hayes,R.D., Munoz-Torres,M., Helt,G., Goodstein,D.M., Elsik,C.G., Lewis,S.E., Stein,L., *et al.* (2016) JBrowse: A dynamic web platform for genome visualization and analysis. *Genome Biol.*, **17**, 66.
35. Lyons,T.W., Reinhard,C.T. and Planavsky,N.J. (2014) The rise of oxygen in Earth's early ocean and atmosphere. *Nature*, **506**, 307–315.
36. Archibald,J.M. (2012) The evolution of algae by secondary and tertiary endosymbiosis. *Adv. Bot. Res.*, **64**, 87–118.
37. Garcia-Pichel,F., Belnap,J., Neuer,S. and Schanz,F. (2003) Estimates of cyanobacterial biomass and its distribution. *Arch. Hydrobiol. Suppl. Algol. Stud.*, **109**, 213–228.
38. Huisman,J., Codd,G.A., Paerl,H.W., Ibelings,B.W., Verspagen,J.M.H. and Visser,P.M. (2018) Cyanobacterial blooms. *Nat Rev Microbiol*, **16**, 471–483.
39. Kettler,G.C., Martiny,A.C., Huang,K., Zucker,J., Coleman,M.L., Rodrigue,S.,

Chen,F., Lapidus,A., Ferriera,S., Johnson,J., *et al.* (2007) Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet.*, **3**, e231.

40. Nakamura,Y., Kaneko,T., Hirosawa,M., Miyajima,N. and Tabata,S. (1998) CyanoBase, a www database containing the complete nucleotide sequence of the genome of *Synechocystis* sp. strain PCC6803. *Nucleic Acids Res.*, **26**, 63–67.
41. Fujisawa,T., Narikawa,R., Maeda,S.I., Watanabe,S., Kanesaki,Y., Kobayashi,K., Nomata,J., Hanaoka,M., Watanabe,M., Ehira,S., *et al.* (2017) CyanoBase: A large-scale update on its 20th anniversary. *Nucleic Acids Res.*, **45**, D551–D554.
42. Sasaki,N. V. and Sato,N. (2010) CyanoClust: comparative genome resources of cyanobacteria and plastids. *Database (Oxford).*, **2010**, Bap025.
43. Kelly,L., Huang,K.H., Ding,H. and Chisholm,S.W. (2012) ProPortal: a resource for integrated systems biology of *Prochlorococcus* and its phage. *Nucleic Acids Res*, **40**, D632–D640.
44. Shukla,A., Biswas,A., Blot,N., Partensky,F., Karty,J.A.A., Hammad,L.A.A., Garczarek,L., Gutu,A., Schluchter,W.M.M. and Kehoe,D.M.M. (2012) Phycoerythrin-specific bilin lyase-isomerase controls blue-green chromatic acclimation in marine *Synechococcus*. *Proc Natl Acad Sci U S A*, **109**, 20136–20141.
45. Goericke,R. and Repeta,D.J. (1992) The pigments of *Prochlorococcus marinus* : the presence of divinyl chlorophyll *a* and *b* in a marine prochlorophyte. *Limnol. Oceanogr.*, **37**, 425–433.
46. Ito,H. and Tanaka,A. (2011) Evolution of a divinyl chlorophyll-based photosystem in *Prochlorococcus*. *Proc Natl Acad Sci USA*, **108**, 18014–18019.
47. Nagata,N., Tanaka,R., Satoh,S. and Tanaka,A. (2005) Identification of a vinyl reductase gene for chlorophyll synthesis in *Arabidopsis thaliana* and implications for the evolution of *Prochlorococcus* species. *Plant Cell*, **17**, 233–240.
48. Islam,M.R., Aikawa,S., Midorikawa,T., Kashino,Y., Satoh,K. and Koike,H. (2008) *slr1923* of *Synechocystis* sp. PCC6803 is essential for conversion of 3,8-divinyl(proto)chlorophyll(ide) to 3-monovinyl(proto)chlorophyll(ide). *Plant Physiol.*, **148**, 1068–1081.
49. Guyet,U., Nguyen,N.A., Doré,H., Haguait,J., Pittera,J., Conan,M., Ratin,M., Corre,E., Le Corguillé,G., Brillet-Guéguen,L., *et al.* Synergic effects of temperature and irradiance on the physiology of the marine *Synechococcus* strain WH7803. *Front Microbiol*.
50. Villar,E., Farrant,G.K., Follows,M., Garczarek,L., Speich,S., Audic,S., Bittner,L., Blanke,B., Brum,J.R., Brunet,C., *et al.* (2015) Environmental characteristics of Agulhas rings affect interocean plankton transport. *Science*, **348**, 1261447.
51. Caputi,L., Carradec,Q., Eveillard,D., Kirilovsky,A., Pelletier,E., Pierella Karlusich,J.J., Rocha Jimenez Vieira,F., Villar,E., Chaffron,S., Malviya,S., *et al.* (2019) Community-Level Responses to Iron Availability in Open Ocean Plankton Ecosystems. *Global Biogeochem. Cycles*, **33**, 391–419.

Supplementary information

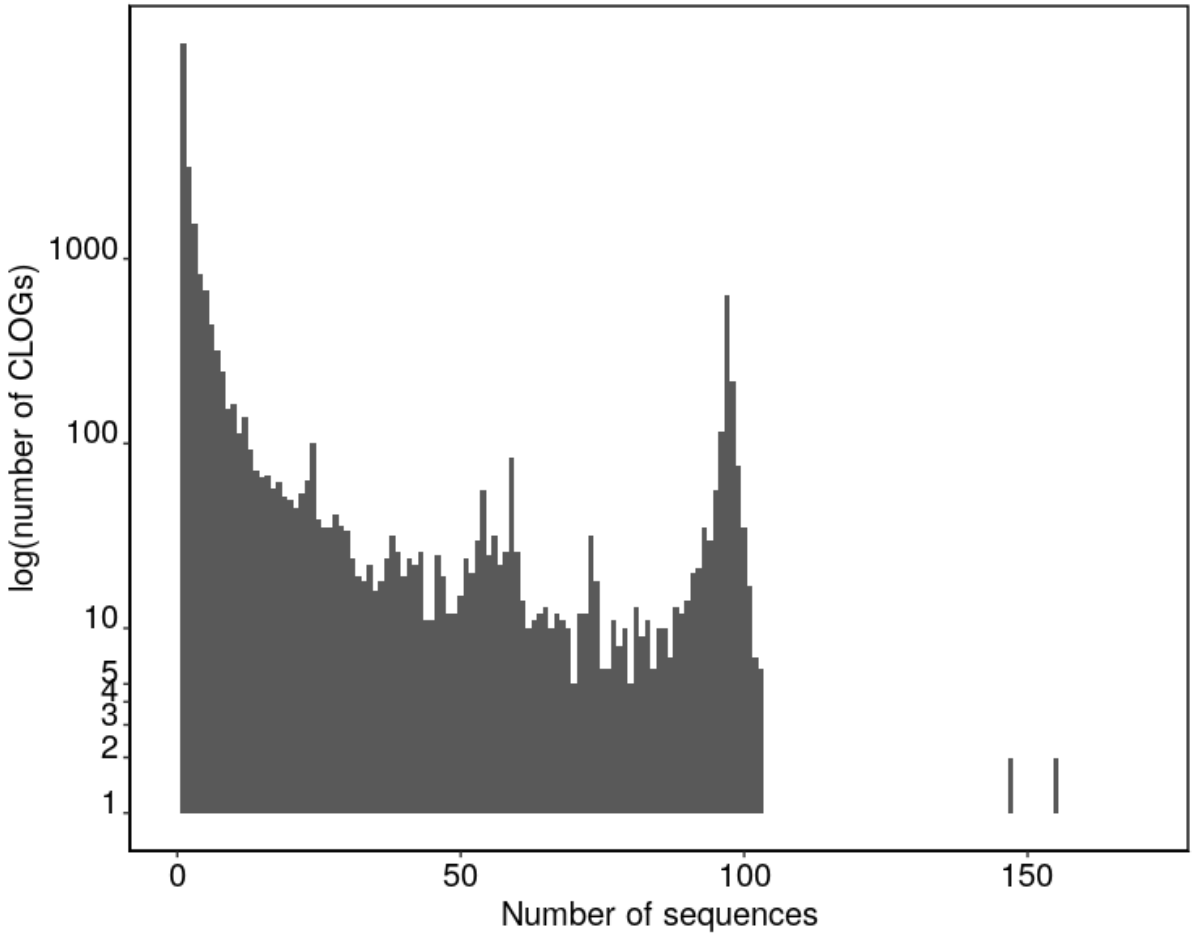


Figure S1: Distribution of the number of sequences per CLOG.

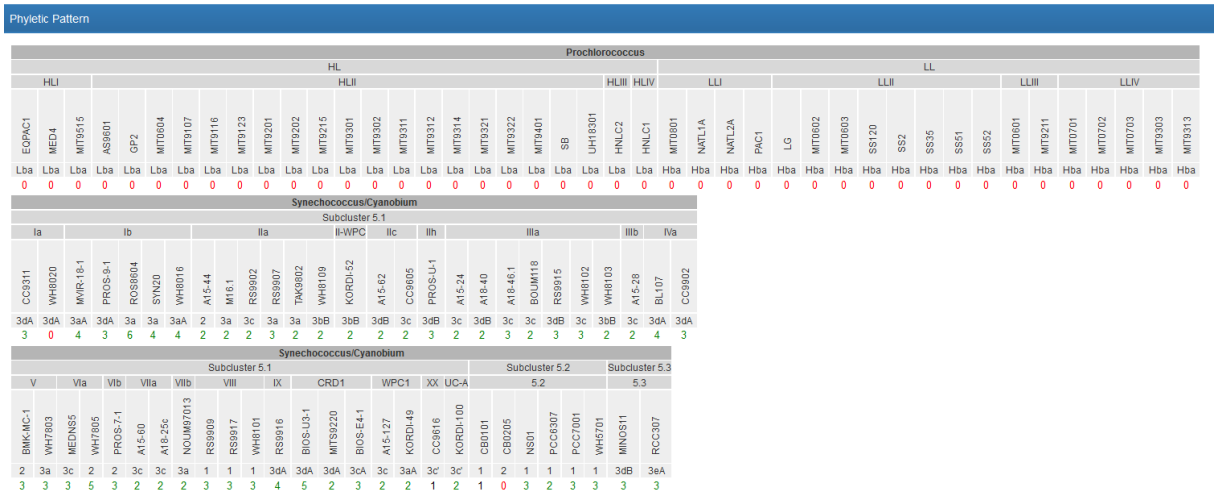


Figure S2: Example of phyletic pattern for the psbA gene copies encoding the D1.2 isoform of the D1 protein of photosystem II. Note that this isoform is absent from all Prochlorococcus strains, while it can be present in up to 6 copies in Synechococcus/Cyanobium strains. No psbA was modelled in CB0205 due to poor assembly of identical gene copies in this genome.

Figure S4: Example of the genomic context of the malQ gene. The genomic context is highly conserved downstream this core gene but more variable upstream: Synechococcus strains possess either cvrA or dvr, while Prochlorococcus possess neither of these gene (see text for explanations).

I.2 Evolutionary mechanisms of long-term genome diversification associated with niche partitioning in marine picocyanobacteria

Résumé de l'article

Les picocyanobactéries marines des genres *Prochlorococcus* et *Synechococcus* sont les organismes photosynthétiques les plus abondants sur Terre, une réussite écologique que l'on pense liée au partitionnement différentiel d'écotypes distincts en niches écologiques spécifiques. Toutefois, les processus sous-jacents qui ont régi la diversification de ces micro-organismes et l'apparition de traits phénotypiques liés à des niches commencent à peine à être élucidés. Ici, en comparant 81 génomes, dont 34 nouveaux *Synechococcus*, nous avons exploré les processus évolutifs qui ont façonné la diversité génomique des picocyanobactéries. L'étalonnage temporel d'un arbre de protéines core a montré que les gains/pertes de gènes se produisaient à un taux étonnamment bas entre les différentes lignées, avec par exemple 5.6 gènes gagnés par million d'années (My) pour la principale lignée de *Synechococcus* (sous-groupe 5.1), parmi lesquels seulement 0,71 / My ont été fixés sur le long terme. Les comparaisons du contenu génétique ont révélé un certain nombre de candidats impliqués dans l'adaptation des nutriments, dont une grande proportion se trouve dans des îlots génomiques partagées entre des souches plus ou moins proches d'un point vue taxonomique, identifiés grâce à une approche originale de construction de réseaux. Il est intéressant de noter que des souches représentatives des différents écotypes coexistant dans les eaux pauvres en phosphore (*Synechococcus* clades III, WPC1 et sous-groupe 5.3) ont montré qu'elles présentaient différentes stratégies d'adaptation à cette limitation. En revanche, nous avons trouvé peu de gènes potentiellement impliqués dans l'adaptation à la température en comparant des thermotypes froids et chauds. En effet, la comparaison des séquences de protéines core a mis en évidence des variants de séquence spécifiques aux thermotypes froids, notamment impliquées dans la biosynthèse des caroténoïdes et la réponse au stress oxydatif, révélant que l'adaptation à long terme aux niches thermiques repose sur des substitutions d'acides aminés plutôt que sur la variation du contenu des gènes. Dans l'ensemble, cette étude ne se contente pas de déchiffrer les rôles respectifs des gains/pertes de gènes et de la variation de séquence, mais présente également de nombreux gènes candidats probablement impliqués dans le partitionnement des niches de deux membres clés du phytoplancton marin.

Evolutionary mechanisms of long-term genome diversification associated with niche partitioning in marine picocyanobacteria

Hugo Doré[1], Gregory K. Farrant[1], Ulysse Guyet[1], Julie Haguait[2], Florian Humily[1], Morgane Ratin[1], Frances D. Pitt[3], Martin Ostrowski[3,4], Christophe Six[1], Loraine Brillet-Guéguen[5,6], Mark Hoebeke[5], Antoine Bisch[5], Gildas Le Corguillé[5], Erwan Corre[5], Karine Labadie[7], Jean-Marc Aury[7], Patrick Wincker[8], Dong Han Choi[9,10], Jae Hoon Noh[9,11], Damien Eveillard[2,12], David J. Scanlan[3], Frédéric Partensky[1], and Laurence Garczarek[1,12*]

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

[2] LS2N, UMR CNRS 6004, IMT Atlantique, ECN, Université de Nantes, 44035 Nantes, France

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

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

[5] CNRS, FR 2424, ABiMS Platform, Station Biologique de Roscoff (SBR), Roscoff, France

[6] Sorbonne Université, CNRS, UMR 8227, Integrative Biology of Marine Models (LBI2M), Station Biologique de Roscoff (SBR), Roscoff, France

[7] Genoscope, Institut de biologie François-Jacob, Commissariat à l'Energie Atomique (CEA), Université Paris-Saclay, Evry, France

[8] Génomique Métabolique, Genoscope, Institut de biologie François Jacob, CEA, CNRS, Université d'Evry, Université Paris-Saclay, Evry, France.

[9] Marine Ecosystem Research Center, Korea Institute of Ocean Science and Technology, Busan 49111, Korea

[10] Ocean Science and Technology School, Korea Maritime and Ocean University, Busan 49112, Korea

[11] Department of Marine Biology, Korea University of Science and Technology, Daejeon 34113, Korea

[12] Research Federation (FR2022) *Tara* Océans GO-SEE, Paris, France

[*] Corresponding author: laurence.garczarek@sb-roscoff.fr

Short title: Genome diversification mechanisms in marine picocyanobacteria

Keywords: Marine cyanobacteria, *Prochlorococcus*, *Synechococcus*, comparative genomics, niche adaptation, amino-acid substitutions, genomic islands, evolution.

SUBMITTED

Abstract

Marine picocyanobacteria of the genera *Prochlorococcus* and *Synechococcus* are the most abundant photosynthetic organisms on Earth, an ecological success thought to be linked to the differential partitioning of distinct ecotypes into specific ecological niches. However, the underlying processes that governed the diversification of these microorganisms and the appearance of niche-related phenotypic traits are just starting to be elucidated. Here, by comparing 81 genomes, including 34 new *Synechococcus*, we explored the evolutionary processes that shaped the genomic diversity of picocyanobacteria. Time-calibration of a core-protein tree showed that gene gain/loss occurred at an unexpectedly low rate between the different lineages, with e.g. 5.6 genes gained per million years (My) for the major *Synechococcus* lineage (sub-cluster 5.1), among which only 0.71/My have been fixed in the long term. Gene content comparisons revealed a number of candidates involved in nutrient adaptation, a large proportion of which are located in genomic islands shared between either closely or more distantly-related strains, as identified using an original network construction approach. Interestingly, strains representative of the different ecotypes co-occurring in phosphorus-depleted waters (*Synechococcus* clades III, WPC1 and sub-cluster 5.3) were shown to display different adaptation strategies to this limitation. In contrast, we found few genes potentially involved in adaptation to temperature when comparing cold and warm thermotypes. Indeed, comparison of core protein sequences highlighted variants specific to cold thermotypes, notably involved in carotenoid biosynthesis and the oxidative stress response, revealing that long-term adaptation to thermal niches relies on amino acid substitutions rather than on gene content variation. Altogether, this study not only deciphers the respective roles of gene gains/losses and sequence variation but also uncovers numerous

gene candidates likely involved in niche partitioning of two key members of the marine phytoplankton.

Introduction

Understanding how phytoplankton species have adapted to the marine environment, a dynamic system through time and space, is a significant challenge, notably in the context of rapid global change [1–4]. Even though these microorganisms might adapt more rapidly than larger organisms to environmental change due to their short generation times and large population sizes, the underlying mechanisms and timescales required for such evolutionary processes to occur remain mostly unknown. One of the best ways to better understand these processes is by deciphering the links between current genomic diversity and niche occupancy of these organisms. Such an approach requires complete genomes with representatives of distinct ecological niches, a resource which remains limited even with the advent of high-throughput sequencing and the multiplication of partial single amplified genomes (SAGs;[5–9]) or metagenomes assembled genomes (MAGs; [10,11]). Due to their ubiquity, their natural abundance *in situ*, the occurrence of well-defined ecotypes and good knowledge of how environmental parameters influence their biogeography, marine picocyanobacteria constitute excellent model organisms to tackle evolutionary processes involved in niche partitioning.

Synechococcus and *Prochlorococcus* are the two most abundant photosynthetic organisms on Earth [12,13]. As major primary producers, they have a pivotal role in CO_2 fixation and carbon export and are key players in marine trophic networks [14–16]. Although these organisms often co-occur in (sub)tropical and temperate waters, *Synechococcus* is present from the equator to sub-polar waters, while the distribution of *Prochlorococcus* is restricted to the latitudinal band between 45°N and 40°S [15,17,18]. This broad distribution

implies that these two microorganisms 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 [13,19–22].

The ability of marine picocyanobacteria to occupy various niches is likely related to the high intrinsic genetic diversity of these taxa. The *Synechococcus*/*Cyanobium* radiation has been split into three main groups, called Sub-Clusters (hereafter SC) 5.1 to 5.3 [23,24]. While members of SC 5.2, currently encompassing strains assigned to both the *Synechococcus* and *Cyanobium* genera, are restricted to near coastal and estuarine areas, SC 5.1 and 5.3 are mainly marine, with SC 5.1 dominating in most oceanic waters and showing the highest genetic diversity currently comprising 18 distinct clades and 40 sub-clades so far described [25,26]. The *Prochlorococcus* genus forms a branch at the base of the *Synechococcus* SC 5.1 radiation and although it includes seven major lineages, usually referred to as 'clades', the whole genus is actually equivalent to a single marine *Synechococcus* clade from a phylogenetic viewpoint [21,24,27]. Lineages thriving in the upper mixed layer, so-called High Light-adapted (HL) clades, are genetically distinct from those occupying the bottom of the euphotic zone, so-called Low Light-adapted (LL) clades. Furthermore, while members of HLI were shown to colonize subtropical and temperate waters, HLII to IV are adapted to higher temperatures [17,28,29], with HLII colonizing N-poor areas and HLIII and IV being restricted to iron(Fe)-limited environments [7,30,31]. For *Synechococcus*, distribution and environmental preferences have only been well characterized for the five 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, thriving in oligotrophic intertropical areas, with clades II and III predominating in N- and P-poor regions, respectively [20–22,32,33]. Finally, members of clade

CRD1 were recently found to be dominant in large Fe-depleted areas of the world Ocean [21,22]. Even though clades globally occupy distinct niches, it was also shown that distinct ecotypes within *Prochlorococcus* and *Synechococcus* clades can display specific distribution patterns [6,25,34], with for instance distinct genetic groups within clades II and CRD1 colonizing different thermal niches [21].

Despite good knowledge of both their genetic diversity and environmental preferences, little is known about how environmental factors influence genome diversity and shape the community structure of marine picocyanobacteria, especially for *Synechococcus*. However, the development of high throughput sequencing techniques now allows such questions to be addressed. In particular, comparative genomics approaches applied to bacteria have revealed the high variability of microbial gene content, even for closely related strains sometimes displaying identical 16S rRNA sequences. They notably led to the definition of i) the core genome, the conserved part of the genome that encompasses genes shared by all strains, and ii) the flexible genome, the content of which is much more variable and dependent on the local biotic and abiotic environment [35,36]. In cyanobacteria, previous studies based on multiple genome comparisons have shown that these organisms still present a so-called 'open pan-genome' [37–39]. Indeed, each newly sequenced genome brings novel genes without diversity saturation, and this holds true for *Prochlorococcus* and *Synechococcus*, for which only 17 [40,41] and 14 genomes, respectively [23,39] have so far been compared. These studies thus highlight that the genomic diversity of natural populations is still mostly under-sampled, which strongly limits the interpretation of comparative genomic analyses. Here, we use a dataset of 81 non-redundant genomes of marine or halotolerant picocyanobacteria, of which 34 are newly sequenced complete *Synechococcus* genomes, to further assess the genomic diversity within these genera and how occupancy of new realized niches has

impacted the evolution of these genomes. Analysis of this unprecedented genome dataset with original bioinformatic tools allowed us to estimate the relative contribution of gene gains/losses and sequence divergence on the diversification of marine picocyanobacteria and to highlight key processes involved in their adaptation to various environmental niches.

Results

Picocyanobacteria exhibit a wide intra-clade genomic diversity

In order to expand the coverage of *Synechococcus* in available marine picocyanobacterial genomes, 34 new strains were sequenced from cultured isolates, resulting in a quasi-doubling of the current number of complete or near-complete genomes publicly available for this genus. Strains were selected to cover the extent of the phylogenetic and pigment diversity of *Synechococcus,* as well as maximize their geographic origin and trophic regimes of their isolation site (Fig. 1, Supplementary Table S1). It should be noted though, that no cultured isolates are available yet for the EnvA and EnvB clades [21,25]. The use of Wisescaffolder [42] allowed us to close 29 out of the 31 genomes sequenced by the Genoscope and the Center for Genomic Research, with only one gap remaining in strains RS9915 and BOUM118 (both in the giant gene *swmB* [43,44]) and three gaps in strain BIOS-E4-1 (two in genes encoding a PQQ enzyme repeat family protein and one in an LVIVD repeat family protein). This high-quality genome dataset constitutes a key asset for comparative genomics analyses. Consistent with the genome streamlining that occurred in most *Prochlorococcus* lineages [23,40,45], average genome size and GC% are expectedly lower in *Prochlorococcus* (1.815 Mb and 34.8%, respectively) than in *Synechococcus*/*Cyanobium* (2.533 Mb and 59.18% respectively), with genome sizes ranging from 1.625 Mb for *Prochlorococcus* HLII strain GP2

to 3.342 Mb for *Cyanobium gracile* PCC 6307 (SC 5.2) and GC% from 30.8% (EQPAC1, MED4 and MIT9515) to 68.7% (PCC 7001 and PCC 6307, Supplementary Table S1). Of note, members of the cold-adapted *Synechococcus* clades I and IV exhibited the lowest GC% values of all *Synechococcus*/*Cyanobium* strains (53.8 ± 0.73%) and this difference is even more marked using GC_3%, i.e. the GC content at the third codon position (56.7 ± 1.25%; Fig. 2; $p < 10^{-8}$ Wilcoxon test for clades I and IV *vs.* all other *Synechococcus*/*Cyanobium*). By contrast, the warm-adapted clades II and III displayed significantly higher values (70.2 ± 1.5%; $p < 10^{-5}$ Wilcoxon test clades II and III *vs.* clades I and IV), while the highest GC_3% was found for members of the brackish strains of *Synechococcus* clade VIII and SC 5.2 (81.1 ± 4.6%; $p<10^{-5}$ Wilcoxon test clade VIII and SC 5.2 *vs.* all other *Synechococcus*). Thus, although the strongest GC_3% variation was associated to genome reduction in *Prochlorococcus*, some of the GC_3% variations might be related to the ecological niches occupied by these organisms and notably to thermal and variable salinity niches.

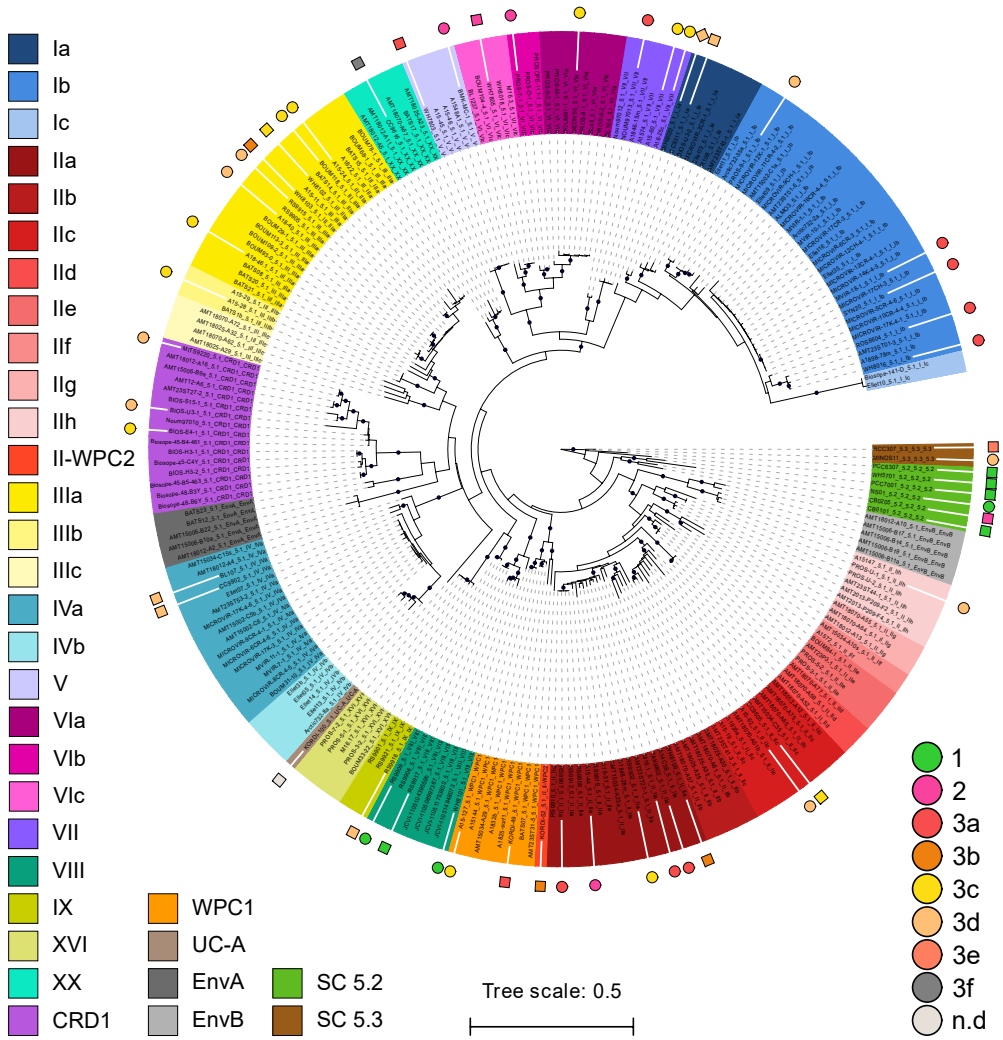


Figure 1: Phylogenetic position of the 53 marine Synechococcus genomes used in this study. A maximum-Likelihood tree was generated based on 231 petB marine Synechococcus sequences from both cultured and environmental samples. Diamonds indicate bootstrap support over 70%. Sequences were named after strain name_sub-cluster_clade_subclade (sub-clade assignments as in Farrant et al. [20] and the background colors correspond to the finest possible taxonomic resolution obtained for each strain using the petB marker gene (left hand side legend). Colored circles surrounding the tree indicate newly sequenced genomes, while squares indicate previously available ones. Note that the WH8020 genome indicated by a diamond was not used in this study due to its poor quality. Symbols are colored according to their pigment type as defined previously ([137–139]; right hand side legend).

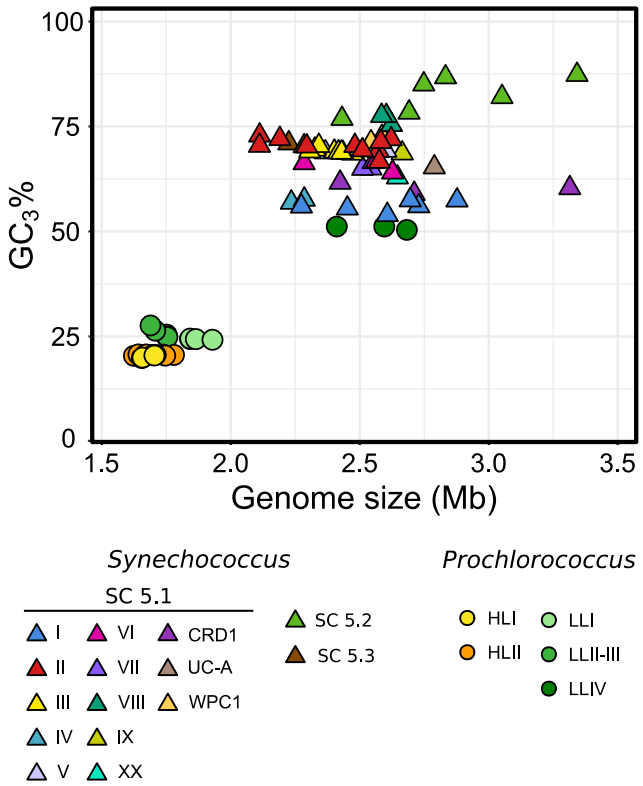


Figure 2: Relationship between genome size and GC3% (GC content at the third codon position). Each symbol corresponds to a different genome, with Prochlorococcus indicated by circles and Synechococcus by triangles. The color of each symbol indicates the clade or SC.

Although they all belong to a monophyletic, long diverged branch within the cyanobacteria radiation [46,47], picocyanobacterial genomes show a tremendous diversity of both nucleotide sequences and gene content. Average nucleotide identity (ANI) and average amino acid identity (AAI) between pairs of picocyanobacterial genomes indeed ranged from 54.1 to 99.9% and 53.16 to 98.9%, respectively and intra-clade ANI and AAI were on average 91.8% and 91.04% (Fig. 3A, Supplementary Fig. S1), showing that each clade and even most sub-clades are at least as distant genetically as microbial species, classically defined by the 95% ANI criterion [48,49]. Interestingly, *Synechococcus* clades I and IV showed a particularly low ANI with other *Synechococcus* strains, while their ANIs with *Prochlorococcus* genomes were higher than for other *Synechococcus-Prochlorococcus* pairs. Since we did not observe this specificity with AAI, it is likely due to the low GC_3% of *Synechococcus* clades I and IV (Fig. 2).

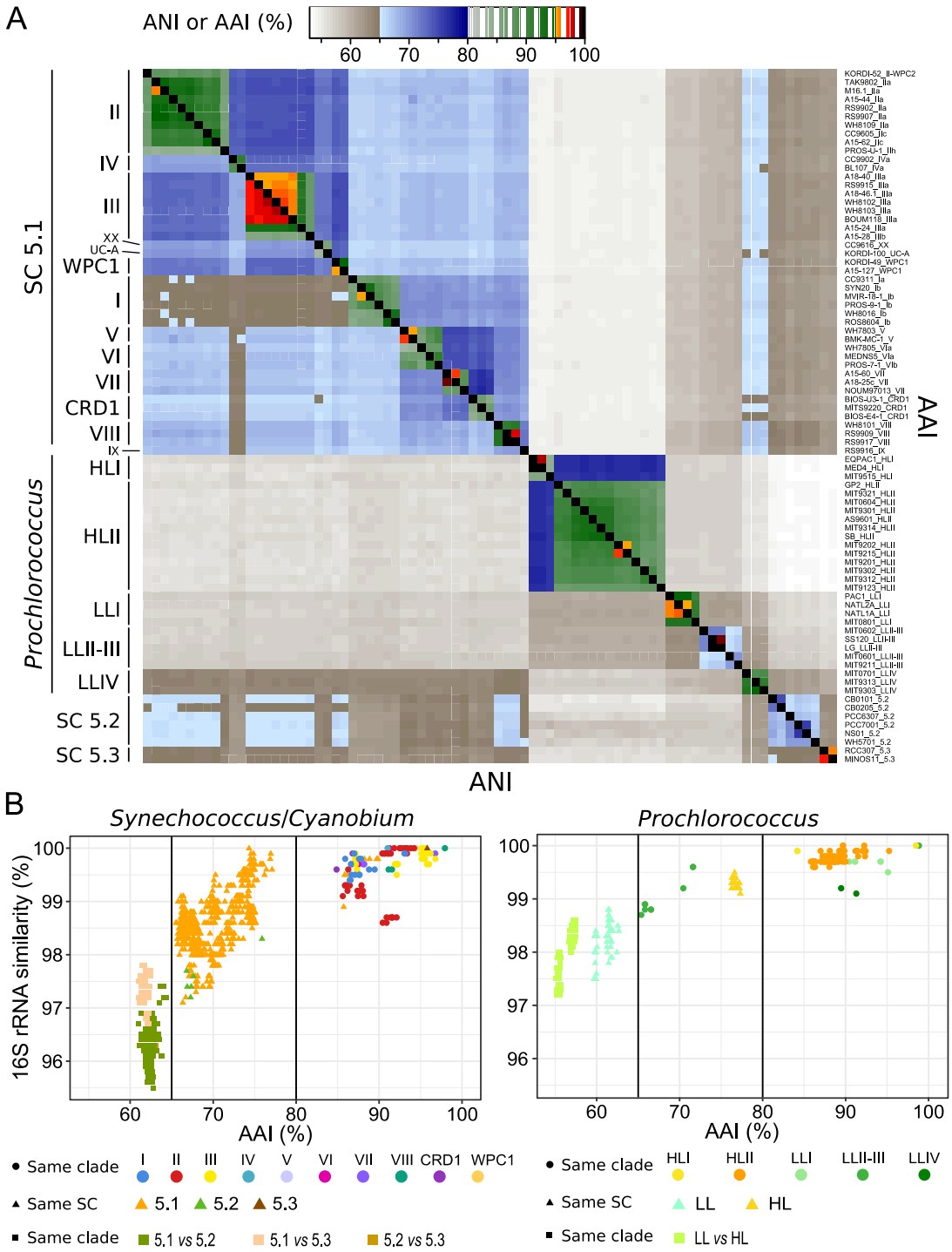


Figure 3: Genomic diversity of marine picocyanobacteria. A. Heatmap of average nucleotide identity (ANI, bottom left triangle) and average amino acid identity (AAI, upper right triangle) between pairs of genomes. Each lane corresponds to a strain, and strains are ordered according to their phylogenetic relatedness. Strains are as labeled as strain_subclade (or higher taxonomic level when no sub-clade has been defined). B. Relationships between 16S rRNA identity, AAI, and taxonomic information for Synechococcus (left panel) and Prochlorococcus (right panel) genomes. Dots correspond to comparisons between pairs of genomes belonging to the same clade, triangles between pairs of genomes belonging to the same SC but different clades and squares between pairs of genomes belonging to different SC.

A plot of the relationship between 16S rRNA identity and AAI for the different pairs of genomes (Fig. 3B) additionally showed two major discontinuities. The first one at 80% AAI discriminated pairs of strains of the same clade from pairs of strains from different clades. Notable exceptions concerned the closely related and globally scarce *Synechococcus* clades V and VI as well as clades XX and UC-A, which fall within the intra-clade divergence level in terms of 16S rRNA identity and AAI, and *Prochlorococcus* clade LLII-III, which showed a divergence level similar to *Synechococcus* intra-SC divergence, suggesting that the gathering of these two clades into a single clade [27,40] should be reconsidered, as suggested by Yan *et al* [50]. The second discontinuity set apart *Synechococcus* strains of the same SC from strains of different SC (<98% 16S rRNA identity and <65% AAI), reflecting a very ancient genomic diversification between the three SC (see below). *Prochlorococcus* strains from different LL clades also fell below the 65% AAI discontinuity, highlighting the large divergence within this group. It is noteworthy that strains within SC 5.2 displayed a particularly low 16S rRNA identity compared to strains within SC 5.1, likely due to the low number of sequenced genomes relative to the wide diversity of this lineage, while in contrast the only two *Synechococcus* SC 5.3 genomes of our dataset were very closely related.

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 v2.1 information system (www.sb-roscoff.fr/cyanorak; Garczarek et al., submitted), in which predicted genes were grouped into clusters of likely orthologous genes (CLOGs) by all-against-all sequence similarity. This clustering allowed us to determine the core genome, i.e. CLOGs present in all strains, and the pan-genome, i.e. all CLOGs present in at least one strain, at various phylogenetic depths [37]. 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. 4B). In contrast, the pan-genome of marine picocyanobacteria, containing 27,376 CLOGs, was still far from saturation, revealing that even with 81 genomes, every newly sequenced picocyanobacterial genome still brought about 192 new genes. This result held true when considering *Prochlorococcus* (7,537 CLOGs) and *Synechococcus* (20,986 CLOGs) independently, indicating that we still missed an essential part of the genetic diversity within both genera that is yet to be sequenced from the field. A major asset brought by the 34 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 (i.e. genus, SC and clades), the accessory genome, i.e. CLOGs shared by at least two strains but not core, and unique genes, i.e. CLOGs present in a single strain (Fig. 4A, Supplementary Table S2). While the proportion of accessory genes was pretty constant between genomes, constituting on average 13 ± 2.4% and 20.7 ± 6.3% of the *Prochlorococcus* and *Synechococcus* genomes respectively, unique genes constituted the most variable part of the genomes, ranging from 0.6% to 21.9% and 1.5% to 31.2% of the *Prochlorococcus* and *Synechococcus* genomes respectively, and were directly related to genome size. The newly sequenced strain BIOS-E4-1 (clade CRD1) contained by far the largest gene number of the genome dataset (4,426 genes), with a large proportion of unique genes (31.2%). Noteworthy, a significant proportion of CLOGs was present in all strains of a given clade (e.g. 335 genes for *Synechococcus* clade III, or 143 genes for *Prochlorococcus* HLI) and could thus potentially be involved in the adaptation of these taxa to specific environmental conditions. However, it should be noted that only a sub-set of these CLOGs were truly specific to each clade (e.g. 32 and 11 genes present in clades III and HLII, respectively; Supplementary Table S3) or ecologically significant taxonomic units (ESTU *sensu* [21]; see below and Supplementary Table S4), i.e. absent from all other clades or ESTUs.

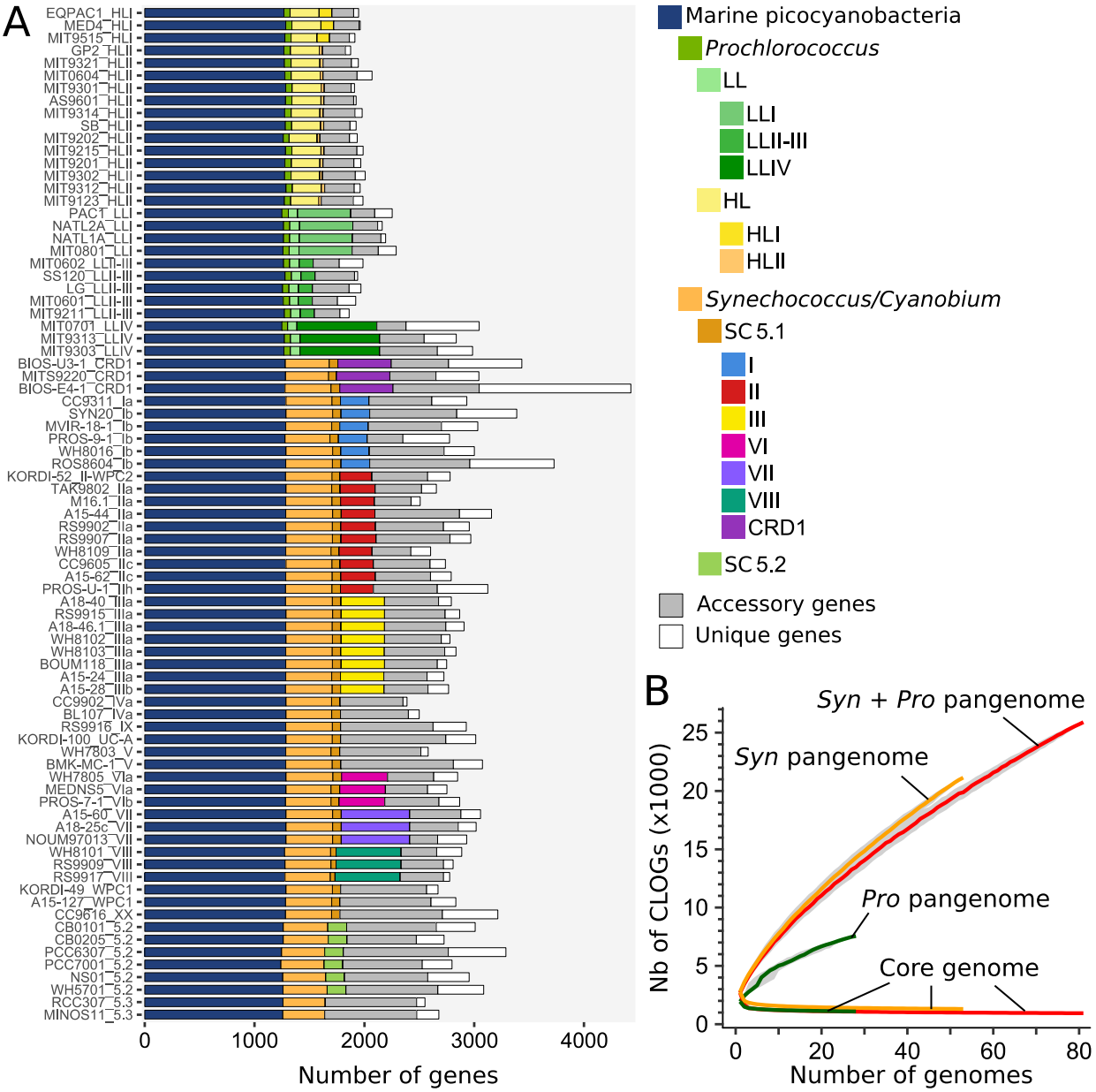


Figure 4: Core, accessory and pan genomes of marine picocyanobacteria. A. Distribution of clusters of likely orthologous genes (CLOGs) in picocyanobacterial genomes. A CLOG is considered as core in a taxonomic group if it is present in ≥ 90% of the strains within this group. Sets of core CLOGS are inferred only for taxonomic groups with more than 3 genomes. Strains are labeled as strain_subclade (or higher taxonomic level when no sub-clade has been defined). B. Evolution of the pan and core genomes for an increasing number of picocyanobacterial genomes (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 1,000 samplings by randomly modifying the order of genome integration.

Dynamics of the evolution of gene content in marine picocyanobacteria

To better understand the evolutionary processes that led to the diversification of gene content within 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 single core proteins (Fig 5). As previously observed [40,45], the gain and loss values obtained for *Prochlorococcus* were consistent with the scenario of a major genome streamlining process that occurred during the evolution of this genus, since an excess of gene loss was observed at the base of this radiation (Fig. 5). 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 ≤378, losses ≤479; not taking into account the genome streamlining at the base of the *Prochlorococcus* radiation), a higher number of events were generally observed on terminal branches as well as an excess of gains compared to losses, with up to 1,662 gained genes on the branch leading to *Synechococcus* BIOS-E4-1 (SC 5.1) and 831 on the one leading to *Prochlorococcus* MIT0701, for 105 and 108 lost genes, respectively.

By using calibration time points from a previous study [47], we estimated that this corresponds to about 0.71 and 4.62 genes gained (1.67 and 1.80 genes lost) per million years (My) on internal and terminal branches of *Synechococcus* SC 5.1, respectively, while internal and terminal branches of *Prochlorococcus* HL gained 1.45 and 4.5 genes (0.87 and 3.72 lost; Table 1). The higher values observed for the terminal branches are related to the high number of strain-specific genes and reflect the fact that most of the variability in gene content occurs at the 'leaves' of the tree. If we assume 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 therefore not represented in our genomic dataset.

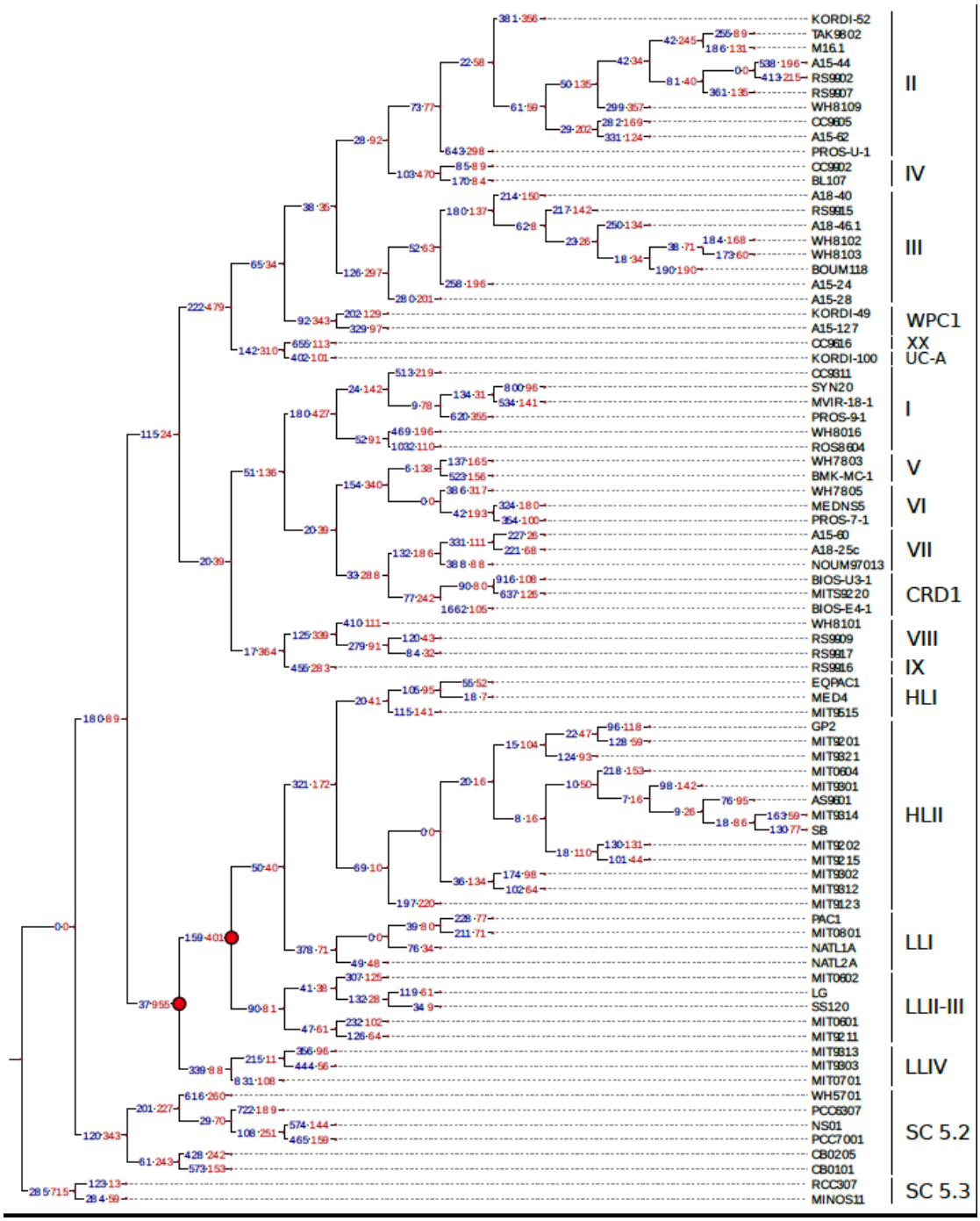


Figure 5: Estimation of the gene gains and losses during the evolution of marine picocyanobacteria. The ancestral state of presence/absence of every cluster of likely orthologous genes (CLOGs) was assessed using Count [130] and used to infer the number of gains and losses of gene families on each branch of the tree using the phylogenetic core protein tree as reference. The number of gained and lost genes is labeled in blue and red, respectively. Nodes highlighted in red correspond to the major genome streamlining events that have occurred in the *Prochlorococcus* radiation.

As genomic islands have been shown to play a key role as repositories of laterally transferred genes potentially involved in niche adaptation in marine picocyanobacteria [23,40,50,52,53], we explored the contents of these islands in all analyzed genomes. Most genomes were too distant to compare genomic islands between strains by whole genome alignment as performed by Coleman *et al.* [52] on *Prochlorococcus*, so here genomic islands were defined in each strain as regions of the genome enriched in gained genes using a similar approach as Kashtan *et al.* [40] but using a threshold to define the limits of the islands in each strain (see Methods; Supplementary Table S5). The number of gained genes located in genomic islands and shared by pairs of strains showed that closely related strains share many more island genes than distantly related ones and that only a few exchanges of genes occur between distantly related clades (Supplementary Fig. S2). These observations are particularly striking for *Prochlorococcus* HL streamlined genomes that share only a low proportion of island genes with *Synechococcus.* A notable exception is *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 these groups co-occur in coastal or estuarine waters of variable salinity [23,54,55]. To further explore how strains share genomic islands, we used an innovative network method based on the partial similarity of gene contents between islands shared by pairs of strains. It allowed us to retrieve islands previously identified either by direct pairwise comparison of *Prochlorococcus* HLI MED4 and HLII MIT9312 strains [52] or by analyzing the deviation in tetranucleotide frequency in individual *Prochlorococcus* and *Synechococcus* genomes [23], demonstrating the validity of our automated approach (Supplementary Fig. S3 and Supplementary Tables S6 and S7). Interestingly, most islands identified by these authors in *Prochlorococcus* HL strains appeared to be shared by all HL strains, forming dense red, knot-shaped modules in the network (e.g., Pro_GI033 = MED4 ISL1; Pro_GI048 = MED ISL2;

Pro_GI028 = MIT9312 ISL5; Pro_GI000 = MED SVR2; Pro_GI015 = MED4 SVR4; Pro_GI041 = MED4 ISL1.1; Pro_GI023 = MED4 ISL2.2; Fig. 6 and Supplementary Table S6). These red knots correspond to genomic regions prone to gene integration that have likely been acquired by the common ancestor of all HL strains, then vertically transferred to all descendants, much like the phycobilisome region that is shared by all *Synechococcus* strains [23]. In contrast, ISL4 island, initially identified in MED4 by Coleman *et al*. [52] and later confirmed both by Dufresne and coworkers [23] and our automated island detection approach (Pro_GI004; Supplementary Fig. S3), does not form a red knot but only a fuzzy network of interconnected islands, each shared by only 2 to 4 strains (Fig. 6), so seemingly corresponds to a more recently acquired island with a highly variable content shared only by a subset of the HL population. Our approach also unveiled previously undescribed islands specifically shared by sets of *Prochlorococcus* LL strains, including Pro_GI027, 039, 044 and 049 specific to LLI strains (several being enriched in *hli* genes, known to be amplified in LLI compared to other LL strains; [56]), Pro_GI010, 018 and 025 specific to LLII/III strains, and Pro_GI002 as well as 13 other modules specific to LLIV strains, including several containing genes encoding lanthipeptides ([57]; Fig. 6 and Supplementary Table S6). In *Synechococcus,* the network included relatively few dense red knots compared to *Prochlorococcus* (Fig. 7). Among the most notable ones are three clade III-specific islands, the first one (Syn_GI013) gathering a gene cluster (*cynA-B-D*) involved in cyanate transport ([58]; Supplementary Table S7), the second one (Syn_GI087) including a specific beta-glycosyltransferase and *swmA*, a protein involved in a special type of motility observed only in members of this clade [59] and the third one (Syn_GI102) that notably encompasses *swmB*, encoding a giant protein also involved in this motility process [44]. Another interesting example is Syn_GI100, which notably encompasses a 3-gene cluster composed of one *nfeD* homolog and two flotillin-like genes that both have similarity to the

floT gene involved in the production of lipid rafts, whose deletion in *Bacillus subtilis* was found to strongly affect cell shape and motility [60]. Interestingly, this gene cluster was found in the only two clade III strains (A15-24 and A15-28) that lack *swmA* and *swmB* as well as in several distantly related strains. Conversely, no *swmA-B*-containing strain was found to possess the *nfeD-floT1-floT2* gene cluster. The network approach also detected quite a few knots containing both red and blue edges. The latter color indicates that strains sharing these islands are distantly related to one another. Thus, knots that are mixing red and blue edges potentially emphasize relatively recent horizontal gene transfers between clades or longer phylogenetic distances. This includes i) Syn_GI022, a module found in many SC 5.1 strains with the notable exception of clade II strains, which encompasses a large gene cluster involved in glycine betaine synthesis (*gbmt1*-2) and transport (*proV-W-X*), located in some strains next to another gene cluster involved in the biosynthesis of glucosylglycerate (*gpgS-gmgG-gpgG*; [32]) and ii) Syn_GI122, a module comprising strains from almost all lineages that encompasses genes encoding uncharacterized cell surface proteins, secreted CHAT domain-containing proteins and/or genes involved in the biosynthesis of cyclic AMP (cAMP), including adenylate cyclases located in the vicinity of cyclic nucleotide-binding proteins, such as the cAMP receptor protein (CRP) or a cAMP-regulated small-conductance mechanosensitive ion channel. Altogether, this network approach nicely complements the detection of genomic islands in single genomes by providing insights about the evolutionary history of these genomic islands.

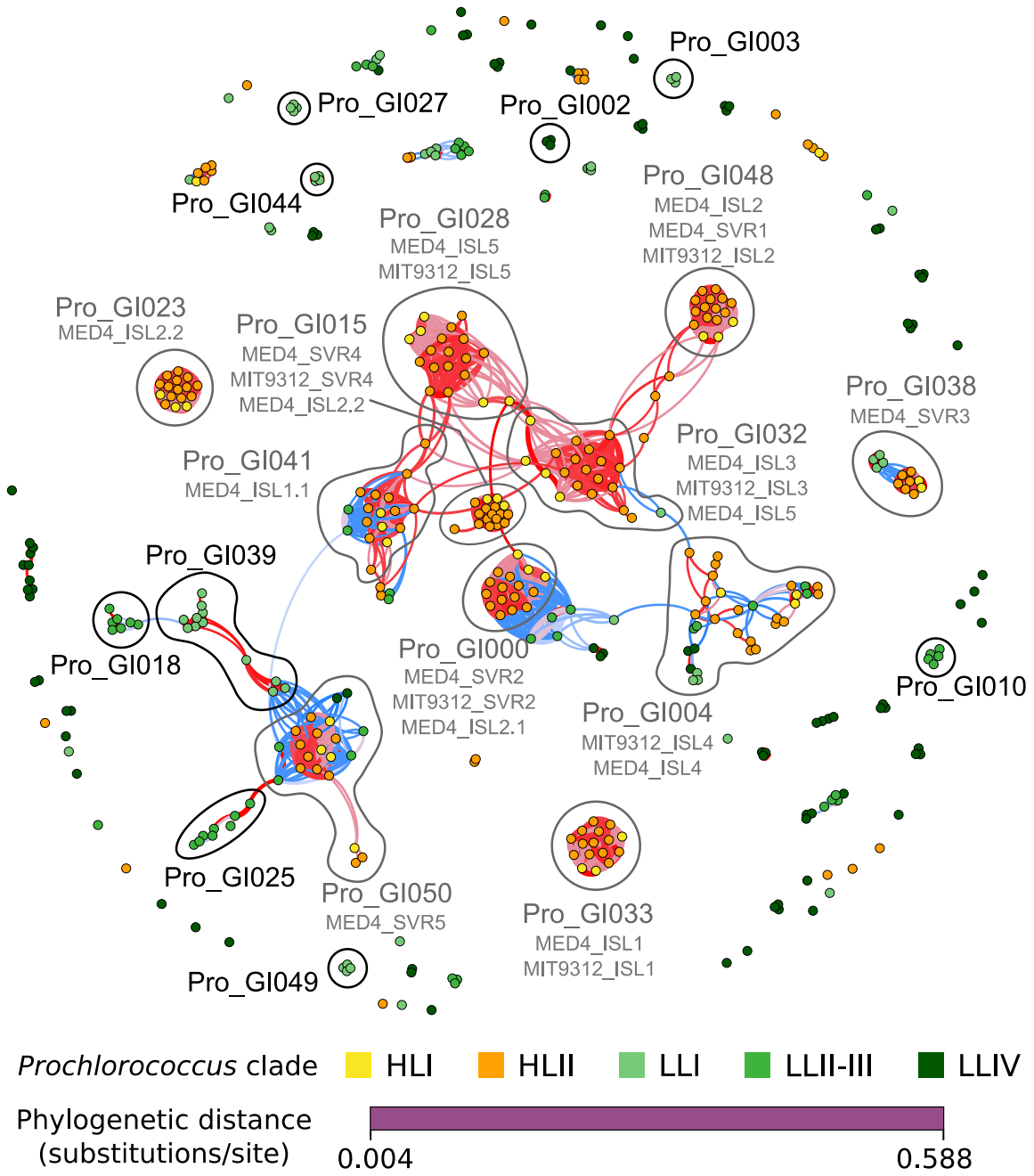


Figure 6: Network of shared gene islands between all *Prochlorococcus* strains analyzed in this study. Each node corresponds to a genomic island in a given strain, the gene content of which is listed in Additional file 1: Table S5. Edges were colored according to the phylogenetic distance between strains, with red indicating closely-related strains and blue more distantly related strains. Edge width corresponds to the Jaccard distance between islands based on gene content. Nodes were colored based on *Prochlorococcus* clade. Modules cited in the text are surrounded with a grey line for those containing islands already described in the literature (subtitled with their names in [51] and [22]) and a black line for new modules described in the present study. The gene and genomic island composition of each module is described in Additional file 1: Table S6.

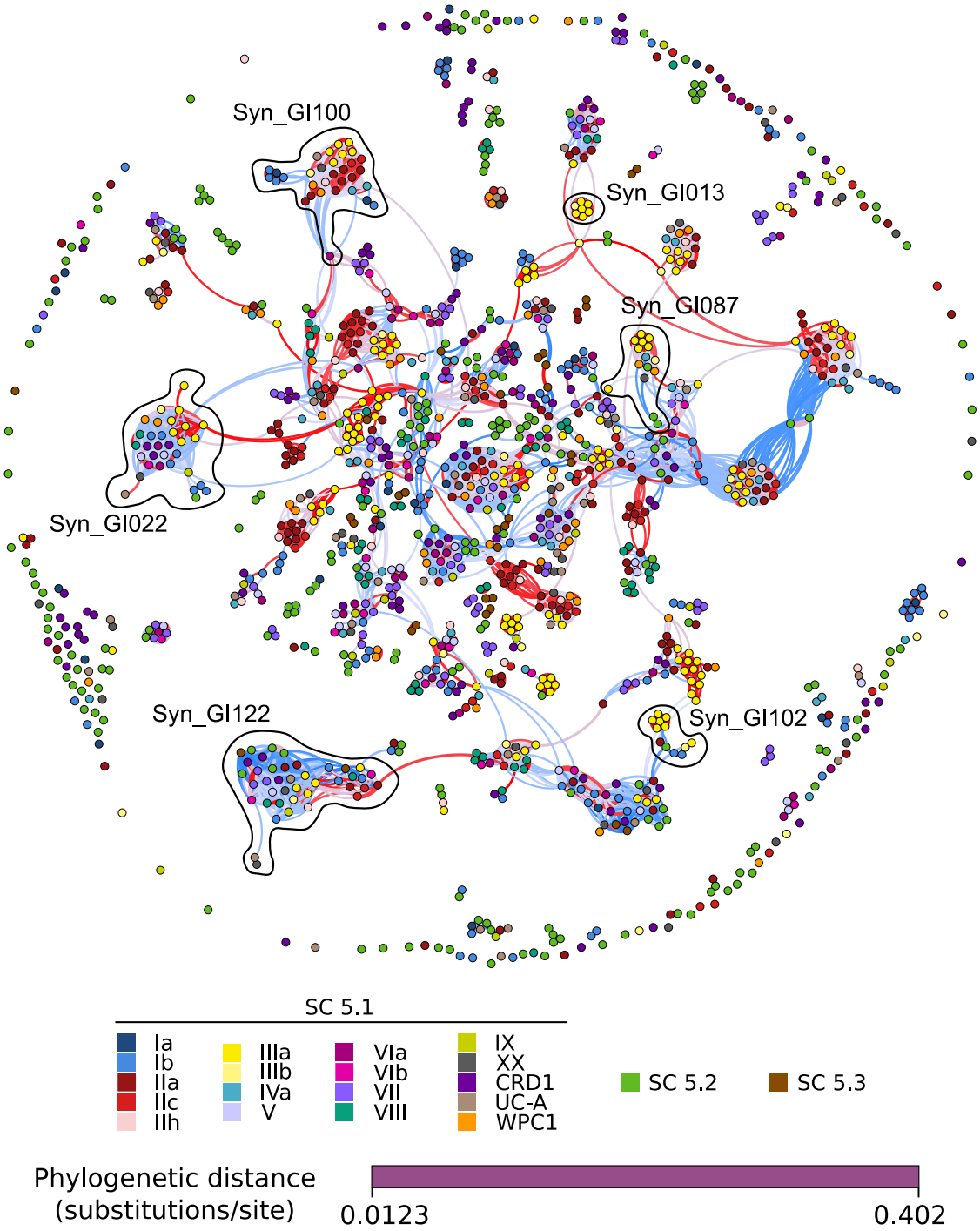


Figure 7: Same as Fig. 6 but for marine *Synechococcus/Cyanobium* strains. The gene and genomic island composition of each module is described in Additional file 1: Table S7.

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

The fairly low rate of gene acquisition evidenced in this study raises the question of the relative weight of gene content variations *vs.* substitutions in the nucleotide sequence in the long-term diversification and adaptation processes of these organisms. Figure 8 compares a phylogenetic tree built with a concatenation of 821 picocyanobacterial core protein sequences 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 fixation of genes and fixation of mutations occurred concomitantly during the evolutionary history of marine picocyanobacteria. Yet, *Synechococcus* clade VIII and SC 5.2 were found to be closely related in the dendrogram based on the phyletic pattern. Indeed, as previously reported in a study using 11 *Synechococcus* genomes [23], these taxa share a fair number of genes, potentially related to their co-occurrence in brackish environments. Interestingly, the closely related clades V and VI cluster together with these two taxa, indicating that they may also share with clade VIII and SC 5.2 some mechanisms of adaptation to low salinity niches (see below). Although the presence of SC 5.3 has been recently documented in freshwater environments [61], the presence of the two marine sequenced strains (RCC307 and MINOS11) at the base of this halotolerant group might instead be due to attraction by SC 5.2.

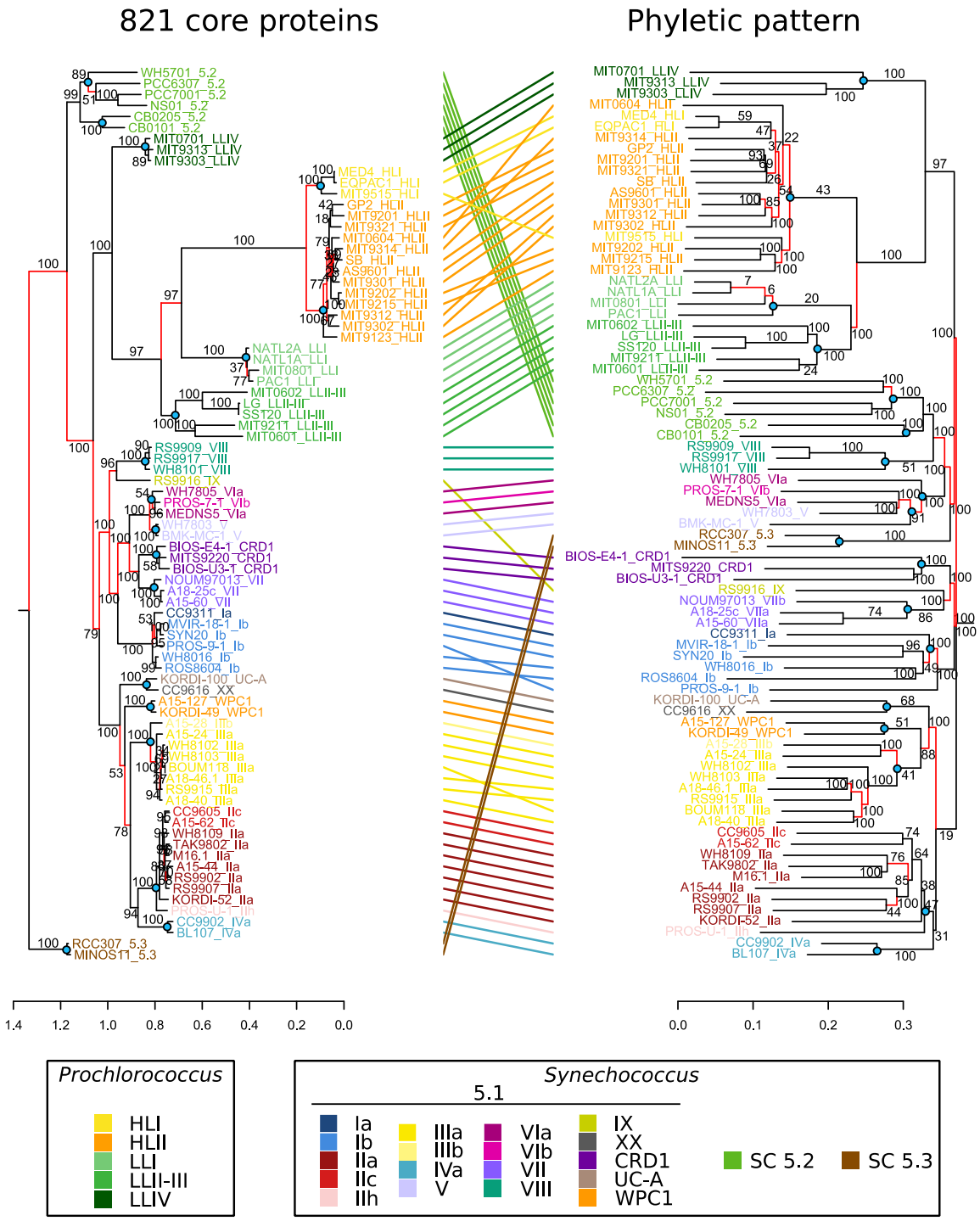


Figure 8: Comparison of phylogenies based on core protein sequences and phyletic patterns of non-core genes. *Left*, Maximum Likelihood tree based on the alignment of 821 concatenated core proteins. *Right*, Neighbor-Joining tree based on the Jaccard distance between the phyletic patterns of 27,376 accessory gene families found in the 81 picocyanobacterial genomes. Labels are colored according to the strain sub-clade. Red branches indicate discrepancies between the topology of the two trees. Nodes located at the base of a clade and highlighted by blue dots were used for branch length comparisons in Additional file 2: Fig. S4.

Among the *Synechococcus* SC 5.1 and *Prochlorococcus* radiations, it is worth noting that some clades were closer in terms of gene content than expected from the core phylogeny, besides a few incongruences within *Synechococcus* clades I, II, III and VI and *Prochlorococcus* HLII that are likely due to the relatively low number of specific genes within these clades. This remark particularly holds for *Synechococcus* clades WPC1, XX and UC-A grouping with clade III in the tree based on the phyletic pattern (Fig. 8). Finally, some clades lost their monophyly in the tree based on phyletic pattern, such as *Synechococcus* clades V and VI that were mixed together or *Prochlorococcus* HLI that was found to be mixed with HLII. This example is particularly interesting, since despite their clearly distinct phylogenetic clustering based on protein sequences and well-known ecological and physiological differences [17,29], these two clades have a quite similar gene content, with only a few genes (29) present in all HLII strains but not in all HL strains (Fig. 4A). Similarly, *Prochlorococcus* clade LLI, which was previously shown to occupy an intermediate niche between HL and strict LL members (LLII-IV) and to share genes with both ecotypes [17,62], actually appeared to share more genes with the LLII-III clade (1,382 genes) than with HL (1,290 genes). Altogether, these two examples show that within *Prochlorococcus*, although HL and LL have different gene contents, differentiation within HL and to a lesser extent within LLI-III rather relies on substitution accumulations than on variation in gene content.

Another major difference between these trees concerned branch lengths. By computing for each node at the base of a clade (blue dots in Fig. 8) 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 <0.0015) in the phyletic pattern tree than in the core tree (Supplementary Fig. S4). This suggests that there were more amino acid

substitutions before the divergence of clades than after, whereas there was more gene content variation between strains of a clade than between clades. In other words, this comparison revealed that most of the changes that were fixed in the long term by evolution are substitutions and not changes in gene content.

In order to quantify more precisely this difference, we compared the estimated number of gene gains and losses per My (Supplementary Fig. S5) to the number of amino acid substitutions in core proteins per My (Supplementary Fig. S6) and results of these comparisons are shown in Table 1. It is important to note that the rates of gene gain/loss and amino acid substitutions calculated this way should only be considered as lower bound estimates for several reasons. First, since we only have access to the present-day genomes and not to ancestral ones, measurements of the rate of genes gained in fact refer to genes gained and successfully retained over time in at least one strain. Second, the amino acid substitution rates were measured on core proteins, whose genes likely undergo a strong purifying selection. This, together with the much longer generation time of picocyanobacteria compared to model bacteria and with their considerable population size [15,45,63], could explain why estimated rates were lower than for other bacterial lineages [64,65]. With this caveat in mind, in *Prochlorococcus* HL, 356x more amino acid substitutions than gene gains were estimated for internal branches per My, and 69.6x for terminal branches, primarily due to a higher rate of gene gain in the latter branches. In *Synechococcus* SC 5.1, a ratio of 164 and 20 was obtained for internal and terminal branches, respectively, the difference between the two genera likely being due to the higher rate of protein sequence evolution observed in *Prochlorococcus* [45].

We also compared at each node the fixation rate of amino acid substitutions in core proteins (i.e. amino acids in the alignment that are identical in all descending strains and

different in all other strains) to the fixation rate of genes (i.e. present in all descending strains and in no other strain). 201x more amino acid variants than genes were fixed per My in *Prochlorococcus* HL (and 116x more for *Synechococcus* SC 5.1). This corresponds to a fixation rate of 78 and 18 amino acid changes in core proteins per My for *Prochlorococcus* HL and *Synechococcus* SC 5.1, respectively, while one gene is fixed once every 2.6 My for *Prochlorococcus* HL and once every 6.3 My for *Synechococcus* SC 5.1. While these numbers show that substitutions played a major role in genomic diversification, the question remains as what part of this diversification is related to an adaptive process.

Role of gene content in the adaptation of *Synechococcus* to specific niches

In contrast to *Prochlorococcus* [27,40,50,53,62], few genomic diversity studies have been conducted so far in *Synechococcus*. In order to reveal whether the presence or absence of genes might be related to *Synechococcus* adaptation to specific niches, we defined sets of clades co-occurring in the field and occupying similar niches, based on assemblages of Ecologically Significant Taxonomic Units (ESTUs) as defined in [21]. We then searched for genes occurring in strains within a given set and absent from other picocyanobacterial strains using a relaxed, niche-related definition of specificity (Supplementary Table S4). These analyses revealed only 18 CLOGs specific to members of both cold thermotypes, clades I and IV, among which 6 had a putative function, though with seemingly no direct relationship with adaptation to low temperature. However, the set of 19 CLOGs specific to clade I includes a particular isoform of the chaperone protein DnaK (DnaK4, CK_00056929; Supplementary Table S3) in addition to the three gene copies present in most *Synechococcus* SC 5.1 strains. This additional copy might be involved in protein folding in cold conditions [66].

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; Supplementary Table S4), 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 [67] and a chromate transporter (ChrA), which was recently suggested to be involved in phosphate acquisition in *Prochlorococcus*, based on its enrichment in P-poor oligotrophic areas [68]. 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 *Escherichia coli* K1*,* responsible for the formation of a polysialic acid extracellular capsule; see Kps cluster in Supplementary Table S4) and another cluster of 7 genes that might be involved in the use of organic nitrogen 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 above-mentioned cyanate transporter genes (*cynABD;* [58]) as well as a phosphate starvation-induced protein (PsiP1;[69]) and a specific alkaline phosphatase (CK_00052500) that potentially hydrolyses extracellular organic phosphates (Supplementary Table S3). Similarly, the two members of SC 5.3 also share a large number of strictly specific genes (215), including a regulator of phosphate uptake (PhoU; CK_00005756; [70]) as well as two putative phosphatases (CK_00005504, CK_00005619) and a putative pyrophosphatase (CK_00005811), in addition to the 4 potential pyrophosphatases present in most picocyanobacterial genomes (CK_00000642, CK_00000654, CK_00000805 and CK_00008108; Supplementary Table S3). Altogether, these results suggest that the occurrence of these genes might contribute to the success of clade III, WPC1 and SC 5.3 cells in oligotrophic, P-depleted environments such as the Mediterranean Sea in summer [21], and indicates that members of these three taxa have adopted partially different strategies to cope with P depletion.

Genes potentially involved in niche adaptation were also found in all three strains of the CRD1 clade, known to dominate in iron-depleted oceanic regions, which share a quite high number of specific CLOGs (81, Supplementary Tables S3 and S4), though most of them have no known function. Among the characterized ones were a second copy of the flavodoxin IsiB, a Cu-containing protein known to replace ferredoxin in iron-depleted conditions [71], the ferrous iron transport protein FeoA, an iron-sulfur cluster biosynthesis family protein (CK_00008433) as well as 3 specific high light-induced proteins (HLIPs) that might provide protection from oxidative stress to photosystems [72].

Finally, in agreement with their clustering in the dendrogram based on phyletic pattern (Fig. 8), clades VIII and SC 5.2 share 28 genes including a few strictly specific genes (Supplementary Table S4), 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 family of ATP-driven pumps known to transport a variety of different ions and phospholipids across membranes [73]. It is also noteworthy 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 [20,21,74] and possibly encompass environments with variable salinity (Supplementary Table S4). This includes a specific small-conductance mechanosensitive ion channel (MscS family) that might be involved in the response to osmotic stress (CK_00056919; [75]) and a bacterial regulatory protein of the ArsR family that besides regulating the efflux of arsenic and arsenite was suggested to participate in salt tolerance in *Staphylococcus aureus* through a Na^+ efflux activity [76]. In addition, members of clade VIII share 22 specific genes, including a second potential mechanosensitive ion channel (MscS*;* CK_00056915), while members of SC 5.2 share 31 specific genes, including another *mscS* gene copy (CK_00003081) as well as genes encoding a putative chloride channel

(CK_00042275) and a NAD-dependent malic enzyme, a protein known to be enhanced under salt stress in plants ([77]; Supplementary Table S3). Despite these few examples, it seems that the number of genes potentially related to the ecological niche occupied by each clade or assemblage of clades is fairly limited and varies depending on the considered niche, with for instance few genes related to thermal niche adaptation. Most of the diversity in gene content therefore relies on differences between individual strains rather than between phylogenetic or ecotype groups, a large proportion of the sparsely distributed genes having yet unknown functions, some potentially being involved in niche adaptation.

Role of substitutions in adaptation

Given our observation that a high number of amino acid substitutions have been fixed in the long term, we also searched for those potentially involved in niche adaptation. We identified “specific variants” as positions in core protein alignments for which a particular amino acid is found in all strains of a given clade, ESTU or set of ESTUs and a different amino acid is found in other strains. In order to reduce the noise due to the accumulation of clade-specific substitutions and to better identify the niche adaptation signal, we focused on variants shared by clades I and IV, which do not form a monophyletic group (Fig. 8, left) but usually co-occur in cold, temperate waters [20–22,29,78,79]. We identified 180 proteins mainly involved in i) energy metabolism, ii) biosynthesis of cofactors, prosthetic groups and carriers, such as pigments and vitamins, iii) protein synthesis and protein fate, and to a lesser extent iv) transport and DNA metabolism (Supplementary Table S8). The first category encompassed proteins responsible for carbon fixation (RuBisCO subunits RbcS and RbcL, carbonic anhydrase CsoSCA, carboxysome proteins CsoS1E and CsoS2, and Calvin cycle enzyme Fbp-Sbp), two photosystem II subunits (the extrinsic PsbU protein and the manganese

cluster assembly protein, Psb27) and a number of proteins involved in electron transport for photosynthesis and/or respiration (CtcAI, CtcEI, NdhA and two ATP synthase subunits: AtpA and AtpD). Furthermore, this set includes six proteins potentially involved in the response to light or oxidative stress: two High Light Inducible Proteins (HLIPs; CK_00001609 and CK_00001414), two peroxiredoxins (PrxQ), a glutaredoxin (CK_00000445) and a flavoprotein involved in the Mehler reaction (Flv1). We also identified a few enzymes involved in sugar metabolism and in particular in the pentose phosphate pathway (Pgl, TalA and Zwf). As concerns the 'protein synthesis' and 'protein fate' categories, this includes six ribosomal proteins and nine amino acid biosynthesis proteins, several tRNA/rRNA modification enzymes and tRNA aminoacyltransferases as well as seven proteins responsible for folding and stabilization of polypeptides. Of particular interest are the proteins belonging to the 'biosynthesis of cofactors, prosthetic groups and carriers' category, including enzymes involved in chlorophyll (HemC, ChlN, ChlB), cobalamin (CobO, cobQ, CobU-CobP) and carotenoid biosynthesis. The latter includes CrtE and GpcE, two enzymes involved in the phytoene biosynthesis pathway and CrtP, CrtQ and CrtL-b, the three enzymes catalyzing all the steps required to transform phytoene into β-carotene. It is also interesting to note that the five proteins displaying the largest number of specific substitutions relative to protein length are a putative ABC multidrug efflux transporter (CK_00008042; 19 positions specific to clades I and IV out of 607 amino acids), lycopene β-cyclase, responsible for the last step of β-carotene synthesis (CrtL-b; 7/347), the bifunctional Calvin cycle enzyme fructose-1,6-biphosphatase/sedoheptulose-1,7-biphosphate phosphatase (7/347), the photosystem II manganese cluster assembly protein Psb27 (3/160) and the ribosomal protein RpmB (1/78). Even though the number of substitutions is not directly correlated to the level of selection

pressure, the high proportion of specific substitutions in these proteins suggests that they have been subjected to positive selection.

Discussion

The availability of 81 complete and closed picocyanobacterial genomes with extensive manually refined annotations, including 34 novel *Synechococcus*, constitutes a key asset for comparative genomics analyses. With regard to previous studies (see e.g. [23,32,40]), sequencing of several strains for most major *Synechococcus* clades revealed that the extent of genomic diversity is tremendous, at all taxonomic levels including within clades and most sub-clades. As previously observed for SAR11 [80,81], ANI and AAI were indeed most often well below the cut-off of 95% (Fig. 3), usually considered to be the limit between bacterial species [48,82,83]. Thus, based on this cut-off, most clades within cluster 5 *sensu* [84] would correspond to one or even several species, as suggested by one research group [85–87]. However, the delineation of so many species in a radiation that mostly exhibits a continuum in terms of within clade sequence identity (ID% range: 84 to 100%; Fig. 3B) would create more confusion than clarification as it would result in most cases into single-strain species, which cannot be clearly differentiated based on their fundamental (see e.g. [33,88]) and/or environmental realized niches [21,22,24,68]. With this caveat in mind, it is clear that besides the *Prochlorococcus* lineage, there are three extremely divergent monophyletic groups within the marine *Synechococcus*/*Cyanobium* radiation [89], which furthermore can be clearly discriminated based on 16S similarity *vs.* AAI plots (Fig. 3B), with an AAI divergence below the 65% limit that has been proposed to discriminate distinct genera [90]. We thus propose to split the marine *Synechococcus* group into three distinct taxonomic groups: *Ca.*

Marinosynechococcus (SC 5.1), *Cyanobium* (SC 5.2) and *Ca.* Juxtasynechococcus (SC 5.3). This proposal notably solves the inconsistency to have a mix of strains named *Cyanobium* spp. and *Synechococcus* spp. within SC 5.2, which should clearly all be called *Cyanobium* spp. For the universal acceptance of the revised taxonomy of this group and cyanobacteria at large [91], both temporary names proposed for SC 5.1 and 5.3 as well as the potential definition of species within each of these radiations await validation by a large panel of cyanobacterial community members. In any case, any creation of new species within this group should likely take into account previously defined monophyletic clades and subclades as these phylogenetic groups have been used in most previous laboratory and environmental studies, whatever the genetic marker used [13,24,25,92–94].

The particularly high degree of genomic divergence occurring within Cyanobacteria Cluster 5 needs to be taken into account when putting results from comparative genomics of marine picocyanobacteria in the context of other highly sequenced bacterial groups such as pathogens and commensals [95–97]. While high divergence and associated low level of synteny somehow limit the application of classical population genetics approaches, such as calculation of recombination rates [65], our dataset is in contrast well suited to study the long-term evolutionary processes that have shaped the genomes of these abundant and widespread organisms in relation to 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 largely relies on differential gene content [23,40,98,99]. Here, a 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 necessary for niche adaptation, even using relaxed stringency criteria (e.g. by selecting genes present in >80 or 90% of strains within a clade/ESTU

assemblage and in <20 or 10% of others; Supplementary Tables S3 and S4). This may be due to the existence of an extended within-taxa microdiversity [6,21,29,100]: the more genomes in 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. 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 took place [23,32,40,50,101–103]. Here, although the higher estimated rate of gene gains on the terminal branches of the phylogenetic tree indicates that most detectable events occurred fairly recently with regard to the long evolutionary history of both genera (Fig. 5, Table 1), adding time calibration to the tree led to an estimation of only 4.5 and 5.6 genes gained per My on terminal branches in *Prochlorococcus* HL and *Synechococcus* SC 5.1 strains, respectively. Thus, gene gains appear to be rather rare events. Even though these rates are approximate due to uncertainties in time calibration and probably underestimated, they are entirely 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 [6]. Furthermore, in accordance with previous studies on other bacterial groups [65,104–106], the fact that rates of gene gain/loss are estimated to be higher on terminal branches of the tree (Supplementary Fig. S4), together with the high number of unique genes in every sequenced strain (Fig. 4A), clearly suggests that most recently acquired genes will not be kept in the long term in both genera. Our calculation indeed gives an approximate value of 1.45 and 0.71 genes gained and subsequently kept per My in *Prochlorococcus* HL and *Synechococcus* SC 5.1, respectively (Table 1). This low fixation

rate suggests that most of the recently gained genes have no or little beneficial effect on fitness in the long term and that we observe them in genomes because purging selection has not deleted them yet [107–109]. Still, these recently gained genes could be involved in more transient adaptation processes at the evolutionary scale such as biotic interactions e.g. resistance to viral attacks, grazing, etc.

Such a result also has important implications for interpreting the role of flexible genomes in the context of adaptation to distinct niches. Indeed, genes conferring adaptation to a specific niche are mixed in the genomes with genes with no or little beneficial effect and are thus difficult to identify – in particular when they have only a putative function. The relatively low gene fixation rate that we observed (Table 1) also implies that flexible genes that are fixed within a clade (i.e. clade-specific genes) were gained tens of millions of 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 shared by *Synechococcus* clade VIII and SC 5.2 suggest that adaptation to low salinity environments was a critical factor in their differentiation from other taxa and the most parsimonious evolutionary scenario would be a lateral transfer of these genes from a SC5.2-like strain to the common ancestor of clade VIII, which might date back to 51.6 My (confidence interval 0-141 My). Similarly, adaptation to phosphorus-depleted oligotrophic areas probably drove the differentiation of *Synechococcus* clade III, as revealed by the relatively large number of P- and other nutrient-uptake genes specific to this clade. Interestingly, co-occurring ESTUs IIIA, WPC1A and SC 5.3A only share a few common genes potentially involved in the adaptation to this limitation. Instead, these ESTUs seem to have independently acquired different sets of genes to improve P-uptake and/or assimilation (see Results and Supplementary Table S4). It is also noteworthy in this context that in *Prochlorococcus*, P metabolism is not clade-related but dependent on within-

clade variability in the gene content of specific genomic islands [101,102], further highlighting the variety of evolutionary paths that led to adaptation to low-P environments in these different lineages.

As proposed recently for other bacterial model organisms [65,110], sequence alterations also appear to play a crucial role in genome diversification of marine picocyanobacteria and to have driven their adaptation to specific environments. Indeed, in the time necessary for one gene to be gained, we found that 20 to 60 amino acid substitutions accumulate in any picocyanobacterial genome (as estimated based on terminal branches of the phylogeny, Table 1). This finding brings new evidence to support the "Maestro Microbe" model of bacterial genome evolution recently proposed by Larkin and Martiny [111], which posits that some phenotypic traits, such as thermal preferences, evolve by progressive fitness optimization of protein sequences rather than gene gains and losses. 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 temperate (HLI) and warm (HLII) environments [40,52,101,111], which were partly mixed on our tree based on gene content despite a clear phylogenetic separation based on core marker genes (Fig. 8). The sequencing of new *Synechococcus* genomes also allowed us to extend the Maestro Microbe hypothesis to *Synechococcus* thermotypes [20,33], since particularly few genes were found to be specific to the cold-adapted clades I and IV (Supplementary Table S4). In contrast, our analysis of *Synechococcus* core genes containing amino acid variants shared exclusively by all members of these cold thermotypes revealed potential candidates for adaptation to cold waters (Supplementary Table S8). The fact that all members of clades I and IV share specific variants of the three proteins involved in the β-carotene synthesis pathway (with e.g. >2% of the protein sequence comprising residues

specific to these clades in CrtL-b) is particularly striking, since physiological experiments have shown that members of clades I and IV were able to maintain or increase their β-carotene:chlorophyll *a* ratio in response to cold stress, while this ratio decreased in strains representative of warm thermotypes [33]. Thus, these substitutions might allow cells of the former clades to maintain β-carotene synthesis in cold conditions, resulting in a reduction of the cold-induced oxidative stress. Additionally, four proteins potentially involved in the response to oxidative stress were found to display variants specific to clades I and IV (Supplementary Table S8). In much the same way, a recent study identified two substitutions in genes encoding the two subunits of phycocyanin in *Synechococcus* between these cold-adapted clades and the warm-adapted clades II and III, which were also thought to be involved in adaptation to distinct thermal niches: RpcA G-43 and RpcB S-42 in the former clades and RpcA A-43 and RpcB N-42 in the latter [112]. It is worth noting that these genes were not detected by the stringent approach used here either because of the absence of the multi-copy *cpcA* gene in the CB0101 genome due to assembly issues or to a single exception among the newly sequenced genomes, the clade I strain PROS-9-1 having an RpcB S-42. Given that clades I and IV have diverged about 425 My ago (confidence interval 308-468 My), the most parsimonious explanations for these many shared substitutions would be either an adaptive convergence or an ancient homologous recombination between ancestors of these clades. In this context, it is interesting to note that mutations were found to arise in just a few generations in a clonal *Prochlorococcus* strain as an adaptation to selective conditions such as UV radiation [113], antibiotics [114] or phage pressure [115], emphasizing the role of such substitutions in short-term adaptation, although only a subset of these are kept in the long term.

Conclusions

Current clades of marine picocyanobacteria might be considered as survivors of a former set of “backbone” populations (as defined by [6]) that appeared hundreds of millions years ago, and then optimized their sequence, while eventually losing most of the genes that initially allowed niche colonization [6,116–118]. 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. [6,21,29,100]). 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. 8) would be the occurrence of periods of rapid diversification, as previously suggested for the occurrence of the different *Synechococcus* clades within SC 5.1 and of the *Prochlorococcus* radiation [23,119] and more extended 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 continuous diversification at a fairly constant rate, with diversity purged during rare but severe extinction events, leaving traces only of the surviving ones. 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 substitutions) to a particular niche. We could advocate achieving this via comparative genomics, but this usually necessitates hundreds to thousands of closely related genomes (for review see [120,121]), as well as a refined phenotypic characterization of the sequenced strains. Alternatively, one could search *in situ* data for genes or substitutions related to a

particular niche or environmental parameter (see e.g. [68,122–124]). Given the wealth of marine metagenomes that are becoming available for a large variety of environmental niches, such an approach should be particularly powerful to unveil niche adaptation processes in the forthcoming years.

Methods

Genome sequencing and assembly

Thirty-four *Synechococcus* strains were chosen for genome sequencing based on their phylogenetic position, pigment content and isolation sites (Fig. 1, Supplementary Table S1). All but the three KORDI strains were retrieved from the Roscoff Culture Collection (RCC; http://roscoff-culture-collection.org/) and transferred three times on 0.3% SeaPlaque Agarose (Lonza, Switzerland) to clone them and reduce contamination by heterotrophic bacteria. A first set of 25 *Synechococcus* genomes (including WH8103) were generated at the Genoscope (CEA, Paris-Saclay, France) by shotgun sequencing of two libraries: a short-insert forward-reverse pair-end (PE) library (50-150 bp) and a long-insert reverse-forward mate-pair library (4-10 kb), both sequenced by Illumina™ technology. Additionally, seven other genomes were sequenced at the Center for Genomic Research (University of Liverpool, UK) by shotgun sequencing of 250 bp reads. Single or PE reads were first assembled into contigs using the CLC AssemblyCell© 4.10 (CLC Bio, Aarhus, Denmark). *Synechococcus* contigs were identified based on their different coverage compared to heterotrophic bacteria, scaffolded using WiseScaffolder and 28 out of 31 genomes were closed by manual finishing as described in [42]. Three genomes (BIOS-E4-1, BOUM118 and RS9915), had only one to three gaps in highly repeated genomic regions. The base numbering of the circularized genomes was set at 174

bp before the *dnaN* start, corresponding approximately to the origin of replication. Automatic structural and functional annotation of the genomes was then realized using the Institute of Genome Science (IGS) Annotation Engine (http://ae.igs.umaryland.edu/cgi/index.cgi; [125]). As concerns KORDI-49, KORDI-52 and KORDI-100 strains, genomes were sequenced from axenic cultures using a 454 GS-FLX Titanium sequencing system (Roche) at Macrogen (Seoul, Korea). The obtained reads were assembled using the Newbler assembler (version 2.3, Roche). To fill contig gaps, additional PCR and primer walking was conducted. Sequences of all new *Synechococcus* genomes were deposited in GenBank under accession numbers CP047931-CP047961 (BioProject PRJNA596899), except *Synechococcus* sp. WH8103 that was previously deposited to illustrate the performance of the pipeline used to assemble, scaffold and manually finish these genomes (Supplementary Table S1).

Clustering of orthologous genes

Protein and RNA sequences retrieved from new genomes were clustered with genomes previously available (Supplementary Table S1) into CLOGs using the OrthoMCL algorithm [126] and were then imported into the custom-designed Cyanorak v2.1 information system (www.sb-roscoff.fr/cyanorak/) for further manual curation and functional annotation, as detailed in [51]. Clustering into CLOGs allowed us to build phyletic patterns (i.e. the number of copies of each gene in each genome per CLOG), 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 three genomes were available for a given taxonomic level (see Supplementary Table S2).

The phyletic pattern was also used to estimate the size of the pan-genome and core genome. The sampling of genome combinations necessary to draw pan-genome curves was

performed with the software PanGP [127] using as parameters 'Totally Random', SR=100 and SS=1000. Pan-genome curves were then drawn with R custom designed scripts (v3.3.1.; [128]). The results of PanGP exponential fits were used as estimates of the asymptotic number of core genes.

ANI/AAI calculation

Whole-genome ANI and percentage of conserved DNA between pairs of genomes (percentage of the genome length aligned by Blast with more than 90% ID) were calculated following the method described in [49]. AAI was calculated following the method described by [82]. When AAI values differed for a given pair of strains depending on which strain was used as a query for BLAST, the highest value was kept.

Phylogeny and tree comparisons

The *petB* phylogenetic tree was built using PhyML 3.1 [129] with the HKY model and by estimating gamma parameters and the proportion of invariant sites, based on a database of 230 *petB* sequences [21,25]. The confidence of branch points was determined by performing bootstrap analyses, including 1000 replicate data sets. Phylogenetic trees were edited using the Archaeopteryx v0.9901 beta program [130]. The tree was drawn using iTOL (http://itol.embl.de; [131]). Additionally, a set of 821 single-copy core proteins were aligned with MAFFT v7.164b [132] and concatenated into a single alignment, resulting in a total of 226,778 amino acids. A phylogenetic tree was built with PhyML 3.1 with the WAG model and estimation of parameters of the gamma distribution and of the proportion of invariant sites. The phylogeny based on gene content was performed as described in [133]: a Jaccard distance matrix was computed from the phyletic pattern with the package *vegan* [134] and the matrix

was then used by the Neighbor-Joining algorithm implemented in the R package *ape* [135] to generate a tree with 100 bootstraps.

The phylogenetic tree based on core proteins was then compared to the tree based on the phyletic pattern using the R package *dendextend* v.1.3.0 [136]. Branch lengths were compared using custom python scripts based on the ete2 toolkit [137]. Briefly, for each node at the base of a clade (highlighted by blue dots in Fig. 8), 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 distribution of ratios of external to internal branch lengths were drawn in R for both trees and a paired Mann-Whitney-Wilcoxon test assessed the difference between the mean ratios.

Estimation of gene gains and losses

The number of gene gains and losses were assessed from phyletic patterns using the software *Count* [138] that implements a Maximum Likelihood method for estimating the ancestral states (presence, absence or multiple copies) of every CLOG in the dataset using the phylogenetic core protein tree as reference and allowing four categories for the gamma distribution of duplications and branch lengths (options -transfer_k 1 -length_k 4 -loss_k 1 -duplication_k 4). Cut-off on posterior probability was set at 90%, which allowed us to obtain 2,921 CLOGs at the root of the tree, a number similar to the average number of CLOGs in present-day *Synechococcus* strains. The state of presence-absence of each gene family was then extracted at each node of the tree, and used to calculate the number of gene gains and losses on every branch.

These estimations of gained genes were also used to predict genomic islands in each strain. A genomic island, starting and finishing with full-length gained genes, was defined from

consecutive sliding windows (size 10,000 bp, interval 100 bp) with a ratio of nucleotides from gained CDS to total coding nucleotides higher than 50%. A network approach was then applied on all predicted islands to compare the gene content of these islands between all strains. Jaccard distances based on shared gene content were calculated between islands and an edge was drawn to connect two islands if their distance was higher than 0.1 (i.e., when two islands shared at least 10% of their pooled gene content). Network modules detection was then performed using the modularity algorithm ([139]; resolution = 0.2) implemented in Gephi version 0.9.2 [140]. Furthermore, in order to take into account the phylogenetic relatedness between strains sharing genomic islands, a distance matrix based on core protein sequences was computed and used to color edges between nodes. Networks were then represented following the “Atlas 2” spatialization implemented in Gephi.

Time calibration of the tree

The core protein phylogeny was used as input for the *reltime* algorithm [141] and the JTT matrix-based model [142], as implemented in MEGA7 [143], with default parameters and SC 5.3 designated as an outgroup. Two calibration points were used, based on [47] and TimeTree [144]: the first calibration point was set on node n2 (Supplementary Fig. S7), i.e. the common ancestor of strains WH5701 and WH8102 estimated to have occurred between 582 and 878 My ago, and the second on node n4 (i.e. the common ancestor of strains CC9311 and WH8102; Supplementary Fig. S7), set between 252 and 486 My. This method allowed us to relate 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* [45]. 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 and the branch to obtain a substitution rate per My.

Estimation of the number of fixed genes and fixed substitutions specific to a taxon or shared between taxa

At a given node of the tree, 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, every position that showed the same amino acid variant in all leaves below a node and another amino-acid in every other strain were considered as fixed variants specific to this node. Terminal branches were not taken into account in these calculations since, by definition, strain-specific amino acids or genes occurring in these branches cannot be considered as fixed.

Additionally, we also looked in *Synechococcus-Cyanobium* core genes for amino acid variants specific to a set of strains corresponding to clades (Supplementary Table S8). A variant was considered as specific to a set of strains if it showed the same amino acid in every strain within the set and any other amino acid in every other strain. To allow comparison between proteins of different lengths, the number of specific variants was normalized by gene length. Given that older clades are expected to have accumulated more substitutions, each set of strains proteins were ranked according to their proportion of specific variants. To identify candidate proteins potentially involved in adaptation to cold conditions in clades I and IV, we took the ratio of the protein rank for the “clades I and IV” set of strains to the median rank for other clades (excluding the clades containing a single strain). We kept only proteins for which this ratio was below 0.33, i.e. proteins with a rank 3 times higher in the “clades I and IV” set than in other clades (Supplementary Table S8).

Data availability

The genome dataset supporting the conclusions of this article is available in the Genbank repository and NCBI accession numbers of each genome is available in Supplementary Table S1.

Ethics declarations

Ethic approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests

Funding

This work was supported by the French “Agence Nationale de la Recherche” Programs SAMOSA (ANR-13-ADAP-0010) and CINNAMON (ANR-17-CE2-0014-01), the Genoscope project METASYN and Natural Environment Research Council grant NE/I00985X/1.

Author contributions

FH, MR, FDP, MO, DHC, JHN and CS purified newly sequenced *Synechococcus* strains, FH, MR, FDP, DHC, JHN and MO extracted the DNA, KL, JMA, PW, FDP, DHC, JHN, FP, LG, MH and GKF participated in sequencing and/or assembly of the genomes, MH, GLC, EC, AB, LBG and GFK developed and run the automatic clustering and annotation pipelines. HD, UG, FP and LG

participated in the expert manual annotation of the genomes. HD, GKF, UG and LG generated and processed the data. UG, HD, JH and DE produced the genomic island networks. HD, GKF, UG, DE, DJS, FP and LG analyzed the results. HD, UG, JH, GKF and LG made the figures. All authors contributed to the preparation of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank the Institute for Genome Sciences Annotation Engine service at the University of Maryland School of Medicine, and in particular Michelle Giglio and Suvarna Nadendla, for providing automatic structural and functional annotation of the sequences, Brian Palenik and Tanja Woike for authorizing us to use the two unpublished *Synechococcus* genomes WH8016 and CC9616 as well as Garance Monier and Théo Sciandra for participating in the curation of the Cyanorak v2.1 database. We warmly thank the Roscoff Culture Collection and Sophie Mazard for maintaining and isolating some of the *Synechococcus* strains used in this study as well as the ABiMS platform for providing computational support for this work. This work is dedicated to our esteemed colleague Christophe Caron, who deceased on May 5th 2018.

References

1. Edwards M, Richardson AJ. Impact of climate change on marine pelagic phenology and trophic mismatch. Nature. 2004;430: 881–884. doi:10.1038/nature02808
2. Sears MW, Angilletta MJ. Introduction to the symposium: Responses of organisms to climate change: A synthetic approach to the role of thermal adaptation. Integr Comp Biol. 2011;51: 662–665. doi:10.1093/icb/icr113
3. Doblin MA, Van Sebille E. Drift in ocean currents impacts intergenerational microbial exposure to temperature. Proc Natl Acad Sci U S A. 2016;113: 5700–5705. doi:10.1073/pnas.1521093113
4. Irwin AJ, Finkel Z V., Müller-Karger FE, Ghinaglia LT. Phytoplankton adapt to changing ocean environments. Proc Natl Acad Sci U S A. 2015;112: 5762–5766. doi:10.1073/pnas.1414752112
5. Berube PM, Rasmussen A, Braakman R, Stepanauskas R, Chisholm SW. Emergence of trait variability through the lens of nitrogen assimilation in prochlorococcus. Elife. 2019;8: e41043. doi:10.7554/eLife.41043
6. Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A, et al. Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. Science (80-). 2014;344: 416–20. doi:10.1126/science.1248575
7. Malmstrom RR, Rodrigue S, Huang KH, Kelly L, Kern SE, Thompson A, et al. Ecology of uncultured *Prochlorococcus* clades revealed through single-cell genomics and biogeographic analysis. ISME J. 2012;7: 184–98. doi:10.1038/ismej.2012.89
8. Nakayama T, Nomura M, Takano Y, Tanifuji G, Shiba K, Inaba K, et al. Single-cell genomics unveiled a cryptic cyanobacterial lineage with a worldwide distribution hidden by a dinoflagellate host. Proc Natl Acad Sci. 2019;116: 15973 LP – 15978. doi:10.1073/pnas.1902538116
9. Stepanauskas R, Sieracki ME. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. Proc Natl Acad Sci. 2007;104: 9052 LP – 9057. doi:10.1073/pnas.0700496104
10. Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust EV. Untangling genomes from metagenomes: Revealing an uncultured class of marine Euryarchaeota. Science (80-). 2012;335: 587–590. doi:10.1126/science.1212665
11. Haro-Moreno JM, López-pérez M, Torre JR De, Picazo A, Camacho A, Rodriguez-Valera F, et al. Fine metagenomic profile of the Mediterranean stratified and mixed water columns revealed by assembly and recruitment. Microbiome. 2018;6: 128. doi:10.1186/s40168-018-0513-5
12. Partensky F, Blanchot J, Vaulot D. Differential distribution and ecology of Prochlorococcus and Synechococcus in oceanic waters: a review. In: Charpy L, Larkum AWD, editors. Marine Cyanobacteria. Monaco: Musée Océanographique; 1999. pp. 457–475.
13. Scanlan DJ. Marine picocyanobacteria. In: Whitton BA,

editor. Ecology of Cyanobacteria II: Their Diversity in Space and Time. Dordrecht: Springer Netherlands; 2012. pp. 503–533. doi:10.1007/978-94-007-3855-3_20

14. Jardillier L, Zubkov M V, Pearman J, Scanlan DJ. Significant CO2 fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean. ISME J. 2010;4: 1180. doi:10.1038/ismej.2010.36
15. Flombaum P, Gallegos JL, Gordillo R a, Rincón J, Zabala LL, Jiao N, et al. Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. Proc Natl Acad Sci U S A. 2013;110: 9824–9. doi:10.1073/pnas.1307701110
16. Guidi L, Chaffron S, Bittner L, Eveillard D, Larhlimi A, Roux S, et al. Plankton networks driving carbon export in the oligotrophic ocean. Nature. 2016;532: 465–470. doi:10.1038/nature16942
17. Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward EMS, Chisholm SW. Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. Science (80-). 2006;311: 1737–1740. doi:10.1126/science.1118052
18. Paulsen ML, Doré H, Garczarek L, Seuthe L, Müller O, Sandaa R-A, et al. *Synechococcus* in the Atlantic Gateway to the Arctic Ocean. Front Mar Sci. 2016;3. doi:10.3389/fmars.2016.00191
19. Bouman HA, Ulloa O, Scanlan DJ, Zwirglmaier K, Li WKW, Platt T, et al. Oceanographic basis of the global surface distribution of *Prochlorococcus* ecotypes. Science (80-). 2006;312: 918–921. doi:10.1126/science.1122692
20. Zwirglmaier K, Jardillier L, Ostrowski M, Mazard S, Garczarek L, Vaulot D, et al. Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. Environ Microbiol. 2008;10: 147–161. doi:10.1111/j.1462-2920.2007.01440.x
21. Farrant GK, Doré H, Cornejo-Castillo FM, Partensky F, Ratin M, Ostrowski M, et al. Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. Proc Natl Acad Sci. 2016;113: E3365–E3374. doi:10.1073/pnas.1524865113
22. Sohm JA, Ahlgren NA, Thomson ZJ, Williams C, Moffett JW, Saito MA, et al. Co-occurring *Synechococcus* ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. ISME J. 2015;10: 1–13. doi:10.1038/ismej.2015.115
23. Dufresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik BP, et al. Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. Genome Biol. 2008;9: R90. doi:10.1186/gb-2008-9-5-r90
24. Huang S, Wilhelm SW, Harvey HR, Taylor K, Jiao N, Chen F. Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. ISME J. 2012;6: 285–97. doi:10.1038/ismej.2011.106
25. Mazard S, Ostrowski M, Partensky F, Scanlan DJ. Multi-locus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*. Environ Microbiol. 2012;14: 372–86. doi:10.1111/j.1462-2920.2011.02514.x
26. Ahlgren NA, Rocap G. 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. 2012;3: 213. doi:10.3389/fmicb.2012.00213

27. Biller SJ, Berube PM, Lindell D, Chisholm SW. *Prochlorococcus*: the structure and function of collective diversity. Nat Rev Microbiol. 2014;13: 13–27. doi:10.1038/nrmicro3378
28. Zinser ER, Johnson ZI, Coe A, Karaca E, Veneziano D, Chisholm SW. Influence of light and temperature on *Prochlorococcus* ecotype distributions in the Atlantic Ocean. Limnol Oceanogr. 2007;52: 2205–2220. doi:10.4319/lo.2007.52.5.2205
29. Martiny AC, Tai APK, Veneziano D, Primeau F, Chisholm SW. Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. Environ Microbiol. 2009;11: 823–32. doi:10.1111/j.1462-2920.2008.01803.x
30. Rusch DB, Martiny AC, Dupont CL, Halpern AL, Venter JC. Characterization of *Prochlorococcus* clades from iron-depleted oceanic regions. Proc Natl Acad Sci U S A. 2010;107: 16184–16189. doi:10.1073/pnas.1009513107
31. West NJ, Lebaron P, Strutton PG, Suzuki MT. A novel clade of *Prochlorococcus* found in high nutrient low chlorophyll waters in the South and Equatorial Pacific Ocean. ISME J. 2011;5: 933–944. doi:10.1038/ismej.2010.186
32. Scanlan DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR, et al. Ecological genomics of marine picocyanobacteria. Microbiol Mol Biol Rev. 2009;73: 249–99. doi:10.1128/MMBR.00035-08
33. Pittera J, Humily F, Thorel M, Grulois D, Garczarek L, Six C. Connecting thermal physiology and latitudinal niche partitioning in marine *Synechococcus*. ISME J. 2014;8: 1221–1236. doi:10.1038/ismej.2013.228
34. Larkin AA, Blinebry SK, Howes C, Lin Y, Loftus SE, Schmaus CA, et al. Niche partitioning and biogeography of high light adapted *Prochlorococcus* across taxonomic ranks in the North Pacific. ISME J. 2016;10: 1555–1567. doi:10.1038/ismej.2015.244
35. Lan R, Reeves PR. Intraspecies variation in bacterial genomes: The need for a species genome concept. Trends Microbiol. 2000;8: 396–401. doi:10.1016/S0966-842X(00)01791-1
36. Cordero OX, Polz MF. Explaining microbial genomic diversity in light of evolutionary ecology. Nat Rev Microbiol. 2014;12: 263–273. doi:10.1038/nrmicro3218
37. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. Proc Natl Acad Sci U S A. 2005;102: 13950–5. doi:10.1073/pnas.0506758102
38. Simm S, Keller M, Selymesi M, Schleiff E. The composition of the global and feature specific cyanobacterial core-genomes. Front Microbiol. 2015;6: 1–21. doi:10.3389/fmicb.2015.00219
39. Baumdicker F, Hess WR, Pfaffelhuber P. The infinitely many genes model for the distributed genome of bacteria. Genome Biol Evol. 2012;4: 443–456.

doi:10.1093/gbe/evs016
40. Kettler GC, Martiny AC, Huang K, Zucker J, Coleman ML, Rodrigue S, et al. Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. PLoS Genet. 2007;3: e231. doi:10.1371/journal.pgen.0030231
41. Biller SJ, Berube PM, Berta-Thompson JW, Kelly L, Roggensack SE, Awad L, et al. Genomes of diverse isolates of the marine cyanobacterium *Prochlorococcus*. Sci Data. 2014;1: 1–11. doi:10.1038/sdata.2014.34
42. Farrant GK, Hoebeke M, Partensky F, Andres G, Corre E, Garczarek L. WiseScaffolder: an algorithm for the semi-automatic scaffolding of Next Generation Sequencing data. BMC Bioinformatics. 2015;16: 281. doi:10.1186/s12859-015-0705-y
43. Brahamsha B. An abundant cell-surface polypeptide is required for swimming by the nonflagellated marine cyanobacterium *Synechococcus*. Proc Natl Acad Sci U S A. 1996;93: 6504–6509. doi:10.1073/pnas.93.13.6504
44. McCarren J, Brahamsha B. SwmB, a 1.12-megadalton protein that is required for nonflagellar swimming motility in *Synechococcus*. J Bacteriol. 2007;189: 1158–1162. doi:10.1128/JB.01500-06
45. Dufresne A, Garczarek L, Partensky F. Accelerated evolution associated with genome reduction in a free-living prokaryote. Genome Biol. 2005;6: R14.1-10. doi:10.1186/gb-2005-6-2-r14
46. Shih PM, Wu D, Latifi A, Axen SD, Fewer DP, Talla E, et al. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. Proc Natl Acad Sci U S A. 2013;110: 1053–8. doi:10.1073/pnas.1217107110
47. Sánchez-Baracaldo P. Origin of marine planktonic cyanobacteria. Sci Rep. 2015;5: 17418. doi:10.1038/srep17418
48. Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for prokaryotes. Proc Natl Acad Sci U S A. 2005;102: 2567–2572. doi:10.1073/pnas.0409727102
49. Goris J, Konstantinidis KT, Klappenbach J a., Coenye T, Vandamme P, Tiedje JM. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol. 2007;57: 81–91. doi:10.1099/ijs.0.64483-0
50. Yan W, Wei S, Wang Q, Xiao X, Zeng Q, Jiao N, et al. Genome rearrangement shapes *Prochlorococcus* ecological adaptation. Elliot MA, editor. Appl Environ Microbiol. 2018;84: e01178-18. doi:10.1128/AEM.01178-18
51. Garczarek L, Guyet U, Doré H, Farrant GK, Hoebeke M, Guéguen L, et al. Cyanorak v2.1, a scalable information system dedicated to the visualization and expert curation of marine and brackish picocyanobacteria genomes. Submitted.
52. Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, Delong EF, et al. Genomic islands and the ecology and evolution of *Prochlorococcus*. Science. 2006;311: 1768–1770. doi:10.1126/science.1122050
53. Delmont TO, Eren AM. Linking pangenomes and metagenomes : the *Prochlorococcus* metapangenome. PeerJ. 2018;6: 1–23. doi:10.7717/peerj.4320
54. Fuller NJ, Marie D, Partensky F,

Vaulot D, Post AF, Scanlan DJ. 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. 2003;69: 2430–2443. doi:10.1128/AEM.69.5.2430-2443.2003

55. Chen F, Wang K, Kan J, Suzuki MT, Wommack KE. Diverse and unique picocyanobacteria in Chesapeake Bay , revealed by 16S-23S rRNA internal transcribed spacer sequences. Appl Environ Microbiol. 2006;72: 2239–2243. doi:10.1128/AEM.72.3.2239
56. Partensky F, Garczarek L. *Prochlorococcus*: advantages and limits of minimalism. Ann Rev Mar Sci. 2010;2: 305–31. doi:10.1146/annurev-marine-120308-081034
57. Tang W, van der Donk WA. Structural characterization of four prochlorosins: A novel class of lantipeptides produced by planktonic marine cyanobacteria. Biochemistry. 2012;51: 4271–4279. doi:10.1021/bi300255s
58. Kamennaya NA, Post AF. Characterization of cyanate metabolism in marine Synechococcus and Prochlorococcus spp. Appl Env Microbiol. 2011;77: 291–301. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=21057026
59. McCarren J, Heuser J, Roth R, Yamada N, Martone M, Brahamsha B. Inactivation of *swmA* results in the loss of an outer Cell layer in a swimming *Synechococcus* strain. J Bacteriol. 2005;187: 224 LP – 230. doi:10.1128/JB.187.1.224-230.2005
60. Dempwolff F, Wischhusen HM, Specht M, Graumann PL. The deletion of bacterial dynamin and flotillin genes results in pleiotrophic effects on cell division, cell growth and in cell shape maintenance. BMC Microbiol. 2012;12: 298. doi:10.1186/1471-2180-12-298
61. Cabello-Yeves PJ, Haro-Moreno JM, Martin-Cuadrado AB, Ghai R, Picazo A, Camacho A, et al. Novel *Synechococcus* genomes reconstructed from freshwater reservoirs. Front Microbiol. 2017;8: 1–13. doi:10.3389/fmicb.2017.01151
62. Partensky F, Garczarek L. *Prochlorococcus*: advantages and limits of minimalism. Ann Rev Mar Sci. 2010;2: 305–31. doi:10.1146/annurev-marine-120308-081034
63. Partensky F, Blanchot J, Vaulot D. Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters: a review. Bull l'institut océanographique. 1999;19: 457–475.
64. Lawrence JG, Ochman H. Molecular archaeology of the *Escherichia coli* genome. Proc Natl Acad Sci U S A. 1998;95: 9413–9417. doi:10.1073/pnas.95.16.9413
65. McDonald BR, Currie CR. Lateral gene transfer dynamics in the ancient bacterial genus Streptomyces. MBio. 2017;8: e00644-17. doi:10.1128/mBio.00644-17
66. Genevaux P, Georgopoulos C, Kelley WL. The Hsp70 chaperone machines of *Escherichia coli*: A paradigm for the repartition of chaperone functions. Mol Microbiol. 2007;66:

840–857. doi:10.1111/j.1365-2958.2007.05961.x

67. Tetu SG, Brahamsha B, Johnson DA, Tai V, Phillippy K, Palenik B, et al. Microarray analysis of phosphate regulation in the marine cyanobacterium *Synechococcus* sp. WH8102. ISME J. 2009;3: 835–849. doi:10.1038/ismej.2009.31
68. Kent AG, Dupont CL, Yooseph S, Martiny AC. Global biogeography of *Prochlorococcus* genome diversity in the surface ocean. ISME J. 2016;10: 1856–1865. doi:10.1038/ismej.2015.265
69. Scanlan DJ, West NJ. Molecular ecology of the marine cyanobacterial genera *Prochlorococcus* and *Synechococcus*. FEMS Microbiol Ecol. 2002;40: 1–12. doi:10.1111/j.1574-6941.2002.tb00930.x
70. diCenzo GC, Sharthiya H, Nanda A, Zamani M, Finan TM. PhoU allows rapid adaptation to high Phosphate concentrations by modulating PstSCAB transport rate in *Sinorhizobium meliloti*. Stock AM, editor. J Bacteriol. 2017;199: e00143-17. doi:10.1128/JB.00143-17
71. Erdner DDL, Anderson DMD. Ferredoxin and flavodoxin as biochemical indicators of iron limitation during open-ocean iron enrichment. Limnol Oceanogr. 1999;44: 1609–1615. doi:10.4319/lo.1999.44.7.1609
72. He Q, Dolganov N, Bjo O, Grossman AR, Natl P, Sci A. The high light-inducible polypeptides in *Synechocystis* PCC6803. J Biol Chem. 2001;276: 306–314. doi:10.1074/jbc.M008686200
73. Axelsen KB, Palmgren MG. Evolution of substrate specificities in the P-type ATPase superfamily. J Mol Evol. 1998;46: 84–101. doi:10.1007/PL00006286
74. Xia X, Guo W, Tan S, Liu H. *Synechococcus* assemblages across the salinity gradient in a salt wedge estuary. Frontiers in Microbiology. 2017. p. 1254. Available: https://www.frontiersin.org/article/10.3389/fmicb.2017.01254
75. Haswell ES, Phillips R, Rees DC. Mechanosensitive channels: What can they do and how do they do it? Structure. 2011;19: 1356–1369. doi:10.1016/j.str.2011.09.005
76. Scybert S, Pechous R, Sitthisak S, Nadakavukaren MJ, Wilkinson BJ, Jayaswal RK. NaCl-sensitive mutant of *Staphylococcus aureus* has a Tn917-*lacZ* insertion in its *ars* operon. FEMS Microbiol Lett. 2003;222: 171–176. doi:10.1016/S0378-1097(03)00312-4
77. Liu S, Cheng Y, Zhang X, Guan Q, Nishiuchi S, Hase K, et al. 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. 2007;64: 49–58. doi:10.1007/s11103-007-9133-3
78. Zwirglmaier K, Heywood JL, Chamberlain K, Woodward EMS, Zubkov M V, Scanlan DJ. Basin-scale distribution patterns of picocyanobacterial lineages in the Atlantic Ocean. Environ Microbiol. 2007;9: 1278–1290. doi:10.1111/j.1462-2920.2007.01246.x
79. Kent AG, Baer SE, Mouginot C, Huang JS, Larkin AA, Lomas MW, et al. Parallel phylogeography of *Prochlorococcus* and

Synechococcus. ISME J. 2019;13: 430–441. doi:10.1038/s41396-018-0287-6

80. Tsementzi D, Wu J, Deutsch S, Nath S, Rodriguez-R LM, Burns AS, et al. SAR11 bacteria linked to ocean anoxia and nitrogen loss. Nature. 2016;536: 179–183. doi:10.1038/nature19068
81. Nayfach S, Rodriguez-Mueller B, Garud N, Pollard KS. An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. Genome Res. 2016;26: 1612–1625. doi:10.1101/gr.201863.115
82. Konstantinidis KT, Tiedje JM. Towards a genome-based taxonomy for prokaryotes. J Bacteriol. 2005;187: 6258–6264. doi:10.1128/JB.187.18.6258-6264.2005
83. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018;9: 5114. doi:10.1038/s41467-018-07641-9
84. Herdman M, Castenholz RW, Waterbury JB, Rippka R. Form-genus XIII. *Synechococcus*. 2nd Editio. In: Boone D, Castenholz R, editors. Bergey's Manual of Systematics of Archaea and Bacteria Volume 1. 2nd Editio. New York: Springer-Verlag; 2001. pp. 508–512.
85. Thompson CC, Silva GGZ, Vieira NM, Edwards R, Vicente ACP, Thompson FL. Genomic taxonomy of the genus *Prochlorococcus*. Microb Ecol. 2013;66: 752–762. doi:10.1007/s00248-013-0270-8
86. Coutinho F, Tschoeke DA, Thompson F, Thomson C. Comparative genomics of *Synechococcus* and proposal of the new genus *Parasynechococcus*. PeerJ. 2016;4: e1522. doi:10.7717/peerj.1522
87. Coutinho FH, Dutilh BE, Thompson CC, Thompson FL. Proposal of fifteen new species of *Parasynechococcus* based on genomic, physiological and ecological features. Arch Microbiol. 2016;198: 973–986. doi:10.1007/s00203-016-1256-y
88. Moore LR, Chisholm SW. Photophysiology of the marine cyanobacterium *Prochlorococcus*: Ecotypic differences among cultured isolates. Limnol Oceanogr. 1999;44: 628–638. doi:10.4319/lo.1999.44.3.0628
89. Sánchez-Baracaldo P, Bianchini G, Di Cesare A, Callieri C, Chrismas NAM. Insights into the evolution of picocyanobacteria and phycoerythrin penes (*mpeBA* and *cpeBA*). Front Microbiol. 2019;10: 45. doi:10.3389/fmicb.2019.00045
90. Konstantinidis KT, Tiedje JM. Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. Curr Opin Microbiol. 2007;10: 504–509. doi:https://doi.org/10.1016/j.mib.2007.08.006
91. Komárek J. A polyphasic approach for the taxonomy of cyanobacteria: principles and applications. Eur J Phycol. 2016;51: 346–353. doi:10.1080/09670262.2016.1163738
92. Palenik B, Toledo G, Ferris M. Cyanobacterial diversity in marine ecosystems as seen by RNA polymerase (*rpoC1*) gene sequences. International Symposium on Marine Cyanobacteria and Related

Organisms. Musée océanographique; 1997. pp. 101–105.

93. Penno S, Lindell D, Post AF. Diversity of *Synechococcus* and *Prochlorococcus* populations determined from DNA sequences of the N-regulatory gene ntcA. Environ Microbiol. 2006;8: 1200–1211. doi:10.1111/j.1462-2920.2006.01010.x
94. Ahlgren NA, Rocap G. 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. 2012;3: 213. doi:10.3389/fmicb.2012.00213
95. Harris SR, Feil EJ, Holden MTG, Quail MA, Nickerson EK, Chantratita N, et al. Evolution of MRSA during hospital transmission and intercontinental spread. Science (80-). 2010;327: 469 LP – 474. Available: http://science.sciencemag.org/content/327/5964/469.abstract
96. Kennemann L, Didelot X, Aebischer T, Kuhn S, Drescher B, Droege M, et al. *Helicobacter pylori* genome evolution during human infection. Proc Natl Acad Sci. 2011;108: 5033–5038. doi:10.1073/pnas.1018444108
97. Mather AE, Reid SWJ, Maskell DJ, Parkhill J, Fookes MC, Harris SR, et al. Distinguishable epidemics of multidrug-resistant *Salmonella typhimurium* DT104 in different hosts. Science (80-). 2013;341: 1514–1517. doi:10.1126/science.1240578
98. Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA, et al. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. Nature. 2003;424: 1042–1047. doi:10.1038/nature01947
99. Palenik B, Ren Q, Dupont CL, Myers GS, Heidelberg JF, Badger JH, et al. Genome sequence of *Synechococcus* CC9311: Insights into adaptation to a coastal environment. Proc Natl Acad Sci USA. 2006;103: 13555–13559. doi:10.1073/pnas.0602963103
100. Larkin AA, Blinebry SK, Howes C, Lin Y, Loftus SE, Schmaus CA, et al. Niche partitioning and biogeography of high light adapted *Prochlorococcus* across taxonomic ranks in the North Pacific. ISME J. 2016;10: 1555–1567. doi:10.1038/ismej.2015.244
101. Martiny AC, Coleman ML, Chisholm SW. Phosphate acquisition genes in *Prochlorococcus* ecotypes: Evidence for genome-wide adaptation. Proc Natl Acad Sci U S A. 2006;103: 12552–12557. doi:10.1073/pnas.0601301103
102. Martiny AC, Huang Y, Li W. Occurrence of phosphate acquisition genes in *Prochlorococcus* cells from different ocean regions. Environ Microbiol. 2009;11: 1340–7. doi:10.1111/j.1462-2920.2009.01860.x
103. Berube PM, Biller SJ, Kent AG, Berta-Thompson JW, Roggensack SE, Roache-Johnson KH, et al. Physiology and evolution of nitrate acquisition in *Prochlorococcus*. ISME J. 2014;9: 1195–1207. doi:10.1038/ismej.2014.211
104. Lerat E, Daubin V, Ochman H, Moran NA. Evolutionary origins of genomic repertoires in bacteria. PLoS Biol. 2005;3: 0807–0814. doi:10.1371/journal.pbio.0030130
105. Ochman H, Lerat E, Daubin V.

Examining bacterial species under the specter of gene transfer and exchange. Proc Natl Acad Sci. 2005;102: 6595–6599. doi:10.1073/pnas.0502035102

106. Nowell RW, Green S, Laue BE, Sharp PM. The extent of genome flux and its role in the differentiation of bacterial lineages. Genome Biol Evol. 2014;6: 1514–1529. doi:10.1093/gbe/evu123
107. Hao W, Golding G. The fate of laterally transferred genes: life in the fast lane to adaptation or death. Genome Res. 2006; 636–643. doi:10.1101/gr.4746406.Freely
108. Abby S, Daubin V. Comparative genomics and the evolution of prokaryotes. Trends Microbiol. 2007;15: 135–141. doi:10.1016/j.tim.2007.01.007
109. Rocha EP. Evolutionary patterns in prokaryotic genomes. Curr Opin Microbiol. 2008;11: 454–460. doi:10.1016/j.mib.2008.09.007
110. Thrash JC, Temperton B, Swan BK, Landry ZC, Woyke T, DeLong EF, et al. Single-cell enabled comparative genomics of a deep ocean SAR11 bathytype. ISME J. 2014;8: 1440–51. doi:10.1038/ismej.2013.243
111. Larkin AA, Martiny AC. Microdiversity shapes the traits, niche space, and biogeography of microbial taxa. Environmental Microbiology Reports. 2017. pp. 55–70. doi:10.1111/1758-2229.12523
112. Pittera J, Partensky F, Six C. Adaptive thermostability of light-harvesting complexes in marine picocyanobacteria. ISME J. 2017;11: 112–124. doi:10.1038/ismej.2016.102
113. Osburne MS, Holmbeck BM, Frias-Lopez J, Steen R, Huang K, Kelly L, et al. 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. 2010;12: 1978–1988. doi:10.1111/j.1462-2920.2010.02203.x
114. Osburne MS, Holmbeck BM, Coe A, Chisholm SW. The spontaneous mutation frequencies of *Prochlorococcus* strains are commensurate with those of other bacteria. Environ Microbiol Rep. 2011;3: 744–749. doi:10.1111/j.1758-2229.2011.00293.x
115. Avrani S, Wurtzel O, Sharon I, Sorek R, Lindell D. Genomic island variability facilitates *Prochlorococcus*-virus coexistence. Nature. 2011;474: 604–608. doi:10.1038/nature10172
116. Lawrence JG. Gene transfer in bacteria: speciation without species? Theor Popul Biol. 2002;61: 449–460. doi:10.1006/tpbi.2002.1587
117. Cohan FM, Koeppel AF. The origins of ecological diversity in prokaryotes. Curr Biol. 2008;18: 1024–1034. doi:10.1016/j.cub.2008.09.014
118. Polz MF, Alm EJ, Hanage WP. Horizontal gene transfer and the evolution of bacterial and archaeal population structure. Trends Genet. 2013;29: 170–175. doi:10.1016/j.tig.2012.12.006
119. Urbach E, Scanlan DJ, Distel DL, Waterbury JB, Chisholm SW. Rapid diversification of marine picophytoplankton with dissimilar light- harvesting structures inferred from sequences of *Prochlorococcus* and *Synechococcus* (cyanobacteria). J Mol Evol. 1998;46: 188–201. doi:10.1007/PL00006294
120. Read TD, Massey RC. Characterizing

the genetic basis of bacterial phenotypes using genome-wide association studies : a new direction for bacteriology. Genome Med. 2014;6: 109. doi:10.1186/s13073-014-0109-z

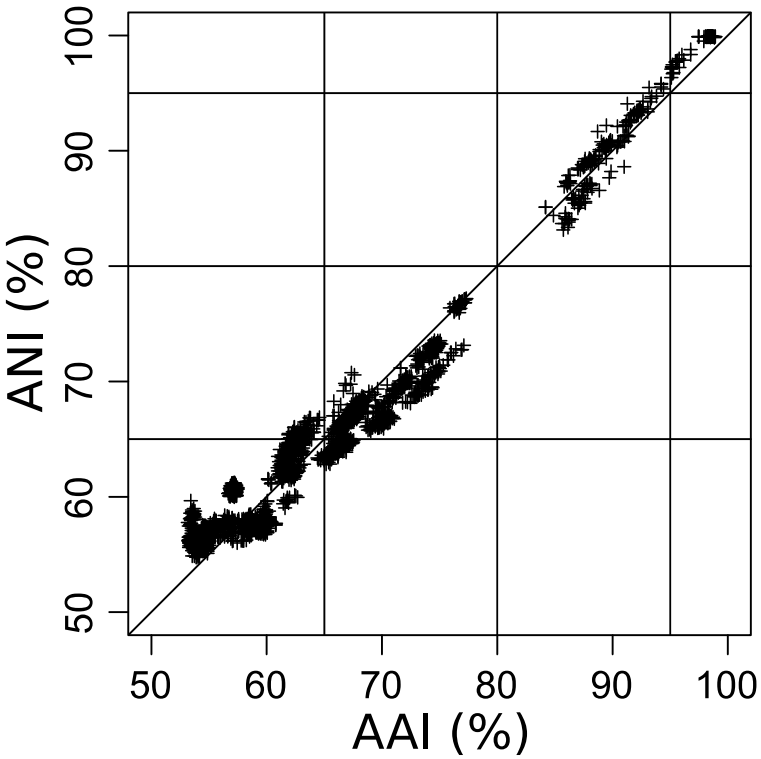
121. Chen PE, Shapiro BJ. The advent of genome-wide association studies for bacteria. Current Opinion in Microbiology. Elsevier Ltd; 2015. pp. 17–24. doi:10.1016/j.mib.2015.03.002
122. Ahlgren NA, Belisle BS, Lee MD. Genomic mosaicism underlies the adaptation of marine Synechococcus ecotypes to distinct oceanic iron niches. Environ Microbiol. 2019; 10.1111/1462-2920.14893. doi:10.1111/1462-2920.14893
123. Garcia CA, Hagstrom GI, Larkin AA, Ustick LJ, Levin SA, Lomas MW, et al. Linking regional shifts in microbial genome adaptation with surface ocean biogeochemistry. Philos Trans R Soc London B. 2020;375: 20190254. doi:10.1098/rstb.2019.0254
124. Grébert T, Doré H, Partensky F, Farrant GK, Boss ES, Picheral M, et al. Light color acclimation is a key process in the global ocean distribution of *Synechococcus* cyanobacteria. Proc Natl Acad Sci. 2018;115: E2010–E2019. doi:10.1073/pnas.1717069115
125. Galens K, Orvis J, Daugherty S, Creasy HH, Angiuoli S, White O, et al. The IGS standard operating procedure for automated prokaryotic annotation. Stand Genomic Sci. 2011;4: 244–251. doi:10.4056/sigs.1223234
126. Li L, Stoeckert CJJ, Roos DS. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Res. 2003;13: 2178–2189. doi:10.1101/gr.1224503.candidates
127. Zhao Y, Jia X, Yang J, Ling Y, Zhang Z, Yu J, et al. PanGP: A tool for quickly analyzing bacterial pan-genome profile. Bioinformatics. 2014;30: 1297–1299. doi:10.1093/bioinformatics/btu017
128. R Core Team. R: A language and environment for statistical computing. Vienna, Austria; 2013. Available: https://www.r-project.org/
129. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by Maximum Likelihood. Syst Biol. 2003;52: 696–704. doi:10.1080/10635150390235520
130. Han M V., Zmasek CM. PhyloXML: XML for evolutionary biology and comparative genomics. BMC Bioinformatics. 2009;10. doi:10.1186/1471-2105-10-356
131. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016;44: W242–W245. doi:10.1093/nar/gkw290
132. Katoh K, Standley DM. MAFFT: Iterative refinement and additional methods. In: D. R, editor. Methods in Molecular Biology (Methods and Protocols). Totowa, NJ: Humana Press; 2014. pp. 131–146. doi:10.1007/978-1-62703-646-7_8
133. Wolf YI, Rogozin IB, Grishin N V, Koonin E V. Genome trees and the Tree of Life. Trends Genet. 2002;18: 472–479. doi:https://doi.org/10.1016/S0168-9525(02)02744-0
134. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, et al. Vegan: Community ecology package. R package version 1.17-2.

2015. Available: https://cran.r-project.org/web/packages/vegan/index.html
135. Paradis E, Claude J, Strimmer K. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics. 2004;20: 289–290. doi:10.1093/bioinformatics/btg412
136. Galili T. dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics. 2015;31: 3718–3720. doi:10.1093/bioinformatics/btv428
137. Huerta-Cepas J, Dopazo J, Gabaldón T. ETE: a python environment for tree exploration. BMC Bioinformatics. 2010;11: 24. doi:10.1186/1471-2105-11-24
138. Csurös M. Count: Evolutionary analysis of phylogenetic profiles with parsimony and likelihood. Bioinformatics. 2010;26: 1910–1912. doi:10.1093/bioinformatics/btq315
139. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech Theory Exp. 2008;2008: P10008. doi:10.1088/1742-5468/2008/10/p10008
140. Bastian M, Heymann S, Jacomy M. Gephi: An open source software for exploring and manipulating networks. Int AAAI Conf Web Soc Media. 2009. Available: https://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154
141. Tamura K, Battistuzzi FU, Billing-Ross P, Murillo O, Filipski A, Kumar S. Estimating divergence times in large molecular phylogenies. Proc Natl Acad Sci U S A. 2012;109: 19333–8. doi:10.1073/pnas.1213199109
142. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. Bioinformatics. 1992;8: 275–282. doi:10.1093/bioinformatics/8.3.275
143. Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33: 1870–1874. doi:10.1093/molbev/msw054
144. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: A resource for timelines, timetrees, and divergence times. Mol Biol Evol. 2017;34: 1812–1819. doi:10.1093/molbev/msx116
145. Humily F, Partensky F, Six C, Farrant GK, Ratin M, Marie D, et al. A gene island with two possible configurations is involved in chromatic acclimation in marine *Synechococcus*. PLoS One. 2014;8: e84459. Available: https://doi.org/10.1371/journal.pone.0084459
146. Xia X, Partensky F, Garczarek L, Suzuki K, Guo C, Cheung SY, et al. Phylogeography and pigment type diversity of *Synechococcus* cyanobacteria in surface waters of the northwestern Pacific Ocean. Environ Microbiol. 2017;19: 142–158. doi:10.1111/1462-2920.13541

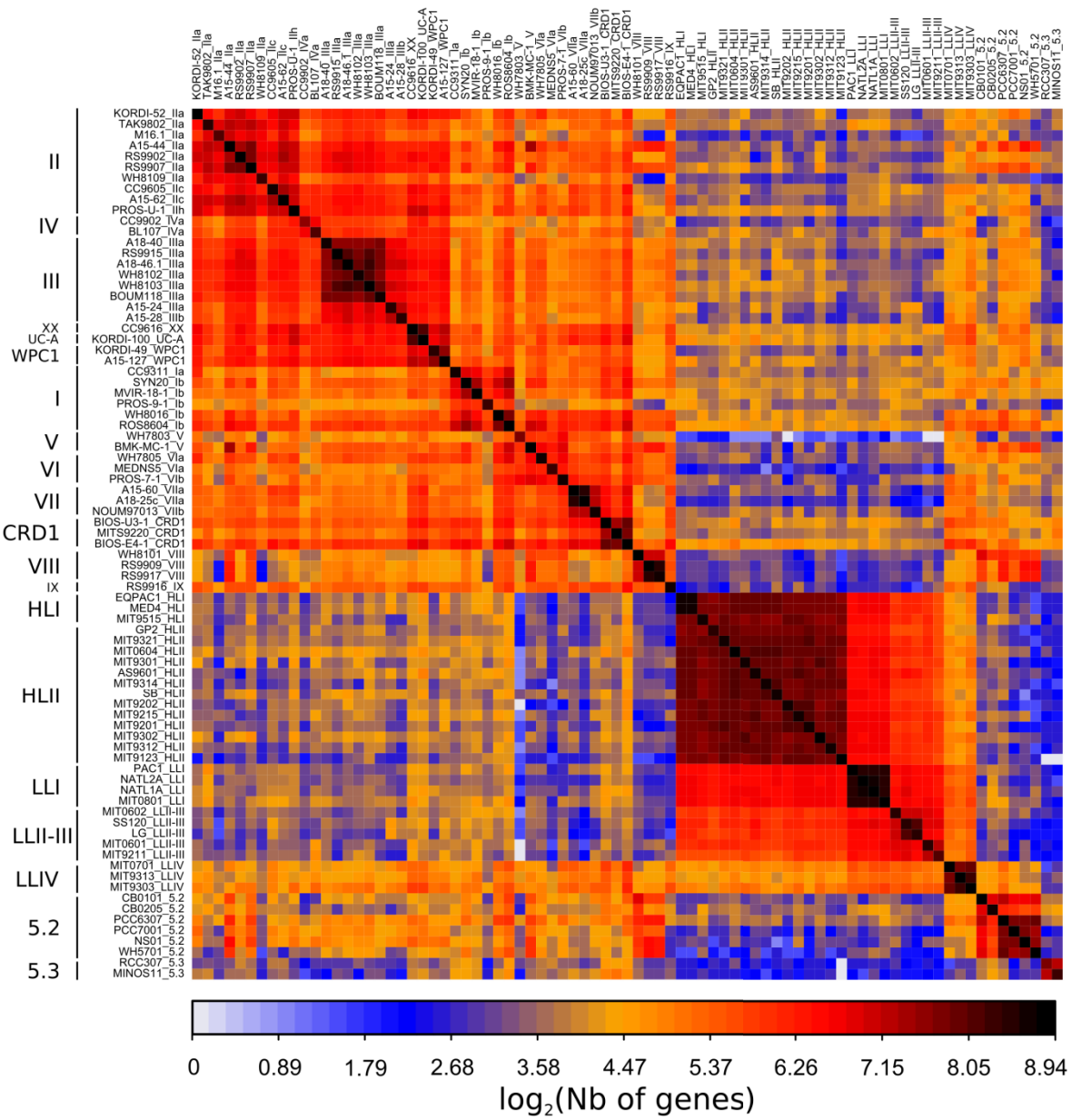
Table 1: Estimation of the number of gained, lost and/or fixed genes per million years (My) as well as total and fixed number of substitutions on internal branches (int. b.) or terminal branches (ter. b.) for *Prochlorococcus* (*Pro*) HL and *Synechococcus* (*Syn*) SC 5.1. SE: standard error, adj. R^2: adjusted R^2.

Rate (per My)		***Pro* HL** int. b.	***Pro* HL** ter. b.	***Syn* SC 5.1** int. b.	***Syn* SC 5.1** ter. b.
Gene gain	value	**1.45**	**4.5**	**0.72**	**4.62**
	SE	0.08	0.52	0.12	0.68
	adj. R^2	0.95	0.83	0.46	0.50
	p-value	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$
Gene loss	value	**0.87**	**3.72**	**1.68**	**1.8**
	SE	0.26	0.44	0.16	0.22
	adj. R^2	0.41	0.82	0.73	0.60
	p-value	4.7×10^{-3}	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$
Specific gene fixation	value	**0.39**	**-**	**0.16**	**-**
	SE	0.03	-	0.06	-
	adj. R^2	0.9	-	0.11	-
	p-value	$< 10^{-5}$	-	0.01	-
Amino acid Substitutions	value	**515.51**	**312.97**	**117.8**	**96.54**
	SE	17.2	9.11	3.64	1.86
	adj. R^2	0.98	0.99	0.96	0.98
	p-value	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$
Specific amino acid fixation	value	**78.1**	**-**	**18.41**	**-**
	SE	5.44	-	0.83	-
	adj. R^2	0.93	-	0.92	-
	p-value	$< 10^{-5}$	-	$< 10^{-5}$	-

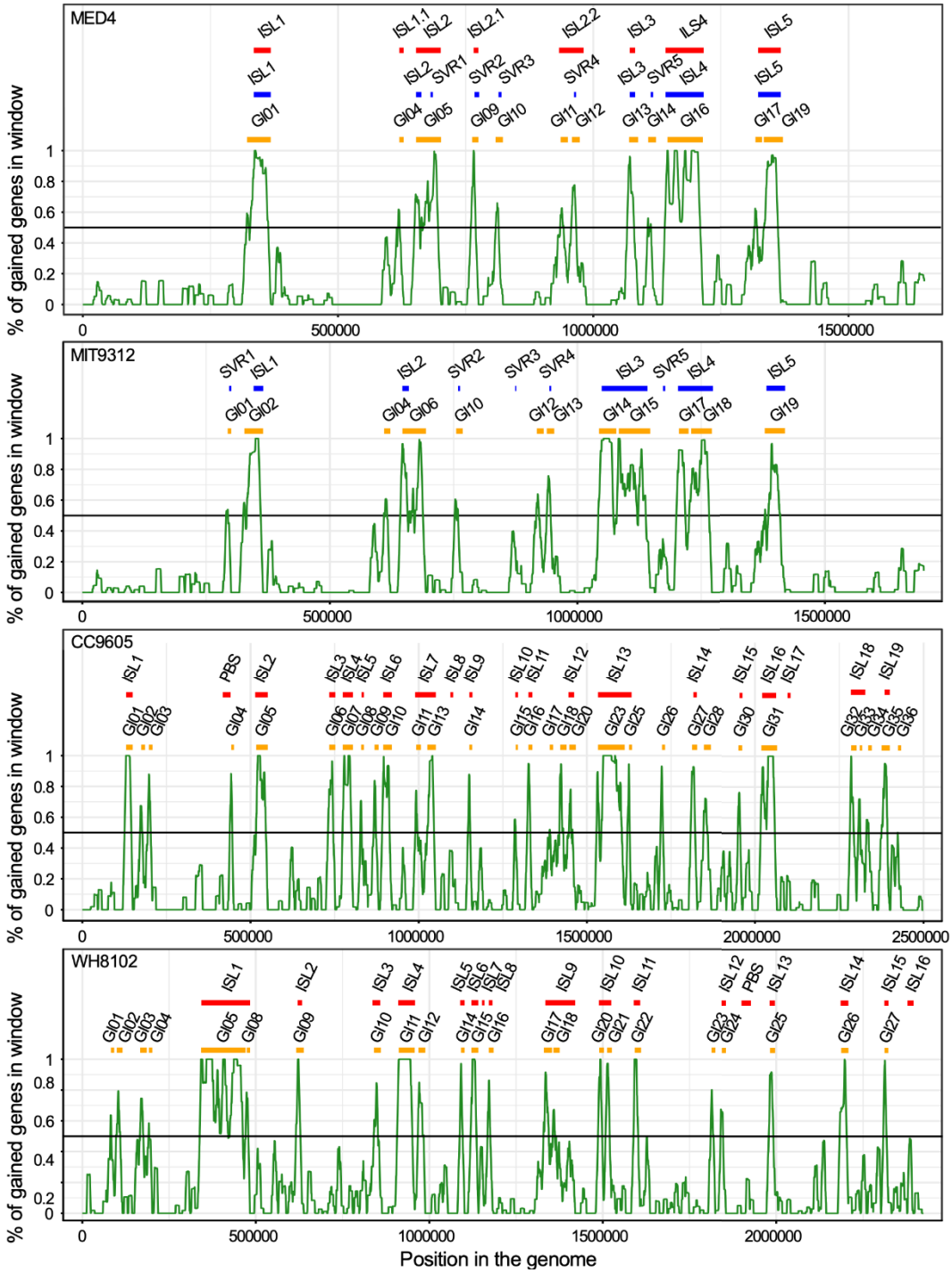
Supplementary Figures



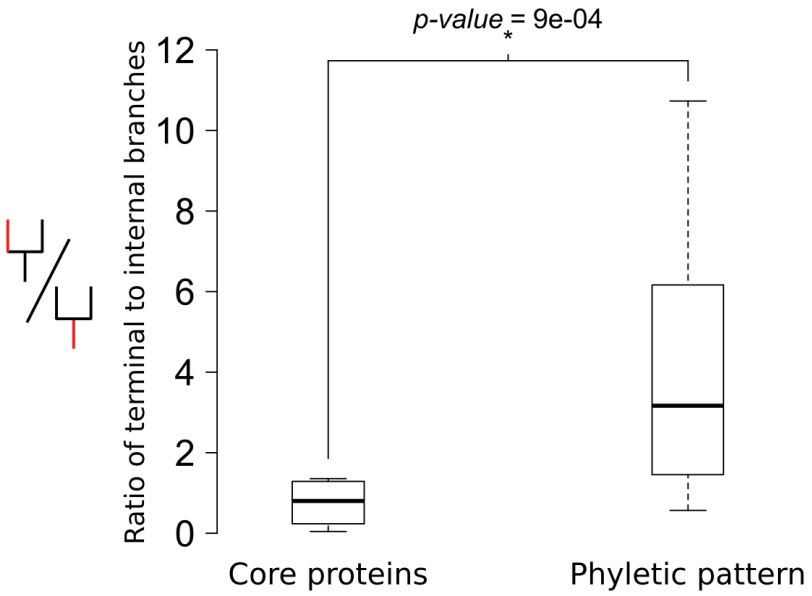
Supplementary Fig. S1: Relationship between Average Amino-acid Identity (AAI) and Average Nucleotide Identity (ANI). ANI and AAI are shown in Fig 3A.



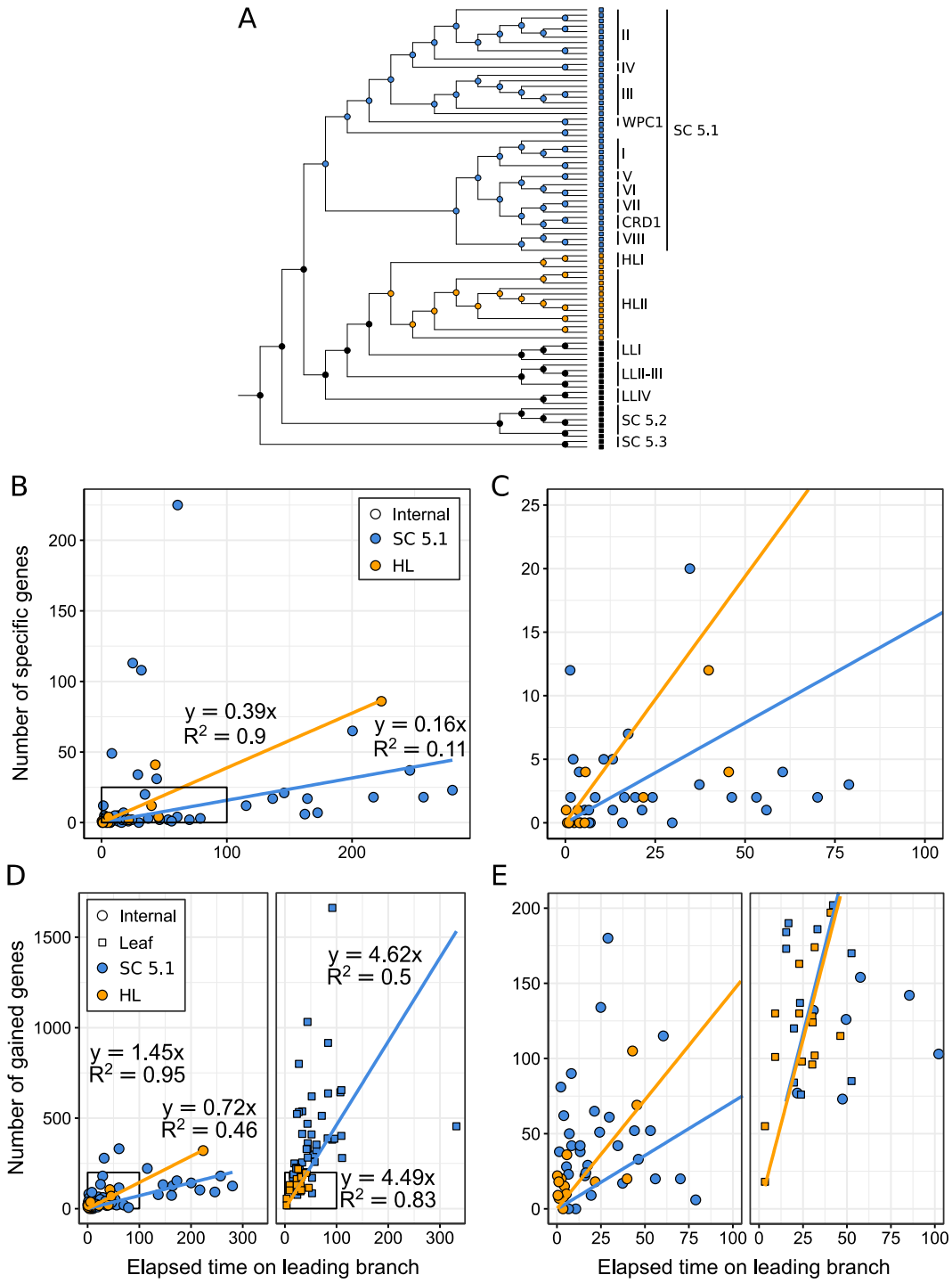
Supplementary Fig. S2: Number of gained genes located in genomic islands for all 81 picocyanobacterial genomes. The color scale indicates the total number of gained genes (log_2) predicted to be located in genomic islands in each pair of genomes. The diagonal color is thus representative of the number of gained genes in genomic islands in each genome. Strains are ordered according to their phylogenetic relatedness.



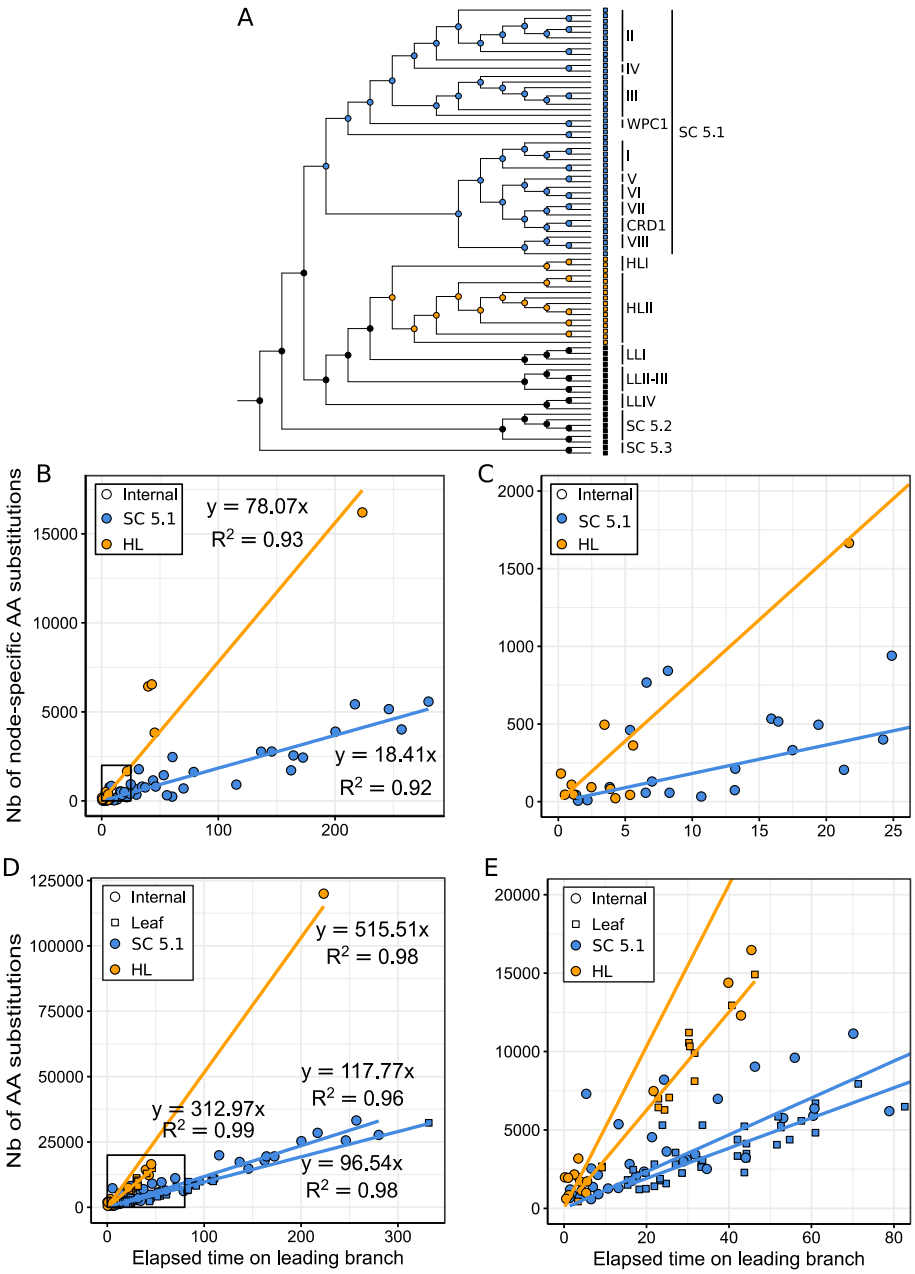
Supplementary Fig. S3: Comparison of the genomic islands delineated in previous and current work for a selection of picocyanobacterial strains. Results are shown for 2 *Prochlorococcus* strains (MED4, HLI and MIT9312, HLII) and 2 *Synechococcus* strains (CC9605, clade II and WH8102, clade III) for which islands were defined in previous studies. The green line indicates the percentage of gained genes in 10 kb windows with a 100 bp step. The black line indicates the 50% cut-off that we applied to delineate genomic islands. The location of islands defined in this study are indicated in orange. The location of islands previously defined in Supplementary Table S3 of [52] and Supplementary Material 5 of [23] are indicated in blue and red, respectively. Abbreviations: ISL and SVR correspond to 'islands' and 'smaller variable regions', respectively as defined in previous work; GI, genomic islands, as defined in the present work.



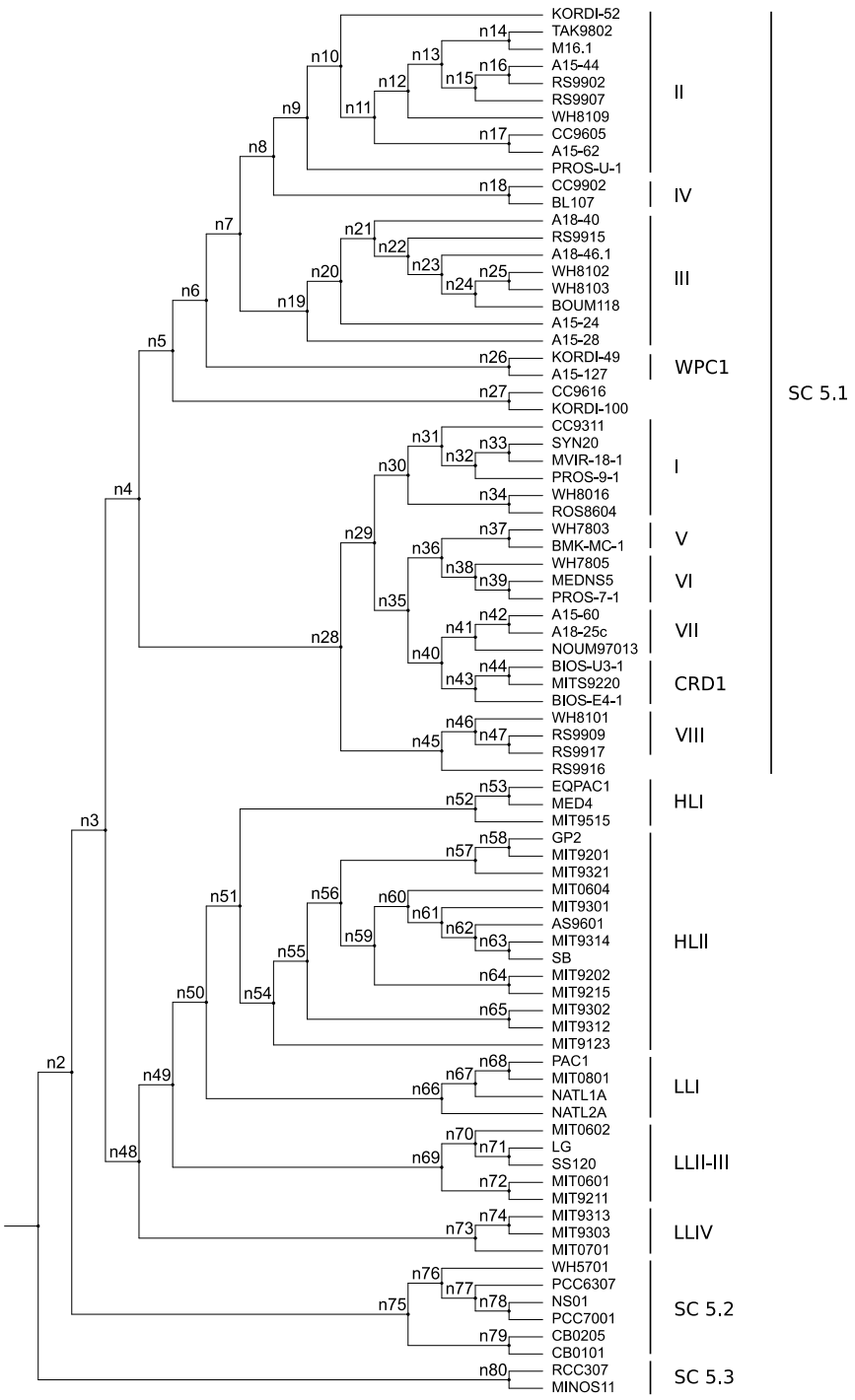
Supplementary Fig. S4: Comparison of within and between clades evolution rates. The boxplots show the distribution of ratios of clade external to internal branch lengths for each node highlighted by blue dots in Fig. 8, as calculated from trees based on core proteins and phyletic patterns, respectively. Differences between the mean ratios were assessed by a paired Mann-Whitney-Wilcoxon test (p-value ≤ 0.0009).



Supplementary Fig. S5: Linear regressions used to calculate the rates of gene gains and the rates of fixation of specific genes. A. Maximum-likelihood tree, only the topology is given. Nodes used to calculate evolutionary rates are colored in blue (SC 5.1) and orange (*Prochlorococcus* HL). Circles indicate internal nodes, and squares indicate leaves. B. The rate of fixation of specific genes is calculated as the slope of the linear regression between the number of specific genes and the time elapsed on the leading branch, for internal nodes of SC 5.1 (blue) and HL (orange). C. A zoom on the black rectangle drawn in panel B. D. The rate of gene gains is calculated as the slope of the linear regression between the number of gained genes per node and the time elapsed on the leading branch, for internal nodes (circles, left panel) and leaves (squares, right panel) of SC 5.1 (blue) and HL (orange). E. A zoom on the black rectangle drawn in panel D. Equations and R^2 are indicated for each regression.



Supplementary Fig. S6: Linear regressions used to calculate the rates of substitution and the rates of fixation of specific substitutions. A. Maximum-likelihood tree, only the topology is given. Nodes used to calculate evolutionary rates are colored in blue (SC 5.1) and orange (*Prochlorococcus* HL). Circles indicate internal nodes, and squares indicate leaves. B. The rate of specific amino-acid fixation is calculated as the slope of the linear regression between the number of node-specific amino-acid substitutions and the time elapsed on the leading branch, for internal nodes of SC 5.1 (blue) and HL (orange). C. A zoom on the black rectangle drawn in panel B. D. The rate of amino-acid substitution is calculated as the slope of the linear regression between the number of amino-acid substitutions and the time elapsed on the leading branch, for internal nodes (circles) and leaves (squares) of SC 5.1 (blue) and HL (orange). E. A zoom on the black rectangle drawn in panel D. Equations and R^2 are indicated for each regression.



Supplementary Fig. S7: 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.

Supplementary Tables

Supplementary Table S1: Accession numbers and characteristics of the genomes used in this study.

Supplementary Table S2: Core and accessory genes at each taxonomic level (phylum, genus, sub-cluster or clade) corresponding to colored boxes in Fig. 4A. A CLOG is considered as core in a taxonomic group if it is present in >= 90% of the strains within this group and only taxonomic groups with more than 3 genomes are considered. Accessory genes are shared by at least 2 strains but are not core at any higher taxonomic level. Unique genes are present only in one strain. Paralogs are considered separately.

Supplementary Table S3: Genes specific to each *Synechococcus* clade. As only a few genes were found to be strictly clade specific, relaxed rules were used (i.e., genes present in at least 80% of strains within a clade and no other *Synechococcus* strain, cf. column B). Clades gathering only one sequenced strain were not considered (i.e., clades IX, XX and UC-A). CK numbers, i.e. CLOG numbers in the Cyanorak v2.1 database, which are cited in the text are highlighted.

Supplementary Table S4: Genes specific to sets of *Synechococcus* clades representative of ecologically significant taxonomic unit (ESTU) co-occurring in the field [21]. CLOGs present in at least 90% of strains within a given assemblage and in less than 10% of strains outside of this set (90-10) were selected. 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 (cf. column B). These relaxed rules allowed us 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. CK numbers, i.e. CLOG numbers in the Cyanorak v2.1 database, which are cited in the text are highlighted.

Supplementary Table S5: Coordinates and composition of every genomic island detected in our dataset. The description of each CLOG (CK_XXXXXXXX) is available in Supplementary Table 6 for *Prochlorococcus* and Supplementary Table 7 for *Synechococcus*.

Supplementary Table S6: Composition of modules of *Prochlorococcus* genomic islands. The strain genomic islands comprised in a module are indicated. Numbers in the table indicate the number of copies of a CLOG that were detected as part of an island of this module. Modules cited in the text and in Fig. 6 are highlighted.

Supplementary Table S7: Composition of modules of *Synechococcus* genomic islands. The strain genomic islands comprised in a module are indicated. Numbers in the table indicate the number of copies of a CLOG that were detected as part of an island of this module. Modules cited in the text and in Fig 7 are highlighted.

Supplementary Table S8: Proportion of clade-specific amino acid variants observed in every *Synechococcus/Cyanobium* core protein. For each clade and CLOG, the percentage of specific variants is shown, followed by the CLOG rank based on this percentage. The percentage of specific variants is also indicated for sets of strains corresponding to ESTUs co-occurring in the field (as defined by [21]). CLOGs are ordered according to the ratio of the rank observed in the “clades I/IV set" to the median rank of other clades. Note that clades indicated by a star

(light grey) contain a single sequenced strain or two quasi-identical strains (e.g. in the case of SC 5.3) and were not considered in the median calculation. CLOGs selected for the “clade I/IV set" are highlighted in yellow (see main text). Gene names cited in the text are highlighted in green.

CHAPITRE II

Variabilité physiologique en réponse à divers stress environnementaux chez la souche modèle *Synechococcus* sp. WH7803

Contexte de l'étude

Les travaux présentés dans le premier chapitre soulignent l'importante diversité génétique des genres *Prochlorococcus* et *Synechococcus* en termes de variations nucléotidiques des séquences et du contenu en gènes et le rôle respectif de chacun de ces deux mécanismes évolutifs dans l'adaptation à la niche écologique. Cette diversité génétique ainsi que les analyses de la distribution *in situ* des picocyanobactéries soulève également des questions concernant la plasticité physiologique des différents écotypes de ces organismes et leur capacité à surmonter plus ou moins efficacement les variations des paramètres environnementaux dans leurs niches écologiques respectives.

Afin de mieux comprendre les mécanismes de réponse de ces organismes aux variations des divers facteurs environnementaux, une seconde partie de mon travail a consisté à étudier l'effet de divers stress modérés (fortes lumières, radiations UV, faibles et fortes températures) couplés à l'acclimatation à deux intensités lumineuses différentes (basses et hautes lumières), sur la physiologie de *Synechococcus* sp. WH7803, la souche la mieux caractérisée d'un point de vue physiologique et qui présent l'avantage supplémentaire d'être axénique. Elle constitue donc une souche modèle pour les *Synechococcus* marins. Des cultures ont également été soumises un cycle jour-nuit, simulant les variations journalières naturelles d'intensité lumineuse, et ce à deux températures différentes, afin d'observer l'impact physiologique de la température sur les variations circadiennes du métabolisme cellulaire. Outre des mesures de (photo)physiologie (croissance, rendement quantique et réparation du PS II, etc.), des transcriptomes ont également été générés à différents temps de stress (ou du cycle nycthéméral) afin d'étudier, au travers des variations d'expression génique, les réponses des cellules de WH7803 au stress, tant d'un point vue métabolique que régulationnel. Ces analyses ont révélé que les cellules précédemment acclimatées aux fortes lumières sont mieux préparées que les cellules acclimatées aux faibles lumières pour supporter un stress lumineux supplémentaire alors qu'un stress froid semble induire un effet synergique avec les stress lumineux. Globalement, cette étude nous a permis d'identifier les réponses communes à tous les stress et celles plus spécifiques à un stress donné, mettant ainsi en évidence des gènes potentiellement impliqués dans l'acclimatation à la niche d'un membre clé du phytoplancton.

<u>Contribution</u>

Etant donné la multiplicité des conditions et des mesures réalisées, ces expériences ont impliqué la plupart des membres de l'équipe pour la mise en place des expériences, la récolte des échantillons, les extractions d'ARN et/ou les optimisations nécessaires à la construction des librairies et à leur séquençage avant mon arrivée en thèse. J'ai pour ma part pris en charge le développement d'un

pipeline d'analyse des 154 transcriptomes générés dans le cadre de cette étude, analysé les données d'expression différentielle obtenues ainsi qu'une partie des données physiologiques et participé à l'interprétation des résultats obtenus.

Synergic effects of temperature and irradiance on the physiology of the marine Synechococcus strain WH7803

Résumé de l'article

Comprendre comment les micro-organismes ajustent leur métabolisme pour maintenir leur capacité à faire face aux variations environnementales à court terme constitue l'un des principaux défis actuels de l'écologie microbienne. Ici, la souche marine de *Synechococcus* la mieux caractérisée physiologiquement, WH7803, a été exposée à des cycles jour/nuit modulés ou acclimatée à une lumière continue haute (HL) ou basse (LL), puis soumises à diverses conditions de stress, y compris un stress de basse (LT) ou haute (HT) température, des radiations HL et des ultraviolets (UV). Les réponses physiologiques ont été analysées en mesurant l'évolution dans le temps du rendement quantique (quantum yield) du photosystème (PS) II, le taux de réparation du PSII, les changements globaux dans l'expression des gènes ainsi qu'en utilisant les rapports de pigmentation et la composition des lipides membranaires publiés précédemment pour les analyses de corrélation. Ces données ont révélé que les cellules précédemment acclimatées au HL sont mieux préparées que les cellules acclimatées au LL pour supporter un stress supplémentaire de lumière ou d'UV, mais pas un stress de LT. En effet, le LT semble induire un effet synergique avec le traitement HL, comme on l'a observé précédemment avec le stress oxydatif. Alors que toutes les conditions de stress testées ont induit la sous-expression de nombreux gènes photosynthétiques, notamment ceux codant pour le PSI, le cytochrome b6/f et les phycobilisomes, le stress UV s'est avéré plus délétère pour le PSII que les autres traitements, et la rétablissement complet du PSII endommagé par le stress UV semble impliquer la néo-synthèse d'un assez grand nombre de sous-unités du PSII et pas seulement le réassemblage de sous-unités préexistantes après le remplacement du D1. En revanche, les gènes impliqués dans la dégradation du glycogène et les voies de biosynthèse des caroténoïdes étaient plus particulièrement sur-exprimés en réponse au LT. En résumé, ces expériences ont permis d'identifier les réponses communes à tous les stress et celles plus spécifiques à un stress donné, mettant ainsi en évidence des gènes potentiellement impliqués dans l'acclimatation à la niche d'un membre clé des écosystèmes marins. Nos données ont également révélé d'importantes caractéristiques spécifiques des réponses au stress par rapport aux cyanobactéries d'eau douce modèles.

Synergic effects of temperature and irradiance on the physiology of the marine *Synechococcus* strain WH7803

Ulysse Guyet[1], Ngoc A. Nguyen[1a], Hugo Doré[1b], Julie Haguait[2], Justine Pittera[1], Maël Conan[3], Morgane Ratin[1], Erwan Corre[4], Gildas Le Corguillé[4], Loraine Brillet-Guéguen[4,5], Mark Hoebeke[4], Christophe Six[1], Claudia Steglich[6], Anne Siegel[3], Damien Eveillard[2], Frédéric Partensky[1] and Laurence Garczarek[1]*

[1]Sorbonne Université, CNRS, UMR 7144 Adaptation and Diversity in the Marine Environment, Station Biologique de Roscoff (SBR), 29688 Roscoff, France; [2]LS2N, UMR CNRS 6004, IMT Atlantique, ECN, Université de Nantes, 44035 Nantes, France; [3]DYLISS (INRIA–IRISA)–INRIA, CNRS UMR 6074, Université de Rennes 1, 35042 Rennes, France; [4] Sorbonne Université, CNRS, FR2424, ABiMS, Station Biologique, F-29680 Roscoff, France. [5]Sorbonne Université, CNRS, UMR 8227 Integrative Biology of Marine Models (LBI2M), Station Biologique de Roscoff (SBR), 29688 Roscoff, France. [6]Faculty of Biology, University of Freiburg, Freiburg, Germany.

[a]Present address: Institute of Biotechnology and Food Technology, Industrial, University of Ho Chi Minh City, Ho Chi Minh city, Vietnam

[b]Present address: Wilbanks Lab, Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, Santa Barbara, CA, United States

Keywords: marine cyanobacteria, *Synechococcus*, transcriptomics, light stress, temperature stress, UV radiations

***Correspondence**: Laurence Garczarek, laurence.garczarek@sb-roscoff.fr

SUBMITTED

Abstract

Understanding how microorganisms adjust their metabolism to maintain their ability to cope with short-term environmental variations constitutes one of the major current challenges in microbial ecology. Here, the best physiologically characterized marine *Synechococcus* strain, WH7803, was exposed to modulated light/dark cycles or acclimated to continuous high-light (HL) or low-light (LL), then shifted to various stress conditions, including low (LT) or high temperature (HT), HL and ultraviolet (UV) radiations. Physiological responses were analyzed by measuring time courses of photosystem (PS) II quantum yield, PSII repair rate, global changes in gene expression as well as using previously published pigment ratios and membrane lipid composition for correlation analyses. These data revealed that cells previously acclimated to HL are better prepared than LL-acclimated cells to sustain an additional light or UV stress, but not a LT stress. Indeed, LT seems to induce a synergic effect with the HL treatment, as previously observed with oxidative stress. While all tested shift conditions induced the downregulation of many photosynthetic genes, notably those encoding PSI, cytochrome b_6/f and phycobilisomes, UV stress proved to be more deleterious for PSII than the other treatments, and full recovery of damaged PSII from UV stress seemed to involve the neo-synthesis of a fairly large number of PSII subunits and not just the reassembly of pre-existing subunits after D1 replacement. In contrast, genes involved in glycogen degradation and carotenoid biosynthesis pathways were more particularly upregulated in response to LT. Altogether, these experiments allowed us to identify responses common to all stresses and those more specific to a given stress, thus highlighting genes potentially involved in niche acclimation of a key member of marine ecosystems. Our data also revealed important specific features of the stress responses compared to model freshwater cyanobacteria.

Introduction

All microorganisms are constrained to adjust their metabolism in order to maintain their ability to survive in constantly changing environments. These variations can occur at different timescales, from seconds to several cell generations. While long-term variations often induce adaptation, whereby the natural selection alters the genetic composition (e.g., by gene loss/gain or substitutions; Galhardo et al., 2007; Doré et al., 2020), short-term variations of the environment can be managed by phenotypic plasticity, involving physiological adjustments that allow to maintain cellular performance across varying environmental conditions (Dudley, 2004; Gabriel, 2005). Bacteria possess numerous mechanisms that enable acclimation to the variety of possible external stresses (Schimel et al., 2007). These range from up or downregulation of a single metabolic process to the activation of complex gene expression networks, e.g. through regulation by two-component sensory systems that translate extracellular signals into intracellular responses (Suzuki et al., 2001; Los et al., 2010). In this context, comparing global gene expression profiles on laboratory strains submitted to different individual stresses and for which conditions can be perfectly controlled, allows not only to uncover stress-specific global transcriptional responses but also to shed light on novel or unsuspected cellular stress responses.

Most transcriptomic studies so far have focused on model microorganisms, such as *Escherichia coli* (Kannan et al., 2008; Wang et al., 2009; Jozefczuk et al., 2010), *Bacillus subtilis* (Mostertz et al., 2004; Nicolas et al., 2012), *Caulobacter crescentus* (da Silva Neto et al., 2013) or *Chlamydomonas reinhardtii* (Nguyen et al., 2008). In the specific case of cyanobacteria, changes in whole transcriptomes have been extensively investigated in the unicellular freshwater strains *Synechocystis* sp. PCC 6803 and *Synechococcus* sp. PCC 7942 in response to various environmental conditions including light (Hihara et al., 2001; Huang et al., 2002), temperature (Suzuki et al., 2001; Mikami et al., 2002; Inaba et al., 2003), oxidative stress (Hihara et al., 2003; Singh et al., 2004), nutrient stresses (Singh et al., 2003; Suzuki et al., 2004; Krasikov et al., 2012; Blanco-Ameijeiras et al., 2017) and/or change in ambient CO_2 levels (Schwarz et al., 2011; Jablonsky et al., 2016). However, none of these organisms are ecologically relevant, especially concerning the marine environment, which is affected by universal but also specific environmental parameters as compared to terrestrial and freshwater ecosystems.

In oceanic waters, the cyanobacterial community is largely dominated by two genera, *Prochlorococcus* and *Synechococcus*, which are only distantly related to their model freshwater counterparts. While *Prochlorococcus* mainly thrives in warm, nutrient-poor oceanic waters, *Synechococcus* can colonize a broader range of ecological niches, extending from the equator to subpolar waters as well as from estuaries to oligotrophic waters of the ocean (Flombaum et al., 2013; Xia et al., 2015; Paulsen et al., 2016). This abundance, ubiquity and the availability of many genomes and strains in culture make *Synechococcus* one of the most relevant model microorganisms to study the response to variations of environmental conditions in the marine ecosystem. Quite a few transcriptomic studies have been conducted on *Prochlorococcus* and *Synechococcus* strains (Tolonen et al., 2006; Lindell et al., 2007; Zinser et al., 2009; Tetu et al., 2009, 2013; Blot et al., 2011; Thompson et al., 2011; Mella-Flores et al., 2012; Reistetter et al., 2013; Voigt et al., 2014; Stazic et al., 2016; Lambrecht et al., 2019), but they have mainly focused on the effect of single environmental factors. Here, the well-characterized marine *Synechococcus* strain WH7803, a warm temperature-adapted ecotype, which also has the advantage of being axenic (Kana and Glibert, 1987a, 1987b; Kana et al., 1988; Garczarek et al., 2008; Blot et al., 2011; Mella-Flores et al., 2012; Pittera et al., 2018), was selected to study the effect of various stress conditions, high light (HL), UV, low (LT) and high (HT) temperatures, on cultures previously acclimated to either low light (LL) or HL conditions, as well as to assess the effect of diel variations, as triggered by a modulated light/dark (L/D) cycle. These parameters are indeed well known to affect the physiology of this organism but the regulatory processes involved remain poorly studied. A comparison of global gene expression profiles allowed us to identify the common and specific transcriptomic responses to the different conditions. This study notably revealed that cells previously acclimated to HL seem to be better prepared than LL-acclimated cells to sustain an additional light stress but not a LT stress, which in contrast seems to induce a synergic effect with the HL treatment, as previously observed for oxidative stress (Blot et al., 2011).

Materials and methods

Culture Conditions

Synechococcus sp. WH7803, an axenic strain retrieved from the Roscoff culture collection (RCC752; http://roscoff-culture-collection.org/), was grown in PCR-S11 culture medium (Rippka et al., 2000) supplemented with 1 mM sodium nitrate and 1 µg.L^{-1} vitamin B12. Cultures were acclimated for at least two weeks at 22°C and 20 µmol photon.m^{-2}.s^{-1} (hereafter LL and µE.m^{-2}.s^{-1}) or 250 µE.m^{-2}.s^{-1} (hereafter HL) provided by multicolor (cool white: 6500K, blue: 470 nm, green: 530 nm) LED systems (Alpheus, France). For each growth irradiance, a 9 L exponentially growing culture was split in 500 mL aliquots into 1 L flasks (Nalgene, St. Louis, MO, USA) and left under their initial growth irradiance overnight before being submitted to light, UV or temperature stress conditions and sampled at different times (Table 1) for various parameters (see below). UVA and UVB radiations, generated from UVA-351 and UVB-313 fluorescent bulbs (Q-Panel Lab products, Cleveland, OH, USA), were measured between 280–320 and 320–400 nm, respectively using a USB2000 spectroradiometer (Ocean Optics, EW Duiven, The Netherlands). In order to estimate the cell recovery capacities, HL and UV stressed cultures were then shifted back to their initial growth condition and sampled after 1h (R1) and 24h (R24). Low (13°C, hereafter LT) and high temperature (30°C, hereafter HT) stresses were performed in temperature-controlled chambers (Liebherr-Hausgeräte, Lienz, Austria).

For diel cycle experiments, run at two temperatures (21 and 27°C), a bell-shaped 12/12 h L/D cycle, triggering a proper synchronization of cell division, was generated using the multicolor LED systems and the Ether software (Alpheus, France). The maximal irradiance (at virtual noon) was set at 669 µE.m^{-2}.s^{-1}. Two replicate cultures per temperature were acclimated to L/D cycles for at least two weeks prior to starting monitoring the different parameters. During the experiment, two replicate cultures were grown with a continuous input of fresh medium, in order to maintain cells in exponential growth throughout the whole sampling period. In order to study the kinetics of the response of cells to light fluctuations, cultures were sampled at virtual 6:00, 9:00, 12:00, 15:00, 18:00, 20:00, 22:00, and 2:00 over three days for measuring a variety of parameters described below. Additionally, for flow cytometric analyses, 200 µL samples were transferred every hour using an Omnicoll Fraction Collector (Lambda, Brno, Czech

Republic) into microtubes maintained at 4°C by Peltier effect and containing 0.25% glutaraldehyde grade II (Sigma Aldrich, St Louis, MO, USA), then stored at -80°C until analysis.

Flow cytometry

Synechococcus cell concentration was measured by flow cytometry as previously described (Marie et al., 1999) using a FACSCanto II flow cytometer (Becton Dickinson, San Jose, CA, USA) with a laser emission set at 488 nm and distilled water as sheath fluid.

***In vivo* fluorescence measurement**

PSII quantum yield (F_V/F_M) was measured upon excitation at 520 nm with a Pulse Amplitude Modulation fluorometer (Phyto-PAM I, Walz, Effeltrich, Germany) equipped with a chart recorder (Vernier, LabPro, Beaverton, OR, USA), as previously described (Mella-Flores et al., 2012). Briefly, after 3 min acclimation in the dark, cultures were exposed to modulated light and basal level fluorescence (F_0) was then measured. The maximal fluorescence levels (F_M) were measured after adding 100 µM of PSII blocker 3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU) and by triggering saturating light pulses (655 nm; 2000 $\mu E.m^{-2}.s^{-1}$). The PSII quantum yield was calculated as:

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

Fluorescence emission spectra were recorded between 545 and 750 nm upon excitation at 530 nm (λ_{max} absorption of phycoerythrobilin) with an LS-50B spectrofluorometer (Perkin-Elmer, Waltham, MA, USA) equipped with a red sensitive photomultiplier as previously described (Six et al., 2004). In order to determine the coupling of phycobiliproteins and to observe any dismantlement of phycobilisome (PBS) rods, fluorescence emission spectra were then used to calculate the phycoerythrin (PE, λ_{max} = 565 to 575 nm) to phycocyanin (PC, λ_{max} = 645 to 655 nm) ratio as well as the PC to PBS terminal acceptor (TA; 680 nm). The latter ratio reflects the energy transfer from phycocyanin to reaction center chlorophylls (Pittera et al., 2017).

PSII repair rate measurements

PSII repair rate was calculated in two different ways depending on stress duration. For short-term stresses (HL and UV stress, 6-hour stress experiments), PSII repair rate was

estimated over the 6 hours of stress. Two aliquots of 20 mL were sampled at T0. While the first one was used as a control, lincomycin (protein synthesis inhibitor) was added to the second one to a final concentration of 0.5 $mg.mL^{-1}$ and the F_V/F_M parameter was monitored at each timepoint. The difference between the coefficients of the exponential curves fitted on F_V/F_M measurements was used as proxy of the PSII repair rate. For the long-term (4-days) experiments (LT and HT stress) and for L/D cycles, six aliquots of 2 mL were taken at each timepoint, placed in 8-wells plates and lincomycin was then added to three of them, as described above. All aliquots were placed in the same conditions as the rest of the culture and F_V/F_M was monitored after 15 min, 30 min and one hour, respectively.

Pigment and lipid analyses

Pigment content and lipid analyses, corresponding to the same experiments, were published elsewhere (Pittera et al., 2018) and used here for correlation analyses with gene expression data (see below).

RNA Extraction

RNA extractions were performed from 150 mL of culture harvested by centrifugation in an Eppendorf 5804R (7 min, 10,414 × g, 4°C) then in an Eppendorf 5417R centrifuge (2 min, 20,817 × g, 4°C). The cell pellet was then resuspended in 1 mL Qiazol (Qiagen, Valencia, CA, USA) and quickly frozen in liquid nitrogen. The total duration of cell harvesting was kept below 15 min. After thawing the tubes at 65°C, followed by incubation and vortexing for 5 min at this temperature, total RNA was extracted using the Direct-Zol kit (Zymo Research Corp, Proteigene) as recommended by the manufacturer. Three successive DNase treatments were performed on the Zymo-Spin™-IIC Column using the Qiagen RNase-free DNase Set (Qiagen), followed by elution from the column in 30 µL DEPC-treated water. RNA quantification and quality check were performed using a NanoDrop® ND-1000 (Thermo Fisher Scientific) and a BioAnalyzer 2100 with the RNA 6000 Nano Kit (Agilent, Santa Clara, CA, USA), respectively. RNA samples were then frozen in liquid nitrogen and stored at -80°C.

Library preparation and sequencing

Ribodepletion was performed at the Genotoul platform (Toulouse, France) from 5 μg of total RNA using the Ribo-Zero Bacteria Magnetic Kits (Illumina) in the presence of RiboGuard RNase Inhibitor (Epicentre) and treated RNA samples were purified and concentrated using RNA Clean & Concentrator™-5 columns (Zymo Research). Ribo-depleted RNA was then was then quantified by Qubit (RNA HS Assay Kit, Thermo Fisher Scientific) and quality check was performed using a Bioanalyzer (RNA Pico kit) or a Labchip GX. Libraries were then prepared from 100 to 400 ng of RNA using the Illumina TruSeq Stranded mRNA kit and sequenced as 150 bp paired-end reads on Illumina HiSeq 3000 System (Illumina, San Diego, CA, USA) at Genotoul (Toulouse, France). $18.5 \pm 4.7 \cdot 10^6$ paired-end reads were generated per sample.

RNA-Seq Analysis

Reads quality check was performed with the FastQC tool (Andrews, 2015). Reads smaller than 50 nt and with a mean quality score lower than 25 were excluded using Prinseq v0.20.4 (Schmieder and Edwards, 2011). Illumina adapters, constituting 23.5 ± 4.5 % of reads, and rRNAs, constituting 5.0 ± 5.9 % of all reads, were removed with Cutadapt v1.8.3 (Martin, 2011) and SortMeRNA v2.1 (Kopylova et al., 2012), respectively. The remaining reads were then mapped to the reference genome of *Synechococcus* sp. WH7803 using Bowtie2 v2.2.9 software (Langmead and Salzberg, 2012) with '--non-deterministic --end-to-end –sensitive' parameters and by selecting only the properly paired reads (i.e., SAM flagged as 99, 147, 163 or 83). Read count tables were obtained using HTSeq Count v0.6.0 (Anders et al., 2015) with following parameters: '--stranded=reverse -a 10 -m intersection-nonempty'. Gene expression levels between stress and control conditions as well as adjusted p-values based on the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995) were determined using SARTools v1.4.0 (Varet et al., 2016) with embedded DESeq2 v1.14.1 (Love et al., 2014). Genes were considered as differentially expressed (DE) if the adjusted p-value was ≤ 0.05 and the absolute value of $\log_2$ fold change ($\log_2$FC) was ≥ 1. The RNA-seq dataset is available as raw and processed data in the SRA and GEO databases, respectively (see supplemental

Tables) and can be visualized using JBrowse (Buels et al., 2016) in the Cyanorak v2 database (www.sb-roscoff.fr/cyanorak).

Co-expression network model analysis

Weighted gene correlation network analysis (WGCNA; Langfelder and Horvath, 2008) was performed on shift experiments to define gene subnetworks, called modules, based on gene expression patterns. A signed adjacency matrix between genes was calculated and Pearson correlations were weighted by taking their absolute value and raising them to the power $\beta = 12$ to optimize the scale-free topology network fit. In order to identify groups of genes, whose expression was correlated to the biological, physical or chemical traits available for this experiment (including previously published pigment and lipid data; Pittera et al., 2018), the pairwise Pearson correlation coefficients between the principal component of each module, referred to as the module *eigenvalue* (ME), and these traits was then calculated using the R package WGCNA (Langfelder and Horvath, 2008). Modules were then filtered using a partial least square (PLS) regression, a dimensionality-reduction method that aims at determining predictor combinations with maximum covariance with the response variable. The predictors were ranked according to their value importance in projection (VIP), as previously described (Guidi et al., 2016). Furthermore, the UpSet R package was used to count and plot the number of genes per module that were DE in a given set of conditions (e.g. LLHT and HLLT) by considering only genes significantly DE in at least one time point of each condition for this set of conditions.

Relationship between genomic features and transcriptomic responses to different stresses

Transcriptomic data were synthesized as a circos plot integrating: i) core genes, as determined based on the comparison of the 81 non-redundant genomes of the cyanorak database (www.sb-roscoff.fr/cyanorak/; Doré et al., 2020), ii) WGCNA module and submodule membership, iii) predicted operons as extracted from ProOpDB (Taboada et al., 2012), iv) 'cyanorons' that we defined as suites of ≥ 4 adjacent genes on the genome that belong to the same WGCNA module and displayed at least half of all time points for a given stress with a similar expression pattern (e.g., all 4 adjacent genes belonging to module Blue and all displaying either a $Log_2FC > 1$ or < 1).

Results

Photophysiological response of *Synechococcus* sp. WH7803 to variations in light and temperature conditions

Cultures pre-acclimated under LL (25 µE.m^{-2}.s^{-1}) or HL (250 µE.m^{-2}.s^{-1}) and 22°C were subjected to a series of light or temperature shifts, as detailed in Table 1. The quantum yield of the PSII reaction center (F_V/F_M), used as a proxy for PSII activity, showed that UV stress induced a much stronger response than HL stress in LL-acclimated cultures of *Synechococcus* sp. WH7803, with a decrease of the F_V/F_M after six hours of stress of 60% and 20%, respectively (Fig. 1A). This F_V/F_M drop was much less pronounced in HL-acclimated culture under UV stress (20%), potentially related to the much faster D1 repair rate measured in control HL- than LL-acclimated cultures (6-fold, Fig. 1B). D1 repair rates were also further enhanced by UV and HL treatments in LL-acclimated conditions and by UV in HL-acclimated cells. In all cases, cultures were capable of recovering most of their PSII activity after 24 h, showing that none of these stresses were lethal for the cultures (Fig. 1A). As concerns thermal stresses, HT shift, induced by an 8°C temperature increase, only had a small effect on PSII activity for cells previously acclimated to both light conditions (Fig. 1C), except at the end of the LL HT experiment when cultures reached the stationary growth phase (data not shown), and these HT conditions led to a fairly high induction of the D1 repair rate compared to the control at T0 (Fig. 1D). In contrast, a shift to LT, corresponding to a 9°C temperature decrease, induced a strong PSII photoinactivation with a drop in F_V/F_M of 45% followed by a stabilization for LL-acclimated cells after one day and a continuous decrease reaching about 70% within a day for HL-acclimated cells (Fig. 1C), and none of these conditions led to a significant change of the D1 repair rate of the cells (Fig. 1D). Thus, in contrast to light stresses, cells previously acclimated to HL do not seem to be better prepared for thermal stress than LL-acclimated cells.

Photoprotective and/or antioxidant pigments only showed low variations in response to light and temperature shifts. The ratios of the β-carotene (β-car), zeaxanthin (zea) and β-cryptoxanthin (β-crypto) to chlorophyll (chl) *a* only showed low variations in response to light and temperature shifts, with i) a slight increase of the two latter ratios and decrease of the first one in response to HL in LL-acclimated cultures, ii) an increase of the sole β-crypto/chl *a* ratio in response to UV and iii) a decrease of β-car/chl *a* in

response to LT in both light acclimated cultures (Fig. S1). Furthermore, variations of the ratios of the fluorescence emission peaks were used to assess the efficiency of the energy transfer within the PBS rods (PE/PC) as well as between the basal part of the rods and the PSII reaction center (PC/TA), as measured by fluorescence leaks between these phycobiliprotein subunits (Pittera et al., 2017; Fig. S2). While HL and UV had a rapid and stronger effect on the base of the PBS rods and/or on the connection of the PBS to the thylakoid membrane (PC/TA) than at the rod level (PE/PC, Fig. S2A-B), LT seemingly induced a progressive disruption of the PBS, both at the rod level and at the core level (Fig. S2C-D). In contrast, the shift to HT had virtually no effect on the PBS structure.

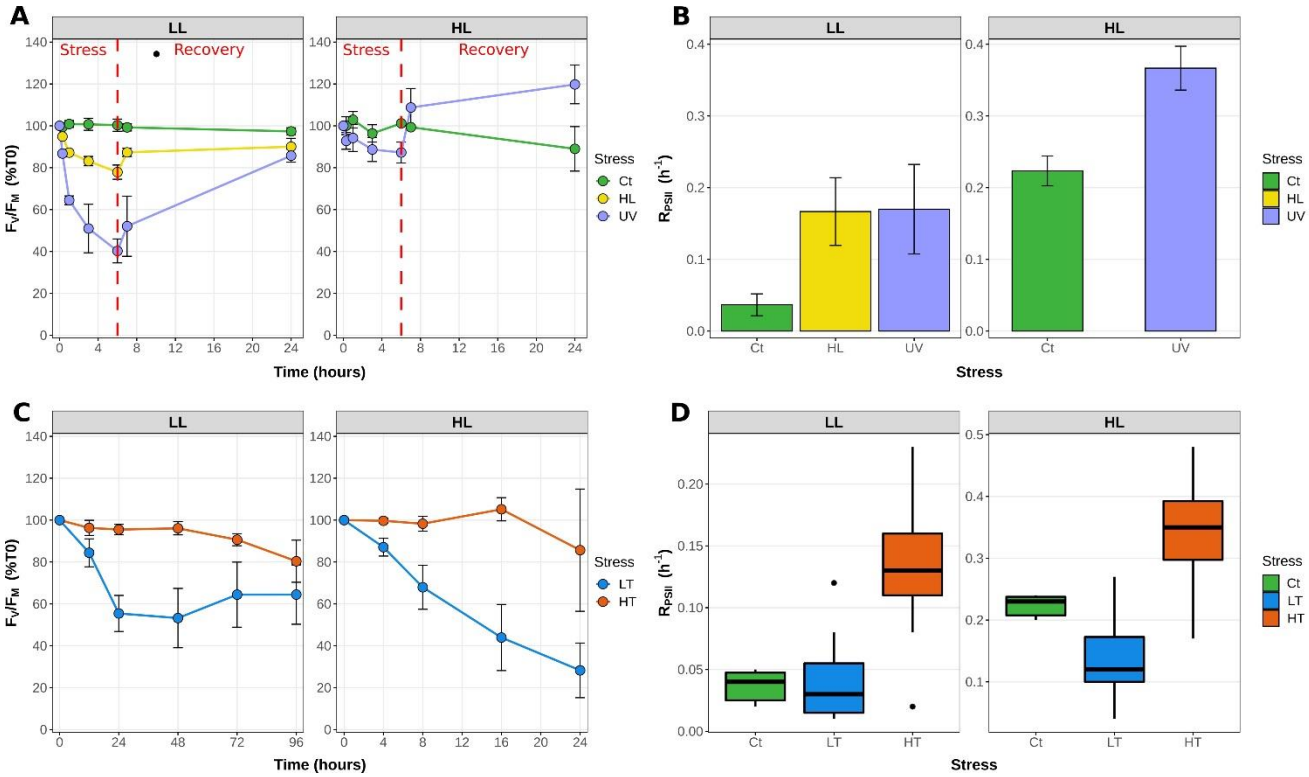


Figure 1. Variation of the photophysiological parameters in response to changes in light or thermal conditions in *Synechococcus* sp. WH7803 acclimated to low or high light. (A,C) PSII maximal photochemical yield (F_V/F_M) expressed as % of initial F_V/F_M after a shift (A) to high light or UV radiations and (C) to low temperature (LT: shift from 22°C to 13°C) or high temperature (HT; shift from 22°C to 30°C). The dashed line indicates the time (6h) at which cultures submitted to light stresses were shifted back to their initial light conditions for recovery. (B, D) Cumulative photosystem II repair rate (RPSII), (B) during the 6 h of light stress and (D) during 1h for low and high temperature shifts. While for panels A, B and C, data represent averages and standard deviations of three biological replicates, for panel D, boxplots represent distribution of all three independent measurements, with their medians and interquartiles. The top grey banners indicate the culture acclimation condition. These data are based on at least three independent experiments. Abbreviations: Ct, control; LL, low light; HL, high light; UV, ultraviolet; LT, low temperature; HT, high temperature.

Comparison of the photophysiological response of *Synechococcus* sp. WH7803 to a L/D cycle at two temperatures

At both temperatures, the F_V/F_M ratio showed a cyclic evolution, reaching a minimum value during the light period. However, while at 21°C the F_V/F_M mirrored the light curve with a minimum at noon and fairly constant higher values during the whole dark period, at 27°C F_V/F_M was already close to the minimal value at 9 a.m. and reached a maximum at the light/dark transition and tended to decrease during the night (Fig. 2). Consistently with the LT and HT stress experiments in continuous light conditions (Fig. 1D), the D1 repair rate was much higher at 27 than at 21°C at all time-points during the light period, reaching a comparable value at 9 a.m. for the 27°C treatment to the one measured at noon for the 21°C treatment.

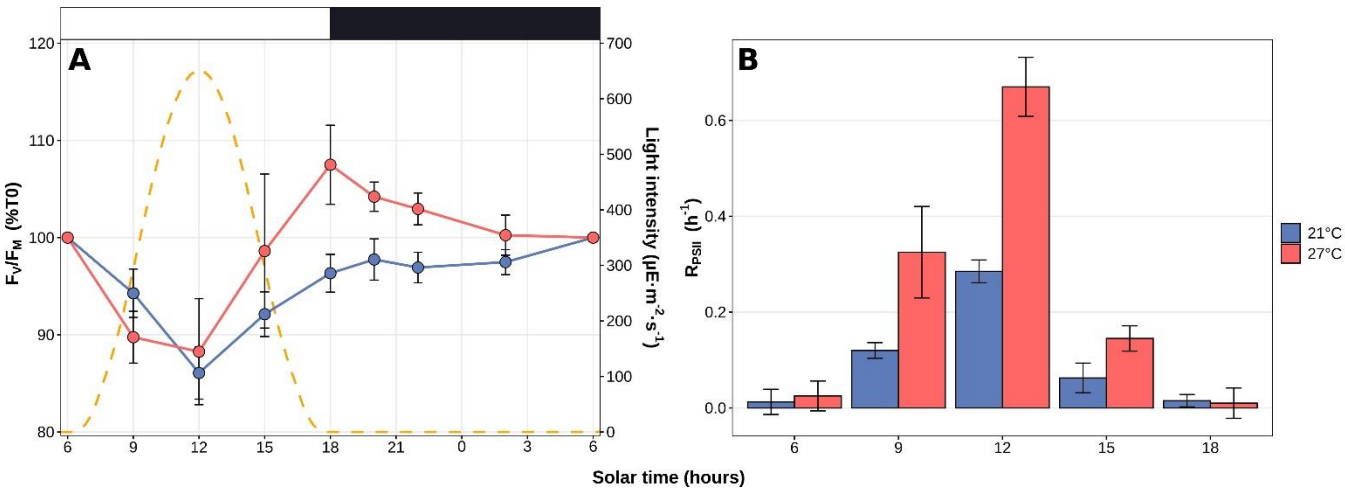


Figure 2. Daily variations of the photosystem II maximal quantum yield (F_V/F_M) and D1 repair rates for *Synechococcus* sp. WH7803 cells acclimated to a modulated 12h/12h L/D cycle at 21 or 27°C. (**A**) PSII maximal photochemical yield (F_V/F_M) expressed as % of F_V/F_M at 6 a.m. (**B**) Photosystem II repair rate (RPSII). Note that this parameter was not measured during the night since a previous study showed that there is no repair during this period (Mella-Flores et al., 2011). Light and dark periods (µE.m^{-2}.s^{-1}) are indicated by white and grey background on the figure and by a dashed line during the light period showing the modulation of light intensity. Error bars indicate standard deviation for four biological replicates.

Global description of transcriptional responses

Transcriptome analyses showed that 97.6 ± 1.2 % of the reads mapped to the reference WH7803 genome, resulting in average in 98.9 % of the genome, including intergenic and gene coding regions, being covered by more than two reads in the different experiments, and 99.99 % when considering all experiments altogether. Global analyses showed a transcriptional response proportional to the stress intensity as assessed from photophysiological measurements, i.e. there were more DE genes at the end of the stress than at the beginning (Fig. 3). While light stresses (HL and UV) led to a high number of DE genes in LL-acclimated cultures (24 and 26 % of all genes at the end of stress, respectively), only a few genes were DE after UV stress in HL-acclimated cultures. Differences between LL- and HL-acclimated cultures were less conspicuous during thermal stress, both responding strongly to LT stress (25 and 29% of all genes were DE at the end of stress, respectively) and only weakly to HT stress (up to 12 and 4% DE genes, respectively). Thus, consistently with PSII activity (Fig. 1A, C), the transcriptomic response showed that LL-acclimated cells were much more sensitive to UV stress than HL-acclimated cells, while LT but not HT shift seemed to constitute a very stressful condition whatever the initial light acclimation condition used.

As concerns the L/D cycle, there were large daily variations of the number of DE genes compared to the 6 a.m. time point, the highest number being reached at the light-dark transition at 21°C, while variations were less pronounced during the day at 27°C with a peak of DE genes at 3 p.m. (Fig. 3B). The direct comparison between 21 and 27°C at each time point showed that differences in number of DE genes between these treatments mainly occurred during the dark period and at the L/D and D/L transitions (Fig. S3).

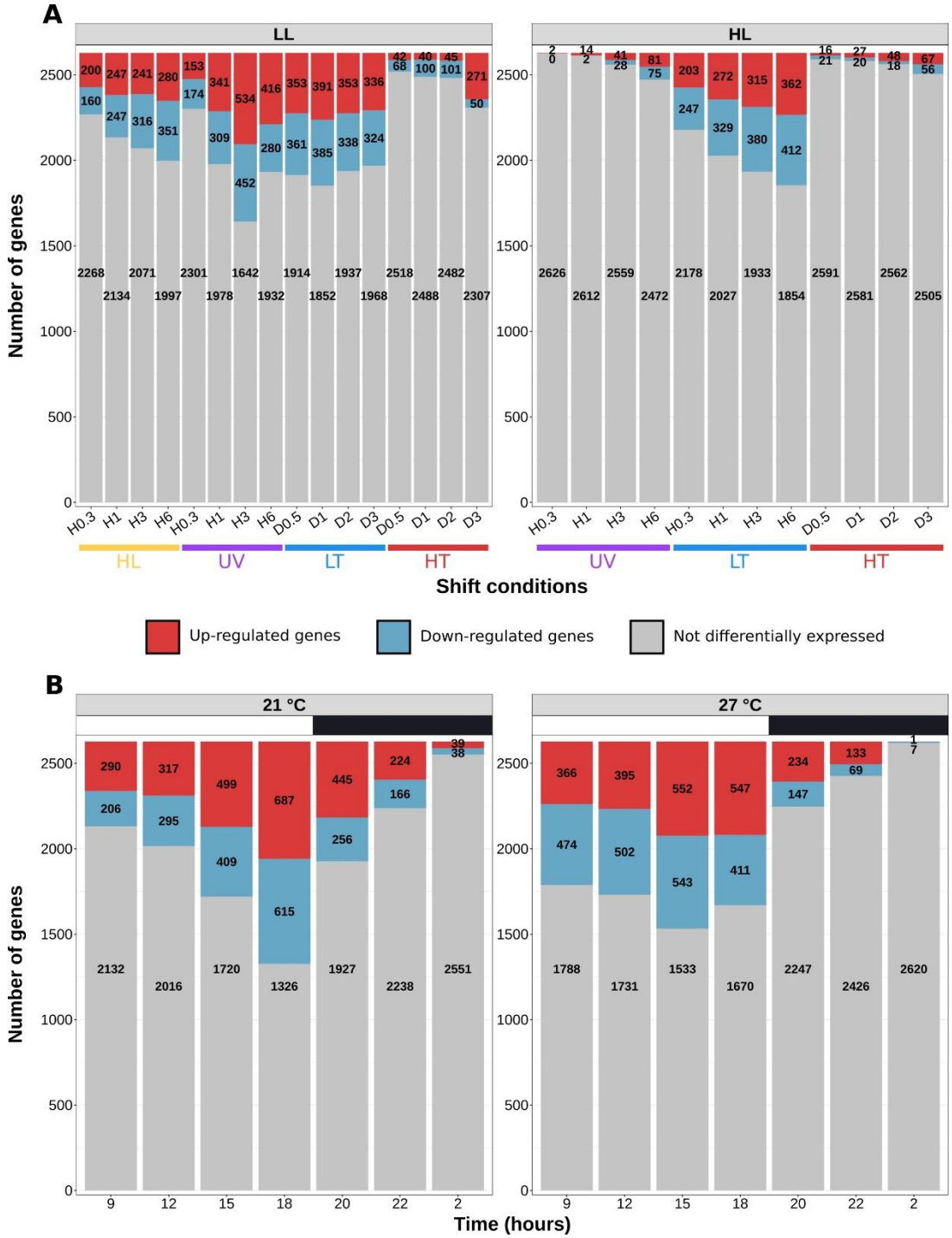


Figure 3. Number of differentially expressed (DE) genes in Synechococcus sp. WH7803 in the different tested conditions. For each condition, the number of induced, repressed and non-DE genes among the 2,634 genes of the Synechococcus sp. WH7803, are shown in red, blue and grey, respectively. (**A**) Shift experiments. The top grey banners indicate the prior acclimation condition of the culture: LL, low light; HL, high light. Data are based on three biological replicates. For each stress condition, numbers in x-axis correspond to 'sampling time' in hours and the colored bars to the 'shift condition' (UV, ultraviolet; LT, low temperature; HT, high temperature). (**B**) Light/dark cycle at 21 and 27°C. The top grey banners above graphs indicate the temperature used for the experiment, while the white and black banners indicate light and dark periods. Data are based on four biological replicates and differential expression was calculated relatively to the 6 a.m. data point.

Correlation between gene expression and biophysical parameters

In order to explore transcriptome-wide responses of WH7803 in all shift conditions, we used a WGCNA approach (Langfelder and Horvath, 2008; Guidi et al., 2016). This analysis delimited three modules and 13 submodules of co-expressed genes within the global co-expression network. Each (sub)module thus contains a subset of DE genes, whose pairwise expression was highly correlated, i.e. expression of genes gathered within a given (sub)module had a high probability of varying similarly over all stress experiments (Fig. 4). To further reduce the complexity of these transcriptomic data, we first focused on i) genes with VIP scores values higher than 1 (hereafter VIP genes), i.e. the minimum set of genes of a module to obtain an association (correlation or anti-correlation) with a given trait similar to that of all genes of the module (Table S1), on ii) genes in each module that were significantly up or down-regulated in specific sets of conditions (hereafter called UpSet genes; Fig. S4). This approach also allowed us to determine the metabolic pathways of WH7803, based on KEGG (https://www.genome.jp/kegg/), which were most affected in response to environmental changes, as detailed in the discussion part.

The yellow module was correlated to the β-car/chl *a* ratios, F_V/F_M, temperature and anti-correlated to the PC/TA ratios, to the glycolipid sulfoquinovosyldiacylglycerol (SQDG) *sn*-1 and *sn*-2 acyl chains lengths, as well as to the number of unsaturations on both *sn*-2 and/or *sn*-1 acyl chains of SQDG and most other glycolipids (Table S1; Fig. 4). Among the 1,520 genes of this module, 102 were VIP and were dominantly photosynthetic genes notably involved in the synthesis of pigments, PSI and PSII, cytochrome b_6/f, plastocyanin as well as PBS subunits and linkers, most of them being downregulated and belonging to the turquoise submodule. Of note, most genes coding for PBS components were gathered into two 'cyanorons', i.e. were co-expressed and adjacent on the genome (Fig. 5). The fact that many genes involved in the light-dependent reactions of photosynthesis were downregulated in HL, UV and LT treatments suggests that these stresses all had a strong inhibitory effect on this central metabolic process. Notable exceptions, all clustered in the brown submodule, were in contrast upregulated in response to most treatments (Table S1). These include a D1.2 encoding *psbA* gene, a *psbD* gene copy, the D1 repair FtsH2 protein, all involved in the turnover of the two main proteins of the PSII reaction center, but also three genes coding for high-light-inducible proteins (HLIPs) as well as *ubiH* involved in plastoquinone biosynthesis. VIP genes of this submodule also included several

chaperones and proteases (GroL1, GroL2, ClpB1, ClpC, DnaK3 and a small heat shock protein, HSP20) as well as three genes coding for sigma factors (*rpoD4, 6 and 7*), all three latter genes also showing a notable diel expression (Table S1).

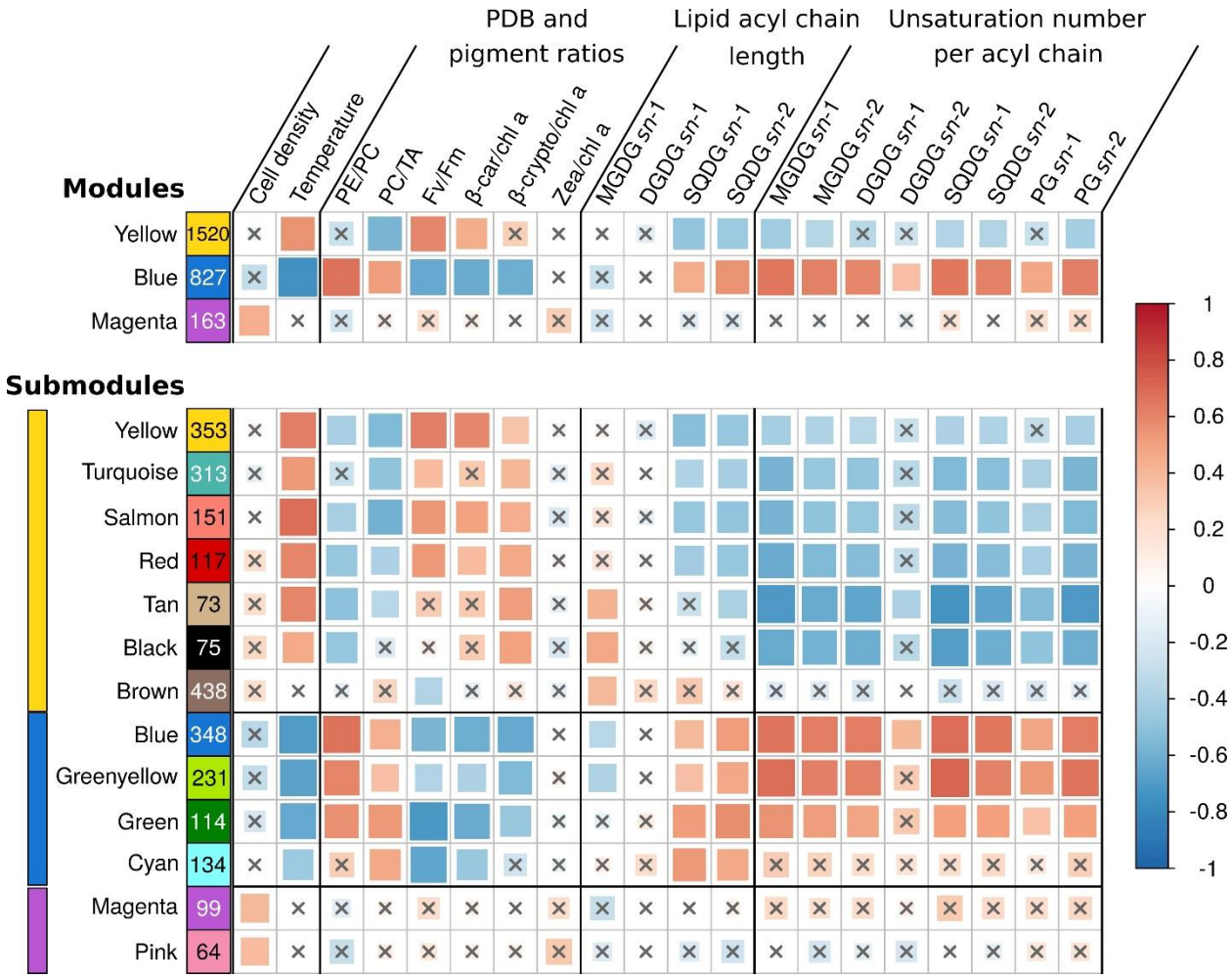


Figure 4. Correlations between WGCNA modules eigengenes and phenotypic traits. Each row represents either a gene module (top four rows) or submodules (bottom 14 rows) and number of genes in the module or submodule are specified in colored rectangles. Colored bars on left hand side indicate the corresponding module. Biophysical parameters are expressed as a percentage of the T0 of each experiment. The color scale reflects the Pearson correlation coefficient (R2). The size of the correlation squares is inversely related to the p-value. Genes that could not be clustered into one of the modules were assigned to the grey module (not shown), and non-significant correlations (Student Adjusted p-value > 10-3) are indicated by a cross. Abbreviations: PC, phycocyanin; PE, phycoerythrin; TA, terminal acceptor; FV/FM, PSII maximal photochemical yield; β-car, β-carotene; chl, chlorophyll; β-crypto, β-cryptoxanthin; zea, zeaxanthin; MGDG, monogalactosyldiacylglycerol; DGDG, digalactosyldiacylglycerol; SQDG, sulfoquinovosyldiacylglycerol; PG, phosphatidylglycerol.

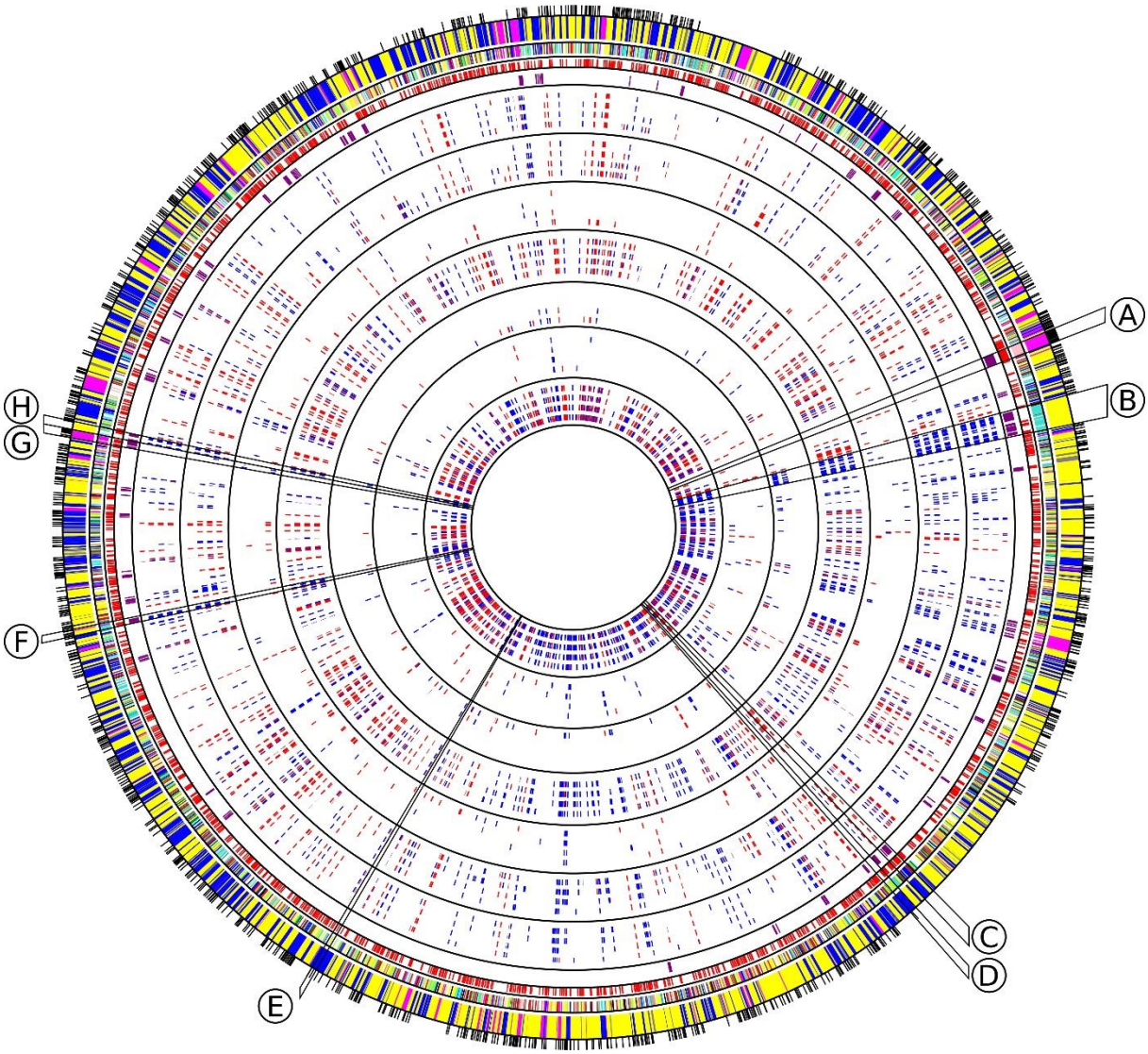


Fig. 5: Circos representation of Synechococcus sp. WH7803 genomic features and transcriptomes in response to different stresses. Circles from outside to inside, separated by black lines, include core genes (black), WGCNA module and submodule membership (cf. Fig. 4 for the corresponding colors), predicted operon using ProOpDB (orange), Cyanorons (purple) and heatmap for differential gene expression [Log2(FC)] in all tested stress conditions displayed in the following order: LLHL (0.3, 1, 3, 6), LLUV (0.3, 1, 3, 6), LLHT (12, 24, 48, 72), LLLT (12, 24, 48, 72), HLUV (0.3, 1, 3, 6), HLHT (4, 8, 16, 24), HLLT (4, 8, 16, 24), with number between brackets indicating 'sampling time' in hours. Slices delineated by black lines indicate group of genes mentioned in the text involved in specific functions: A: ribosomal proteins, B: phycobilisome rod biosynthesis, C: CO2 fixation (carboxysomes and RuBisCO), D: multiprotein Na+ antiporter subunits, E: ABC-type Mn2+/Zn2+ transport system, E: exodeoxyribonuclease V (RecBCD complex), F: cytochrome c oxidase subunits, G: ATP synthase, H: Allophycocyanin.

When looking at the 1,039 UpSet genes of this module, one can see that 149 were induced (brown submodule, 110 genes) or repressed (turquoise submodule, 39 genes) specifically in LLUV and 60 more were mostly induced in LLUV and LLHT, the latter gene set being upregulated mainly at the last data point of the high temperature stress, corresponding to cells in stationary phase (Fig. S4). These categories notably included genes coding for five ATP-dependent Clp proteases (ClpP1-4, ClpS), the stress-inducible DNA-binding protein DpsA, as well as seven out the 20 subunits of the NADH dehydrogenase I (Table S1). Furthermore 42 UpSet genes were DE in all stress conditions except HT treatments, while 84 others were DE in all but HT and HLUV, the latter conditions previously showed to be the least stressful conditions in this experimental setup (Fig. 3A). Both sets of genes were mainly downregulated and essentially encompassed genes involved in the abovementioned photosynthetic light reactions, but also a few genes involved in heme/vitamin biosynthesis (5 genes), cell division (4 genes) as well as the universal stress family protein UspG.

The blue module, gathering 827 genes, exhibited a mirrored correlation pattern compared to the yellow one, except that a significant correlation was found for several parameters, which were anticorrelated with the yellow module but not significantly (Fig. 4). Although the WGCNA analysis resulted in four submodules within the blue module, their analysis did not reveal any clear trend with regard to the different treatments. Among the 135 VIP genes of this module, most of them were upregulated and numerous genes were involved in central metabolism, including DNA replication and repair, protein fate, oxidative stress response, fatty acid biosynthesis as well as sugar catabolism (Fig. 6). Also notable in this module were several two-component systems, three genes of the SUF (sulfur utilization factor) system (Banerjee et al., 2017), as well as several members of the DNA/RNA helicase family, which by rearranging the secondary structure of nucleic acids, could play a role in the response to variations of environmental conditions (see e.g., Chamot et al., 1999; Chamot and Owttrim, 2000; Owttrim, 2006). This includes CrhR, an ortholog of Slr0083, known to be upregulated by cold stress in *Synechocystis* sp. PCC 6803 (Georg et al., 2019), which was indeed also the case for HL-acclimated WH7803 cells.

Consistently with the fact that the blue module was strongly anticorrelated to temperature, UpSet genes of this module gathered 55 genes seemingly specific to LT stresses (LLLT and HLLT), while 41 more genes were found only in LLLT and 49 only in HLLT

conditions (Fig. S4 and Table S1). Genes among these categories were mostly upregulated and notably included *ctpA,* encoding the carboxyl-terminal processing protease involved in the maturation of the PSII core subunit D1, the lycopene β-cyclase encoding gene *crtL-b,* several genes involved in DNA replication recombination and repair, numerous glycosyltransferases, potentially involved in cell wall biosynthesis as well as genes involved in biosynthesis and regulation of vitamins, including biotin (B7; Rodionov et al., 2002) and pseudocobalamin (B12; Helliwell et al., 2016). Also notable were several transporters constituting 'cyanorons' on the circos representation (Fig. 5): an ABC-type Mn^{2+}/Zn^{2+} transporter as well as several subunits of the multiprotein Na^{+}/H^{+} antiporter (*mrpCDF)*, which like genes coding for the four other subunits of the latter transporter, also in the blue module, suddenly shift from being mostly downregulated during the day to being upregulated during the night. Interestingly, the blue module also encompassed 49 genes differentially regulated specifically in LL-acclimated cultures submitted to UV stress. While half of them were downregulated, including *ftsZ* involved in ring formation during cell division, the other half were upregulated and included a few photosynthetic genes and several ABC transporters (Table S1).

The magenta module, which encompasses 163 genes, including 12 VIP, was significantly correlated with only one parameter, namely the variation of cell density compared to T0. VIP genes notably include two genes involved in the response to oxidative stress, two coding for subunits of the DNA-directed RNA polymerase (the other two subunits being also clustered in the magenta module) as well as *rbcL*, encoding the large subunit of the RuBisCO, the latter being upregulated in response to the LLHL treatment but downregulated in LLUV, LLLT and HLLT. Additionally, a large number of UpSet genes (24), mainly found in the pink submodule, seemed to be upregulated significantly in LLHL but non-significantly in other conditions (Table S1). This included genes coding for 19 out of the 53 ribosomal proteins, while 19 others were also in the pink submodule, all these genes constituting the second largest 'cyanoron' of the WH7803 genome (Fig. 5). In contrast, the magenta submodule essentially gathered genes upregulated in LLHL but downregulated or not DE in most other shift conditions. It contains i) not only the abovementioned *rbcL* gene but also genes encoding the second subunit of the RuBisCO (*rbcS*) and most carboxysome shell proteins, altogether constituting another large 'cyanoron', ii) all genes coding for the different subunits of the

ATP synthase as well as iii) numerous genes involved in nitrogen uptake (*focA*, *amt1*, *nrtP*), assimilation (*cynH, S, glnA, nirA, narB, mobA*) and regulation (*ntcA*). Thus altogether, it seems that all critical pathways involved in CO_2 fixation, ATP production and nitrogen metabolism were specifically activated in response to HL treatment (and showed a peak of expression at 9 a.m. in L/D), while being downregulated in response to other treatments.

Discussion

General response of marine *Synechococcus* to various stress conditions

Understanding how marine phytoplankton will react to variations in the physico-chemical properties of the marine ecosystem, some of which are triggered by the ongoing global changes, constitutes one of the major current challenges in microbial ecology. For this reason, we investigated the response to various ecologically realistic stresses as well as L/D cycles in WH7803, one of the best physiologically characterized marine *Synechococcus* isolates (Kana and Glibert, 1987a, 1987b; Kana et al., 1988; Garczarek et al., 2008; Blot et al., 2011; Mella-Flores et al., 2012; Pittera et al., 2018). Physiological responses to these conditions showed that a shift from 22°C to 30°C (LLHT and HLHT) actually did not constitute a stress for this strain, as shown by a moderate change in F_V/F_M (Fig. 1C) and the low number of DE genes (Fig. 3A). In contrast, a shift from 22°C to 13°C induced a strong stress response, whatever the initial light acclimation conditions, although the cells were seemingly able to maintain ~60% of their F_V/F_M over the time course of the experiment when they were initially acclimated to LL, while HL-acclimated cells appear to be more strongly affected, with a sharp and continuous decrease of their F_V/F_M. Consistent with a previous study of the oxidative stress response of WH7803 cells after a pre-acclimation at LL or HL (Blot et al., 2011), a synergic effect seems to occur between light history and low temperature stress. While this behavior was attributed to the generation of reactive oxygen species (ROS), inducing both direct damages to the reaction centers and inhibition of their repair cycles, the deleterious effect of ROS could be even amplified by the slowing down of the metabolism, and thus of the D1 turnover, at LT. In contrast, cells acclimated to HL seem to be better prepared to sustain UV stress than LL-acclimated cells since a much lower number of genes were DE at all time points

of UV stress in HL- than in LL-acclimated cultures (Fig. 3A) and there was a much smaller drop in F_V/F_M in HL conditions, partially attributable to a sharp increase in the D1 repair capacity of the cells (Fig. 1A). As concerns the L/D cycle, the F_V/F_M ratio varied strongly during the day, as previously described (Mella-Flores et al., 2012). The comparison between 21 and 27°C revealed that the amplitude of variations of both F_V/F_M and D1 repair rate was more pronounced at higher temperature (Fig. 2). In contrast, the number of DE genes varied much more at 21°C with a peak at the L/D transition (18:00), while a large number of genes were DE during the whole light period at 27°C (Fig. 3B), which tends to support the synergic effect of light and temperature observed in continuous light conditions. Analyses of the 154 transcriptomes also revealed the common and specific responses to the different tested conditions and the respective roles of various metabolic pathways (based on KEGG) either particularly represented in the WGCNA modules based on TIGR roles and Cyanorak functional categories or previously known to be involved in stress responses. These pathways are detailed in the following sections.

Chaperones

Chaperone- and protease-encoding genes were among the most strongly induced genes in response to all stressful conditions (HL, UV and LT) but also at the end of the exponential growth phase (LLHT-D3). Among these, the two *clpB1* gene copies as well as genes encoding the GroEL/ES system (*groS, L1, L2*), the small heat shock protein (HSP20), HtpG, three out of seven DnaJ (*dnaJ1*, *2* and *5*), one out of three DnaK (*dnaK3*) as well as the serine endoprotease *degQ,* were activated in all stressful conditions, while *clpP1*-4 and *clpC* seem to be more specific to UV. Consistent with previous studies, the expression of *clpX*, encoding a regulatory ATPase/chaperone interacting with ClpP, was repressed under LLHL, LLUV and HLLT (Schelin et al., 2002; Blot et al., 2011).

Photosynthesis light reactions

As expected from the central role of photosynthesis in the metabolism of cyanobacteria, photosynthetic genes were among the most DE in the various stresses. Most genes encoding PSI, cytochromes b_6/f and phycobilisomes were strongly downregulated in LL-acclimated cells in response to HL, UV and LT stresses, with the notable exception of the four phycocyanin-specific phycobilin lyases *cpcS* and *rpcE-F-T*

(Six et al., 2007), which were slightly upregulated by UV stress. In contrast, genes involved in PSII biosynthesis displayed a much more variable response to stress, likely due to the particular sensitivity of PSII reaction centers to photodamage (Table S1). Indeed, two competing processes are at work in the transcriptional response of PSII genes to stress: i) the *de novo* assembly of all PSII subunits and its cofactors and ii) the replacement of damaged D1 proteins, so-called 'PSII repair', which involves partial disassembly and subsequent rebuilding of PSII (Nickelsen and Rengstl, 2013; Mabbitt et al., 2014; Fig. S5). Like for other components of the photosynthetic apparatus, the neo-biogenesis of PSII components seems to be repressed in response to all stresses, as suggested by the downregulation of *psbB-I-K-M-O-T-U-V-X-Y-Z, cyanoP-Q* as well as *psb30* and *psb32* genes. Yet, the expression pattern of all other PSII genes was seemingly influenced by the D1 repair process. This includes the three *psbA* gene copies encoding the D1.2 isoform, which were strongly induced in all treatments in contrast to the sole gene encoding the more photosynthetically efficient but less stress-resistant D1.1 isoform (Garczarek et al., 2008; Kós et al., 2008). Similarly, as previously observed for *Synechococcus* sp. PCC 7942 (Bustos and Golden, 1992) and *Synechocystis* sp. PCC 6803 (Viczián et al., 2000), the two genes encoding D2 proteins displayed a distinct DE pattern. Indeed, while the *psbD* copy in operon with *psbC* was weakly but specifically induced by UV stress, the monocistronic *psbD* gene copy was in contrast almost as strongly upregulated as the pool of D1.2-encoding *psbA* genes under LLHL, LLUV and LLHT conditions. This suggests that D2 needs to be synthesized at a similar rate as D1.2 in order to maintain the PSII active in LL-acclimated cells exposed to HL or UV radiations (Fig. S5). Of note, none of the *psbD* copies were DE in LLLT and the polycistronic copy was even downregulated in HLLT conditions, while all three *psbA* genes encoding the D1:2 isoforms were upregulated in these conditions, suggesting that these stresses induced different damages to the D1 and D2 subunits. Consistently, the expression pattern of the *ctpA* gene, known to encode the D1 protein carboxyl-terminal processing peptidase, mimicked that of D1:2 but not D2 encoding genes, though being much less DE.

A number of other PSII genes, including *psbH-J-L* and *psbE-F1* encoding cyt b_{559}, were specifically upregulated in the LLUV treatment, like the *psbCD* operon, but more faintly. Interestingly, this parallels the expression of several genes coding for PSII assembly factors. This includes *ALB3*, involved in the integration of precursor D1 proteins (pD1) into

the thylakoid membrane (Ossenbühl et al., 2006), *ycf39* that encodes a component of a chl-binding protein complex also including HliC and HliD, which is involved in the delivery of chl *a* to newly synthesized D1 (Knoppová et al., 2014), *ycf48* that encodes a factor participating in both i) the stabilization of newly synthesized D1 precursors and their subsequent binding to D2–cyt b_{559} pre-complexes and ii) in selective replacement of damaged D1 during PSII repair (Komenda et al., 2012), as well as *psbN* that, by analogy with its homolog in plants, is assumed to be involved in both PSII dimerization during early biogenesis and in PSII repair (Plöchinger et al., 2016; Fig. S5). Additionally, *deg1*, involved in the degradation of the D1 protein during repair after photoinhibition (Kapri-Pardes et al., 2007) as well as three genes encoding a complex also specifically involved in PSII repair, *ftsH2-3 and psb29* (Bečková et al., 2017), were all specifically upregulated in the LLUV treatment. Altogether, these data suggest that, in our experimental conditions, UV stress is more deleterious for PSII than the other tested treatments in WH7803, and that the full recovery of damaged PSII from UV stress might involve the neo-synthesis of a fairly large number of PSII subunits and not just the reassembly of pre-existing subunits after D1 replacement. In agreement with this hypothesis, several genes involved in key steps of the biosynthesis of chl *a*, which binds at very early steps of PSII assembly and that is required for stabilizing PSII pre-complexes (Komenda et al., 2012), were also more particularly upregulated in response to UV stress. This includes i) two of the three genes encoding the protoporphyrin IX Mg-chelatase complex (*chlD* and *H*), which catalyzes the insertion of Mg^{2+} into protoporphyrin IX, ii) *cycI1* that transforms Mg-protoporphyrin IX monomethyl ester into divinyl protochlorophyllide as well as iii) *cvrA* that transforms divinyl-chlorophyll *a* precursor into monovinyl chl *a* (Islam et al., 2008). Furthermore, the strong UV-induced upregulation of the *pao* gene encoding the pheophorbide *a* oxygenase, responsible for opening of the chlorin macrocycle of pheophorbide a, likely indicates that an active chl breakdown occurs in this condition, potentially allowing cells to eliminate molecules of this potentially phototoxic pigment associated to damaged PSII proteins (Hörtensteiner and Kräutler, 2011).

Another stress-specific response related to photosynthesis is the significant induction of genes coding for RuBisCo, most carboxysome subunits as well as all ATPase subunits only in response to HL stress, while being downregulated in response to all other stress conditions. This suggests that, although cells seemingly actively respond to the increase

in irradiance by reducing the number of photosystems and their light harvesting capacity, as demonstrated by the downregulation of most genes involved in biosynthesis of phycobilisomes, PSI as well as PSII subunits not involved in the maintenance of the D1/D2 turnover, they are able to use the extra photons provided by the increase in light intensity to enhance carbon and energy production. Although less striking, a few photosynthetic genes also seem to be more DE in response to LT stress, including genes coding for the three subunits of NADH dehydrogenase specifically involved in CO_2 fixation (*cupB*, *ndhD4*/F4) as well as *ndhM* and V, the two latter subunits being essential for cyclic electron transport around PS I in *Synechocystis* sp. PCC 6803 (Gao et al., 2016; He et al., 2016), the expression pattern of these five genes differing from that of the other Ndh subunits.

Glucose and glycogen metabolism

As previously observed in response to nitrogen deprivation in *Synechocystis* sp. PCC 6803 (Azuma et al., 2011), sugar catabolism pathways were also strongly affected by the different stresses tested in this study. The two glycogen degradation pathways, either through the glycogen phosphorylase (GlgP) or through the isoamylase (GlgX), were indeed both strongly upregulated in LLLT and also slightly ($0.5<\log_2 FC\leq 1$) in the LLHL, LLUV and/or HLLT treatments (Fig. 6). Conversely, all genes involved in glycogen biosynthesis (*glgA-B-C*) were downregulated in all these conditions, while the two copies of phosphoglucomutase (*pgmA1/A2*), a metabolic branch point between storage and utilization of carbohydrates was also downregulated under UV and LT stresses. Furthermore, most genes of the upper glycolysis (including *pgi*, *fbaA* and *tpiA*) and the oxidative pentose phosphate (OPP) pathways were slightly downregulated by the different treatments. Conversely, the *eda* gene, recently shown to be essential for the Entner-Doudoroff pathway in *Synechocystis* (Chen et al., 2016) was found to be strongly upregulated in LLHL, LLUV, LLLT and HLLT treatments (Fig. 6, Table S1). While the absence of activation of the OPP pathway is expected since this pathway is mostly active in the dark and in photomixotrophic conditions (Takahashi et al., 2008), these results also tend to support Chen et al., 2016's recent hypothesis that the absence in all marine picocyanobacteria genomes (including WH7803) of orthologs for *Synechocystis* phosphofructokinase genes *pfkA* and *pfkB* likely implies that these organisms are not able

to perform the upper part of the glycolysis pathway that transforms glucose-6-phosphate (G6P) into glyceraldehyde-3P (GA3P). The *gap1* gene encoding glyceraldehyde-3-phosphate dehydrogenase type I, the first enzyme of the lower part of the glycolysis pathway, which is shared between the three glucose degradation pathways, was strongly upregulated in both LLHL and LLLT treatments, while its anabolic counterpart *gap2* was unaffected in LLHL and downregulated in LLLT (Fig. 6). Moreover, the upregulation of *xfp* that encodes a xylulose 5-phosphate/fructose 6-phosphate (Xu5P/F6P) phosphoketolase that catalyzes the conversion of F6P and inorganic phosphate (Pi) to erythrose 4-phosphate (E4P) and acetyl phosphate (AcP) and/or Xu5P and Pi to glyceraldehyde 3-phosphate (GA3P) and AcP, might provide an additional source of GA3P. Altogether, these results indicate that all stress conditions favored the degradation of glycogen and then glucose into pyruvate mostly via the Entner-Doudoroff pathway, with the concomitant production of ATP, NAD(P)H and biosynthetic precursors for amino acids, nucleotides and fatty acids.

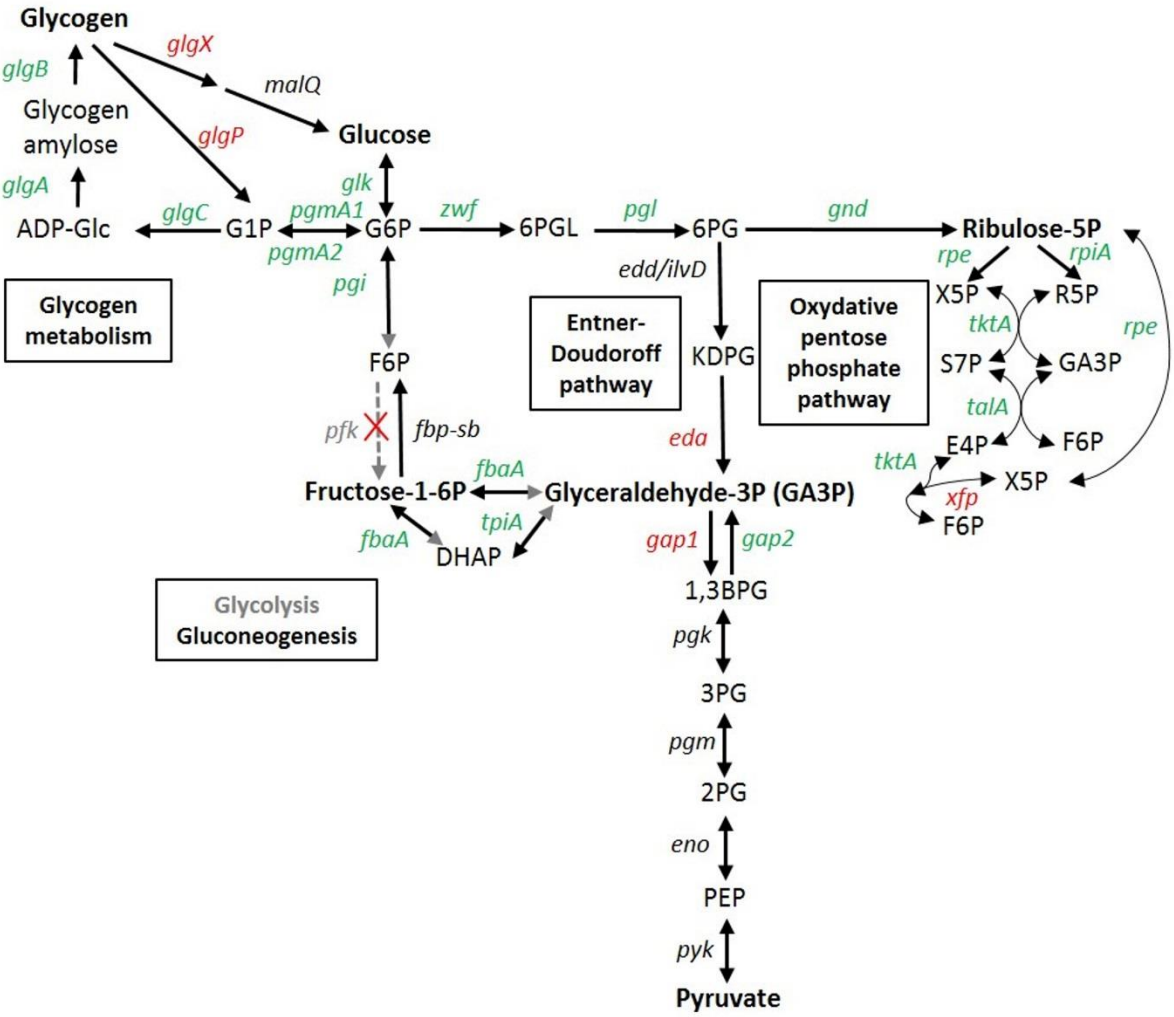


Fig. 6: Transcriptomic response of glycogen and glucose metabolism genes in Synechococcus sp. WH7803. In red, genes induced in at least one stress condition (usually LLLT); in green, genes repressed in at least one stress condition; in black, genes not significantly differentially expressed whatever the condition; in grey, no obvious candidate in the genome for this function. Gene products are the following: glgA, glycogen synthase; glgB, 1,4-alpha-glucan branching enzyme; glgC glucose-1-phosphate adenylyltransferase; glgP, glycogen phosphorylase; glgX, glycogen isoamylase; malQ, 4-alpha-glucanotransferase; pgmA, phosphoglucomutase; glk, glucokinase; pgi, glucose-6-phosphate isomerase; fbp, fructose-1,6-bisphosphatase; pfk, phosphofructokinase; fbaA (a.k.a. cbbA), fructose-bisphosphate aldolase; tpiA, glycogen isoamylase; zwf, glucose-6-phosphate dehydrogenase; pgl (a.k.a devB) 6-phosphogluconolactonase; edd/ilvD (potential bifunctional 6-phosphogluconate dehydratase/dihydroxy-acid dehydratase); eda, 2-keto-3-deoxygluconate-6-phosphate aldolase; gnd, 6-phosphogluconate dehydrogenase; rpiA, ribose-5-phosphate isomerase; rpe, pentose-5-phosphate-3-epimerase; talA, transaldolase; tktA, transketolase, xfp: xylulose 5-phosphate/fructose 6-phosphate phosphoketolase; gap1, glyceraldehyde-3-phosphate dehydrogenase (catabolic reaction) gap2, glyceraldehyde-3-phosphate dehydrogenase (anabolic reaction); pgk phosphoglycerate kinase; pgm (=gpmB), phosphoglycerate mutase; eno, enolase; pyk, pyruvate kinase. Abbreviations for substrates: ADP-Glc, ADP-glucose; G1P, glucose-1-phosphate; G6P, glucose-6-phosphate; F6P, fructose-6-phosphate; DHAP, Dihydroxyacétone phosphate; GA3P, glyceraldehyde-3-phosphate; 6PGL. 6-phosphogluconolactone; 6PG, 6-phosphogluconate; KDPG, 2-keto-3-deoxygluconate-6-phosphate aldolase; X5P, xylulose-5-phosphate; R5P, ribose-5P; S7P, sedoheptulose-7-phosphate; E4P, erythrose-4-phosphate; 1,3 BPG, 1,3, bisphosphoglycerate; 3PG, 3-phosphoglycerate; 2PG, 2-phosphoglycerate; PEP, phosphoenolpyruvate. Adapted from Osanai et al. (2007), Chen et al. (2016) and the KEGG database (www.genome.jp/kegg/pathway.html).

Lipid metabolism

Variations in the relative content of glycolipids are well known to be crucial for cell acclimation to temperature changes by modulating membrane fluidity. Thinner membranes (i.e. with a lower fatty acid average length) and/or highly unsaturated membranes, both favoring their fluidity, are indeed commonly observed in cold-adapted organisms (Chintalapati et al., 2007; Iskandar et al., 2013) and in response to cold stress, notably in marine *Synechococcus* (Varkey et al., 2016; Pittera et al., 2018; Breton et al., 2019). Here, although strong correlations and anticorrelations were found between length and saturation levels of the different lipids and blue and turquoise modules, respectively (Fig. 4), most genes involved in fatty acid chain biosynthesis, in their insertion into membranes as well as in polar head biosynthesis were not significantly or only slightly DE. This suggests that these processes are likely co-regulated with other temperature-dependent response mechanisms and/or post-transcriptionally regulated. Still, the fatty acyl-ACP reductase, *aar*, is specifically induced in response to LT in HL-acclimated cultures. AAR is involved in alkane biosynthesis, a two-step pathway during which the acyl-acyl carrier protein (ACP) is first reduced to aldehyde by AAR, then the aldehyde is oxidized to alkane by the aldehyde-deformylating oxygenase (ADO; Schirmer et al., 2010). The AAR/ADO pathway was recently shown to be involved in cold stress resistance in *Synechocystis* sp. PCC 6803, since the lack of this pathway provoked an increased cyclic electron flow around PSI in the mutant and a significant growth rate reduction (Berla et al., 2015). Alkanes have thus an essential role in the regulation of the redox balance and reductant partitioning under cold stress.

The most striking observation in this functional category concerns the fatty acid desaturases, which by inserting unsaturations into acyl chains, were shown to play a crucial role in temperature stress response through adjustment of thylakoid fluidity (Ludwig and Bryant, 2012). Consistent with the data of Pittera et al. (2018) showing that after a shift from 22 to 13°C, the level of double unsaturations of the MGDG *sn*-1 chain increased, the delta-12 fatty acid desaturase gene *desA3* was found to be upregulated in response to LT, while it was downregulated in the opposite thermal shift (LLHT). The gene encoding the delta-9-desaturase DesC4, which is absent from warm *Synechococcus* thermotypes (clades II and III; Varkey et al., 2016; Pittera et al., 2018; Breton et al., 2019), also appeared strongly DE at LT but also in other stress conditions. At last, genes coding

for DesA2, a delta-12 desaturase mostly found in warm thermotypes (Pittera et al., 2018) and DesC3, a core delta-9 desaturase, were in contrast mostly responsive to UV stress, even though *desA2* was also upregulated in HLLT conditions. Thus, all four WH7803 desaturase genes displayed very different expression patterns in response to the various tested conditions, which might be related to the necessity to finely tune the fluidity of membranes, constituting the matrix for the photosynthetic machinery. Indeed, the membrane unsaturation level plays a role in the assembly and functional regulation of the PSII complex after photoinactivation, notably by influencing the maturation process of the D1 precursor (Kanervo et al., 1997; Mizusawa and Wada, 2012).

DNA repair

Many stresses are known to cause DNA damages to bacteria (including cyanobacteria) either directly (e.g. chemicals, gamma or UV-C radiations, etc.) or indirectly through the generation of ROS (e.g. induced by light, UV-A and B radiations, or temperature stress; He and Häder, 2002). Like other marine *Synechococcus*, WH7803 possesses a broad set of DNA repair proteins (Cassier-Chauvat et al., 2016). Some of them are involved in direct DNA damage reversal, such as DNA photolyases or members of the Odt/Ogt family (e.g. YbaZ, a 6-O-alkylguanine DNA alkyltransferase likely involved in G:C to A:T transversions; Mazon et al., 2009). WH7803 possesses four distinct putative photolyases, one of them being synthesized by an operon of two small genes, the first one coding for the FAD-binding domain (CK_00001541) and the second one for the photolyase domain (CK_00001540). Three of these photolyases were strongly upregulated in response to most stresses and more specifically to LLUV, while the fourth one (CK_00001460), whose C-terminal domain is only distantly related to the typical DNA photolyase domain, showed only a faint response in LLUV and HLLT conditions. In contrast, the *ybaZ* gene was most strongly upregulated in response to LLHL and LLLT. Altogether, it seems that the WH7803 genes involved in direct DNA damage reversal are activated in response to distinct stresses.

WH7803 also possesses many genes potentially involved in SOS response (Table S1; Fig. S6; see also Blot et al., 2011; Cassier-Chauvat et al., 2016). In *E. coli*, this pathway constitutes an essential response to stress-induced DNA damages, starting by the formation by either RecBCD or RecFOR complexes of single-strand DNA (ssDNA),

recognized by RecA, which then catalyzes the auto-proteolysis of the SOS regulon repressor LexA (see Baharoglu and Mazel, 2014) for a review). In contrast, in *Synechocystis* sp. PCC 6803, the latter protein has been shown to regulate carbon assimilation and cell motility but surprisingly not DNA repair (Domain et al., 2004; Kizawa et al., 2016). Accordingly, this strain lacks several DNA repair genes present not only in *E. coli* but also in marine *Synechococcus* spp. (namely *mutY* and *recB*, *C*, *J* and *Q*). It is therefore quite possible that marine *Synechococcus* display an SOS response, regulated by LexA, a hypothesis supported by the fact that in WH7803, both *recA* and *lexA* were strongly upregulated in response to UV shifts (both LLUV and HLUV) and more faintly to low temperature (either LLLT or HLLT) and high light (LLHL) shifts. Interestingly, the biosynthesis of the RecBCD and RecF(O)R complexes, recognizing double-strand breaks and gaps in one DNA strand, respectively, did not seem to occur in the same conditions in WH7803. While the former encoding genes were slightly upregulated at LT, the latter were mostly upregulated during the LLUV shift, suggesting that these stresses induce different DNA damages. Of note, *recO* displayed a different expression pattern than *recF* and *recR,* questioning whether RecO is truly part of the same complex as RecF-R in marine *Synechococcus*, especially since WH7803 RecO-like sequence is less similar to *E. coli* RecO than the corresponding RecF and RecR counterparts. It is also interesting to note that *dprA*, which displaces SSB, loads RecA onto ssDNA, protects ssDNA from nucleases (Hovland et al., 2017) and was previously suggested to belong to the LexA regulon in WH7803 (Blot et al., 2011), is specifically and strongly induced under UV stress.

Additionally, the nucleotide excision repair (NER) pathway, driven by UvrABC, which allows the repair of lesions in double-stranded DNA (Sancar and Rupp, 1983), was also activated in WH7803 (Fig. S6). Genes coding for the UvrA-B complex, which recognizes the DNA lesion, were most strongly upregulated in response to LLUV while UvrC, which makes incisions on both sides of the lesion before the UvrD helicase removes the ssDNA carrying the lesion (Kumura et al., 1985), also responded to the HLLT shift. Surprisingly, the gene encoding UvrD*,* which removes the ssDNA carrying the lesion, showed no differential expression whatever the stress, suggesting a post-transcriptional regulation. The combined action of the DNA polymerase I and a ligase then fills the gap (Husain et al., 1985). In WH7803, both genes were upregulated by both UV and LT stresses.

The last pathway, involving translesion synthesis (TLS) polymerases, is known to be activated only when the SOS-inducing signal persists and allows highly mutagenic bypass of DNA lesions that are not effectively passed through by the standard replicative DNA polymerase Pol III (Patel et al., 2010). In WH7803, the only polymerase potentially involved in this pathway is polymerase V, encoded by the *umuCD* operon, which like *recA* was most strongly upregulated under UV conditions and also responded to the HLLT shift. This suggests that UV and HLLT constitute the most damaging stresses for DNA that can still be repaired but at the expense of an increased frequency of spontaneous mutations.

Compatible solutes and osmoregulation

Besides their role as osmolytes in salt-stressed cells, compatible solutes are also known to directly protect enzymes and membranes against denaturation caused by other environmental stresses and notably high or low temperatures (Diamant et al., 2003; Hincha and Hagemann, 2004; Pade and Hagemann, 2015). *Synechococcus* sp. WH7803 possesses all genes (*ggpP*, *S* and *ggtA-D*) involved in the synthesis and uptake of the heteroside glucosylglycerol (GG), and its ability to accumulate GG, considered as the main compatible solute in marine picocyanobacteria, has been experimentally verified in this strain (Scanlan et al., 2009). Still, all of these genes were downregulated in all stress conditions and showed minimal variations during the L/D cycle, except *ggpP*, displaying upregulation during the day with a peak at noon. Although WH7803 was found to accumulate a considerable amount of sucrose, also known to be involved in osmoregulation, the gene coding for the sucrose phosphate synthase and phosphatase fusion protein (*sps-sds*) was also downregulated in stress conditions and the same is true for most genes involved in synthesis (*gbmt1-2*) and transport/re-uptake (*proW-X*) of glycine betaine (GB), which was shown to be accumulated by WH7803 cells (Scanlan et al., 2009). In contrast, the gene coding for an additional mono-subunit GB uptake system (*betP*) was specifically upregulated under LT (LLLT and HLLT), while both genes (*gpgP*, *S*) potentially involved in glucosyl glycerate (GGA) synthesis, a negatively charged compatible solute, were specifically upregulated in response to HL stress, potentially explaining why it was not detected in extracts of WH7803 grown in standard conditions (Scanlan et al., 2009). Interestingly, genes involved in osmolyte synthesis and transport were also quite differently regulated by the L/D cycle with most genes involved in GG

synthesis and transport showing only faint diel variations, GGA genes being most strongly expressed during the day, while GB genes displayed the most robust diel oscillation with a peak at the L/D transition and still a strong expression in the first part of the dark period.

In summary, although WH7803 is able to accumulate GG, possibly GGA and to take up GG, sucrose and GB, these compounds appear to be involved in osmoregulation at different times of the cell cycle and to respond to distinct environmental conditions.

Oxidative stress response and photoprotection

General oxidative stress response

One of the particularities of photosynthetic organisms is that in response to various environmental stresses, the photosynthetic electron transport can sometimes outpace the rate of electron consumption by CO_2 fixation, leading to a rapid increase of intracellular ROS, which can produce deleterious effects on the cellular machinery, including DNA, lipids, proteins and notably PSII reaction centers (Nishiyama et al., 2006; Takahashi and Murata, 2008; Latifi et al., 2009). In WH7803, a different set of genes involved in ROS protection and detoxification mechanisms was DE in response to distinct stresses, suggesting that like in freshwater cyanobacteria, the different systems may function under a particular condition (Perelman et al., 2003). Among the two WH7803 genes coding for superoxide dismutase (SOD), which catalyzes the dismutation of the superoxide (O_2^-) radical into O_2 and hydrogen peroxide (H_2O_2), only *sodB* encoding the Fe-SOD was significantly upregulated in LLHL and LLUV, while *sodC*, encoding the Cu/Zn-SOD was only faintly upregulated in HLLT. The gene coding for the glutathione peroxidase (Gpx), catalyzing the decomposition of H_2O_2 to water and O_2, was downregulated in response to LT stress, while for comparison the two glutathione peroxidases present in *Synechocystis* sp. PCC 6803 were shown to be upregulated in response to HL and high salinity and to be essential for the protection of membranes against lipid peroxidation (Gaber et al., 2004; Cameron and Pakrasi, 2010). Similarly, the catalase-peroxidase, KatG, only present in a few picocyanobacterial strains and also acting on H_2O_2, was also downregulated in WH7803 but in all stressful conditions. In contrast, three out of the four peroxiredoxins (*ahpC*, *prxQ1* and *prxQ2*), also potentially involved in H_2O_2 detoxification, were upregulated, the latter two responding to most stressful conditions, while *ahpC,* encoding a 2-Cys peroxiredoxin, seemed to be more specific to HL. The differential

expression of these genes may be in part because catalases mainly detoxify high levels of H_2O_2, while peroxiredoxins were shown to rather scavenge low levels of this compound (Stork et al., 2005).

Thioredoxins (TRX) and glutathiones (GSH) are also known to play a critical role in the maintenance of the redox homeostasis and protection from ROS in cyanobacteria (Cameron and Pakrasi, 2010; Sánchez-Riego et al., 2016). Here, among the genes involved in the synthesis of these thiol molecules, the two gene copies encoding glutathione S-transferases (*gst*), were only transiently upregulated in response to LLUV stress, while two out of the six *gst* genes present in *Synechocystis* sp. PCC 6803, *sll1545* and *slr0236*, both homologs of CK_00000203, have been shown to operate in the protection against stresses such as high light and H_2O_2 (Kammerscheit et al., 2019). Additionally, the *Synechococcus* core *trxB* gene, involved in the NADPH-dependent thioredoxin reductase C (NTRC) system (Scanlan et al., 2009), was the only thiol-encoding gene to be significantly upregulated in all stressful conditions. Interestingly, NTRC as well as sulfiredoxin, which in WH7803 displayed a similar gene expression pattern, have been shown in plants and *Anabaena* sp. PCC 7120 to be involved in the protection of the photosynthetic apparatus against oxidative stress damages through an effective reduction of 2-Cys Prx (Sánchez-Riego et al., 2016). Altogether, it seems that at least in the tested conditions, the ROS protection and detoxification systems in WH7803 resemble more to the ones described in PCC 7120 and in plants, formed by NTRC, 2-Cys Prx and sulfiredoxin, than to the ones of PCC 6803 that lacks both NTRC coding gene and sulfiredoxin and which strategy mainly relies on a high peroxidase/catalase activity (Sánchez-Riego et al., 2016). However, we cannot exclude that in other (e.g. more stressful) conditions, the peroxidase/catalase system could be activated for ROS detoxification in WH7803.

Another more unusual antioxidant system detected in this study is the cystein desulfurase SUF system. It is responsible for Fe-S cluster biosynthesis in conditions of Fe limitation or oxidative stress in many bacteria (Nodop et al., 2008; Bolstad and Wood, 2010; Dai and Outten, 2012) and constitutes the major pathway for Fe-S cluster assembly in cyanobacteria (Ayala-Castro et al., 2008; Banerjee et al., 2017). In marine picocyanobacterial, the *sufBCD* operon, which is highly conserved in cyanobacteria, is adjacent to the *sufR* transcriptional repressor and all sequenced picocyanobacterial genomes also possess a group I *sufS* (CK_00000030), homologous to the *Synechocystis* sp.

PCC 6803 group I NifS-like cysteine desulfurase, Slr0387 (Tirupati et al., 2004). Here, the gene coding for the SufE protein, involved in the sulfur release from cysteine and that can stimulate the cysteine desulfurase activity of SufS, was found to be quite strongly upregulated in LT, while *sufBCDS* were mostly upregulated in LLLT but also in LLHL and late LLUV stresses. Thus, by analogy with *Anabaena* sp. PCC 7120, in which the SufS protein was suggested to enhance oxidative stress tolerance by decreasing the intracellular ROS (Banerjee et al., 2017), the SUF system could also be an important antioxidant system in marine picocyanobacteria in particular under LT stress. In this context, it is also worth noting that genes coding for two manganese transporters, the high-affinity MntABC transport system and another ABC-type Mn^{2+}/Zn^{2+} transport system (SynWH7803_0988-90) were particularly upregulated under LT stresses, suggesting that like in *Anabaena* sp. PCC 7120, Mn could also play a role in protection against ROS due to its antioxidative properties or involvement as a cofactor in several antioxidative enzymes (Kaushik et al., 2015).

Photoprotection

In cyanobacteria several mechanisms are involved in the dissipation of excess light as heat in order to limit the production of ROS. While WH7803 lacks the iron stress-induced protein IsiA, potentially involved in light energy dissipation (Havaux et al., 2005), as well as a plastoquinol terminal oxidase (PTOX), an oxidase which could be involved in the removal of electrons from the intersystem photosynthetic electron transport chain (Bailey et al., 2008), this strain possesses several other mechanisms potentially involved in photoprotection and conserved in most cyanobacteria, as detailed below.

Although the exact role of high-light-inducible proteins (HLIPs) is not fully understood, the four *hli* genes of *Synechocystis* sp. PCC 6803 seem to respond differentially to various stresses and their products are thought to prevent photodamages to chl protein complexes by quenching deleterious singlet excited states of chl *a* and/or singlet oxygen (Hirayama et al., 1994; Huang et al., 2002). More specifically, two of them, HliC and D, were recently shown to bind chl and β-car in *Synechocystis* sp. PCC 6803 that could be used for chl recycling (Niedzwiedzki et al., 2016; Shukla et al., 2018). Among the nine *hli* genes present in WH7803, two were highly expressed under most stress conditions, four others (including CK_00000050, the closest homolog of PCC 6803 HliC)

were more specifically induced by UV and HL treatments, two others were only faintly DE in response to stress, while the last one (CK_00001058, corresponding to the closest homolog of PCC 6803 HliD) did not respond to any monitored stress (Table S1). Interestingly, while most *hli* genes were upregulated during the day and downregulated at night, the latter copy, a core and quite well-conserved picocyanobacterial gene, was specifically induced during the night. This somehow recalls the L/D expression patterns of *Prochlorococcus* MED4 *hli4* and *hli11*, which were also found to peak during the night (Zinser et al., 2009).

Photoprotective responses in cyanobacteria also involve the thermal dissipation of excess light energy captured by the light harvesting antenna by the orange carotenoprotein (OCP; Wu and Krogmann, 1997; Kirilovsky and Kerfeld, 2016). The *ocp* gene is part of an operon of three genes, also including the β-carotene ketolase (*crtW*), involved in the synthesis of the keto-carotenoid bound to the OCP protein and the fluorescence recovery protein (FRP), responsible for acceleration of the detachment of the OCP from the PBS complex and subsequent deactivation of the OCP (Thurotte et al., 2017). In *Synechocystis* sp. PCC 6803, the OCP was found to be strongly induced by various stresses that exacerbate photodamage and/or prolongate the excited state of antenna chromophores, notably HL (Hihara et al., 2001), LT (Kerfeld et al., 2017) and oxidative stresses (Shrivastava et al., 2016), but seems to be constitutively expressed and to remain at low concentration in non-stressed conditions (Gwizdala et al., 2011). Here, these genes were indeed only moderately induced during the first part of the light period but were among the most strongly induced genes in response to light stress (HL and UV) in LL-acclimated cultured, while they were in contrast more upregulated by LT in HL- than in LL-acclimated cultures. Altogether, the expression pattern of this operon is thus consistent with a potential synergic effect between HL and LT, as discussed above.

Besides the abovementioned beta-carotene ketolase, several genes involved in carotenoid biosynthesis pathways were also upregulated in response to several stresses. This included *crtPQ* (LLHL and LLUV), involved in the biosynthesis of lycopene, and *crtL*-b (LLLT and HLLT) encoding a cyclase enzyme, which turns lycopene into β-carotene (Fig. S7). The latter pigment, which is mainly located in the molecular neighborhood of the chl *a* molecules bound to reaction centers and acts as an antioxidant (Telfer, 2005), has previously been shown to be involved in the cold stress response (Pittera et al., 2014).

Since the β-carotene to chl *a* ratio decreased in these conditions (Fig. S1D), the upregulation of *crtL*-b suggests that the rate of β-carotene breakdown by ROS exceeded its synthesis rate during the stress period. Interestingly, the *crtL-b* gene was recently shown to display the highest number of substitutions specific of the cold-adapted clades I and IV (Doré et al., 2020), and could thus play a crucial role in the adaptation to cold temperature in marine *Synechococcus*. The likely increase of β-carotene synthesis can also be linked to its role of precursor in the *Synechococcus* xanthophyll pathway. Indeed, *crtR*, coding for an hydroxylase involved in zeaxanthin biosynthesis, was upregulated in response to HL and UV stress, consistent with the increase of the ß-cryptoxanthin and zeaxanthin to chl *a* ratios in these conditions (Fig. S1B-C). Although the role of zeaxanthin has never been formerly demonstrated, it is likely that it affords photoprotection during stress periods.

At last, a gene coding for a carotenoid cleavage dioxygenase (*diox1*), catalyzing the oxidative cleavage of apo-carotenoids was strongly upregulated in all stressful conditions (Cui et al., 2012). As previously suggested for freshwater cyanobacteria (Hihara et al., 2003; Scherzinger et al., 2006), these apo-carotenoids could arise from photodestruction processes mediated by ROS, and cleavage products (retinal or retinal-like compounds) could be involved in stress signaling in WH7803.

In conclusion, this work constitutes the first comprehensive study of the response of a ubiquitous marine cyanobacterium to various stresses, taking into account the light history of the cells. This allowed us to identify the common and specific responses of this organism to a variety of ecologically relevant environmental conditions and to highlight significant differences compared to well-studied freshwater model cyanobacteria and in particular to *Synechocystis* sp. PCC 6803. The availability of this unique set of 154 transcriptomes, notably including L/D cycles at two temperatures, should be particularly useful to better interpret the ever-increasing metatranscriptomic data generated from a variety of ecological niches in the marine ecosystem and at different times of the day.

Data Availability Statement

The transcriptomic data supporting the conclusions of this article are available as raw and processed data in the SRA and GEO databases, respectively, and sample descriptions and accession numbers are available in Tables S2-S4. Differential expression levels and associated statistics as well as genomic features of the *Synechococcus* sp. WH7803 strain, such as core genes and predicted cyanorons, are available in Table S1.

Ethics Statement

Not applicable.

Author Contributions

CSi, MR and LG designed the experiments, NAN, HD, JP, MR, CSi, FP and LG collected the samples and performed physiological measurements, NAN and MR extracted the RNA, NAN, MR, CSt and LG developed the protocol for RNA library preparation, UG, HD, LG, GLC and EC developed and ran the RNAseq analysis pipeline. LBG, MH and UG integrated the genome and transcriptome data into the JBrowse genome browser available through the Cyanorak v2 information system. MC, AS, FP, LG, UG manually annotated genes and pathways. JH, UG, HD, LG and DE performed comparative transcriptomic analyses. UG, JH, FP and LG made the figures. UG, HD, FP and LG interpreted the results. All authors contributed to the preparation of the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the French “Agence Nationale de la Recherche” Programs SAMOSA (ANR-13-ADAP-0010) and CINNAMON (ANR-17-CE2-0014-01).

Conflict of Interest

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.

Acknowledgements

We would like to thank Nathalie Marsaud and Marie Gislard from the GeT-PlaGe as well as the ABiMS platforms for building and sequencing the RNAseq libraries and providing computational support for this work, respectively. We warmly thank Adriana Alberti for useful discussions about ribosomal depletion and RNAseq library preparation, Léna Gouhier, Martin Gachenot and Priscillia Gourvil from the Roscoff Culture Collection for providing the WH7803 strain, Garance Monier and Théo Sciandra for participating in the curation of the Cyanorak v2 database as well as Théophile Grébert for his help with sampling. This work was supported by the French “Agence Nationale de la Recherche” Programs SAMOSA (ANR-13-ADAP-0010) and CINNAMON (ANR-17-CE2-0014-01).

References

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi:10.1093/bioinformatics/btu638.

Andrews, S. (2015). FASTQC a quality control tool for high throughput sequence data. *Babraham Inst.* Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Ayala-Castro, C., Saini, A., and Outten, F. W. (2008). Fe-S Cluster assembly pathways in bacteria. *Microbiol. Mol. Biol. Rev.* 72, 110–125. doi:10.1128/mmbr.00034-07.

Azuma, M., Osanai, T., Hirai, M. Y., and Tanaka, K. (2011). A response regulator Rre37 and an RNA polymerase sigma factor SigE represent two parallel pathways to activate sugar catabolism in a cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Cell Physiol.* 52, 404–412. doi:10.1093/pcp/pcq204.

Baharoglu, Z., and Mazel, D. (2014). SOS, the formidable strategy of bacteria against aggressions. *FEMS Microbiol. Rev.* 38, 1126–1145. doi:10.1111/1574-6976.12077.

Bailey, S., Melis, A., Mackey, K. R., Cardol, P., Finazzi, G., van Dijken, G., et al. (2008). Alternative photosynthetic electron flow to oxygen in marine Synechococcus. *Biochim. Biophys. Acta* 1777, 269–276. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18241667.

Banerjee, M., Chakravarty, D., and Ballal, A. (2017). Molecular basis of function and the unusual antioxidant activity of a cyanobacterial cysteine desulfurase. *Bioch. J.* 474, 2435–2447. doi:10.1042/BCJ20170290.

Bečková, M., Yu, J., Krynická, V., Kozlo, A., Shao, S., Koník, P., et al. (2017). Structure of Psb29/Thf1 and its association with the Ftsh protease complex involved in photosystem II repair in cyanobacteria. *Phil.Trans. R. Soc. B Biol. Sci.* 372, 20160394. doi:10.1098/rstb.2016.0394.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x.

Berla, B. M., Saha, R., Maranas, C. D., and Pakrasi, H. B. (2015). Cyanobacterial alkanes modulate photosynthetic cyclic electron flow to assist growth under cold stress. *Sci. Rep.* 5, 14894. doi:10.1038/srep14894.

Blanco-Ameijeiras, S., Cosio, C., and Hassler, C. S. (2017). Long-term acclimation to iron limitation reveals new insights in metabolism regulation of *Synechococcus* sp. PCC7002. *Front. Mar. Sci.* 4, 247. doi:10.3389/fmars.2017.00247.

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, 1934–1954. doi:10.1104/pp.111.174714.

Bolstad, H. M., and Wood, M. J. (2010). An in vivo method for characterization of protein interactions within sulfur trafficking systems of *E. coli*. *J. Proteome Res.* 9, 6740–6751. doi:10.1021/pr100920r.

Breton, S., Jouhet, J., Guyet, U., Gros, V., Pittera, J., Demory, D., et al. (2019).

Unveiling membrane thermoregulation strategies in marine picocyanobacteria. *New Phytol.*, DOI: 10.1111/nph.16239. doi:10.1111/nph.16239.

Buels, R., Yao, E., Diesh, C. M., Hayes, R. D., Munoz-Torres, M., Helt, G., et al. (2016). JBrowse: A dynamic web platform for genome visualization and analysis. *Genome Biol.* doi:10.1186/s13059-016-0924-1.

Bustos, S. A., and Golden, S. S. (1992). Light-regulated expression of the *psbD* gene family in *Synechococcus* sp. strain PCC 7942: evidence for the role of duplicated *psbD* genes in cyanobacteria. *Mol. Gen. Genet.* 232, 221–230. doi:10.1007/BF00280000.

Cameron, J. C., and Pakrasi, H. B. (2010). Essential role of glutathione in acclimation to environmental and redox perturbations in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Physiol.* 154, 1672–1685. doi:10.1104/pp.110.162990.

Cassier-Chauvat, C., Veaudor, T., and Chauvat, F. (2016). Comparative genomics of DNA recombination and repair in cyanobacteria: Biotechnological implications. *Front. Microbiol.* 7, 1809. doi:10.3389/fmicb.2016.01809.

Chamot, D., Magee, W. C., Yu, E., and Owttrim, G. W. (1999). A cold shock-induced cyanobacterial RNA helicase. *J. Bacteriol.* 181, 1728–1732. doi:10.7939/R36689011.

Chamot, D., and Owttrim, G. W. (2000). Regulation of cold shock-induced RNA helicase gene expression in the cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* 182, 1251–1256. doi:10.1128/JB.182.5.1251-1256.2000.

Chen, X., Schreiber, K., Appel, J., Makowka, A., Fähnrich, B., Roettger, M., et al. (2016). The Entner-Doudoroff pathway is an overlooked glycolytic route in cyanobacteria and plants. *Proc. Natl Acad. Sci. U.S.A.* 113, 5441–5446. doi:10.1073/pnas.1521916113.

Chintalapati, S., Prakash, J. S. S., Singh, A. K., Ohtani, S., Suzuki, I., Murata, N., et al. (2007). Desaturase genes in a psychrotolerant *Nostoc* sp. are constitutively expressed at low temperature. *Biochem. Biophys. Res. Commun.* 362, 81–87. doi:10.1016/j.bbrc.2007.07.150.

Cui, H., Wang, Y., and Qin, S. (2012). Genomewide analysis of carotenoid cleavage dioxygenases in unicellular and filamentous cyanobacteria. *Comp. Funct. Genomics*, 164690. doi:10.1155/2012/164690.

da Silva Neto, J. F., Lourenço, R. F., and Marques, M. V. (2013). Global transcriptional response of *Caulobacter crescentus* to iron availability. *BMC Genomics* 14, 549. doi:10.1186/1471-2164-14-549.

Dai, Y., and Outten, F. W. (2012). The *E. coli* SufS-SufE sulfur transfer system is more resistant to oxidative stress than IscS-IscU. *FEBS Lett.* 586, 4016–4022. doi:10.1016/j.febslet.2012.10.001.

Diamant, S., Rosenthal, D., Azem, A., Eliahu, N., Ben-Zvi, A. P., and Goloubinoff, P. (2003). Dicarboxylic amino acids and glycine-betaine regulate chaperone-mediated protein-disaggregation under stress. *Mol. Microbiol.* 49, 401–410. doi:10.1046/j.1365-2958.2003.03553.x.

Domain, F., Houot, L., Chauvat, F., and Cassier-Chauvat, C. (2004). Function and regulation of the cyanobacterial genes *lexA*, *recA* and *ruvB*: LexA is critical to the survival of cells facing inorganic carbon starvation. *Mol.*

Microbiol. 53, 65–80. doi:10.1111/j.1365-2958.2004.04100.x.

Doré, H., Farrant, G. K., Guyet, U., Haguait, J., Humily, F., Ratin, M., et al. (2020). Evolutionary mechanisms of genome diversification in marine picocyanobacteria. *Genome Biol.* submitted.

Dudley, S. (2004). "Plasticity and the functional ecology of plants," in *Phenotypic Plasticity – Functional and Conceptual Approaches*, eds. T. J. Dewitt and S. M. Scheiner (Oxford, UK: Oxford University Press), 151–172.

Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincon, J., Zabala, L. L., 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, 9824–9829. doi:10.1073/pnas.1307701110.

Gaber, A., Yoshimura, K., Tamoi, M., Takeda, T., Nakano, Y., and Shigeoka, S. (2004). Induction and functional analysis of two reduced nicotinamide adenine dinucleotide phosphate-dependent glutathione peroxidase-like proteins in *Synechocystis* PCC 6803 during the progression of oxidative stress. *Plant Physiol.* 136, 2855–2861. doi:10.1104/pp.104.044842.

Gabriel, W. (2005). How stress selects for reversible phenotypic plasticity. *J. Evol. Biol.* 18, 873–883. doi:10.1111/j.1420-9101.2005.00959.x.

Galhardo, R. S., Hastings, P. J., and Rosenberg, S. M. (2007). Mutation as a stress response and the regulation of evolvability. *Crit. Rev. Bioch. Mol. Biol.* 42, 399–435. doi:10.1080/10409230701648502.

Gao, F., Zhao, J., Wang, X., Qin, S., Wei, L., and Ma, W. (2016). NdhV is a subunit of NADPH dehydrogenase essential for cyclic electron transport in *Synechocystis* sp. strain PCC 6803. *Plant Physiol.* 170, 752–760. doi:10.1104/pp.15.01430.

Garczarek, L., Dufresne, A., Blot, N., Cockshutt, A. M., Peyrat, A., Campbell, D. A., 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. doi:10.1038/ismej.2008.46.

Georg, J., Rosana, A. R. R., Chamot, D., Migur, A., Hess, W. R., and Owttrim, G. W. (2019). Inactivation of the RNA helicase CrhR impacts a specific subset of the transcriptome in the cyanobacterium *Synechocystis* sp. PCC 6803. *RNA Biol.* 16, 1205–1214. doi:10.1080/15476286.2019.1621622.

Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., et al. (2016). Plankton networks driving carbon export in the oligotrophic ocean. *Nature* 532. doi:10.1038/nature16942.

Gwizdala, M., Wilson, A., and Kirilovsky, D. (2011). In vitro reconstitution of the cyanobacterial photoprotective mechanism mediated by the orange carotenoid protein in *Synechocystis* PCC 6803. *Plant Cell* 23, 2631–2643. doi:10.1105/tpc.111.086884.

Havaux, M., Guedeney, G., Hagemann, M., Yeremenko, N., Matthijs, H. C. P., and 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. doi:10.1016/j.febslet.2005.03.021.

He, Y. Y., and Häder, D. P. (2002). Reactive oxygen species and UV-B: Effect on cyanobacteria. *Photochem. Photobiol. Sci.* 10, 729–736.

doi:10.1039/b110365m.

He, Z., Xu, M., Wu, Y., Lv, J., Fu, P., and Mi, H. (2016). NdhM subunit is required for the stability and the function of NAD(P)H dehydrogenase complexes involved in CO2 uptake in *Synechocystis* sp. strain PCC 6803. *J. Biol. Chem.* 291, 5902–5912. doi:10.1074/jbc.M115.698084.

Helliwell, K. E., Lawrence, A. D., Holzer, A., Kudahl, U. J., Sasso, S., Kräutler, B., et al. (2016). Cyanobacteria and eukaryotic algae use different chemical variants of vitamin b12. *Curr. Biol.* 26, R319–R337. doi:10.1016/j.cub.2016.02.041.

Hihara, Y., Kamei, A., Kanehisa, M., Kaplan, A., and Ikeuchi, M. (2001). DNA microarray analysis of cyanobacterial gene expression during acclimation to high light. *Plant Cell* 13, 793–806. doi:10.1105/tpc.13.4.793.

Hihara, Y., Sonoike, K., Kanehisa, M., and Ikeuchi, M. (2003). DNA microarray analysis of redox-responsive genes in the genome of the cyanobacterium *Synechocystis* sp. strain PCC 6803. *J. Bacteriol.* 185, 1719–1725. doi:10.1128/JB.185.5.1719-1725.2003.

Hincha, D. K., and Hagemann, M. (2004). Stabilization of model membranes during drying by compatible solutes involved in the stress tolerance of plants and microorganisms. *Biochem J.* 383, 277–283. doi:10.1042/BJ20040746.

Hirayama, O., Nakamura, K., Hamada, S., and Kobayasi, Y. (1994). Singlet oxygen quenching ability of naturally occurring carotenoids. *Lipids* 29, 149–150. doi:10.1007/BF02537155.

Hörtensteiner, S., and Kräutler, B. (2011). Chlorophyll breakdown in higher plants. *Biochim. Biophys. Acta Bioenerg.* 1807, 977–988. doi:10.1016/j.bbabio.2010.12.007.

Hovland, E., Beyene, G. T., Frye, S. A., Homberset, H., Balasingham, S. V., Gómez-Muñoz, M., et al. (2017). DprA from *Neisseria meningitidis*: Properties and role in natural competence for transformation. *Microbiol. U.K.* 163, 1016–1029. doi:10.1099/mic.0.000489.

Huang, L., McCluskey, M. P., Ni, H., and LaRossa, R. A. (2002). Global gene expression profiles of the cyanobacterium *Synechocystis* sp. strain PCC 6803 in response to irradiation with UV-B and white light. *J. Bacteriol.* 184, 6845–6858. doi:10.1128/JB.184.24.6845-6858.2002.

Husain, I., Chaney, S. G., and Sancar, A. (1985). Repair of cis-platinum-DNA adducts by ABC excinuclease in vivo and in vitro. *J. Bacteriol.* 163, 817–823.

Inaba, M., Suzuki, I., Szalontai, B., Kanesaki, Y., Los, D. A., Hayashi, H., et al. (2003). Gene-engineered rigidification of membrane lipids enhances the cold inducibility of gene expression in *Synechocystis*. *J. Biol. Chem.* 278, 12191–12198. doi:10.1074/jbc.M212204200.

Iskandar, A., Taha, B. H. M., Ahmed, R. Z., Motoigi, T., Watanabe, K., and Kurasawa, N. (2013). “Lipids in cold-adapted microorganisms,” in *Cold-adapted microorganisms*, ed. I. Yumoto (Poole, UK: Caister Academic Press), 189–214.

Islam, M. R., Aikawa, S., Midorikawa, T., Kashino, Y., Satoh, K., and Koike, H. (2008). *slr1923* of *Synechocystis* sp. PCC6803 is essential for conversion of 3,8-divinyl(proto)chlorophyll(ide) to 3-monovinyl(proto)chlorophyll(ide). *Plant Physiol.* 148, 1068–1081.

doi:10.1104/pp.108.123117.

Jablonsky, J., Papacek, S., and Hagemann, M. (2016). Different strategies of metabolic regulation in cyanobacteria: From transcriptional to biochemical control. *Sci. Rep.* 6, 33024. doi:10.1038/srep33024.

Jozefczuk, S., Klie, S., Catchpole, G., Szymanski, J., Cuadros-Inostroza, A., Steinhauser, D., et al. (2010). Metabolomic and transcriptomic stress response of *Escherichia coli*. *Mol. Syst. Biol.* 6, 364. doi:10.1038/msb.2010.18.

Kammerscheit, X., Chauvat, F., and Cassier-Chauvat, C. (2019). First in vivo evidence that glutathione-s-transferase operates in photo-oxidative stress in cyanobacteria. *Front. Microbiol.* 10, 1889. doi:10.3389/fmicb.2019.01899.

Kana, T. M., and Glibert, P. M. (1987a). 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, 479–485. doi:10.1016/0198-0149(87)90001-X.

Kana, T. M., and Glibert, P. M. (1987b). Effect of irradiances up to 2000 µE m-2 s-1 on marine *Synechococcus* WH7803 - II. Photosynthetic responses mechanisms. *Deep Sea Res.* 34, 497–516. doi:10.1016/0198-0149(87)90002-1.

Kana, T. M., Glibert, P. M., Goericke, R., and Welschmeyer, N. A. (1988). Zeaxanthin and ß-carotene in *Synechococcus* WH7803 respond differently to irradiance. *Limnol. Ocean.* 33, 1623–1627. doi:10.4319/lo.1988.33.6part2.1623.

Kanervo, E., Tasaka, Y., Murata, N., and Aro, E. M. (1997). Membrane lipid unsaturation modulates processing of the photosystem II reaction-center protein D1 at low temperatures. *Plant Physiol.* 114, 841–849. doi:10.1104/pp.114.3.841.

Kannan, G., Wilks, J. C., Fitzgerald, D. M., Jones, B. D., BonDurant, S. S., and Slonczewski, J. L. (2008). Rapid acid treatment of *Escherichia coli*: Transcriptomic response and recovery. *BMC Microbiol.* 8, 37. doi:10.1186/1471-2180-8-37.

Kapri-Pardes, E., Naveh, L., and Adam, Z. (2007). The thylakoid lumen protease Deg1 is involved in the repair of photosystem II from photoinhibition in *Arabidopsis*. *Plant Cell* 19, 1039–1047. doi:10.1105/tpc.106.046573.

Kaushik, M. S., Srivastava, M., Verma, E., and Mishra, A. K. (2015). Role of manganese in protection against oxidative stress under iron starvation in cyanobacterium *Anabaena* 7120. *J. Basic Microbiol.* 55, 729–740. doi:10.1002/jobm.201400742.

Kerfeld, C. A., Melnicki, M. R., Sutter, M., and Dominguez-Martin, M. A. (2017). Structure, function and evolution of the cyanobacterial orange carotenoid protein and its homologs. *New Phytol.* 215, 937–951. doi:10.1111/nph.14670.

Kirilovsky, D., and Kerfeld, C. A. (2016). Cyanobacterial photoprotection by the orange carotenoid protein. *Nat. Plants* 2, 16180. doi:10.1038/nplants.2016.180.

Kizawa, A., Kawahara, A., Takimura, Y., Nishiyama, Y., and Hihara, Y. (2016). RNA-seq profiling reveals novel target genes of LexA in the cyanobacterium *Synechocystis* sp. PCC 6803. *Front. Microbiol.* 7, 193. doi:10.3389/fmicb.2016.00193.

Knoppová, J., Sobotka, R., Tichý, M., Yu, J., Konik, P., Halada, P., et al. (2014). Discovery of a chlorophyll binding protein complex involved in the early

steps of Photosystem II assembly in *Synechocystis*. *Plant Cell* 26, 1200–1212. doi:10.1105/tpc.114.123919.

Komenda, J., Knoppová, J., Kopečná, J., Sobotka, R., Halada, P., Yu, J., et al. (2012). The Psb27 assembly factor binds to the CP43 complex of photosystem II in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Physiol.* 158, 476–486. doi:10.1104/pp.111.184184.

Kopylova, E., Noé, L., and Touzet, H. (2012). SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28, 3211–3217. doi:10.1093/bioinformatics/bts611.

Kós, P. B., Deák, Z., Cheregi, O., and Vass, I. (2008). Differential regulation of *psbA* and *psbD* gene expression, and the role of the different D1 protein copies in the cyanobacterium *Thermosynechococcus elongatus* BP-1. *Biochim. Biophys. Acta Bioenerg.* 1777, 74–83. doi:10.1016/j.bbabio.2007.10.015.

Krasikov, V., Aguirre von Wobeser, E., Dekker, H. L., Huisman, J., and Matthijs, H. C. P. (2012). Time-series resolution of gradual nitrogen starvation and its impact on photosynthesis in the cyanobacterium *Synechocystis* PCC 6803. *Physiol. Plant.* 145, 426–439. doi:10.1111/j.1399-3054.2012.01585.x.

Kumura, K., Sekiguchi, M., Stemum, A. L., and Seeberg, E. (1985). Stimulation of the UvrABC enzyme-catalyzed repair reactions by the UvrD protein (DNA helicase II). *Nucleic Acids Res.* 13, 1483–1492. doi:10.1093/nar/13.5.1483.

Lambrecht, S. J., Kanesaki, Y., Fuss, J., Huettel, B., Reinhardt, R., and Steglich, C. (2019). Interplay and targetome of the two conserved cyanobacterial sRNAs Yfr1 and Yfr2 in *Prochlorococcus* MED4. *Sci. Rep.* 9, 14331. doi:10.1038/s41598-019-49881-9.

Langfelder, P., and Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559. doi:10.1186/1471-2105-9-559.

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Meth.* 9, 357–359. doi:10.1038/nmeth.1923.

Latifi, A., Ruiz, M., and Zhang, C. C. (2009). Oxidative stress in cyanobacteria. *FEMS Microbiol. Rev.* 33, 258–278. doi:10.1111/j.1574-6976.2008.00134.x.

Lindell, D., Jaffe, J. D., Coleman, M. L., Futschik, M. E., Axmann, I. M., Rector, T., et al. (2007). Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* 449, 83–86. doi:10.1038/nature06130.

Los, D. A., Zorina, A., Sinetova, M., Kryazhov, S., Mironov, K., and Zinchenko, V. V. (2010). Stress sensors and signal transducers in cyanobacteria. *Sensors* 10, 2386–2415. doi:10.3390/s100302386.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8.

Ludwig, M., and Bryant, D. A. (2012). *Synechococcus* sp. strain PCC 7002 transcriptome: Acclimation to temperature, salinity, oxidative stress, and mixotrophic growth conditions. *Front. Microbiol.* 3, 354. doi:10.3389/fmicb.2012.00354.

Mabbitt PD, Wilbanks SM, Eaton-Rye JJ. 2014 Structure and function of the

hydrophilic photosystem II assembly proteins: Psb27, Psb28 and Ycf48. Plant Physiol Biochem. 81:96-107. doi: 10.1016/j.plaphy.2014.02.013

Marie, D., Partensky, F., Vaulot, D., and Brussaard, C. (1999). Enumeration of phytoplankton, bacteria, and viruses in marine samples. *Curr. Protoc. Cytom.* 10, 11.11.1-11.11.15. doi:10.1002/0471142956.cy1111s10.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.J*. doi:https://doi.org/10.14806/ej.17.1.200.

Mazon, G., Philippin, G., Cadet, J., Gasparutto, D., and Fuchs, R. P. (2009). The alkyltransferase-like ybaZ gene product enhances nucleotide excision repair of O6-alkylguanine adducts in E. coli. *DNA Repair (Amst).* 8, 697–703. doi:10.1016/j.dnarep.2009.01.022.

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, 285. doi:10.3389/fmicb.2012.00285.

Mikami, K., Kanesaki, Y., Suzuki, I., and Murata, N. (2002). The histidine kinase Hik33 perceives osmotic stress and cold stress in *Synechocystis* sp. PCC 6803. *Mol. Microbiol.* 46, 905–915. doi:10.1046/j.1365-2958.2002.03202.x.

Mizusawa, N., and Wada, H. (2012). The role of lipids in photosystem II. *Biochim. Biophys. Acta Bioenerg.* 1817, 194–208. doi:10.1016/j.bbabio.2011.04.008.

Mostertz, J., Scharf, C., Hecker, M., and Homuth, G. (2004). Transcriptome and proteome analysis of *Bacillus subtilis* gene expression in response to superoxide and peroxide stress. *Microbiol. U.K.* 150, 497–512. doi:10.1099/mic.0.26665-0.

Nguyen, A. V., Thomas-Hall, S. R., Malnoë, A., Timmins, M., Mussgnug, J. H., Rupprecht, J., et al. (2008). Transcriptome for photobiological hydrogen production induced by sulfur deprivation in the green alga *Chlamydomonas reinhardtii*. *Eukaryot. Cell* 7, 1965–1979. doi:10.1128/EC.00418-07.

Nickelsen, J., and Rengstl, B. (2013). Photosystem II assembly: from cyanobacteria to plants. *Annu. Rev. Plant Biol.* 64, 609–635. doi:10.1146/annurev-arplant-050312-120124.

Nicolas, P., Mäder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., et al. (2012). Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science (80-.).* 335, 1103–1106. doi:10.1126/science.1206848.

Niedzwiedzki, D. M., Tronina, T., Liu, H., Staleva, H., Komenda, J., Sobotka, R., et al. (2016). Carotenoid-induced non-photochemical quenching in the cyanobacterial chlorophyll synthase-HliC/D complex. *Biochim. Biophys. Acta Bioenerg.* 1857, 1430–1439. doi:10.1016/j.bbabio.2016.04.280.

Nishiyama, Y., Allakhverdiev, S. I., and 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. doi:10.1016/j.bbabio.2006.05.013.

Nodop, A., Pietsch, D., Höcker, R., Becker, A., Pistorius, E. K., Forchhammer, K., et al. (2008). Transcript profiling reveals new insights into the acclimation of the

mesophilic fresh-water cyanobacterium *Synechococcus elongatus* PCC 7942 to iron starvation. *Plant Physiol.* 147, 747–763. doi:10.1104/pp.107.114058.

Ossenbühl, F., Inaba-Sulpice, M., Meurer, J., Soll, J., and Eichacker, L. A. (2006). The *Synechocystis* sp. PCC 6803 Oxa1 homolog is essential for membrane integration of reaction center precursor protein pD1. *Plant Cell* 18, 2236–2246. doi:10.1105/tpc.106.043646.

Owttrim, G. W. (2006). RNA helicases and abiotic stress. *Nucleic Acids Res.* 34, 3220–3230. doi:10.1093/nar/gkl408.

Pade, N., and Hagemann, M. (2015). Salt acclimation of cyanobacteria and their application in biotechnology. *Life* 5, 25–49. doi:10.3390/life5010025.

Patel, M., Jiang, Q., Woodgate, R., Cox, M. M., and Goodman, M. F. (2010). A new model for SOS-induced mutagenesis: How RecA protein activates DNA polymerase V. *Crit. Rev. Biochem. Mol. Biol.* 45, 171–184. doi:10.3109/10409238.2010.480968.

Paulsen, M. L., 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, 191. doi:10.3389/fmars.2016.00191.

Perelman, A., Uzan, A., Hacohen, D., and Schwarz, R. (2003). Oxidative stress in *Synechococcus* sp. strain PCC 7942: Various mechanisms for H2O2 detoxification with different physiological roles. *J. Bacteriol.* 185, 3654–3660. doi:10.1128/JB.185.12.3654-3660.2003.

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.

Pittera, J., Jouhet, J., Breton, S., Garczarek, L., Partensky, F., Maréchal, É., et al. (2018). Thermoacclimation and genome adaptation of the membrane lipidome in marine *Synechococcus*. *Environ. Microbiol.* 20, 612–631. doi:10.1111/1462-2920.13985.

Pittera, J., Partensky, F., and Six, C. (2017). Adaptive thermostability of light-harvesting complexes in marine picocyanobacteria. *ISME J.* 11, 112–124. doi:10.1038/ismej.2016.102.

Plöchinger, M., Schwenkert, S., Von Sydow, L., Schröder, W. P., and Meurer, J. (2016). Functional update of the auxiliary proteins PsbW, PsbY, HCF136, PsbN, TerC and ALB3 in maintenance and assembly of PSII. *Front. Plant Sci.* 7, 423. doi:10.3389/fpls.2016.00423.

Reistetter, E. N., Krumhardt, K., Callnan, K., Roache-Johnson, K., Saunders, J. K., Moore, L. R., et al. (2013). Effects of phosphorus starvation versus limitation on the marine cyanobacterium *Prochlorococcus* MED4. II: Gene expression. *Environ. Microbiol.* 15, 2129–2143. doi:10.1111/1462-2920.12129.

Rippka, R., Coursin, T., Hess, W., Lichtlé, C., Scanlan, D. J., Palinska, K. A., et al. (2000). *Prochlorococcus marinus* Chisholm et al. 1992 subsp. *pastoris* subsp. nov. strain PCC 9511, the first axenic chlorophyll *a*2/*b*2-containing cyanobacterium (Oxyphotobacteria). *Intl J. Syst. Bacteriol J. Syst. Evol. Microbiol.* 50, 1833–1847. doi:10.1099/00207713-50-5-1833.

Rodionov, D. A., Mironov, A. A., and Gelfand, M. S. (2002). Conservation of the biotin regulon and the BirA regulatory signal in eubacteria and archaea. *Genome*

Res. 12, 1507–1516. doi:10.1101/gr.314502.

Sancar, A., and Rupp, W. D. (1983). A novel repair enzyme: UVRABC excision nuclease of *Escherichia coli* cuts a DNA strand on both sides of the damaged region. *Cell* 33, 249–260. doi:10.1016/0092-8674(83)90354-9.

Sánchez-Riego, A. M., Mata-Cabana, A., Galmozzi, C. V., and Florencio, F. J. (2016). NADPH-thioredoxin reductase C mediates the response to oxidative stress and thermotolerance in the cyanobacterium *Anabaena* sp. PCC7120. *Front. Microbiol.* 7, 1283. doi:10.3389/fmicb.2016.01283.

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.

Schelin, J., Lindmark, F., and Clarke, A. K. (2002). The clpP multigene family for the ATP-dependent Clp protease in the cyanobacterium *Synechococcus*. *Microbiol. U.K.* 148, 2255–2265. doi:10.1099/00221287-148-7-2255.

Scherzinger, D., Ruch, S., Kloer, D. P., Wilde, A., and Al-Babili, S. (2006). Retinal is formed from apo-carotenoids in *Nostoc* sp. PCC7120: In vitro characterization of an apo-carotenoid oxygenase. *Biochem. J.* 398, 361–369. doi:10.1042/BJ20060592.

Schimel, J., Balser, T. C., and Wallenstein, M. (2007). Microbial stress-response physiology and its implications for ecosystem function. *Ecology* 88, 1386–1394. doi:10.1890/06-0219.

Schirmer, A., Rude, M. A., Li, X., Popova, E., and Del Cardayre, S. B. (2010). Microbial biosynthesis of alkanes. *Science (80-.).* 329, 559–562. doi:10.1126/science.1187936.

Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27, 863–864. doi:10.1093/bioinformatics/btr026.

Schwarz, D., Nodop, A., Hüge, J., Purfürst, S., Forchhammer, K., Michel, K. P., et al. (2011). Metabolic and transcriptomic phenotyping of inorganic carbon acclimation in the cyanobacterium *Synechococcus elongatus* PCC 7942. *Plant Physiol.* 155, 1640–1655. doi:10.1104/pp.110.170225.

Shrivastava, A. K., Pandey, S., Yadav, S., Mishra, Y., Singh, P. K., Rai, R., et al. (2016). Comparative proteomics of wild type, An+ahpC and AnΔahpC strains of *Anabaena* sp. PCC7120 demonstrates AhpC mediated augmentation of photosynthesis, N2-fixation and modulation of regulatory network of antioxidative proteins. *J. Proteomics* 140, 81–99. doi:10.1016/j.jprot.2016.04.004.

Shukla, M. K., Llansola-Portoles, M. J., Tichý, M., Pascal, A. A., Robert, B., and Sobotka, R. (2018). Binding of pigments to the cyanobacterial high-light-inducible protein HliC. *Photosynth. Res.* 137, 29–39. doi:10.1007/s11120-017-0475-7.

Singh, A. K., Li, H., and Sherman, L. A. (2004). Microarray analysis and redox control of gene expression in the cyanobacterium *Synechocystis* sp. PCC 6803. *Physiol. Plant.* 120, 27–35. doi:10.1111/j.0031-9317.2004.0232.x.

Singh, A. K., McIntyre, L. M., and Sherman, L. A. (2003). Microarray analysis of the genome-wide response to iron deficiency and iron reconstitution in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Physiol.* 132, 1825–1839. doi:10.1104/pp.103.024018.

Six, C., Thomas, J.-C., Brahamsha, B.,

Lemoine, Y., and Partensky, F. (2004). Photophysiology of the marine cyanobacterium *Synechococcus* sp. WH8102, a new model organism. *Aquat. Microb. Ecol.* 35, 17–29. doi:10.3354/ame035017.

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. doi:10.1186/gb-2007-8-12-r259.

Stazic, D., Pekarski, I., Kopf, M., Lindell, D., and Steglich, C. (2016). A novel strategy for exploitation of host RNase E activity by a marine cyanophage. *Genetics* 203, 1149–1159. doi:10.1534/genetics.115.183475.

Stork, T., Michel, K. P., Pistorius, E. K., and Dietz, K. J. (2005). Bioinformatic analysis of the genomes of the cyanobacteria *Synechocystis* sp. PCC 6803 and *Synechococcus elongatus* PCC 7942 for the presence of peroxiredoxins and their transcript regulation under stress. *J.Exp. Bot.* 56, 3193–3206. doi:10.1093/jxb/eri316.

Suzuki, I., Kanesaki, Y., Mikami, K., Kanehisa, M., and Murata, N. (2001). Cold-regulated genes under control of the cold sensor Hik33 in *Synechocystis*. *Mol. Microbiol.* 40, 235–244. doi:10.1046/j.1365-2958.2001.02379.x.

Suzuki, S., Ferjani, A., Suzuki, I., and Murata, N. (2004). The SphS-SphR two-component system is the exclusive sensor for the induction of gene expression in response to phosphate limitation in *Synechocystis*. *J. Biol. Chem.* 279, 13234–13240. doi:10.1074/jbc.M313358200.

Taboada, B., Ciria, R., Martinez-Guerrero, C. E., and Merino, E. (2012). ProOpDB: Prokaryotic operon database. *Nucleic Acids Res.* 40, D627-31. doi:10.1093/nar/gkr1020.

Takahashi, H., Uchimiya, H., and Hihara, Y. (2008). Difference in metabolite levels between photoautotrophic and photomixotrophic cultures of *Synechocystis* sp. PCC 6803 examined by capillary electrophoresis electrospray ionization mass spectrometry. *J. Exp. Bot.* 59, 3009–3018. doi:10.1093/jxb/ern157.

Takahashi, S., and Murata, N. (2008). How do environmental stresses accelerate photoinhibition? *Trends Plant Sci.* 13, 178–182. doi:10.1016/j.tplants.2008.01.005.

Telfer, A. (2005). Too much light? How beta-carotene protects the photosystem II reaction centre. *Photochem Photobiol Sci* 4, 950–956. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16307107.

Tetu, S. G., Brahamsha, B., Johnson, D. A., 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. doi:10.1038/ismej.2009.31.

Tetu, S. G., Johnson, D. A., Varkey, D., Phillippy, K., Stuart, R. K., Dupont, C. L., et al. (2013). Impact of DNA damaging agents on genome-wide transcriptional profiles in two marine *Synechococcus* species. *Front Microbiol* 4, 232. doi:10.3389/fmicb.2013.00232.

Thompson, A. W., Huang, K., Saito, M. A., and Chisholm, S. W. (2011). Transcriptome response of high- and low-light-adapted *Prochlorococcus* strains to changing iron availability. *ISME J.* 5, 1580–1594. doi:10.1038/ismej.2011.49.

Thurotte, A., Bourcier de Carbon, C., Wilson, A., Talbot, L., Cot, S., López-Igual, R., et al. (2017). The cyanobacterial Fluorescence Recovery Protein has two distinct activities: Orange Carotenoid Protein amino acids involved in FRP interaction. *Biochim. Biophys. Acta Bioenerg.* 1858, 308–317. doi:10.1016/j.bbabio.2017.02.003.

Tirupati, B., Vey, J. L., Drennan, C. L., and Bollinger, J. M. (2004). Kinetic and structural characterization of Slr0077/SufS, the essential cysteine desulfurase from *Synechocystis* sp. PCC 6803. *Biochemistry* 43, 12210–12219. doi:10.1021/bi0491447.

Tolonen, A. C., Aach, J., Lindell, D., Johnson, Z. I., Rector, T., Steen, R., et al. (2006). Global gene expression of *Prochlorococcus* ecotypes in response to changes in nitrogen availability. *Mol. Syst. Biol.* 2, 53. doi:10.1038/msb4100087.

Varet, H., Brillet-Guéguen, L., Coppée, J. Y., and Dillies, M. A. (2016). SARTools: A DESeq2- and edgeR-based R pipeline for comprehensive differential analysis of RNA-Seq data. *PLoS One* 11, e0157022. doi:10.1371/journal.pone.0157022.

Varkey, D., Mazard, S., Ostrowski, M., Tetu, S. G., Haynes, P., and Paulsen, I. T. (2016). Effects of low temperature on tropical and temperate isolates of marine *Synechococcus*. *ISME J.* 10, 1252–1263. doi:10.1038/ismej.2015.179.

Viczián, A., Máté, Z., Nagy, F., and Vass, I. (2000). UV-b induced differential transcription of psbD genes encoding the D2 protein of Photosystem II in the cyanobacterium *Synechocystis* 6803. *Photosynth. Res.* 64, 257–266. doi:10.1023/A:1006444932137.

Voigt, K., Sharma, C. M., Mitschke, J., Joke Lambrecht, S., Voss, B., Hess, W. R., et al. (2014). Comparative transcriptomics of two environmentally relevant cyanobacteria reveals unexpected transcriptome diversity. *ISME J* 8, 2056–2068. doi:10.1038/ismej.2014.57.

Wang, S., Deng, K., Zaremba, S., Deng, X., Lin, C., Wang, Q., et al. (2009). Transcriptomic response of *Escherichia coli* O157:H7 to oxidative stress. *Appl. Environ. Microbiol.* 75, 6110–6123. doi:10.1128/AEM.00914-09.

Wu, Y. P., and Krogmann, D. W. (1997). The orange carotenoid protein of *Synechocystis* PCC 6803. *Biochim. Biophys. Acta Bioenerg.* 1322, 1–7. doi:10.1016/S0005-2728(97)00067-4.

Xia, X., Vidyarathna, N. K., Palenik, B., and Lee, P. (2015). Comparison of the seasonal variations of *Synechococcus* assemblage structures in estuarine waters and coastal waters of Hong Kong. 81, 7644–7655. doi:10.1128/AEM.01895-15.Editor.

Zinser, E. R., Lindell, D., Johnson, Z. I., Futschik, M. E., Steglich, C., Coleman, M. L., et al. (2009). Choreography of the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph, *Prochlorococcus*. *PLoS One* 4, e5135. doi:10.1371/journal.pone.0005135.

Table 1: Description of the experimental conditions

For shift experiments, cells were acclimated to medium temperature (22°C) and either low light (LL, 25 µE.m^{-2}.s^{-1}) or high light (HL, 250 µE.m^{-2}.s^{-1}) conditions and submitted to low temperature (LT, 13°C), high temperature (HT, 30°C) or ultraviolet radiations (UV-A, 3 W.m^{-2}; UV-B, 0.3 W.m^{-2}). Sampling times between brackets correspond to recovery time points and were not used for transcriptomic analyses. For 12h/12h L/D cycle experiments, sampling times correspond to the time of the day, with light period starting at 6:00 and dark period at 18:00 and the four biological replicates include two replicate cultures and two consecutive days for each temperature condition.

Acclimation condition	Stress condition	Sampling times	Number of biological replicates
LL	HL	0, 0.3, 1, 3, 6, (7, 24)	3
LL	LLUV	0, 0.3, 1, 3, 6, (7, 24)	3
LL	LLLT	0, 12, 24, 36, 48	3
LL	LLHT	0, 12, 24, 36, 48	3
HL	HLUV	0, 0.3, 1, 3, 6, (7, 24)	3
HL	HLLT	0, 4, 8, 12, 16, 24	3
HL	HLHT	0, 4, 8, 12, 16, 24	3
L/D 21°C	n.a.	6, 9, 12, 15, 18, 20, 22, 2	4
L/D 27°C	n.a.	6, 9, 12, 15, 18, 20, 22, 2	4

Supplementary Material

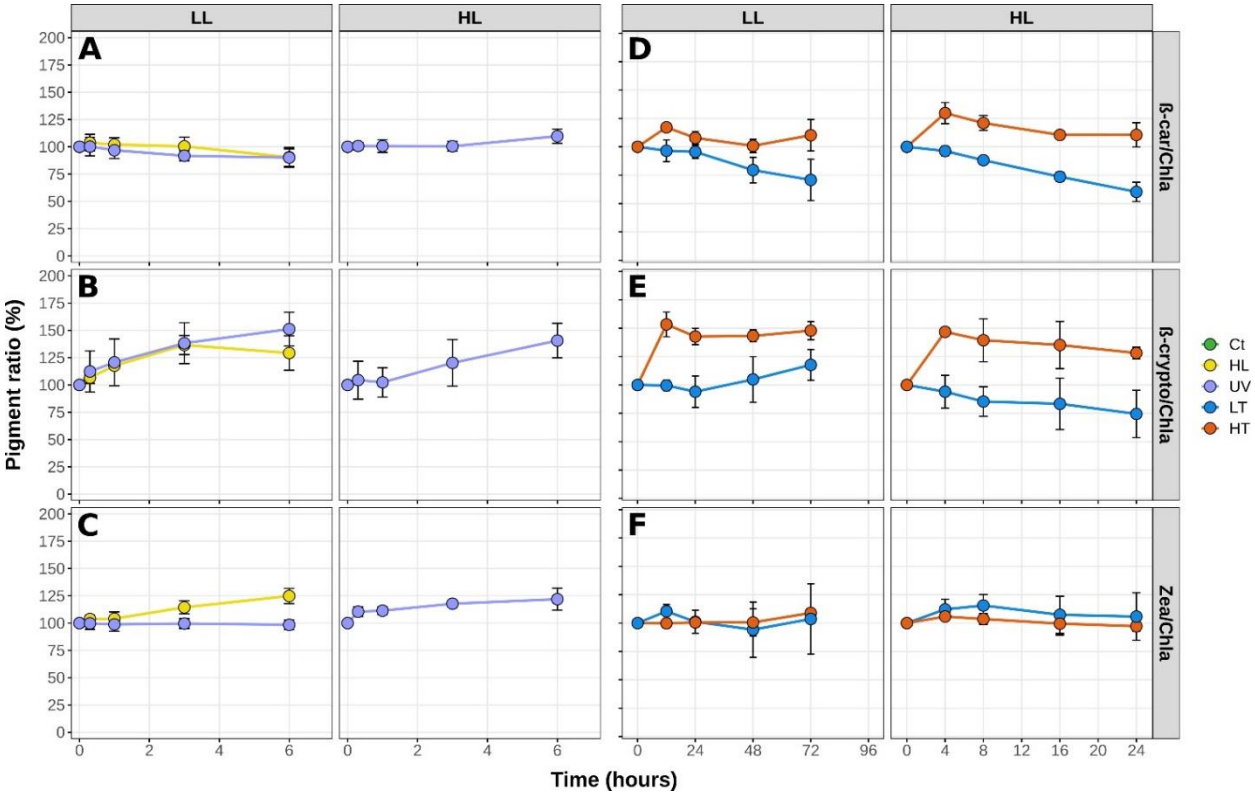


FIGURE S1 | Variations of the major liposoluble pigment ratios in Synechococcus sp. WH7803 acclimated to low (LL, 25 µE.m-2.s-1 or high light (HL, 250 µE.m-2.s-1). (**A,D**) β-carotene to chlorophyll (chl) a, (**B, E**) β-cryptoxanthin to chl a and (**C, F**) Zeaxanthin (zea) to chl a ratios in response to shifts to (**A-C**) HL, ultraviolet radiations (UV-A 3 W.m-2, UV-B 0.3 W.m-2), (**D-F**) low temperature (LT: shift from 22°C to 13°C) or high temperature (HT; shift from 22°C to 30°C). The top grey banners indicate the light acclimation condition. These data are based on at least three independent experiments.

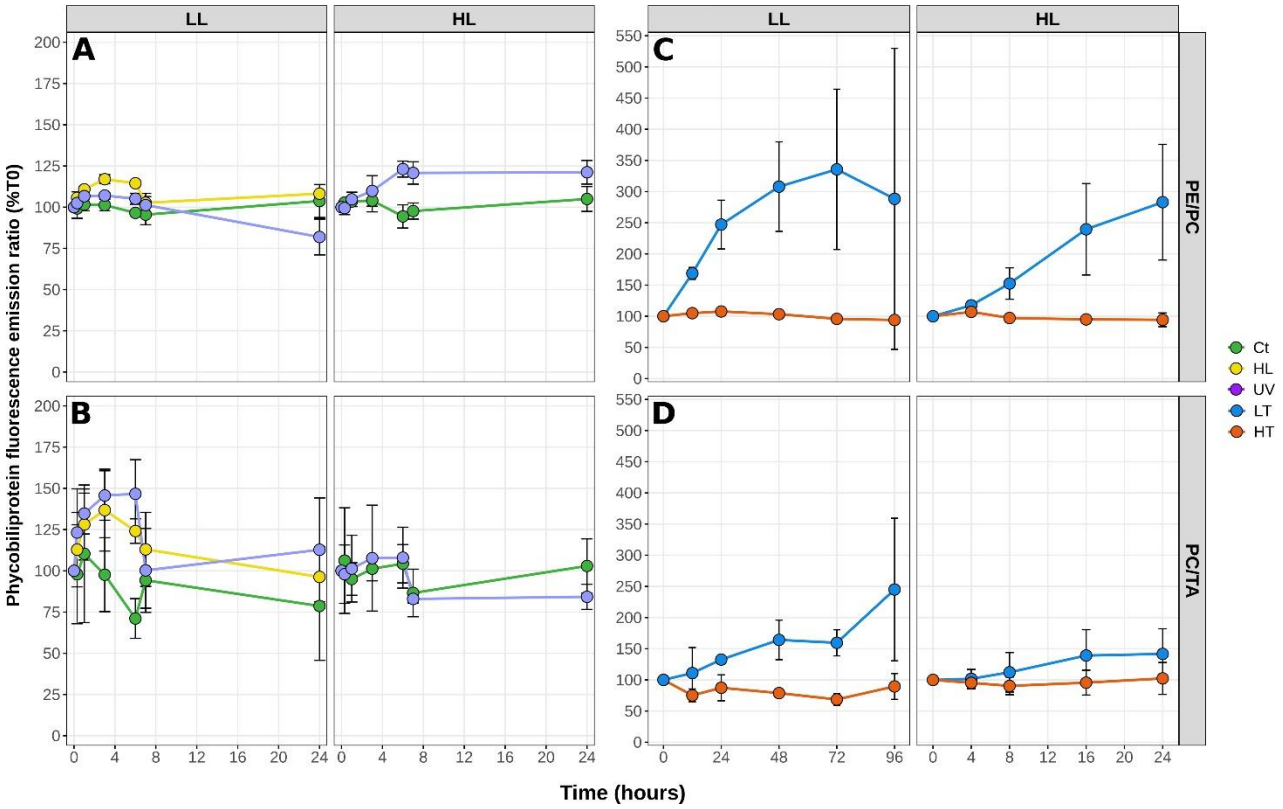


FIGURE S2 | Variations of phycobiliprotein fluorescence emission ratio in Synechococcus sp. WH7803 acclimated to low (LL, 25 µE.m-2.s-1 or high light (HL, 250 µE.m-2.s-1). (**A,C**) Phycoerythrin (PE) to phycocyanin (PC) ratio after a shift (**A**) to high light or ultraviolet radiations (UV-A 3 W.m-2, UV-B 0.3 W.m-2), and (**C**) to low temperature (LT: shift from 22°C to 13°C) or high temperature (HT; shift from 22°C to 30°C). (**B,D**) Same for the PC to terminal acceptor (TA) ratio. The dashed line indicates the time (6h) at which cultures submitted to light stresses were shifted back to their initial light conditions to study recovery. The top grey banners indicate the culture acclimation condition. These data are based on at least three independent experiments. Other abbreviation: Ct, control.

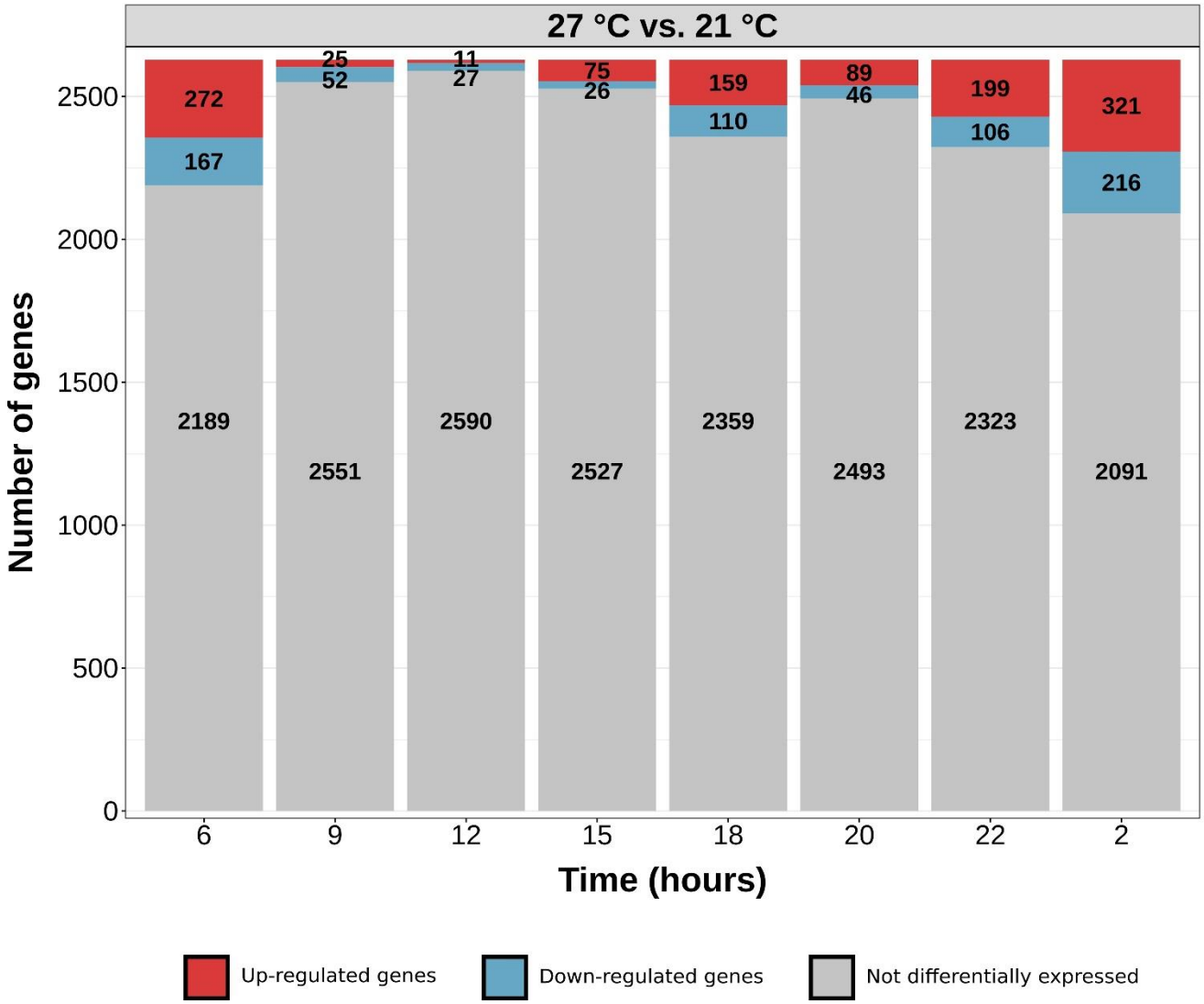


FIGURE S3 | Number of differentially expressed (DE) genes in Synechococcus sp. WH7803 at 27°C as compared to 21°C. The top white and black banners indicate light and dark periods. Data are based on four biological replicates. Abbreviation in x-axis correspond to 'sampling time'. Red and blue colors indicate up- and down-regulated genes, respectively.

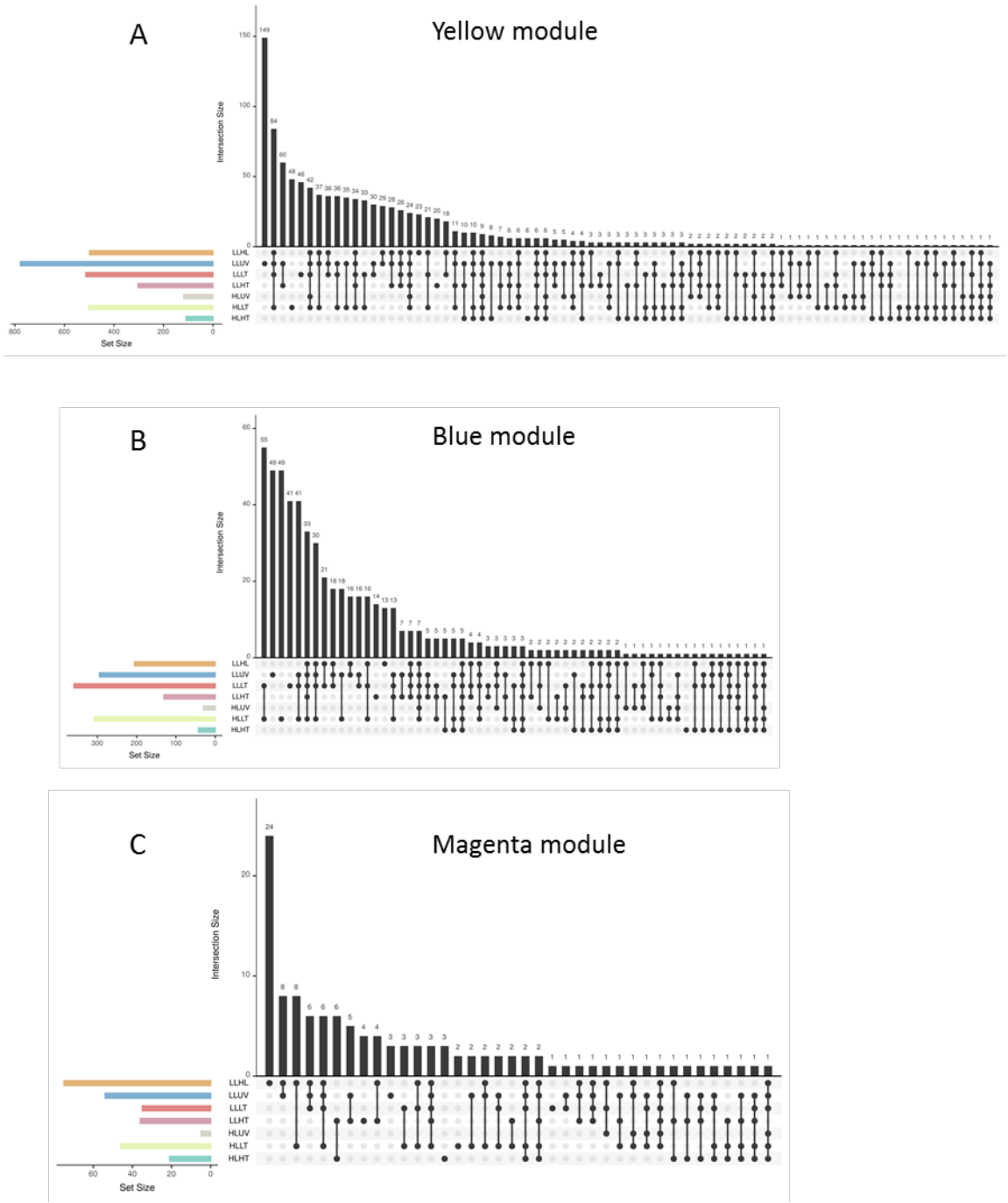


FIGURE S4 | Distribution of genes in the different shift conditions for the three modules as determined using the UpSet R package. Modules (**A**) Yellow, (**B**) Blue, (**C**) Magenta. Only genes with a |Log2FC| ≥ 1 and padj ≤ 0.05 in at least one condition have been taken into account.

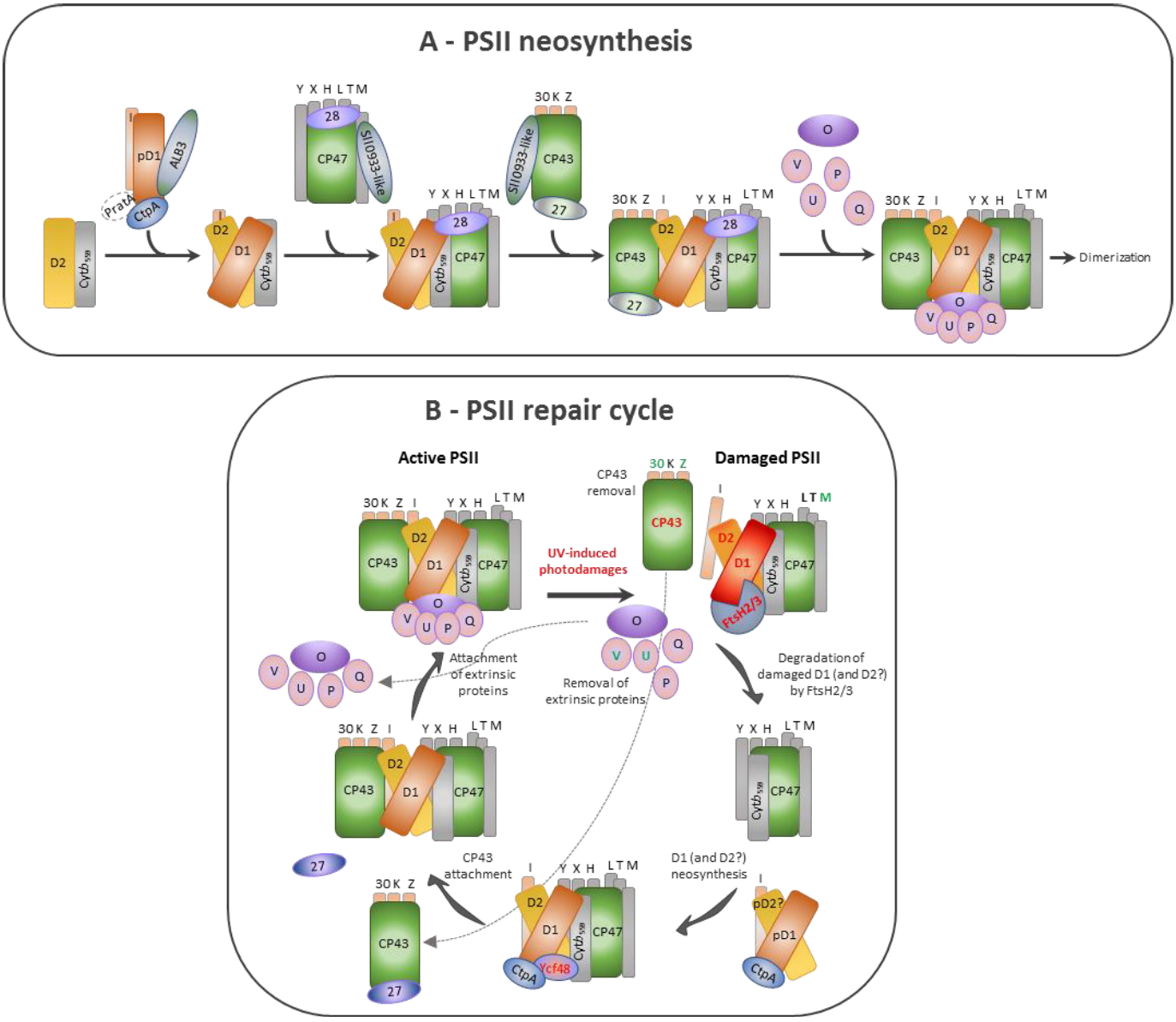


FIGURE S5 | Photosystem II neosynthesis and repair. A. Neosynthesis occurs in a stepwise manner and requires several assembly factors (ALB3, Psb28, Psb27, Sll0933-like protein) as well as the D1 protein carboxyl-terminal processing peptidase CtpA (ovals), which transiently interact to mediate the various PSII assembly steps. Note that Synechococcus sp. WH7803 possesses an homolog of Synechocystis sp. PCC 6803 Sll0933 (CK_00000788) but no homolog of PratA, so the latter factor is probably absent, as indicated by the dotted oval. **B.** Repair after UV photodamage. When the active photosystem is photodamaged the Mn4CaO5 cluster (not shown) and the extrinsic proteins (PsbO, PsbU, PsbV, CyanoP and CyanoQ) as well as the CP43 complex leave the complex. The damaged D1 (and D2?) protein(s) is then removed by the FtsH2/3 hetero-oligomeric complex. The Ycf48 protein binds to PS II, whilst the CtpA protease processes the carboxyl terminal extension of pD1. The Psb27 protein associates with CP43, which attaches to PSII after the replacement D1 (and D2?) proteins. Psb27 leaves the complex and the Mn4CaO5 cluster and extrinsic proteins in turn bind to the PSII monomer, before its dimerization (not shown). PSII subunits or assembly factors, which genes were found to be significantly upregulated (log2FC ≥ 1) or downregulated (log2FC ≤ -1) in response to the LLUV treatment (see Table S1), are indicated in red and green, respectively, while others are shown in black. PSII subunits are only designated by letters or numbers at the ends of their respective Psb protein names (e.g., "I" and "30" for PsbI and Psb30, respectively), except for the core PSII subunits, which are designated by their common names (D1, D2, CP43 and CP47). pD1 and pD2 are the precursors forms of D1 and D2, respectively. Note that D2 and pD2 are indicated with a question mark because it is usually considered that only D1 is actively replaced during the PSII repair process. Adapted from Nickelsen and Rengstl (2013) and Mabbitt et al. (2014).

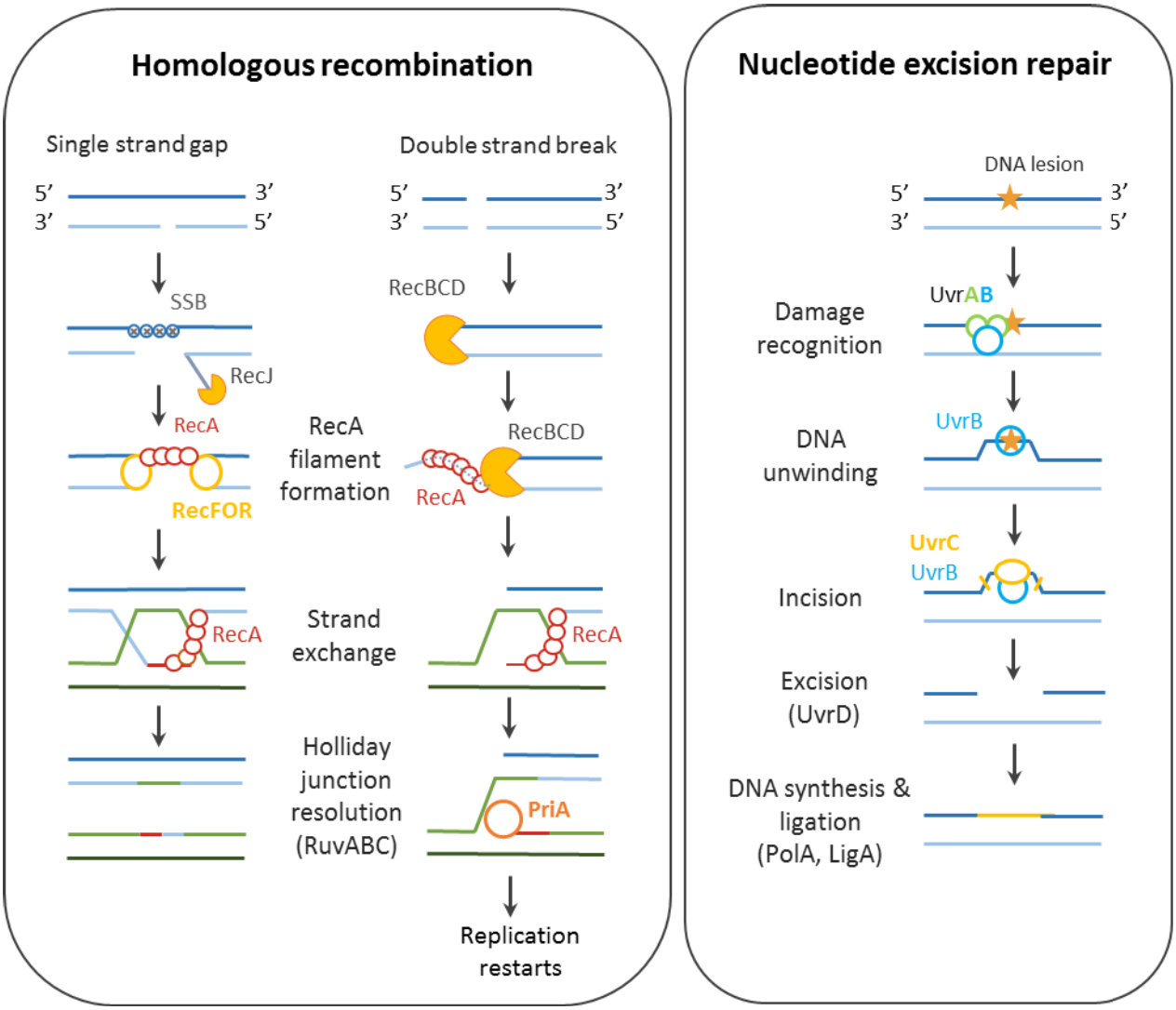


FIGURE S6 | Representations of the main DNA repair pathways induced during the SOS response. Yellow stars represent DNA lesions. The full description of each protein is available in Table S1.

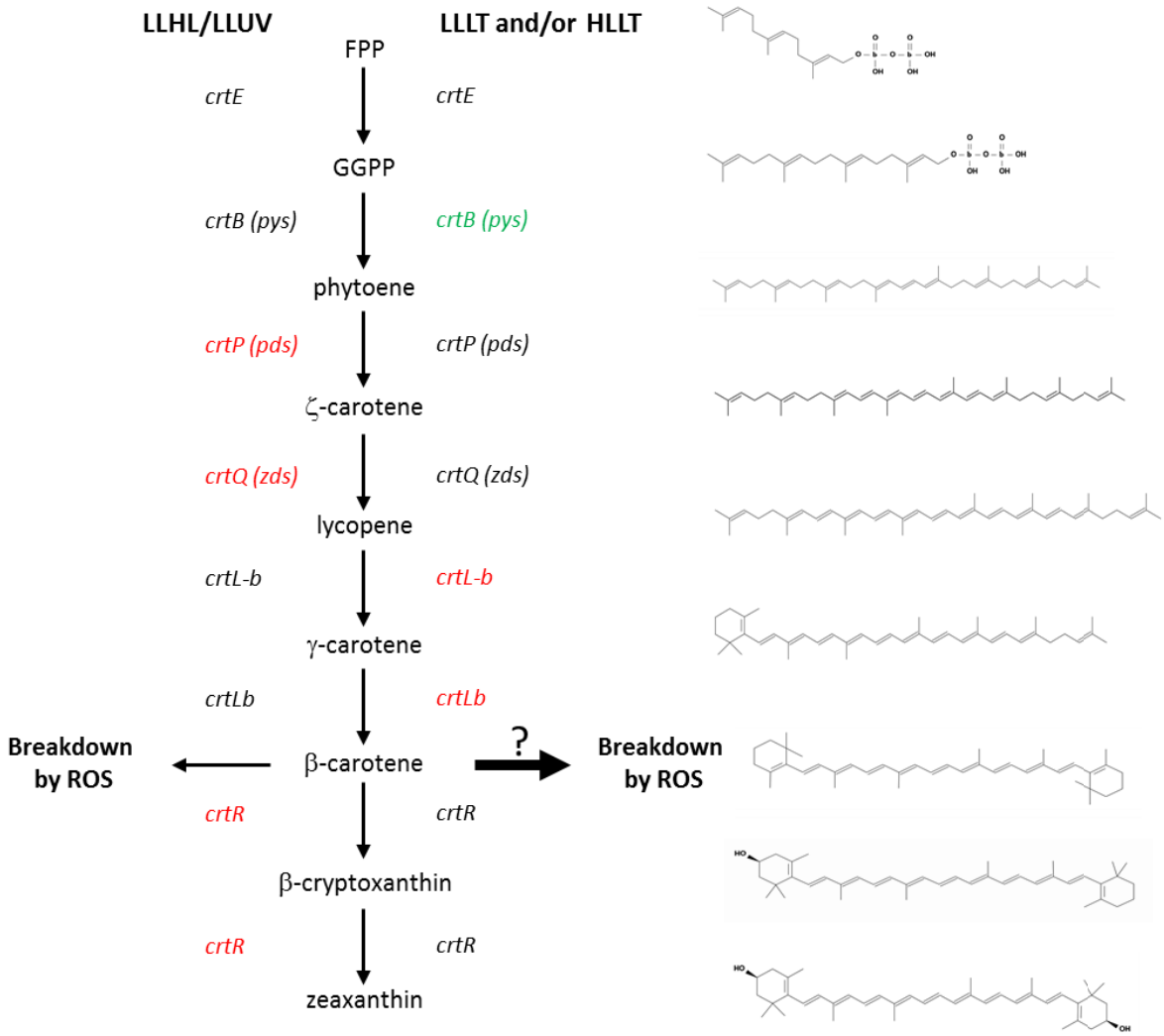


FIGURE S7 | Expression of Synechococcus sp. WH7803 genes involved in carotenoid biosynthesis in either LLHL and LLUV (left) or LLLT and/or HLLT (right). Upregulated genes are indicated in red, downregulated ones are in green, while not significantly differentially expressed genes are in black. Product abbreviations: GGPP, Geranylgeranyl pyrophosphate; FPP, Farnesyl pyrophosphate. Enzyme names: crtE, GGPP synthase, crtB, phytoene synthase; crtP, phytoene dehydrogenase; crtQ, 9 9'-di-cis-zeta-carotene desaturase; crtL-b, lycopene beta cyclase; crtR, beta-carotene hydroxylase.

TABLE S1 | Inventory and description of *Synechococcus* sp. WH7803 genes and their expression patterns in the different tested conditions: HL (yellow background), UV (purple background), LT (blue background), HT (red background), L/D cycle 21°C vs. 6 h (light green), L/D cycle 27°C vs. 6 h (medium green), L/D cycle 27 vs. 21°C (dark green). Gene expression values are shown as heatmap with in red the upregulated genes and in green, the downregulated genes. The first column corresponds to the clustering based on expression values within each WGCNA module. Abbreviations: CK, Cyanorak; WGCNA, Weighted correlation network analysis; PC, phycocyanin; PE, phycoerythrin; TA, terminal acceptor; F_V/F_M, PSII maximal photochemical yield ; β-car, β-carotene; chl, chlorophyll; β-crypto, β-cryptoxanthin; zea, zeaxanthin; MGDG, monogalactosyldiacylglycerol; DGDG, digalactosyldiacylglycerol; SQDG, sulfoquinovosyldiacylglycerol; PG, phosphatidylglycerol; VIP, value importance in projection; kTotal, connectivity of each gene within the WGCNA network; kWithin, connectivity of each gene within its WGCNA module; kOut, connectivity of each gene outside its WGCNA module; kDiff, difference between kWithin and kOut.

TABLE S2 | Description of the transcriptomic samples analyzed in this study

TABLE S3 | Available processed data including normalized counts, log2FC and their associated pvalues and adjusted pvalues. Samples analysed in each file are described in Table S2.

TABLE S4 |Accession numbers of the raw transcriptomic data used in this study. *Please note that the accession numbers were not yet available at the time of the submission of the manuscript and will be added in the revised version.*

CHAPITRE III

Variabilité écotypique des mécanismes d'adaptation et d'acclimatation aux variations des conditions environnementales chez les picocyanobactéries marines

Contexte de l'étude

Les études concernant la biogéographie des picocyanobactéries ont permis de distinguer des communautés de *Synechococcus* et *Prochlorococcus* présentant différents *preferenda* vis-à-vis des facteurs abiotiques et occupant des niches écologiques distinctes. Dans ce contexte, Farrant et al. (2016) ont pu distinguer 8 communautés majeures de *Synechococcus* et 3 de *Prochlorococcus*, chacune dominée par une ou plusieurs ESTUs. Néanmoins, comme nous avons pu le voir dans le chapitre I, la comparaison du contenu en gènes de l'ensemble des génomes disponibles, même avec plusieurs génomes par clade et/ou par ESTU, n'a permis de mettre en évidence que peu de gènes spécifiques d'écotypes et/ou de niches, susceptibles d'être impliqués dans les mécanismes d'adaptation à divers facteurs environnementaux. Cette observation pourrait être en partie dû à un bruit de fond important dans les comparaisons génomiques, du fait que de nombreux gènes accessoires ont été acquis relativement récemment et que ceux ayant peu ou pas d'effets bénéfiques pour l'adaptation à une niche donnée n'ont pas encore été purgés par la sélection naturelle. Une solution alternative pour identifier les gènes potentiellement impliqués dans l'adaptation aux divers facteurs environnementaux caractéristiques d'une niche donnée consiste à identifier les gènes spécifiquement présents ou absents dans cette niche, directement *in situ*. Dans ce contexte, la première étude présentée dans ce chapitre vise à analyser la distribution des gènes de picocyanobactéries à partir des métagénomes de *Tara* Oceans, en utilisant comme référence les 97 génomes de Cyanorak v2.1. Cette approche de recrutement de lectures des metagénomes a été appliquée, d'une part pour étudier sans à priori la distribution de l'ensemble des gènes de ces organismes dans les différentes niches (Doré et al., en prép.), et d'autre part en ciblant un ou plusieurs gènes d'intérêt, correspondant à une fonction donnée, comme cela a été le cas pour les gènes de désaturases afin de mieux comprendre les bases physiologiques de la spécialisation des différents thermotypes (ou écotypes thermiques) chez *Synechococcus* (Breton et al., 2019 en ANNEXES1).

Enfin, afin de mieux comprendre les mécanismes d'acclimatation des différents écotypes de *Synechococcus* aux variations des facteurs environnementaux, il est nécessaire de comparer la physiologie de souches représentatives de ces écotypes. Au cours du chapitre II, l'étude de l'impact de divers stress environnementaux directement sur la physiologie de la souche modèle *Synechococcus* sp. WH7803 a permis de mettre en évidence des réponses communes et spécifiques à ces différents stress. Néanmoins, cette souche appartient au clade V, dont les membres ne sont que peu abondants en milieu océanique. Pour cette raison, la dernière étude présentée dans ce chapitre a donc visé à comparer la physiologie de souches représentatives des quatre clades les plus abondants dans l'environnement (clades I à IV), soumises à des stress haute lumière et basse température, et à analyser leur transcriptome afin d'identifier i) les réponses communes aux différents stress et/ou spécifiques d'un stress donné et ii) les réponses communes à l'ensemble des souches et/ou spécifiques d'une souche donnée, ces dernières pouvant

potentiellement être reliées aux différentes niches écologiques occupées par ces écotypes. Ces résultats très préliminaires ont ensuite été en partie replacés dans un contexte environnemental en les comparant aux résultats des analyses des métatranscriptomes de *Tara* Oceans.

Contribution :

Dans la première étude décrivant la recherche de gènes de niches *in situ*, j'ai participé au recrutement sur les génomes de référence des lectures de picocyanobactéries à partir de l'ensemble des métagénomes de *Tara* Oceans ainsi qu'aux analyses des résultats obtenus. J'ai également réalisé l'intégralité des analyses de la distribution des gènes de désaturases ayant servi à l'article Breton et al. (2019), qui porte sur le rôle de ces enzymes intervenant dans la modulation de la fluidité des membranes photosynthétiques, dans l'adaptation à la température des différents thermotypes de *Synechococcus*. Dans la troisième étude de ce chapitre, j'ai adapté le pipeline qui avait servi pour l'analyse des données de transcriptomique de WH7803 afin d'analyser les 108 transcriptomes des souches représentatives des clades I-IV, puis j'ai interprété les données d'expression différentielle ainsi que les données de photophysiologie comparative liées à cette expérience. Enfin j'ai réalisé l'intégralité des analyses des métatranscriptomes de picocyanobactéries de l'expédition Tara *Oceans*.

III.1 Global distribution of *Synechococcus* and *Prochlorococcus* gene repertoires reveals adaptive strategies of picocyanobacterial communities

Introduction

Marine picocyanobacteria of the *Prochlorococcus* and *Synechococcus* genera are the two most abundant members of the phytoplanktonic communities and have a key role in the biological cycles of the major elements (Partensky *et al.,* 1999, Scanlan *et al*. 2009). Although these ubiquitous organisms usually co-occur in most warm and temperate environments, *Prochlorococcus* is restricted to the 40°S-50°N latitudinal band, while *Synechococcus* distribution extends from the equator to subpolar waters (Flombaum *et al*., 2013). Their ability to colonize such a wide range of ecological niches is seemingly related to their large genetic and genomic diversity with three deep branching genetic groups within *Synechococcus*, called sub-clusters (hereafter SC) 5.1 to 5.3, SC 5.1 being by far the most abundant in oceanic waters but also the most diversified (22 clades and about 30 sub-clades), while the monophyletic *Prochlorococcus* lineage is divided into two major phototypes, constituting about 11 clades (Ahlgren and Rocap, 2012; Kent et al., 2019; Mazard et al., 2012a; Sohm et al., 2016; Zwirglmaier et al., 2008). The delineation of Ecologically Significant Taxonomic Units (ESTUs), i.e., genetic groups within clades occupying a given ecological niche, at the global scale using *Tara* oceans metagenomic data has notably allowed one to discriminate distinct ESTUs assemblages for each of the two genera, each dominated by one or a few ESTUs and displaying distinct *preferenda* with regard to abiotic factors (Farrant *et al*., 2016). In the case of *Prochlorococcus*, three major assemblages were identified in surface waters, whose distribution was found to be mainly driven by temperature and iron availability: (i) the first one dominated by HLIA ESTU in temperate waters (above 35°N and 32°S), (ii) the second one dominated by HLIIA in warm, iron-replete waters located between 30°S and 30°N, and (iii) cooccurrence of HLIIIA and IVA in warm, iron-poor, high nutrients-low chlorophyll (HNLC) areas of the word ocean. As for *Synechococcus*, eight distinct ESTUs assemblages were identified along the *Tara* Oceans transect depending on three main environmental parameters (temperature, iron and phosphate availability): i) assemblage 1, dominated by ESTU IIA was by far the most abundant in warm, iron-replete, nitrogen-depleted environments, ii) assemblage 2, co-dominated by ESTUs IIIA, 5.3A and WPC1A were present in warm, P-depleted areas such as the Mediterranean Sea, iii) assemblages 4, 5 and 8, where ESTUs IA, IVA and IIB were the major contributors in nutrient rich, cold and mixed waters at high latitude and at last iv) assemblages 6 and 7, co-dominated by members of CRD1 and EnvB clades in iron-depleted waters but gathering distinct ESTUs occupying different thermal niches. These include CRD1B/EnvBB in cold waters, CRD1C/EnvBC in warm waters and CRD1A and EnvBA in both thermal niches, suggesting that members of the latter ESTUs would be able to colonize a much larger range of temperature than the two others (Farrant *et al*., 2016).

Several genomics studies have been conducted in order to decipher the genetic bases of these niche specificities (Rocap *et al*., 2003; Kettler *et al*., 2007, Dufresne *et al*., 2008, Scanlan *et al*., 2009; Doré *et al*., submitted). Still, relatively few genes were identified as being specific of the different ecotypes and thus potentially involved in the adaptation to these conditions, except for genes involved in nitrogen and phosphate uptake and assimilation, which are enriched in genomic islands specifically in *Prochlorococcus* populations inhabiting areas where these nutrients are limiting or in strains isolated from such environments (Martiny *et al*., 2006, 2009; Kettler *et al*., 2007; Doré *et al*., submitted). Various reasons can explain the difficulties to identify niche specific genes by a mere comparative genomics approach. This includes the still quite low number of genomes available for both genera considering their extensive genomic diversity even despite the recent doubling of *Synechococcus* genomes available (Doré *et al*., submitted), our limited knowledge of the physiological characteristics of sequenced strains as well as the imprecise delineation of ecotypes and of the limits of their ecological niches especially for lineages present in low relative abundance in the field. Here, we applied a whole genome recruitment approach using as references the 97 well-annotated genomes available within the Cyanorak v2.1 database, complemented by 40 newly released complete, to analyze the distribution of picocyanobacterial genes at the global scale along the *Tara* Oceans transect, as well as their relationship with the ecotypic composition of picocyanobacterial communities and the physico-chemical characteristics of the sampling sites.

Material and Methods:

Tara Oceans dataset

A total of 109 bacterial-size (0.2-1.6 µm for stations TARA_004 to TARA_052 and 0.2-3µm for TARA_056 to TARA_152) metagenomes, collected from 66 stations along the *Tara* Oceans expedition transect and corresponding to a subset of the data presented in (Sunagawa *et al.*, 2015), were used in this study. Briefly, water samples were collected at two depths (surface and Deep Chlorophyll Maximum; DCM) and all metagenomes were sequenced as Illumina overlapping paired reads of 100-108 bp. As described in Farrant *et al*. (2016), paired reads were merged and trimmed based on quality, resulting in 100-215 bp fragments. All environmental parameters were retrieved from PANGAEA https://doi.pangaea.de/10.1594/PANGAEA.840718) except for iron and ammonium concentrations that were modeled as well as iron limitation index Φ_{sat} that was calculated from satellite data, as described in Farrant *et al.* (2016). Picocyanobacterial ESTUs and ESTU assemblages used to describe the genetic diversity of *Prochlorococcus* and *Synechococcus* along the *Tara* Oceans transect were retrieved from Farrant *et al.* (2016).

Taxonomic and functional assignment of metagenomic reads

Relative abundance of picocyanobacterial ESTUs based on the *petB* marker gene at each station was determined using a similar approach as used in Farrant *et al.* (2016) but using a Ward instead of average clustering method, which proved to be more efficient for the large datasets used for whole genome recruitments. For the latter, metagenomic reads mapping on picocyanobacterial genomes were assigned to clusters of orthologous genes (CLOGs) defined in the information system Cyanorak v2.1 (www.sb-roscoff.fr/ Cyanorak; Garczarek *et al*., in prep.). In order to reduce the amount of reads before assignment, metagenomic reads were first recruited against the 97 picocyanobacterial genomes of the Cyanorak v2.1 database (www.sb-roscoff.fr/cyanorak; Garczarek et al., in prep.) 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). Using the same BLASTN options, resulting reads were then mapped to an extended database of 819 genomes, including all genomes of the Cyanorak database complemented with 722 outgroup cyanobacterial genomes downloaded from NCBI ftp on May 4, 2017. Reads mapping to outgroup sequences or having less than 90% of their sequence aligned were filtered out and remaining reads were taxonomically assigned to either *Synechococcus* or *Prochlorococcus* according to their best hit. Reads were then functionally assigned to a cluster of orthologous genes (CLOGs) defined in the information system Cyanorak v2.1 based on the position of their BLAST match on the genome, the coordinates of which corresponding to a particular gene in the database. More precisely, a read was assigned to a gene if at least 75% of its size was aligned to the reading frame of 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 dividing read counts by L-l+1, where L represents the average gene length of the CLOG and l the mean length of recruited reads.

Only environmental 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, respectively. After this filtration step, a CLOG was kept if it showed a gene-length normalized abundance higher than 1 (i.e., a gene coverage of 1) in at least 2 of the selected environmental samples. Then, CLOGs that were found in all retained stations were removed to keep only flexible genes (i.e., genes that are differentially distributed). The resulting abundance profiles were used to perform co-occurrence analyses by weighted genes correlation network analysis, as detailed below (WGCNA; Zhang and Horvath, 2005). At last, in order to cluster stations displaying similar CLOG abundance patterns, abundance of a given CLOG in a sample was divided by the total CLOG abundance in this sample to obtain relative abundance profiles per sample, which were then used to calculate Bray-Curtis similarities for clustering analysis of *Tara* Oceans stations. The same normalisation method was applied to ESTUs. 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, a mantel test was used between the Bray-Curtis distance matrices (based on normalized relative abundance of ESTUs or genes), as implemented in the R package *vegan* v2.4 with 9,999 permutations (Oksanen et al., 2015).

Gene co-occurrence network analysis

A data-reduction approach based on WGCNA, as implemented in the R package WGCNA v1.51 (Langfelder and Horvath, 2008), was used 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., considering correlated and anti-correlated CLOGs separately), by using 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 setting the minimal number of genes in each module to 100 and 50 for *Synechococcus* and *Prochlorococcus,* respectively, and by forcing every gene to be included in a module. The *eigengene* value of each module (representative of the relative abundance of genes of a given module at each *Tara* Oceans station) was then correlated to environmental parameters and to the relative abundance of ESTUs. Finally, genes in each module with the highest correlation to the *eigengene* (a measurement called 'membership'), were extracted in order to identify genes that were the most representative of each module.

Results

Distinct picocyanobacterial communities have distinct gene repertoires

Tara Oceans metagenomic reads corresponding to the bacterial size fraction were recruited against the 97 picocyanobacterial genomes of the Cyanorak v2.1 database, complemented with a selection of 40 WGS. This yielded a total of 1.07 billion recruited reads, of which 87.7% mapped onto *Prochlorococcus* genomes and 12.3% onto *Synechococcus* genomes. In average, *Prochlorococcus* and *Synechococcus* reads represented 6.37 % and 1.08 % of all reads in a given sample, respectively. A total of 96.3% of the recruited reads were successfully assigned to a Cyanorak CLOG, most unassigned reads corresponding to those mapping on intergenic regions.

In order to identify picocyanobacterial genes potentially involved in niche adaptation, we analyzed the distribution of flexible genes across the oceans, by removing core genes i.e. CLOGs detected in every samples. *Tara* Oceans stations were first clustered according to their relative abundance of flexible genes,

resulting in three well-formed clusters for *Prochlorococcus* (left tree in **Fig. 1A**). This clustering matched quite well that obtained when stations were clustered according to the relative abundance of *Prochlorococcus* ESTUs, as assessed using *petB* as a taxonomic marker gene (right tree in **Fig. 1A**). Indeed, three main *Prochlorococcus* ESTU assemblages were also discriminated using *petB*, as previously reported by Farrant *et al.* (2016). Only a few discrepancies can be observed between the two trees, including stations TARA-070 that display one of the most disparate ESTU composition and TARA-094, dominated by the rare HLID ESTU (Fig. 1A). For *Synechococcus*, there was also a good consistency between dendrograms obtained from flexible genes abundance and from relative abundance of ESTUs (Fig. 1B). Of the eight assemblages of stations discriminated based on the latter (Fig. 1B), most were retrieved in the clustering based on flexible gene abundance, except for a few intra-assemblage switches between stations dominated by ESTU IIA and for station TARA-125, which belongs to assemblage 3, dominated by CRD1C, but contains a large proportion of IIA, an ESTU dominating in assemblage 1 (Fig. 1B). Interestingly, despite these few variations between *Synechococcus* trees based on flexible gene content and ESTU composition, four major clusters can be clearly delineated in both trees, corresponding to four broadly defined ecological niches, namely i) cold, nutrient-rich, coastal environments (blue and light red in Fig. 1B), ii) iron-limited environments (purple and grey), iii) temperate, P-depleted, iron-replete areas (yellow) and iv) warm, N-depleted, iron-replete regions (dark red). 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; dataset 1). Altogether, this indicates that distinct picocyanobacterial communities, as assessed based on a single taxonomic marker, also display distinct gene repertoires.

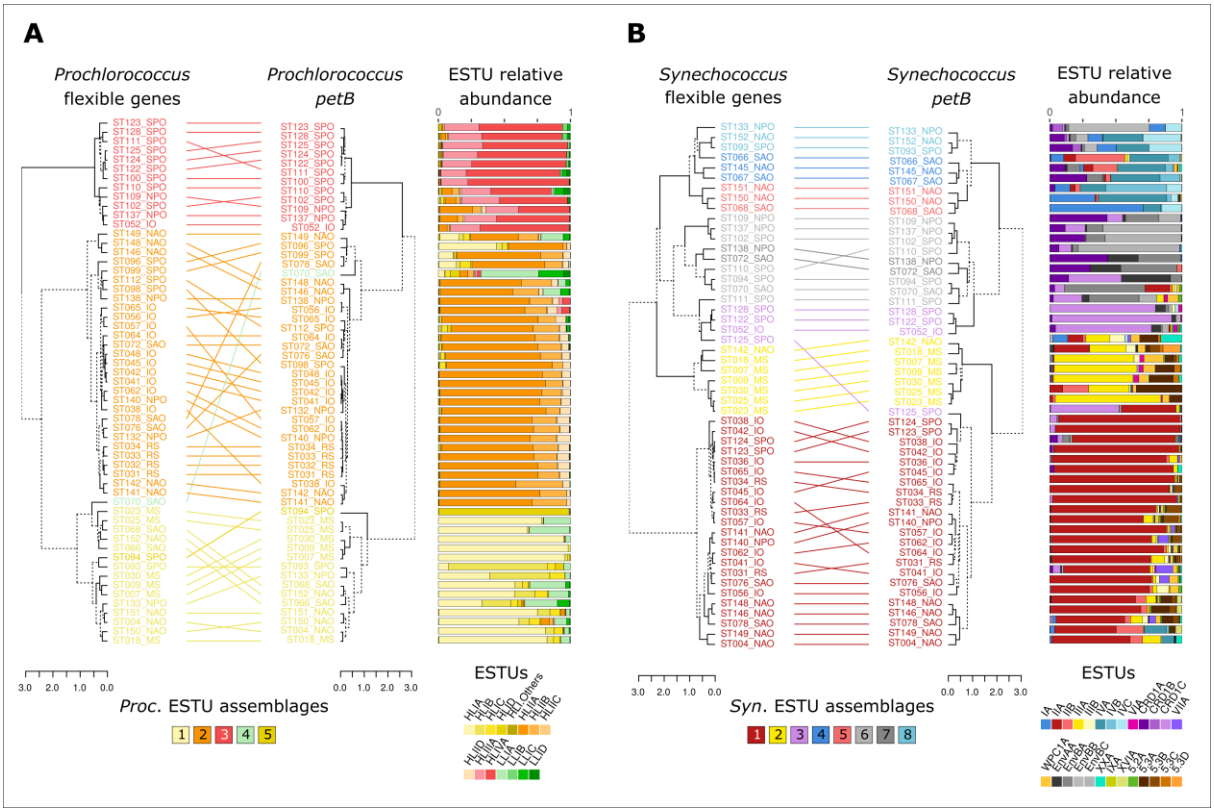


Figure 1. Comparison of clustering based on relative abundance profiles of ecologically significant taxonomic units (ESTUs) and of flexible genes for both picocyanobacteria. A. *Prochlorococcus*. B. *Synechococcus*. Leaves of the trees correspond to stations along the *Tara Oceans* transect that are colored according to the code shown at the bottom of the trees, corresponding to ESTU assemblages as determined by Farrant et al. (2016) by clustering stations exhibiting similar ESTU relative abundance profiles shown here on the right of each tree. ESTUs were colored according to the palette below each panel. Dotted lines in dendrograms indicate discrepancies between trees. Flexible genes correspond to genes that were not detected in a least one sample. Of note, due to a slightly different clustering method (cf. materials and methods), assemblage 7 (dark grey stations in 1B), which was discriminated from assemblage 6 in the Farrant et al. (2016) is now included into this assemblage. 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.

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

Prochlorococcus

In order to reduce the amount of data and better interpret the global distribution of picocyanobacterial gene content, a correlation network of genes was built for both genera based on relative abundance profiles of genes across *Tara* Oceans samples (Fig. S1A-B). For *Prochlorococcus*, this resulted into four main modules of genes and six submodules, which were found to be correlated to different environmental parameters (Fig. 2A). The *brown* module was strongly correlated to nutrient concentrations, in particular nitrates and phosphates, and strongly anti-correlated with iron availability. This module thus corresponds to genes preferentially found in iron-limited HNLC areas. Indeed, the distribution profile at the different stations of the brown module *eigengene* (Fig. S1A), representative of

the abundance profiles of genes of this module, showed that they are particularly abundant at stations TARA-100 to 125, localized in the South and North Pacific Ocean, as well as at TARA-052, a station located close to the northern coast of Madagascar and likely influenced by the Indonesian throughflow originating from the tropical Pacific Ocean (Song *et al*., 2004; Farrant *et al*., 2016). Furthermore, correlation of these modules with the relative abundance of ESTUs at each station showed that the brown module is also strongly associated with the presence of HLIIIA and HLIVA (Fig. 2B), previously shown to constitute the dominant *Prochlorococcus* ESTUs in low-iron environments (West et al., 2010; Rush et al., 2010; Farrant et al., 2016) but also the LLIB ESTU, found to dominate the LLI population in surface iron-limited HNLC areas (Farrant *et al.*, 2016). The blue module is associated to warm, low-chlorophyll oligotrophic regions with low N and P concentrations and high Fe availability (Fig. 2A), where ESTUs HLIIA dominate the *Prochlorococcus* community and HLIIB-D were also present at lower abundance (Fig. 2B, Fig. S1A).

The turquoise module seems to correspond to genes present in cold, chlorophyll-rich waters, colonized by LLIA ESTUs, and to a lower extent to LLIC and LLID, but anti-correlated with HLII ESTUs. The *turquoise* module gathers stations TARA-070, where LLIA is dominating (Fig. 1A) as well as stations dominated either by HLIA and the coldest stations dominated by HLIIA ESTUs (TARA-0146 and 149), the common point between all these stations being a strong relative abundance of LLIA (Figs. S1A). At last, the *red* module seems to be characteristic of cold, Fe-rich, N- and P-depleted waters, and strongly correlated to HLIA and anti-correlated to HLII-IV and LLIB ESTUs (Fig. 2A-B), corresponding to assemblages mainly found at the most septentrional stations of the *Tara* Oceans transect (TARA_066, 068, 093, 094, 133, 150, 151, 152) as well at all stations of the Mediterranean sea (Fig. S1A).

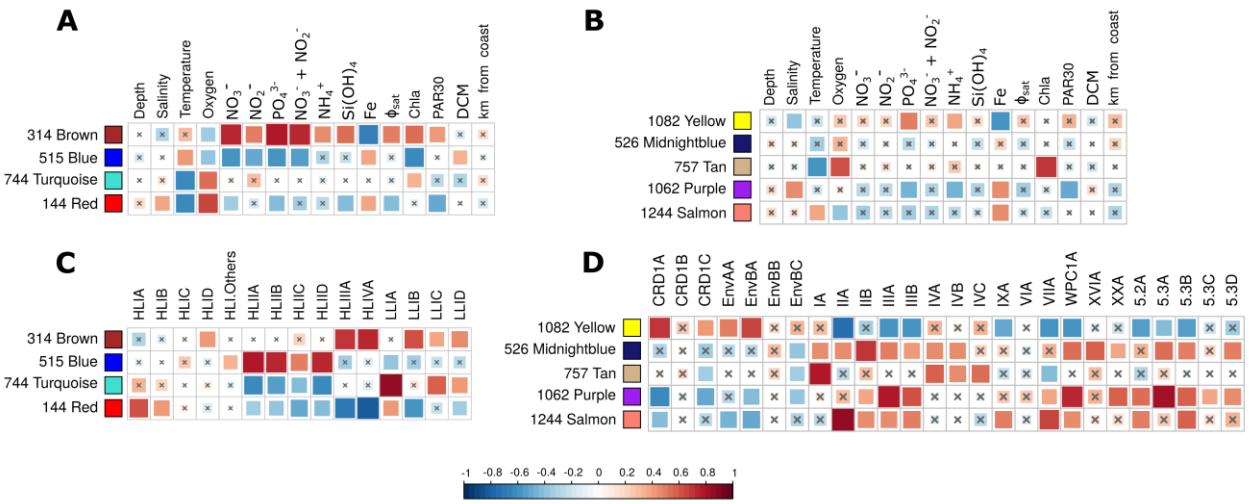


Figure 2. Correlation of picocyanobacterial module eigengenes to physico-chemical parameters and ESTU abundance. A, B. Correlation of module eigenges to physico-chemical parameters for Prochlorococcus (A) and Synechococcus (B) . C, D. Correlation of module eigenges to relative abundance profiles of ESTUs sensu (Farrant et al., 2016). Pearson (A, B) and Spearman (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. Φsat : 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.

Synechococcus

The same analysis was performed on *Synechococcus* genes and the resulting correlation network was divided into five main modules (Fig. 2C-D) that were further split into 18 sub-modules (Fig. S3), each of them being abundant in a different set of stations (Fig. S2).

The *yellow* module is correlated to phosphate and ammonium concentrations and strongly anti-correlated to iron availability, and thus corresponds to genes found in HNLC areas. Accordingly this module is correlated to ESTUs CRD1A, CRD1C, EnvAA and EnvBA, previously reported to dwell in iron-depleted areas (Sohm et al., 2016; Farrant et al., 2016; Figs. 2C-D). Although the *midnightblue* module has no significant correlation with environmental parameters, it is most strongly associated with ESTU IIB (Fig. 2C-D), a minor ESTU from clade II with an atypical distribution, since it is restricted to fairly cold, mixed waters and to co-occur with ESTUs IVA and -B (Farrant et al., 2016). It is also the most strongly correlated module with the scarce ESTU XVIA, the niche of which being still poorly known. In terms of distribution, genes of this module are mainly found at the northernmost Atlantic stations, TARA_149-151 and at the entrance of the Mediterranean sea (TARA_004), where the highest abundance of ESTU IIB is indeed observed (Figs. 1 and S2).

The *tan* module was found in cold, chlorophyll-rich waters with a high relative abundance of ESTUs IA and IVA-C (Fig. 2C-D) and in the most strongly mixed waters of the *Tara* Oceans dataset, notably the

upwelling stations (TARA_067, 093, 133), as well as TARA_145, a cold station sampled in winter, North of the Gulf stream. The purple module is found in waters with high salinity, iron-rich, P-depleted waters, and associated with IIIA/B, WPC1A and all SC 5.3 ESTUs, known to co-occur in low P areas of the world ocean (Fig. 2C-D). Consistently, it was specifically found in Mediterranean sea and the only station of the Gulf of Mexico (TARA_142, Fig. S2). At last, the *salmon* module was associated to warm, iron-rich waters. This module was most strongly associated to ESTU IIA and also to a lower extent to the fairly rare ESTUs VIIA and 5.3B and its eigengene is accordingly found at stations dominated by ESTU IIA (Fig. S2).

Correlations of submodules with environmental parameters and ESTUs were globally similar to that of the corresponding module. Still, sub-modules were generally correlated to a lower number of environmental parameters and/or to a more specific sets of ESTUs than the corresponding module and correlations were sometimes more significant at the level of submodules, in particular for *Synechococcus* for which correlations at the module level were globally lower than for *Prochlorococcus* (see below). For instance, among the 1,244 genes of the *salmon* module, the 170 genes of the *salmon* submodule are likely more relevant to better understand adaptation to Fe-replete conditions, the 713 genes of the *turquoise* submodule to decipher adaptation to high temperature, while in the *lightcyan* submodule (116 genes), we could expect to identify a number of genes specifically present in the still poorly known ESTU VIIA, for which only one genome is available so far (A18-25c). Similarly, specific genes of the two environmental ESTUs EnvAA and EnvBA are likely enriched in submodules *green* and *brown*, respectively, and more particularly at stations TARA_111 for EnvAA and TARA_070 for EnvBA, where these ESTUs reach high relative abundance (Fig. 1B) and the eigengene value of these submodules was also high (Fig. S2).

Identification of genes and genomic islands potentially involved in niche partitioning

In order to identify flexible genes related to particular environmental conditions and to specific ESTU assemblages, we correlated relative abundance profiles of each gene to the *eigengene* of its corresponding (sub)module, a method that allowed us to identify the most representative genes of each (sub)module and thus genes specifically present (or absent) in a given set of stations (Dataset 1, Figs. 3 and 4). Most genes retrieved this way are coding for proteins of unknown function (91.9% of 6,306 genes). Still, among the genes with a functional annotation (Datasets 2 and 3), a large fraction seems to have a function related to their realized environmental niche *sensu* (Pearman et al., 2008). For instance, many genes involved in transport and assimilation of nitrites and nitrates (*nirA*, *moaA-C, moaE, moeA, narB, M, nrtP*, all part of the same genomic island: Pro_GI004; Doré et al. 2020) as well cyanate, an organic form of nitrogen (*cynA, B, D, S*; part of Pro_GI033), are enriched in the *Prochlorococcus blue* module, which is correlated to HLII ESTUs and to low inorganic nutrient levels (Fig. 2A-B). This is consistent with previous

studies showing that while few *Prochlorococcus* strains in culture possess the *nirA* gene and even less the *narB* gene, natural *Prochlorococcus* populations inhabiting N-poor areas did possess one or both of these genes (Martiny et al. 2009; Berube et al., 2015; Berube et al. 2019). This result therefore constitutes a proof of concept that our network analysis was able to retrieve niche-related genes from metagenomic data. Furthermore, a distribution map of these or other inorganic N-related genes (Fig. 5A) shows that they are present in all environments in a various fractions of the *Prochlorococcus* population, but are absent in low-Fe environments generally dominated by HLIIIA-IVA ESTUs or, in the case of stations TARA_093 and 094, dominated by HLIB and HLIC ESTUs, respectively. Genes involved in the cyanate transport have a similar distribution but are usually found in a even lower fraction of the population (Fig. S5B). This suggests a differential abundance of nitrogen metabolism genes in N-depleted areas, nitrates and nitrite being the most current form of nitrogen assimilated by *Prochlorococcus* in these areas, followed by cyanate, a strategy possibly translating various degrees of adaptation of natural *Prochlorococcus* populations to nitrogen deficiency. Although not previously reported as a possible adaptation to low nitrogen in marine picocyanobacteria, the presence in the blue module of a 7-gene cluster (within Pro_GI032) also found in *Synechocystis* sp. PCC 6803 (Podar et al., 2005) and encompassing a nitrilase gene (*merR*), is also worth noting. Indeed, nitrilases hydrolyze organic nitriles to carboxylic acids and ammonia, potentially providing an additional source of NH_4^+ in N-poor areas. In *Synechocystis* grown under nitrogen starvation conditions, three genes of this cluster, *sll0783* encoding a protein of unknown function (ortholog of CK_00002257), *merR* and *slr0801* that encodes a flavoprotein involved in K^+ transport (ortholog of CK_00002258), were highly induced after 6 h of nitrogen starvation, consistent with the presence of NtcA-binding sites in the upstream region of *sll0783* (Schlebusch and Forchhammer, 2010). Although only present in the HLII strain MIT9301, genes of this cluster seem to be present in a large part of the population dominated by HLI and HLII ESTUs but absent from the same station as for the other N-related genes (Fig. S5A-6A). Also worth noting is the presence of a 4-gene cluster encompassing three genes involved in the biosynthesis of pyrimidines (*pyrB2*, *asnB* and *pydC*) and a yet uncharacterized conserved protein with similarity to aminohydrolases (CK_00001893). Interestingly, the distribution of this gene cluster is restricted to a few environments, including the Red Sea, where it is present in the whole *Prochlorococcus* population, as well as the northern and southwestern stations of the Indian Ocean, the southwestern Atlantic Ocean and the central North Pacific ocean (TARA_132), where it occurs only in a fraction of the population (Fig. S7A).

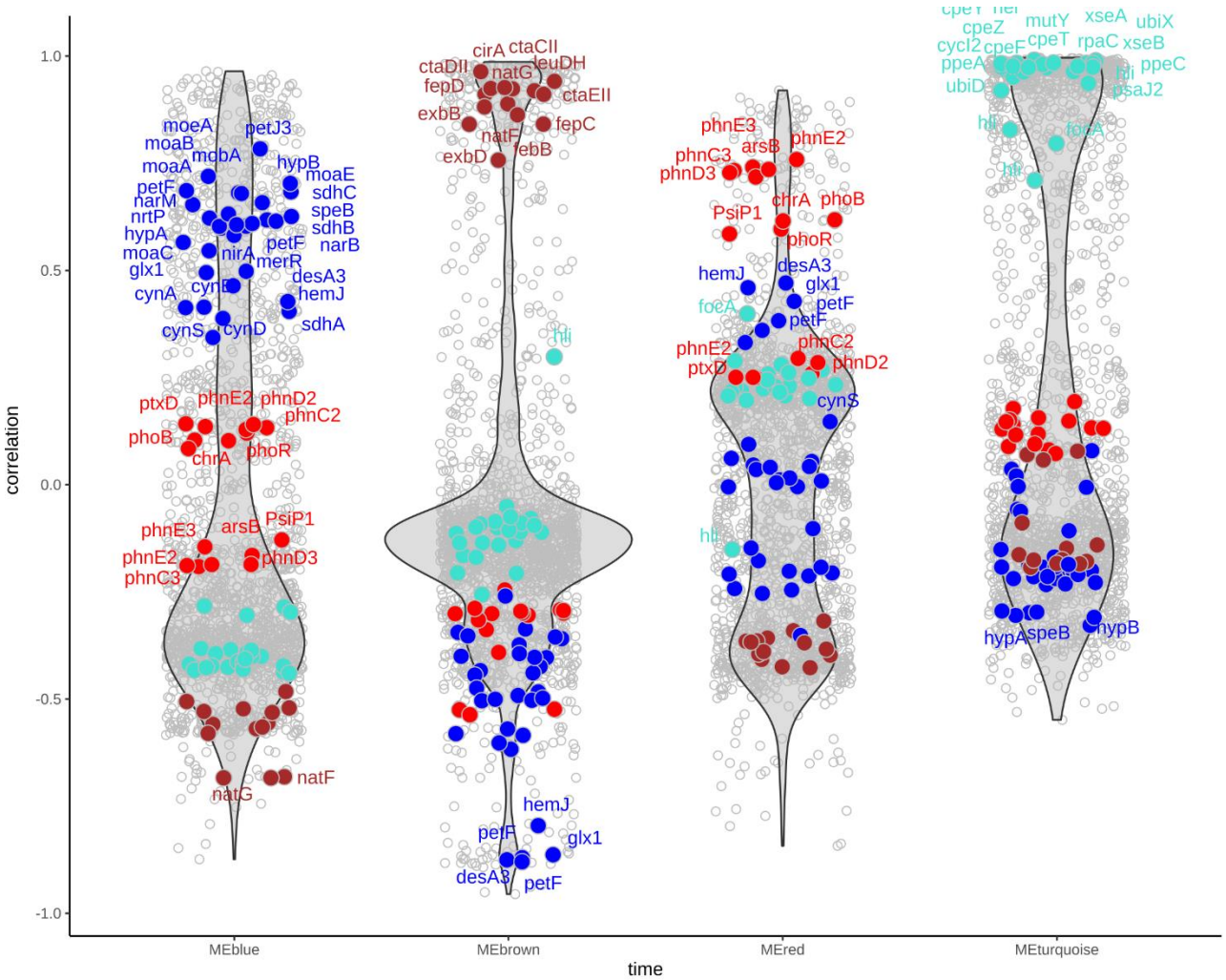


Figure 3. Violin plots highlighting the most representative genes of each Prochlorococcus module. For each module, each gene is represented as a dot positioned according to its correlation with the eigengene of each module, the most representative genes being localized on top of each violin plot. Genes mentioned in the text have been colored according to the color of the corresponding module, indicated by a colored bar above each module. Text between brackets indicates the most significant environmental parameter(s) and/or ESTU(s), as derived from Fig. 2.

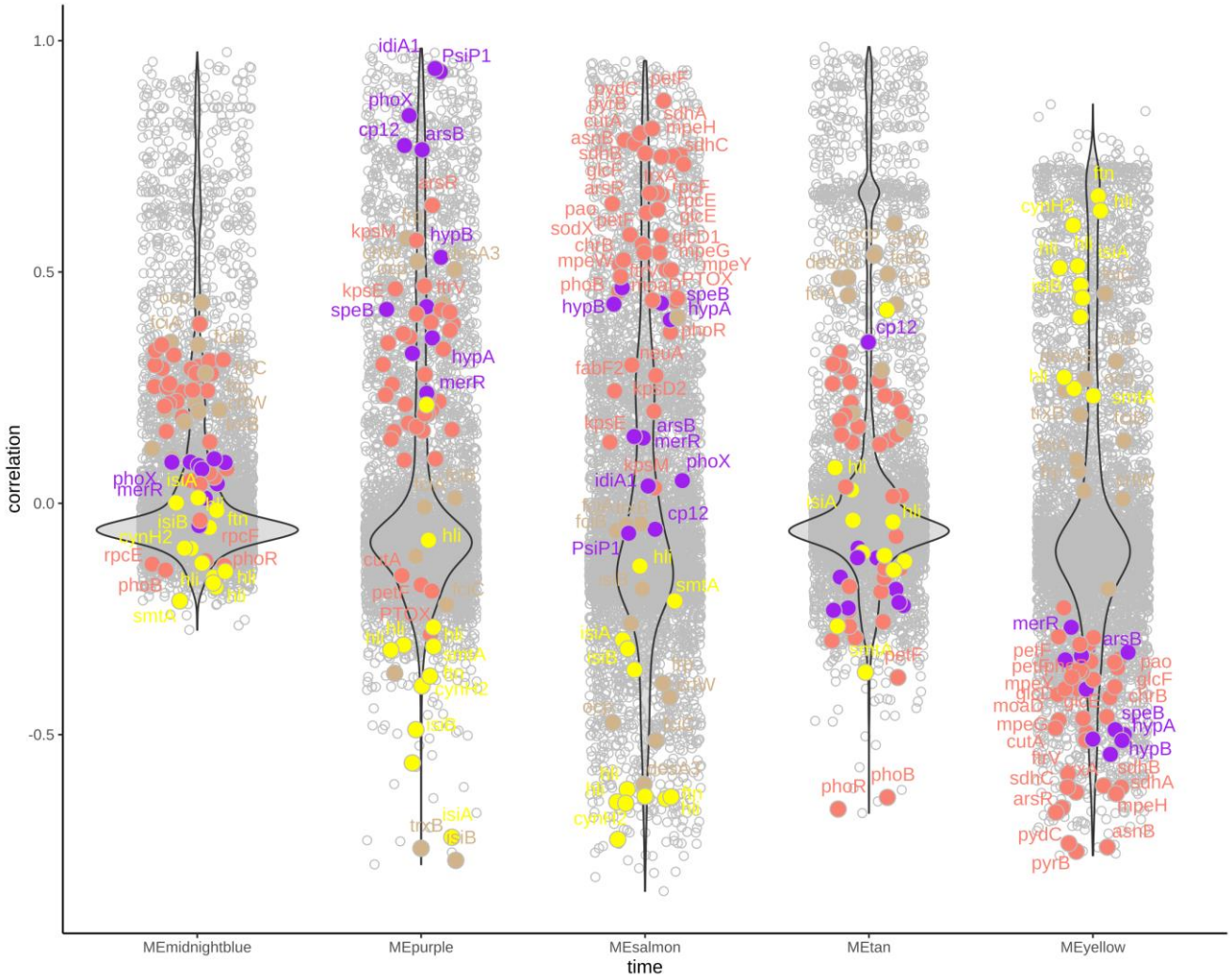


Figure 4. Same as Fig. 3 for *Synechococcus*.

In agreement with the correlation of the *blue* module with iron availability (Fig. 2A), this module was also found to contain several genes coding for iron-containing proteins (or complexes), which are thus specifically present in Fe-rich areas of the world ocean. This includes the abovementioned *nirA* encoding the ferredoxin-nitrite reductase, which in *Synechococcus elongatus* PCC 7942 utilizes six molecules of ferredoxin (Luque et al., 1993), two 2Fe-2S ferredoxins (*petF*), the *hemJ* gene encoding protoporphyrinogen IX oxidase, the *sdhA-C* operon coding for a Fe-S cluster-containing succinate dehydrogenase complex, a member of the 2OG-Fe(II) oxygenase family, several genes involved in the oxidative stress response (*gsx*, *gst, glx1*) as well as the delta-12 fatty acid desaturase DesA3, characterized by the presence of three conserved histidine-rich boxes that are presumed to constitute the Fe-binding active centers of the enzymes (Los and Murata, 1998). All these genes are seemingly absent from HLIII and HLIV MAGs (Dataset 2). Interestingly, the *blue* module also includes another 7-gene cluster, potentially involved in the transport and assimilation of amino acids, possibly L-arginine or derivatives such as taurine, which was recently found to be a major carbon and energy source for prokaryotes in the North Atlantic Ocean (Clifford et al., 2019). Indeed, this cluster encompasses i) the *speB2* gene, coding for a Fe(II)-dependent agmatinase that, together with the core decarboxylase SpeA, is likely involved in the synthesis of urea and putrescine from L-arginine, ii) two homologs of *Synechocystis* sp. PCC 6803 *hypA2* and *hypB2* which, despite their homology with *hypA1* and *B1* encoding the hydrogenase nickel incorporation complex in the same organism, are not involved in hydrogenase activity and therefore have a yet unknown function (Hoffmann et al., 2006) and iii) three genes coding for an ABC transporter with high similarity to the *E. coli* taurine transporter TauA-B-C components. The last gene of the *speB2* cluster, encoding a putative Rieske [2Fe-2S] cluster-containing protein, is surprisingly not found in the corresponding gene cluster in *Synechocystis* (*sll1077-sll1082*) and *Cyanobium* sp. NS01, nor elsewhere in their genomes. Whether it has any role in the metabolism of amino acids, as suggested by its presence in this gene cluster remains unclear. While this gene cluster is only present in a single HLII strain (UH18301), a distribution map of this cluster shows that is widespread in all regions except again in low-Fe regions and it is generally present in lower relative abundance at high latitude (Fig. S8A).

The second *Prochlorococcus* module (*brown*) also encompasses a number of genes and gene clusters that, contrary to the previous ones, appear to be highly correlated to Fe-poor HNLC environments, dominated by HLIII/IV ESTUs (Figs. 2A-B). This includes the *fepDC-cirA-febB-tonB-exbBD* gene cluster potentially involved in siderophore-mediated iron uptake and that was previously observed in HLIV SAGs as well as in the HLII MIT9201 and MIT9202 genomes (Ahlgren et al., 2020; Garcia et al., 2020; Malmstrom et al., 2013; see also Dataset 2). This gene cluster is indeed systematically present in a significant part of the population in all low-Fe areas but also unexpectedly in most stations of the Indian Ocean, dominated by HLIIA ESTU (Fig. S9A).

A more original finding is the presence in this module of a 4-gene cluster encompassing the *ctaC2-D2-E2* operon, which codes for an alternative respiratory terminal oxidase (ARTO) and that is also specifically enriched in F-poor areas (Fig. S9B, Dataset 2). This cytochrome oxidase has been suggested to be part of a minimal respiratory chain in the cytoplasmic membrane and could provide an additional electron sink under iron-deprived conditions, in which the relative amount of cytochrome b_6/f and PSI complexes are reduced (Lea-Smith et al., 2013). Furthermore, a *Synechocystis* mutant in which the *ctaD2* and *ctaE2* genes had been inactivated was found to display markedly impaired iron reduction and uptake rates as compared to wild-type cells, suggesting that ARTO is involved in Fe(III) reduction into Fe(II) prior to its transport through the plasma membrane via the Fe(II) transporter FeoB (Kranzler et al., 2014). The last gene of the ARTO cluster codes for a member of the glucose-methanol-choline oxidoreductase family protein, showing some similarity to choline dehydrogenase, which catalyzes the first biosynthesis step of glycine betaine from choline (Landfald and Strøm, 1986). It is worth noting that the glycine betaine transporter gene *betP* is also highly correlated to this module. Its genome context differs between the only two *Prochlorococcus* genomes that contains this gene, while in *Synechococcus betP* is always the last gene of a cluster encompassing a S-adenosyl-L-methionine (SAM)-dependent methyltransferase, *leuDH* encoding the leucine dehydrogenase and *soxA* encoding a monomeric sarcosine oxidase. This gene cluster, which is possibly in part involved in the metabolism of glycine betaine or choline, is also present (without *betP*) in *Prochlorococcus* and is highly correlated to the brown module. Another interesting gene cluster of the *brown* module comprises three (*natF-G-bgtA*) out of four genes coding for an ABC-type uptake transporter for acidic and neutral polar amino acids similar to that described in *Nostoc* sp. PCC 7120 (Pernil et al., 2008). The fourth gene (*natH*) was missed by our approach since it was detected at all stations and notably found at very low levels at two stations where the other three genes went undetected. Yet, the phyletic pattern is the same for all four genes in our reference genome database (Dataset 2), i.e. the whole 4-gene cluster is present in five out of 25 *Prochlorococcus* HL strains, including a HLIV strain, while it is present in most LL *Prochlorococcus* and *Synechococcus* strains. Thus, this operon is clearly part of the flexible genome in *Prochlorococcus* HL cells and its distribution suggests that its presence might also be favorable to life in HNLC areas (Fig. S9C). Globally, most of these genes were displaying very high correlation values, suggesting that Fe availability constitute a very selective condition for the presence of these genes.

The *Prochlorococcus red* module is highly correlated to P-depleted waters dominated by HLIA ESTU (Fig. 2A-B). Consistently, this module contains numerous flexible genes involved in P transport and metabolism. Contrary to the abovementioned gene clusters, these genes are split between different genomic regions whose gene content and organization are highly variable, not only between strains isolated from P-replete and P-poor regions, the latter generally having more P-related genes than the

former, but also among strains isolated from low-P regions (Martiny et al. 2006; Martiny et al. 2009). Several of these genes are located in a region of *Prochlorococcus* genomes which, in the MED4 strain, starts with the *phoB-R* operon encoding a P-stress responsive two-component system and ends with the core *pstABC* operon, encoding a phosphate ABC transporter (see Fig. 4 in Martiny et al. 2006). Besides P metabolism genes, this region also encompasses genes involved in the efflux of toxic analogues of phosphate (*arsB*, an arsenate transporter) and sulfate (*chrA*, a chromate transporter). A global map of the distribution of several genes of this genomic region coding for key P-related proteins, including PhoB-R, a PhoA-like alkaline phosphatase (Kathuria and Martiny, 2011), a paralog of the general phosphate regulator PtrA and an uncharacterized conserved secreted protein (CK_00003429 corresponding to PMM0707 in MED4) whose gene is located in-between *phoR* and *phoA*, show that these genes are specifically enriched in P-depleted areas of the ocean (Figs. S10A), consistently with previous reports (Martiny et al. 2009; Garcia et al., 2020). While all these genes are present at equal relative abundance as the core *pstA* gene in the most P-limited areas of the world ocean, namely the Gulf of Mexico, the Panama channel, the eastern North Atlantic Ocean and the northern red Sea, *phoA*, CK_00003429 and/or the *ptrA*-like gene were either absent or only present in a fraction of the *Prochlorococcus* population, notably in the South Atlantic and central North Pacific oceans (Fig. S10A). The *red* module also contains a distant homolog of the alkaline phosphatase PhoX (CK_00056879) present in the HLII strain MIT9314, as well as some genes present in another genomic island of HL strains, called ISL5 by Coleman et al. (2006), and previously reported to contain several genes strongly upregulated in response to P starvation, such as *psiP1* and a DUF3303 domain containing protein (Martiny et al. 2009; Scanlan et al. 2009). Interestingly, our approach also retrieved two flexible gene clusters involved in the transport and/or assimilation of phosphonates, i.e. C–P bonds-containing compounds constituting up to 25% of the high-molecular-weight dissolved organic P pool in the open ocean (Clark et al., 1998). While most *Prochlorococcus* strains contains a core phosphonate ABC transporter (*phnD1-C1-E1*), only a few HLII strains also possess genes coding for another ABC transporter, which can be either *phnC2-D2-E2* or *phnC3-D3-E3-E4* both distantly related to the core version (Feingersch et al., 2012), the first one being also associated with genes involved in phosphonate utilization *phnY-Z* (Martinez et al., 2010; McSorley et al., 2012) and a phosphonate dehydrogenase (*ptxD*). Interestingly, all 7 genes of the *phnC2-D2-E2* gene cluster are most often found in low-P areas but in very low proportion of the population colonizing these niches (Fig. S10B). As concerns the third phosphonate region (*phnC3-D3-E3-E4*), it is found in an even higher number of stations, all four genes being present in the whole population in the most P-limited areas mentioned above, while the variation in the relative proportions of *phnE3* and *phnE4* with regard to *phnC3* and *phnD3,* noticed in the southeastern Atlantic and central North Pacific, seems to indicate that the populations colonizing these areas have a phosphonate transport system composed of the ATPase PhnC3, the substrate binding component PhnD3

and either the PhnE3 or PhnE4 permease subunits (Fig. S10C). Of note, the *red* module is also highly anti-correlated to temperature and three genes of this module were indeed previously identified as being strongly associated to cold waters, namely *pyrC2*, a putative dihydroorotase, *pudP* encoding a pseudouridine 5'-phosphatase and CK_00000399 encoding a protein of the 2OG-Fe(II) oxygenase superfamily (Kent et al., 2016). although the adaptive potential of these genes with regard to temperature remains to be confirmed.

The *Prochlorococcus turquoise* module is highly correlated to low temperature and gathers stations characterized by a strong vertical mixing, where the *Prochlorococcus* LLIA ESTU is often a major component and sometimes the dominant ESTU (Figs. 1, 2 and S1). Consistently, the phyletic pattern of most genes present in this module is characterized by their presence in *Prochlorococcus* LLI and absence in HL strains (Dataset 2). While some of these genes are truly specific of LLI strains such as several *hli* gene copies, the endonuclease VIII *nei*, two NAD-dependent DNA ligases as well as *psaJ2* encoding a divergent copy of the core photosystem II PsaJ subunit, others are shared with *Synechococcus* and/or strictly LL *Prochlorococcus* strains (clades LLII-IV), the latter being generally absent from *Tara* Oceans surface samples (Fig. 1A). This gene set notably includes several genes of the phycoerythrin-III biosynthesis gene cluster (*ppeA, C, cpeF, T, Y, Z*), which is much shorter in HL than LL strains (Hess et al., 2001), *rpaC* involved in the regulation of state transitions (Emlyn-Jones et al., 1999), *ubiX* and *D* involved in plastoquinone synthesis, a number of genes involved in DNA repair (*xseA, B, mutY*) as well as the nitrite transporter *focA* and *cnaT*, involved in controlling nitrite assimilation (Frías et al., 2003).

As expected from the fact that *Synechococcus* spp. possess a much larger and diversified genomes than *Prochlorococcus* HLI-IV and LLI and have consequently a more extensive gene repertoire, the same network approach applied to *Synechococcus* retrieved comparatively fewer large gene clusters potentially related to niche adaptation. Alternatively, this might be a bias due to the use of too stringent thresholds and the analysis probably needs to be redone using less stringent values. Yet, the *Synechococcus salmon* module, whose genes are mostly associated with warm, Fe-replete environments dominated by ESTU IIA, contains a fairly large number of genes coding for Fe-rich proteins. A distribution map of *sdhA-B-C* and *gclD1-E-F,* encoding respectively the Fe-S cluster-containing (photo)respiratory complexes succinate dehydrogenase/fumarate reductase (Cooley and Vermaas, 2001) and glycolate oxidase (Eisenhut et al., 2008), showed that both protein complexes were equally abundant, though at a slightly lower relative abundance than the core gene *petB*, over a large part of the *Tara* Oceans transect, with the notable exception of Fe-depleted areas, including several stations of the southeastern and equatorial Pacific ocean, TARA_052 off Madagascar and TARA_070 in the South Atlantic ocean (Fig. S11A), all these stations being dominated by CRD1, EnvA and/or EnvB ESTUs (Fig. 1). Of note, *sdhA-B-C* genes seem to be at lower abundance than *gclD1-E-F* and sometimes even absent from the northernmost stations of the Atlantic

Ocean, and they are only present in a small fraction of the *Synechococcus* population at the upwelling station TARA_133, consistent with their absence from all cold-adapted clade I strains (Dataset 3). Also interesting is the distribution of *ftrC and ftrV* genes, which respectively encode the [4Fe–4S] cluster-containing catalytic subunit and the variable subunit of the ferredoxin-thioredoxin reductase, an enzyme that receives electrons from ferredoxins and then reduces a thioredoxin through a disulfide–dithiol interchange system (Dai et al., 2004; Fig. S11A). While both genes are found in all Fe-replete areas, *ftrC* is found at similar relative abundance as the core gene *petB,* whereas *ftrV* is often about twice as abundant. Even though all sequenced *Synechococcus* strains of our reference database contain a single copy of the *ftrV* gene, this result suggests that natural *Synechococcus* populations dwelling in Fe-replete areas might encode two *ftrV* gene copies, since we did not find any *ftrV*-related gene in *Synechococcus* genomes that could have been recruited by our approach. In contrast, only a very low fraction of the population colonizing HNLC areas possess *ftrC* and *V* genes and most often a higher proportion of the former. This is in agreement with the fact that MITS9220 (CRD1-A) has no *ftrV* and only a degenerated *ftrC* gene, while BIOS-U3-1 (CRD1-B) has neither of them (Dataset 3). So, we hypothesize that *Synechococcus* populations dwelling in HNLC areas might possess such a FtrC remnant, though whether it has retained any function related to electron transport is unlikely since the sequence of MITS9220 misses most of the cysteinyl sites necessary to bind the [4Fe–4S] cluster (Dai et al. 2004). The *salmon* module also contains two non-core ferredoxins*,* including *petF*-CK_00008098 that is restricted to Fe-replete areas, consistent with its absence from all three reference CRD1 strains, while *petF*-CK_00057141 displays a more variable pattern, being notably present at low abundance in Fe-depleted areas and absent at high latitude, in agreement with its absence from the genomes of cold-adapted clades I and IV strains (Fig. S11B and Dataset 3). The *salmon* module also includes several other genes encoding Fe-containing protein genes that are also restricted to Fe-rich waters, including the pheophorbide *a* oxygenase (*pao*), the plastoquinol terminal oxidase (*ptox*) and a 2-oxoglutarate/Fe(II)-dependent dioxygenase domain-containing protein (Fig. S11C). The differential distribution of all the abovementioned genes is probably indicating a reduction of Fe quotas in *Synechococcus* communities thriving in Fe-depleted regions through the loss of genes interacting with Fe, as was suggested for *Prochlorococcus* populations belonging to clades HLIII/IV (Rusch *et al.*, 2010). Also worth noting in this module are i) a gene cluster already found in the *Prochlorococcus blue* module (see above and Fig. S7B) and encompassing *pyrB2*, *asnB, pydC* and an aminohydrolase-like gene, as well as genes coding for a putative urea-proton symporter (*ureP*) and a small uncharacterized conserved membrane protein that are absent from all reference *Prochlorococcus* genomes, and ii) several genes of the *kps* gene cluster involved in the biosynthesis and export of capsular polysaccharides from the cytoplasm to the bacterial cell surface (Silver et al., 2001), which could serve as receptors for bacteriophages and are also involved in the protection against a variety of stress conditions, as capsular

polysaccharides constitute a physical barrier (Kehr and Dittmann, 2015). In *Synechococcus* genomes, the set of *kps* genes is split into two sub-clusters, i.e. a 11-gene region extending from *kpsD2* to *neuB* and a smaller region encoding the ABC transporter subunits KpsT-M-E. Although proteins encoded by these two gene clusters do not contain iron, all of them are in low relative abundance or even absent from low-Fe environments but also at high latitude (Figs. S7 and S12), consistent with the fact that, like many genes of the *salmon* cluster, they are absent specifically from CRD1 strains, possibly because these processes are too energy-consuming for this peculiar niche. Nevertheless, the *asnB* cluster is comparatively present in more niches than *Prochlorococcus*, notably in the Mediterranean Sea and Indian Ocean.

Most genes that are correlated with the *Synechococcus salmon* module are also almost as strongly *anti*-correlated to the *yellow* module, and reciprocally. Indeed the latter module is associated with low-Fe, P-replete areas and CRD1A as the main ESTUs and consistently contains a number of genes potentially related to adaptation to Fe deficiency. Among the most notable genes of this module are the iron-stress induced genes *isiA*, *isiB* as well as *ftn* CK_0001204, encoding one of the two ferritin forms existing in *Synechococcus*. Although the chlorophyll-binding protein IsiA, forming an additional antenna around PSI (Fig. S13A; Michel and Pistorius, 2004), and IsiB encoding flavodoxin, a non-iron electron-transfer protein known to replace the Fe-containing ferredoxin (Fig. S11B; Erdner and Anderson, 1999; Kranzler et al., 2013), have both been associated with iron-deficiency (Geiß et al., 2001), their distributions show that they are present in most oceanic niches, with the notable exception of the Mediterranean sea but also at high latitudes, where only *isiA* and a second form of isiB (CK_00001833) are present, consistent with their presence in cold-adapted strains of clades I and IV. As concerns iron storage, the ferritin CK_00001204 appears to be widespread but while it is in single copy in the Red Sea and Indian Ocean, it is seemingly present in two to three copies at high latitude and in Fe-depleted areas, in agreement with its presence in 3-6 copies in CRD1 genomes. By comparison, we also plotted the distribution of the second ferritin (CK_00033189) and of bacterioferritin present in *Synechococcus* genomes, which seems to have a quite complementary distribution, with *bfr* being present in a significant proportion of the population mostly in low-P area, while CK_00033189 is present at low abundance or absent in these areas (Fig. S13A). Also well represented in this *yellow* module are several *hli* gene copies, which seem to be essentially present in CRD1 strains as well as genes encoding proteins of unknown function containing a nif11-like leader peptide domain (CK_00050707, CK_00001868, CK_00008566, CK_00002199, CK_00056823), one of which being present in high copy number in CRD1 genomes (up to 53 in BIOS-E4-1, CK_00050707) and accordingly present in multiple copies in Fe-depleted areas. Thus, these genes are likely important for survival of *Synechococcus* communities thriving in these Fe-depleted areas (Fig. S13B-C). Finally, a second copy of the cyanate hydratase (*cynH2*), specific to CRD1 genomes, and a sulfiredoxin, also present in these genomes and known to be involved in the protection of the photosynthetic apparatus against oxidative stress

damages through an effective reduction of 2-Cys Prx (Findlay et al., 2005; Sánchez-Riego et al., 2016), are also present in this module.

The *purple* module is associated to high salinity, low-P, Fe-replete, and mainly to ESTUs IIIA, WPC1A and 5.3A, typical of the Mediterranean sea waters (Fig. 2C-D). Consistently a number of genes involved in the transport and/or assimilation of phosphate or its arsenate analog are found in this module, including *phoX*, *sphX*, *psiP1* as well as *arsB* (Dataset 3). Plotting their distribution with regard to the core *petB* gene, shows that there are specifically present in low-P regions with *phoX* and *sphX* present in most cells and *psiP1* only in a subset of the population (Fig. S14A). Comparatively, the arsenate efflux pump-encoding *arsB* gene is present at significant relative abundance in more regions, including the Indian ocean and South Atlantic, while all these genes are completely absent from low-Fe areas. The comparison of their distribution with that of the other P or As-related genes retrieved in the *salmon* module, namely *phoB, phoR, phnZ* and *arsR*, show that the latter are found in a much larger range of environmental conditions, covering both low-P and low-N regions, consistent with the fact that have a much larger phyletic pattern, including both clades II and III strains (Fig. S14B). The *purple* module also contains two 7-gene clusters previously identified in the *Prochlorococcus blue* module and encompassing agmatinase (*speB2*) and nitrilase (*merR*) (see above). As for *Prochlorococcus*, these whole clusters are found in both low-P and low-N niches and quasi absent in low-Fe niches, but they are generally present in a larger fraction of the population for *Synechococcus* and this proportion tends to decrease with latitude (Figs. S6B and S8B).

The *tan* module is associated with cold, chlorophyll-rich environments dominated by ESTUs IA and IVA. It contains a number of genes involved in photosynthesis, in particular those involved in the CA4-A form of type IV chromatic acclimation (*fciA-C*), confirming that *Synechococcus* CA4-A cells are mostly found in temperate or cold environments (Grébert *et al.,* 2018; Xia *et al.*, 2017; Dataset 3). This module also contains a few genes related to iron metabolism, including one isoform of the flavodoxin type (*isiB* CK_0001833) and of the ferredoxin gene (CK_00008099). Also noteworthy in this context, is the presence in this module of the 3-gene *ocp* operon (*ocp, crtW, frp*), encoding an important photoprotection system that protects photosystem II from excess photon energy by increasing thermal dissipation of the energy absorbed by the antenna (Boulay et al., 2008). Here the distribution map clearly show that this system is absent both from N and Fe-depleted areas, notably, the Red Sea, Indian Ocean and South Pacific, while it is systematically present in the Mediterranean sea and at high latitude (Fig. S15A). At last the *tan* module also encompasses a fairly high number of two-component systems genes, including two sensor histidine kinases and three response regulators, one of the latter (CK_00002009) displaying a very similar distribution to the ocp operon (Fig. S15B). Altogether, these results support the hypothesis drawn from experimental studies that coastal strains exhibit a higher tolerance to fluctuating environmental conditions than oceanic strains (Mackey et al., 2015; Stuart et al., 2009).

At last the *midnight* blue module is the smallest *Synechococcus* module and is mostly correlated to the minor IIB ESTU, which was found to be restricted to fairly cold, mixed waters, while the dominant ESTU of this clade (IIA) is in contrast found in warm waters. This module mainly contains only genes with no functional annotation that are present in one or a few strains among the reference genomes, majoritarily PROS-U-1 (ESTU IIB), WH8020 (IA), CC9616 (ESTU XXA) or A15-127 (WPC1A). A network approach based on the similarity of phyletic patterns in reference genomes and in field populations and taking into account the proximity of genes in the reference genomes will be developed in the near future to try to identify gene clusters specific of these populations.

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

Discussion

The unique combination of deeply sequenced metagenomes collected along the transect of the *Tara* Oceans expedition, which crossed a vast range of oceanic regions, and of a manually annotated reference genome database (Cyanorak v2.1) covering much of the wide genetic diversity existing within both *Synechococcus* and *Prochlorococcus* (Garczarek et al. in prep.), allowed us to considerably refine our current understanding of the molecular mechanisms used by these ubiquitous and abundant marine picocyanobacteria to colonize the different oceanic niches. Even though we only detected genes that are present in our reference database, and thus recruited only a part of the whole picocyanobacterial pangenome (see Chapter I and 'Conclusions and Perspectives'), a striking correspondence was found between the distribution of ESTUs, as assessed by a single, high-resolution, core marker gene *petB* (Farrant et al. 2016), and the distribution of genes constituting the flexible genome of these populations in the field (Fig. 1). As previously suggested for *Prochlorococcus* (Kent et al., 2016), this strong correlation between taxonomy and gene content strengthens the idea that, in both genera, genome evolution mainly occurs by vertical transmission, with a fairly low extent of lateral gene transfer. The network approach used here further demonstrated that many flexible genes in both genera do not have a random distribution in the environment but are rather associated to specific ecological niches and often to specific ESTUs. In the case of *Prochlorococcus,* analysis of the functions of flexible genes associated with the different niches did not only confirm previous observations about the presence of individual N-, P- or Fe-uptake genes specifically in areas depleted in these nutrients (Ahlgren et al., 2020; Berube et al., 2014; Garcia et al., 2020; Kent et

al., 2016; Martiny et al., 2006, 2009c), but also provided novel insights about possible adaptive strategies adopted by the ESTUs to colonize these peculiar niches. Indeed, by combining metagenome analyses with information about CLOG phyletic patterns and comparison of the genomic context of individual genes between genomes of our reference database, we managed to identify a large number of gene clusters, including many that were not previously reported in the literature. Several of them are likely involved in the uptake and/or incorporation of complex organic forms of nutrients, which might either be degraded into elementary N, P or Fe molecules or possibly directly used by the cells for e.g. the biosynthesis of proteins or DNA.

In N-depleted areas, besides being able to enhance their capacity for uptake and assimilation of inorganic forms of nitrogen by acquiring or retaining the well-known gene cluster encompassing a nitrate transporter (*nrtP*), nitrite (*nirA*) and nitrate reductases (*narB*) and numerous cofactors (Martiny et al., 2009a), *Prochlorococcus* cells also appear capable to incorporate i) amino acids or derivatives, for which we identified transporters (*natF-G-H-bgtA*, *tauA-B-C*) and possible assimilation pathways (e.g. the agmatinase gene cluster), ii) cyanate (*cynA-B-D-S*), and also possibly iii) other nitrogen sources in the forms of nitriles or cyanides, which could be metabolized by the nitrilase gene cluster, though the availability of such C≡N-containing compounds (other than cyanate) remains poorly documented so far in oceanic waters, besides heavily polluted areas. However, we cannot exclude that the latter gene cluster is in fact devoted to the detoxification of cyanides incorporated by cells, like it is the case for arsenic and chromate (Saunders & Rocap 2016), since these compounds have been found to be toxic to other marine phytoplanktonic organisms (Pablo et al., 1997). Also noteworthy in the N-depleted niche is the presence of a cluster of genes encompassing *asnB, prB2* and *pydC*, which could contribute to an alternative pyrimidine biosynthesis pathway and thus provide another way for cells to recycle complex nitrogen forms. Interestingly, as was previously reported for the *narB* gene cluster (Berube et al., 2019; Martiny et al., 2009b), it is frequent that only one or very few sequenced *Prochlorococcus* strains possess these niche-related gene clusters (e.g. the agmatinase gene cluster is only present in *Prochlorococcus marinus* UH18301), whereas their distribution in the field show that they are present in a large part of (and sometimes the whole) *Prochlorococcus* population in N-depleted areas of the world ocean. Whether isolation steps initially select for the scarce cells missing these gene clusters or whether these clusters are secondarily lost due to a weakening of selective constraints during cultivation in nutrient-replete media remain to be determined. As concerns *Synechococcus*, it is interesting to note that although strains exhibit a more extended gene repertoire to deal with inorganic nutrients compared to the streamlined *Prochlorococcus* genomes (Scanlan *et al.*, 2009), as shown by the presence in 50 out of 54 *Synechococcus/Cyanobium* strains of the whole *narB* gene cluster (Dataset 2), gene clusters involved in organic nitrogen are often shared by less strains and metagenomic analyses of their global distribution

indeed show that a number of them, such as the agmatinase and nitrilase clusters, are, like for *Prochlorococcus*, specific to given niches. For both genera, as mentioned in Kent et al. (2016), most of these N-related genes are specifically absent from Fe-depleted areas, including *nrtP-nirA-narB*, agmatinase, nitrilase and *ureP-pyrB2-pydC-asnB* gene clusters. Thus, the main sources of nitrogen for populations dwelling in these areas seem to be ammonium and amino acids. Indeed, the ammonium transporter *amt1* is core in both genera and a second transporter *amt2* as well as an uptake transporter for polar amino acids (*natF-G-H-bgtA*) and *leudH*, involved in the catabolism of leucine, are specifically present in *Prochlorococcus* populations colonizing HNLC areas (*brown* module) and in one HLI strain isolated from equatorial Pacific Ocean (MIT9515) and a HLIV metagenome, while most of these genes are core in oceanic *Synechococcus*.

Adaptation to P-depleted has already been well documented in *Prochlorococcus* via the acquisition by lateral transfer in P-limited areas of individual genes into two main genomic regions, the *phoB-R* region (Martiny et al., 2006, 2009) and ISL5, encompassing *psiP1* as well as several uncharacterized genes strongly differentially expressed in response to P-depletion (Coleman et al. 2006). Furthermore, Feingersch et al. (2012) also found that, in addition to the core phosphonate ABC transporter (*phnD1-C1-E1*), an accessory transporter (*phnC2D2E2*) was specific of low-P regions. Yet, we show here that the *phnC2-D2-E2* gene cluster, which also encompasses the *ptxD-phnY-Z* genes involved in the metabolism of phosphonates, is present in a very low proportion of the *Prochlorococcus* population colonizing these niches, while a third previously unreported phosphonate operon, *phnC3-D3-(E3)-(E4),* which encompasses either both or only one of the two permease subunits (*phnE3-E4*), can in contrast be found in the whole population in some areas like the Mediterranean sea and is globally more widespread than the other phosphonate gene region. Altogether, these environmental data seem to indicate that in the most P-limited areas, the whole *Prochlorococcus* population possess both the PhoB-R/PhoA/PtrA and the *phnC3-D3-E3-E4*, while only a small proportion of the population also possess the *phnC2-D2-E2* system, which is consistent with the fact that no *Prochlorococcus* strain sequenced to date possesses all three sets of genes (Dataset 2). In *Synechococcus,* comparatively few genes specific of P-depleted areas, including *phoX*, *sphX* and *psiP1* have been detected, while more P-related genes (*phoB, phoR, phnZ* and *arsR*) are found in a larger range of environmental conditions, covering both low-P and low-N areas. Thus, we hypothesize that ESTUs IIIA, WPC1A and 5.3A that colonize low-P areas possess both sets of genes, while ESTU IIA, dominating low-N areas, possesses only the second one. As for N-related genes, P-related genes are quasi absent in low-Fe niches, with the notable exception of *phoB-R*, and are generally present in a larger fraction of the *Synechococcus* than *Prochlorococcus* population and this proportion tends to decrease with latitude.

In contrast to macronutrients, it has been hypothesized that survival of *Prochlorococcus* and *Synechococcus* in low-Fe regions was made possible through the elimination from the genomes of genes

coding for proteins that contain iron as a cofactor as well as the acquisition of genes involved in Fe-storage, such as ferritin, and/or Fe-scavenging, such as the TonB-dependent siderophore uptake operon (Malmstrom *et al.*, 2012; Rusch *et al.*, 2010; Kent *et al.*, 2016; Garcia et al. 2020; Ahlgren et al. 2020). Accordingly, many proteins interacting with iron were indeed found in the present study to be anti-correlated to HNLC regions either in both genera (succinate deshydrogenase), only in *Synechococcus* (glycolate oxidase, pheophorbide *a* oxygenase, 2-OG/Fe(II)-dependent dioxygenase domain-containing protein, PTOX and the typical [4Fe-4S]-containing FtrC/V complex) or in *Prochlorococcus* (the metal-dependent fructose-1-6-bisphosphate aldolase class II FbaA) and populations dwelling on these areas were also found to be enriched in ferritin for *Synechococcus* and iron-scavenging genes (*fepDC-cirA-febB-tonB-exbBD*) for *Prochlorococcus*. One novelty of our study is that most *Prochlorococcus* cells thriving in low-Fe regions possess the *ctaC2-D2-E2* operon, also found in 85% of all *Synechococcus* sequenced to date and notably in all CRD1 strains, which encode the cytochrome *c* oxidase ARTO that could provide an additional electron sink under iron-deprived conditions (Lea-Smith et al., 2013) and was also suggested to be involved in the reduction of Fe(III) into Fe(II) (Kranzler et al., 2014). Also well represented in *Synechococcus* population colonizing Fe-depleted areas are numerous nif11-like leader peptide domain, one of which being present in 53 copies in the CRD1 strain BIOS-E4-1, as well as several genes potentially involved in the protection of the photosynthetic apparatus against oxidative stress damages, notably numerous *hli* gene copies but also sulfiredoxin, which was found to be highly differentially regulated in response to all stress conditions in WH7803 (Findlay et al., 2005; Guyet et al., submitted). This suggests that the reduction of the metabolism when cyt b_6f and PSI are limited under Fe-deprived conditions could also increase the cell damages due to reactive oxygen species in these areas.

Besides genes involved in nutrient acquisition and metabolism, we also found in both *Prochlorococcus* and *Synechococcus* several gene modules associated to low temperature and/or to ecotypes colonizing cold ecological niches (ESTU LLIA and HLIA for *Prochlorococcus* and IA, IVA and IIB for *Synechococcus*). Yet, besides *desA3*, encoding a fatty acid Δ12-desaturase that we recently showed to be present in cold and temperate waters (cf. part II.2 of this chapter), it is not clear whether any annotated genes of these modules could be directly related to adaptation to low temperature. This observation 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 (Kent et al., 2016; Larkin and Martiny, 2017; see chapter I part 1). Still, a number of genes of these modules are 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, 2017; Varkey et al., 2016). We also observed a decrease of the proportion of the population displaying some gene clusters across a latitudinal gradient, such as the agmatinase, the succinate

deshydrogenase or the *phoB-R* gene clusters, suggesting that populations colonizing these niches have developed alternative strategies to cope with cold waters or other environmental factors linked to low temperature. In contrast, our study clearly show that populations colonizing well-mixed waters at high latitudes or in upwelling areas all possess the *ocp-crtW-frp* operon, while it was so far unclear why this important photoprotection system was absent from a number of marine *Synechococcus* strains (Boulay et al. 2008). These environments are also characterized by a high number of specific two-components systems that might also be important to adapt to rapidly changing conditions characterizing these niches.

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, either inherited from ancestral cyanobacteria and retained specifically in some lineages, e.g. for the agmatinase of nitrilase gene clusters, or acquired by more recent lateral transfers like is the case for a number of genes involved in adaptation to P-depletion. In the latter case, genes might sweep into populations independently of the host genome and behave as entities having their own ecological niche (Polz et al., 2013). Such examples was less frequent in *Synechococcus*, in which each module (and sub-module) was found to be highly 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 prone to acquisition of beneficiary genes by lateral transfers. 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 somewhat 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 benefit for the cell. The situation seems to be different in *Prochlorococcus,* in which genome streamlining likely frees more rapidly the genome from such genomic burdens. Although this analysis of the niche partitioning of genes along the *Tara* Oceans transect clearly brought some important new insights about niche adaptation in picocyanobacteria, forthcoming studies should certainly focus on the distribution of sequence variants, since there are clearly critical to the adaptation to cold niches, but also 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 in the future.

Supplementary information

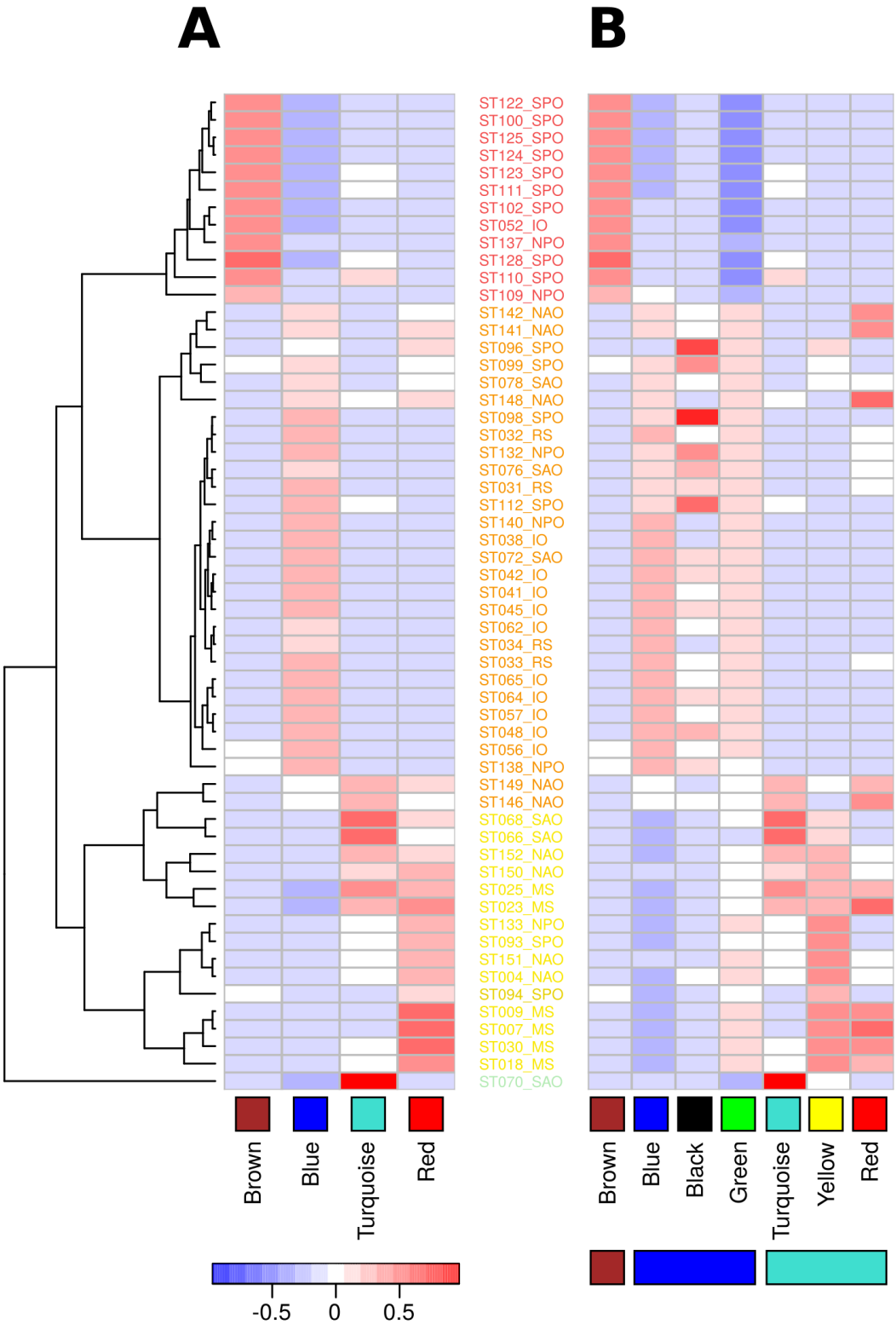


Figure S1. Distribution of the eigengene of each *Prochlorococcus* (sub)module at the different stations along the *Tara* Oceans transect. Eigengenes of modules (A) and submodules (B), designated by color names indicated below each heatmap, represent a consensus of the normalized relative abundance of genes at *Tara* Oceans stations. Station names are colored according to ESTU assemblages defined in Farrant et al. (2016) and specify the oceanic region of each station as follows: SAO, South Atlantic Ocean; MS, Mediterranean Sea; NAO, North Atlantic Ocean; IO, Indian Ocean; RS, Red Sea; SPO, South Pacific Ocean; NPO, North Pacific Ocean.

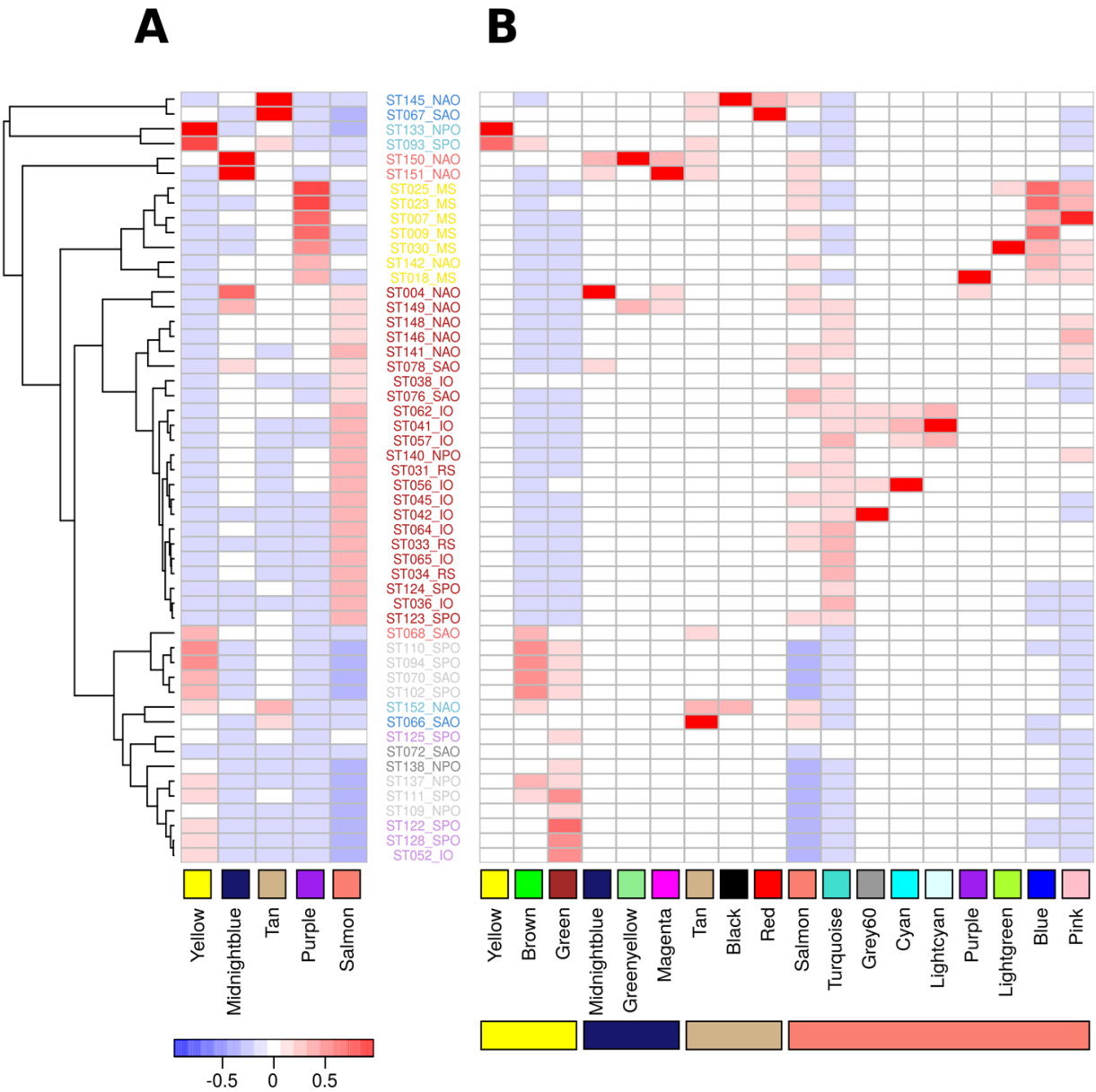


Figure S2. Same as Fig. S1 for *Synechococcus*.

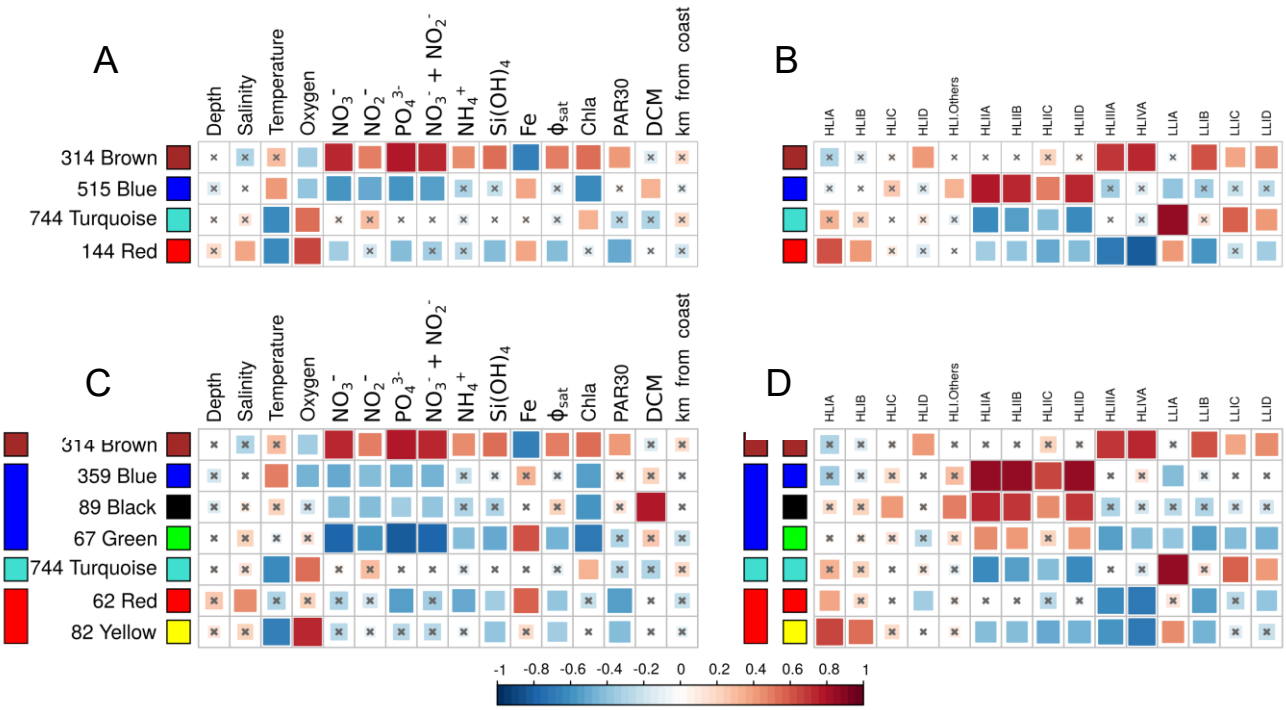


Figure S3. Correlation of *Prochlorococcus* submodule *eigengenes* to physico-chemical parameters and ESTU abundance. A, C. Correlation of eigenges to physico-chemical parameters for modules (**A**) and submodules (**C**). **B, D.** Correlation of eigenges to relative abundance profiles of *Prochlorococcus* ESTUs for modules (**B**) and submodules (**D**). Pearson (**A, C**) and Spearman (**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.

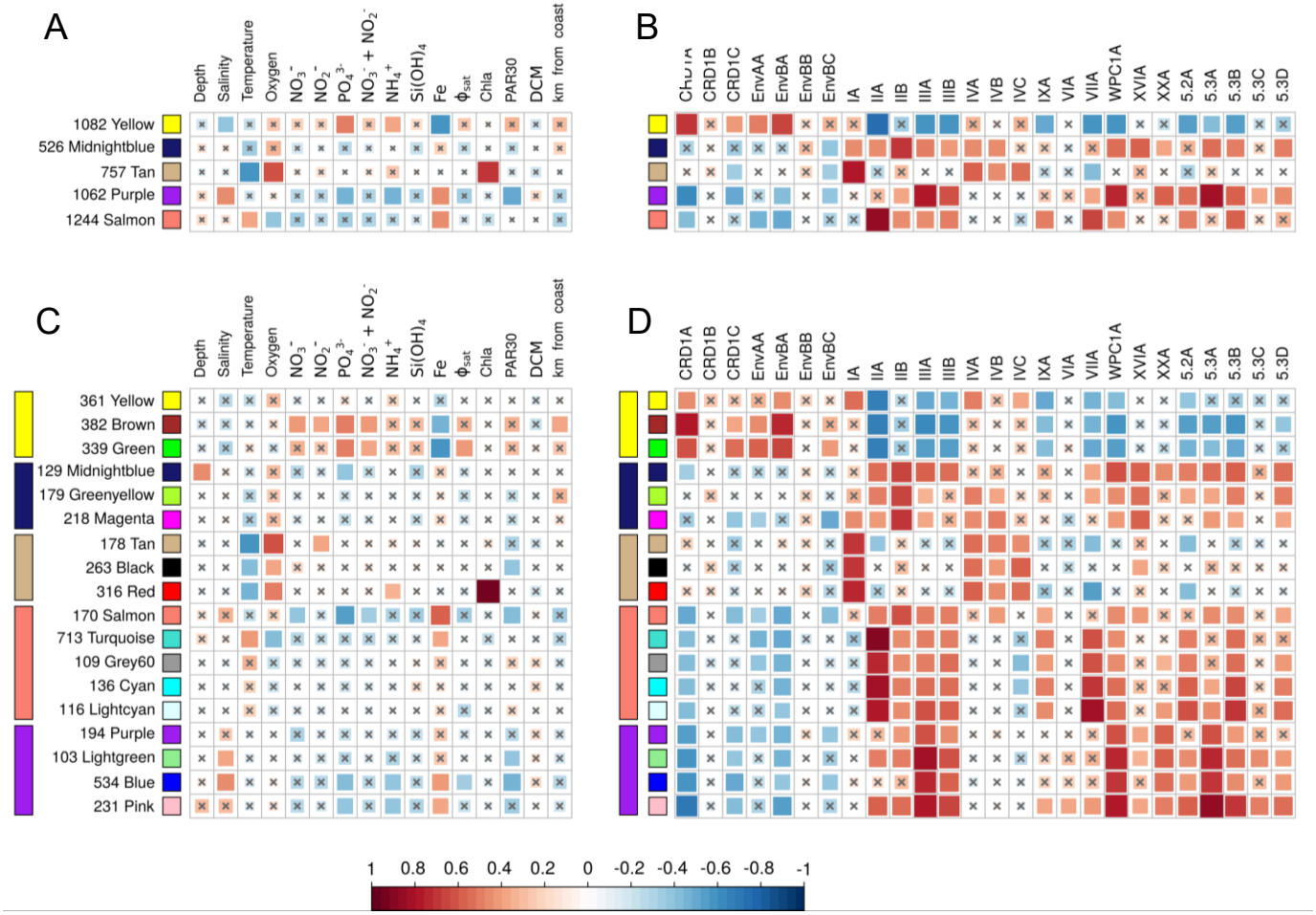


Figure S4. Same as Fig. S3 for *Synechococcus*

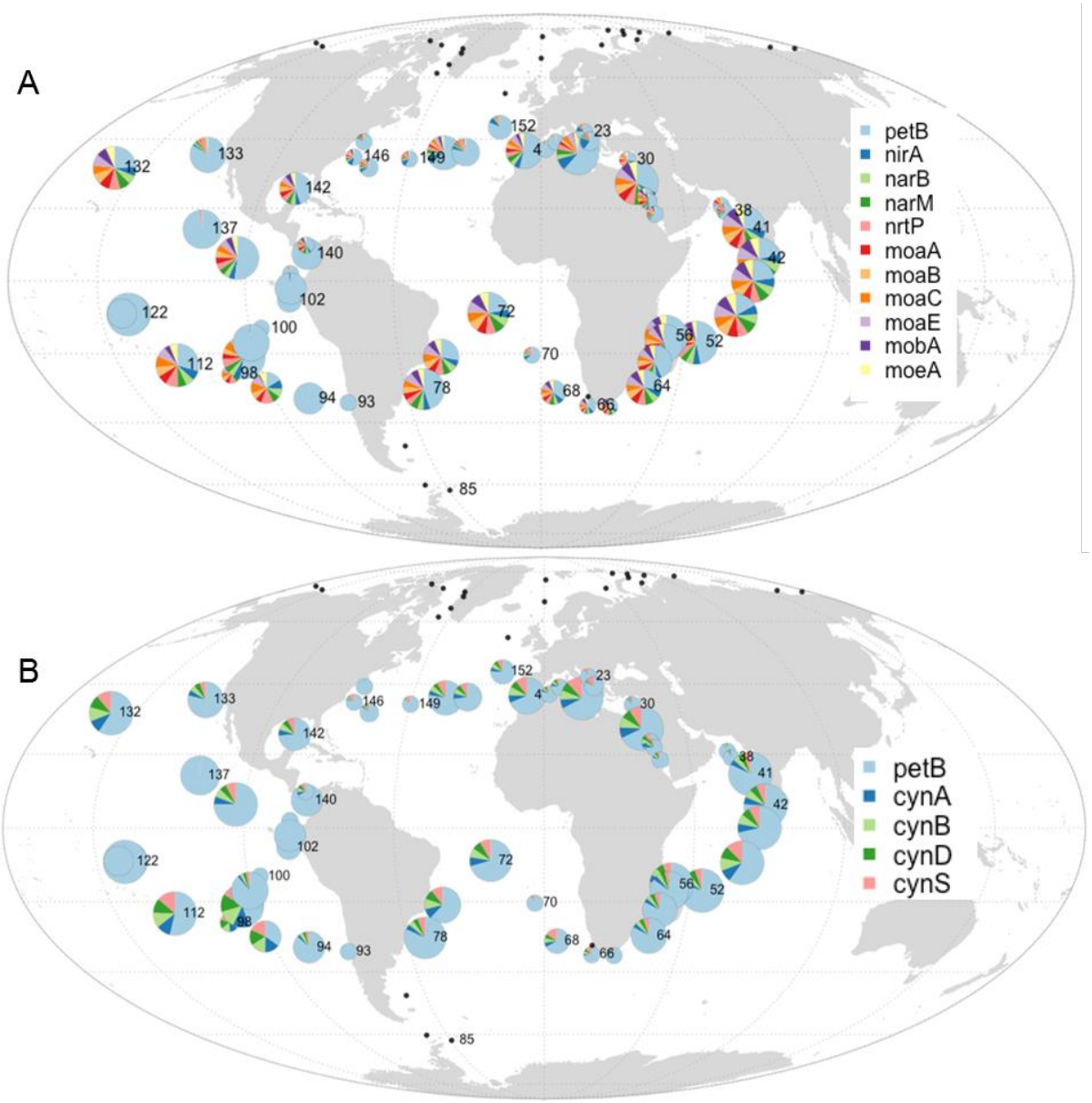


Figure S5 : Global distribution map of the *Prochlorococcus* blue module gene clusters putatively involved in transport and assimilation of nitrogen. The size of the circle is proportional to relative abundance of *Prochlorococcus* as estimated based on the *petB* gene and this gene was also used to estimate the relative abundance of other genes in the population. **A.** Nitrites and nitrates , **B.** Cyanate.

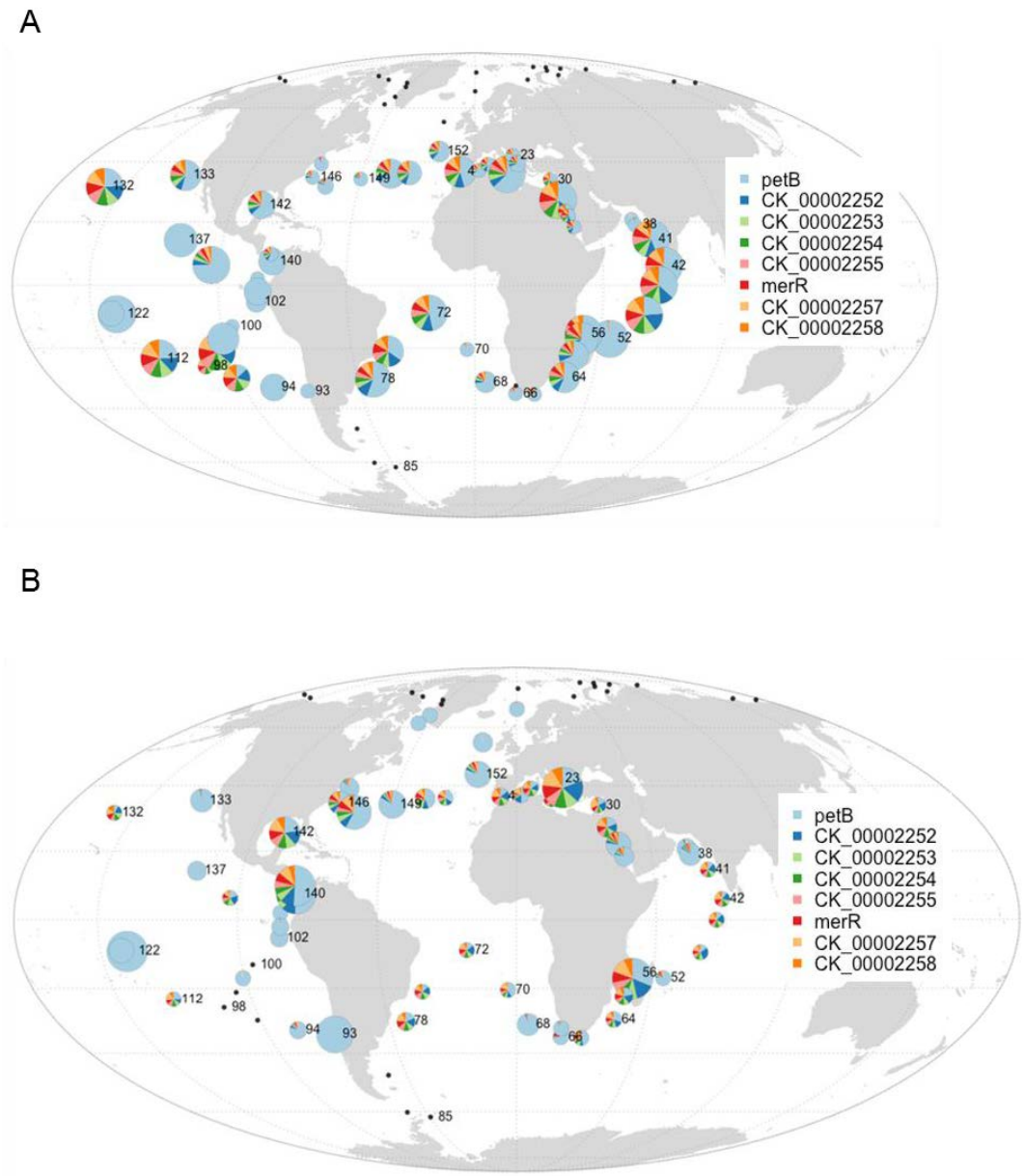


Figure S6 : Global distribution map of the nitrilase gene cluster. The size of the circle is proportional to relative abundance of each genus as estimated based on the *petB* gene and this gene was also used to estimate the relative abundance of other genes in the population. **A.** *Prochlorococcus*, **B.** *Synechococcus*.

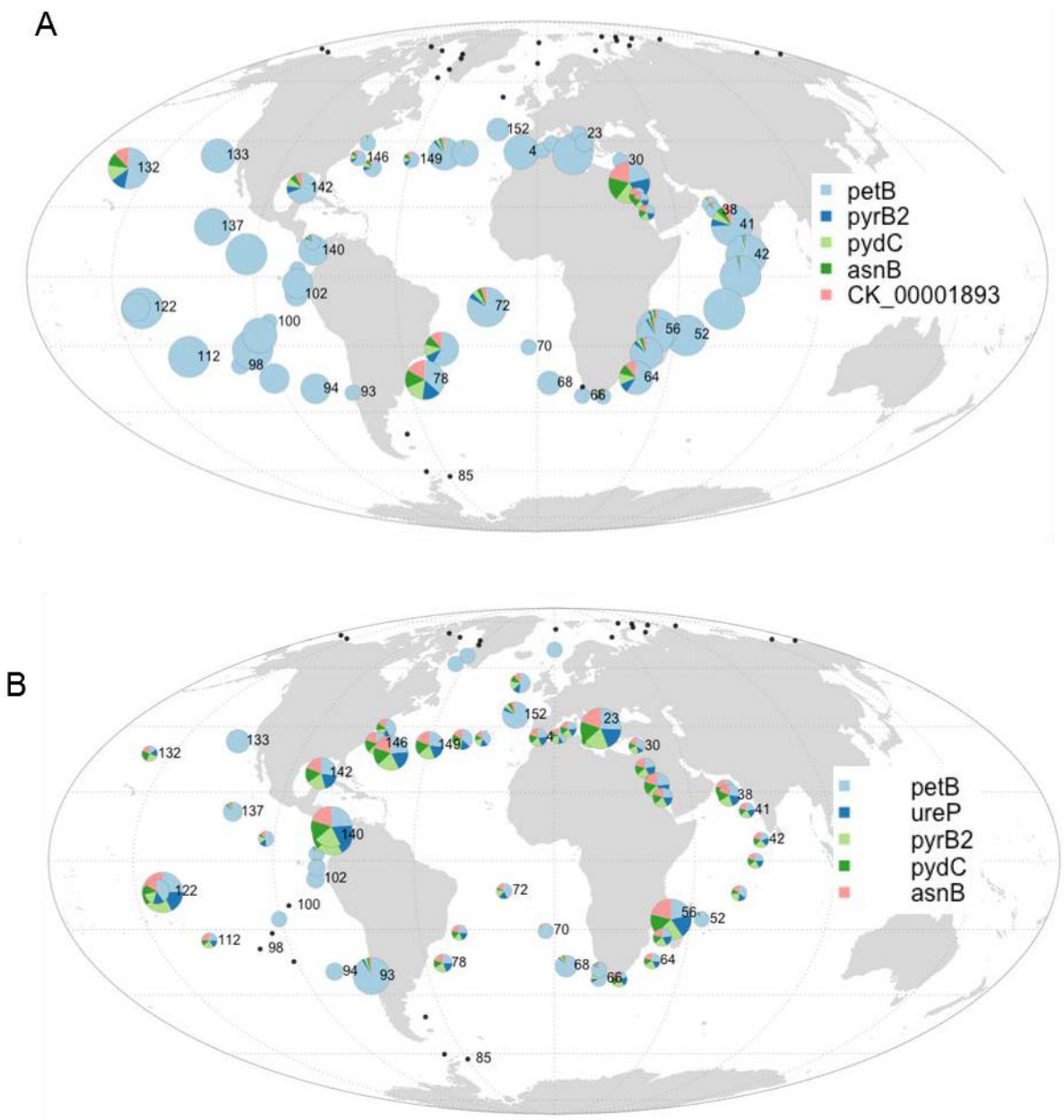


Figure S7 : Global distribution map of picocyanobacterial genes involved in the biosynthesis of pyrimidines The size of the circle is proportional to relative abundance of each genus as estimated based on the *petB* gene and this gene was also used to estimate the relative abundance of other genes in the population. **A.** *Prochlorococcus*, **B.** *Synechococcus*

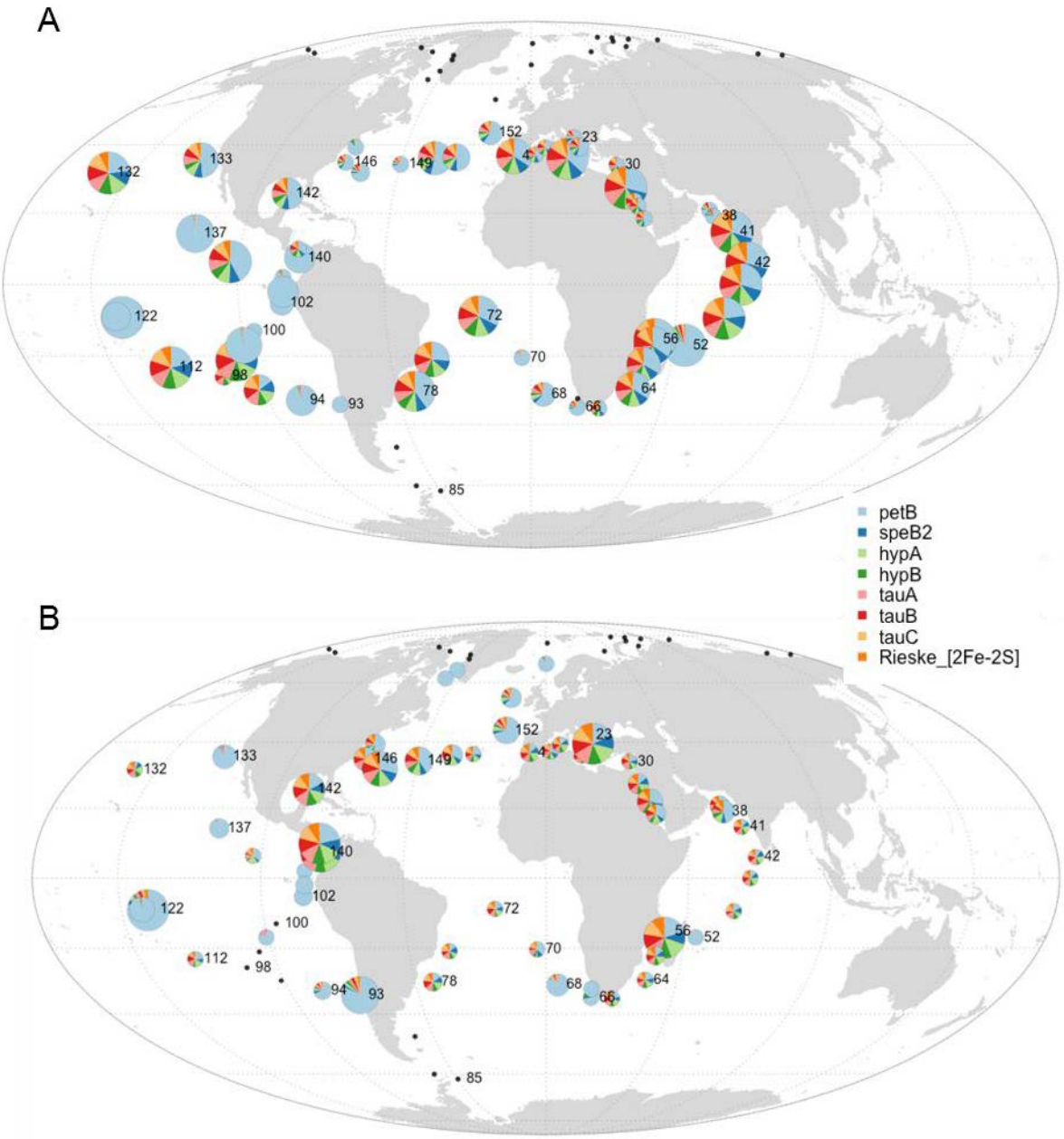


Figure S8 : Global distribution map of the agmatinase gene cluster. The size of the circle is proportional to relative abundance of each genus as estimated based on the *petB* gene and this gene was also used to estimate the relative abundance of other genes in the population. **A.** *Prochlorococcus*, **B.** *Synechococcus*.

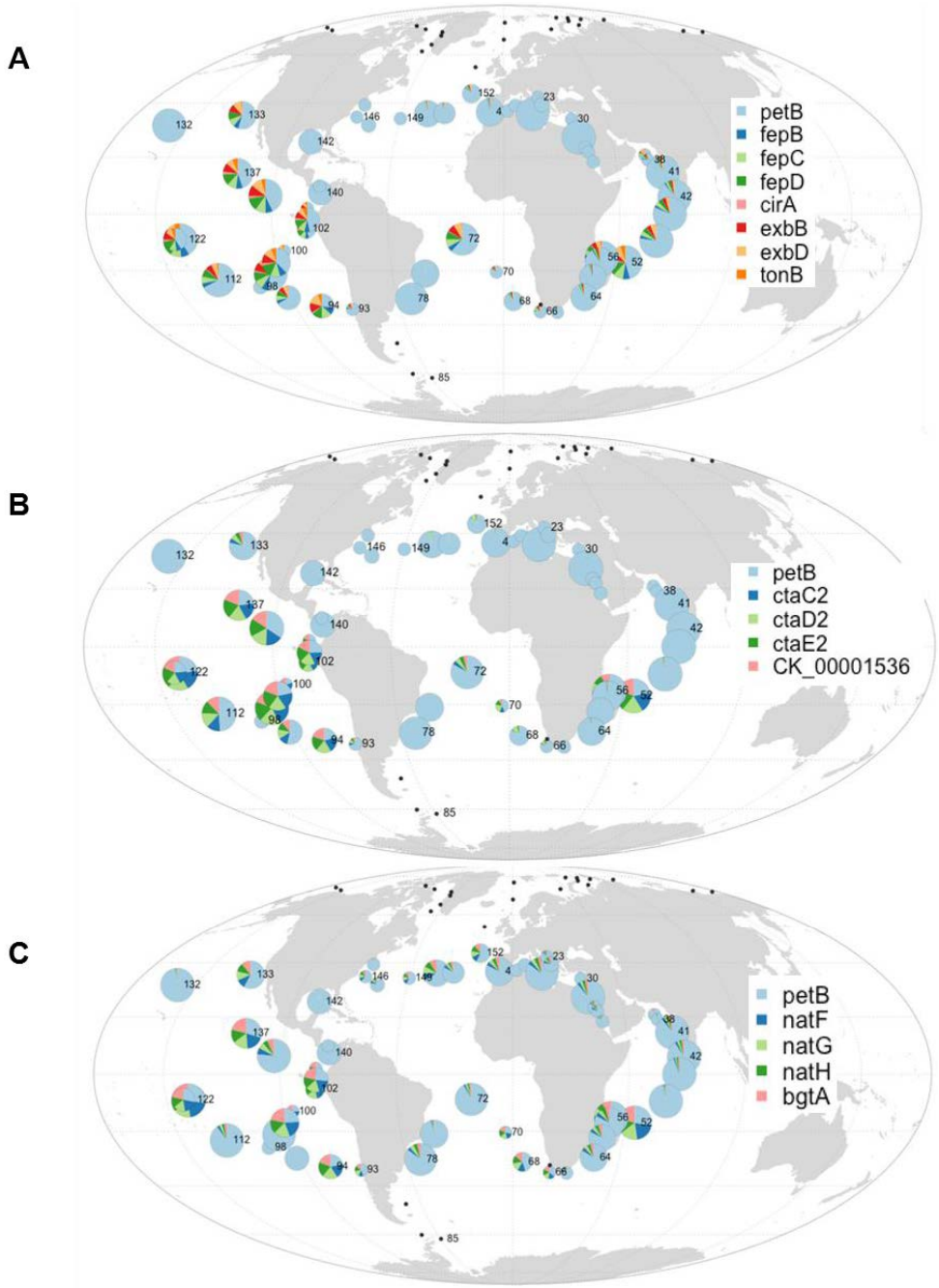


Figure S9 : Global distribution map of the *Prochlorococcus brown* module gene clusters putatively involved in adaptation to low-Fe. The size of the circle is proportional to relative abundance of each genus as estimated based on the *petB* gene and this gene was also used to estimate the relative abundance of other genes in the population. **A.** *siderophores*, **B.** alternative respiratory terminal oxidase (ARTO), **C**. ABC-type uptake transporter for acidic and neutral polar amino acids.

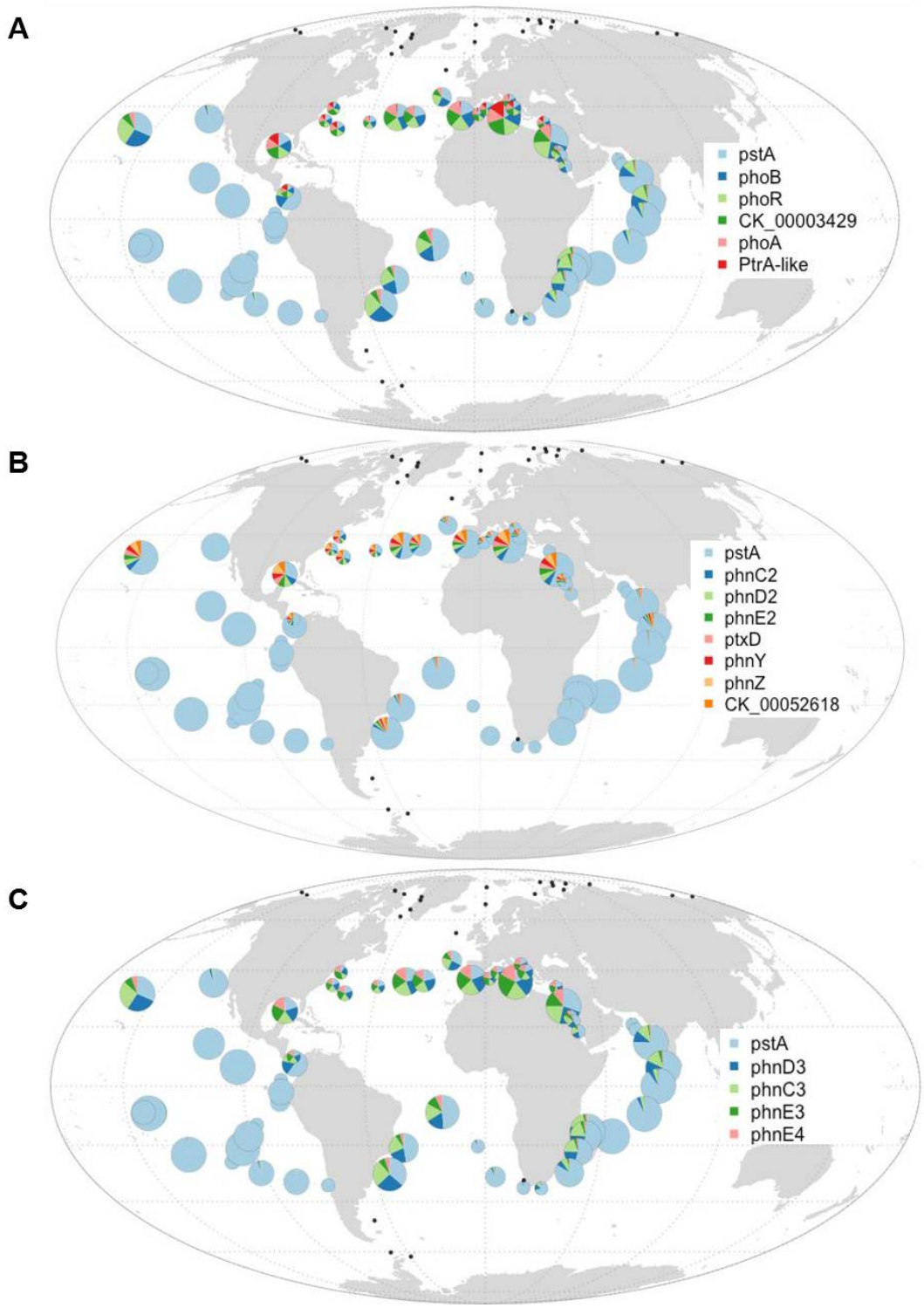


Figure S10 : Global distribution map of the *Prochlorococcus red* module gene clusters putatively involved in adaptation to P-depletion. The size of the circle is proportional to relative abundance of *Prochlorococcus* as estimated based on the *petB* gene, while the core gene *pstA* was used to estimate the relative abundance of red modules genes in the *Prochlorococcus* population. **A.** Phosphate acqquisition genes, **B.** Phosphonate acquisition gene cluster phnC2/D2/E2, **C.** Phosphonate acquisition gene cluster phnD3/C3/E3/E4.

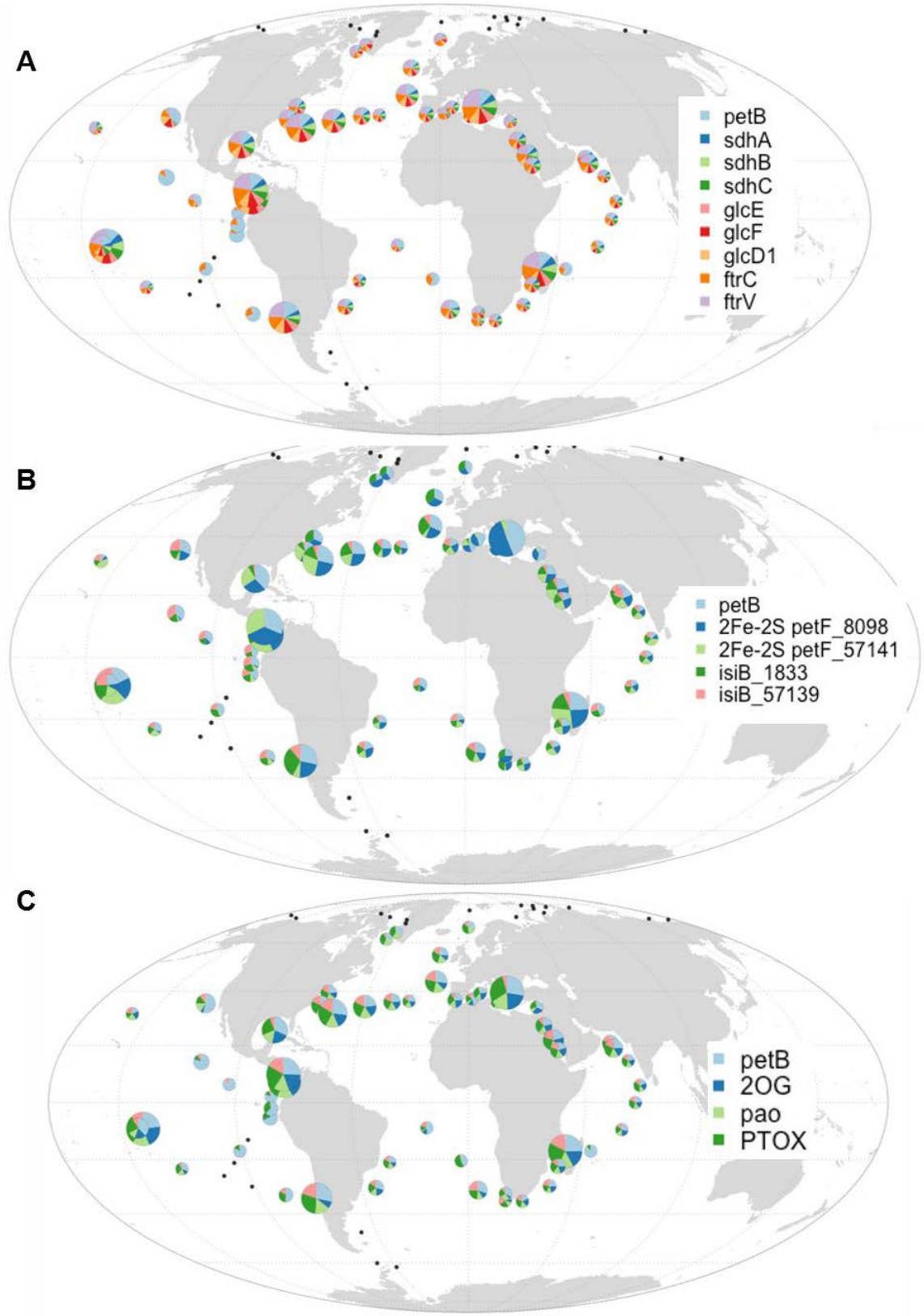


Figure S11 : Global distribution map of the selected *Synechococcus* Fe-containing proteins or proteins using alternative metals. The size of the circle is proportional to relative abundance of *Synechococcus* as estimated based on the *petB* gene and this gene was also used to estimate the relative abundance of other genes in the *Synechococcus* population. **A.** Fe-rich respiratory proteins, **B.** Ferredoxins *vs.* flavodoxins, **C.** Other Fe-rich proteins**.**

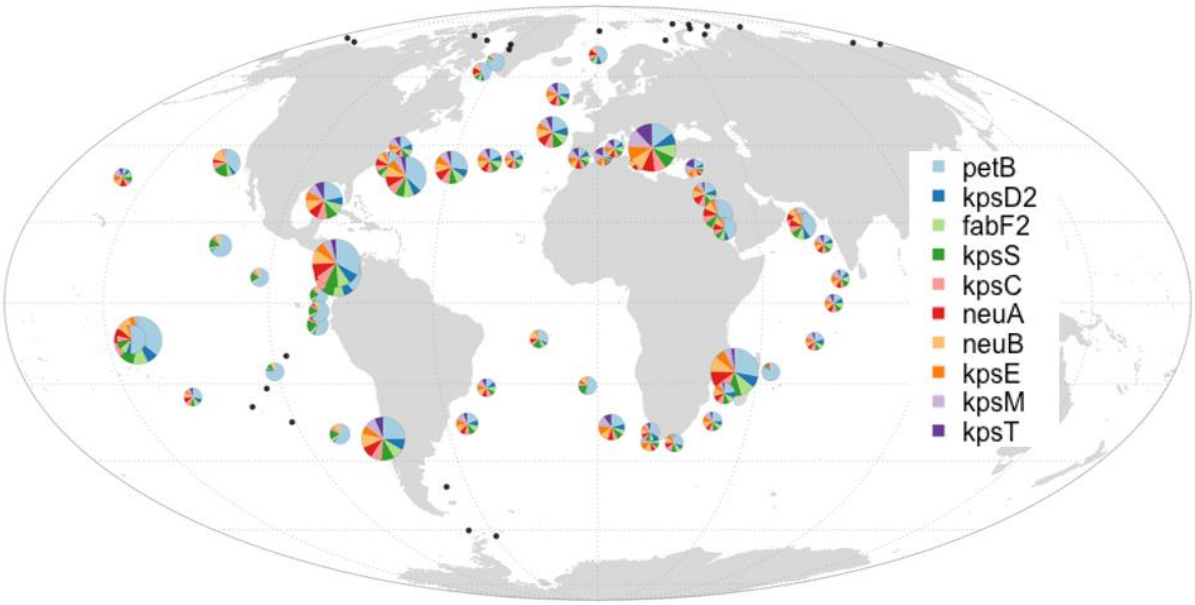


Figure S12 : Global distribution map of *Synechococcus* proteins involved in biosynthesis and export of capsular polysaccharides. The size of the circle is proportional to relative abundance of *Synechococcus* as estimated based on the *petB* gene and this gene was also used to estimate the relative abundance of other genes in the *Synechococcus* population.

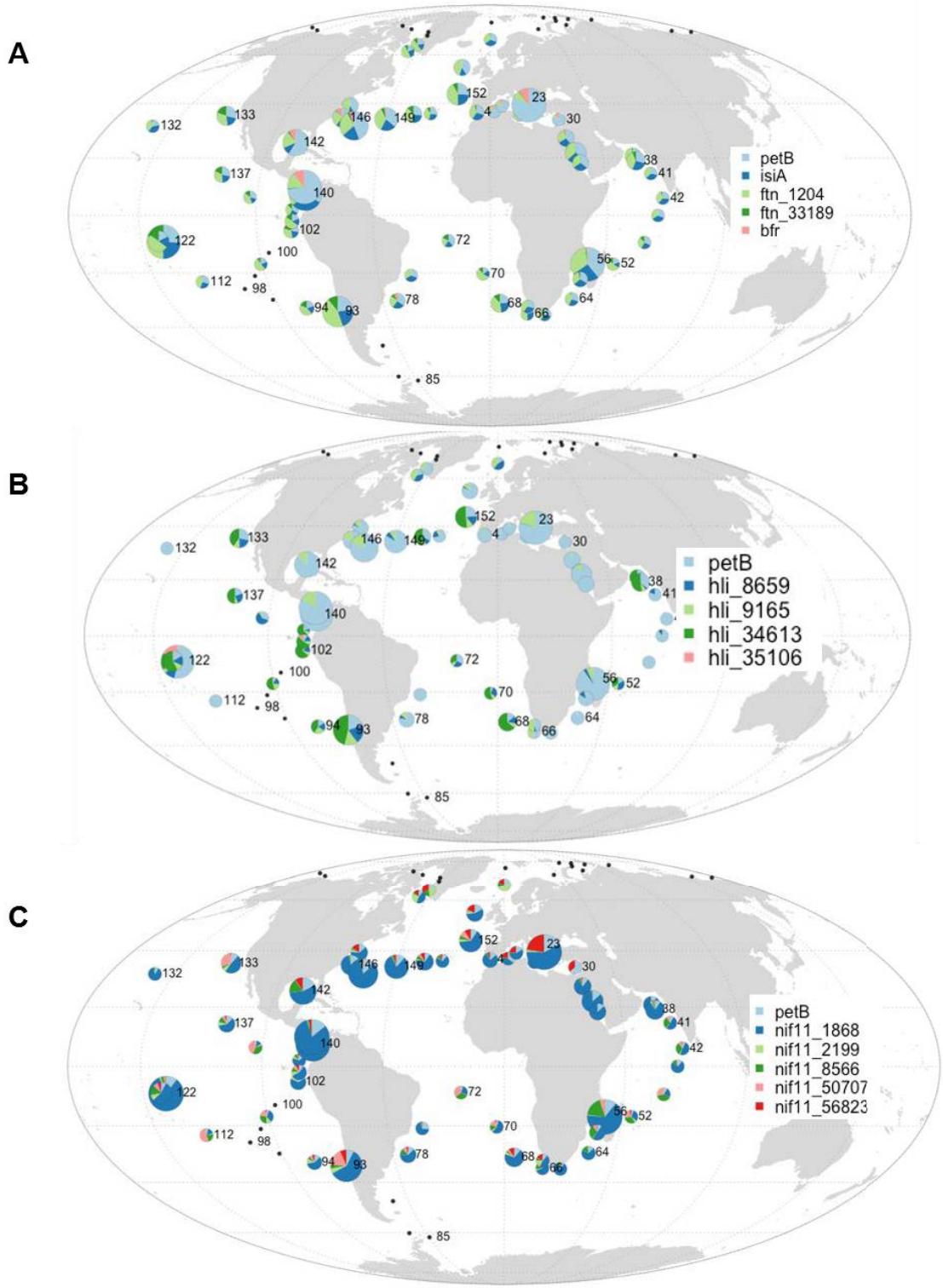


Figure S13 : Global distribution map of other selected *Synechococcus* proteins specifically enriched in Fe-replete environments. The size of the circle is proportional to relative abundance of *Synechococcus* as estimated based on the *petB* gene and this gene was also used to estimate the relative abundance of other genes in the *Synechococcus* population. **A.** Genes involved in biosynthesis of amino acids and pyrimidines, **B.** Export of capsular polysaccharides.

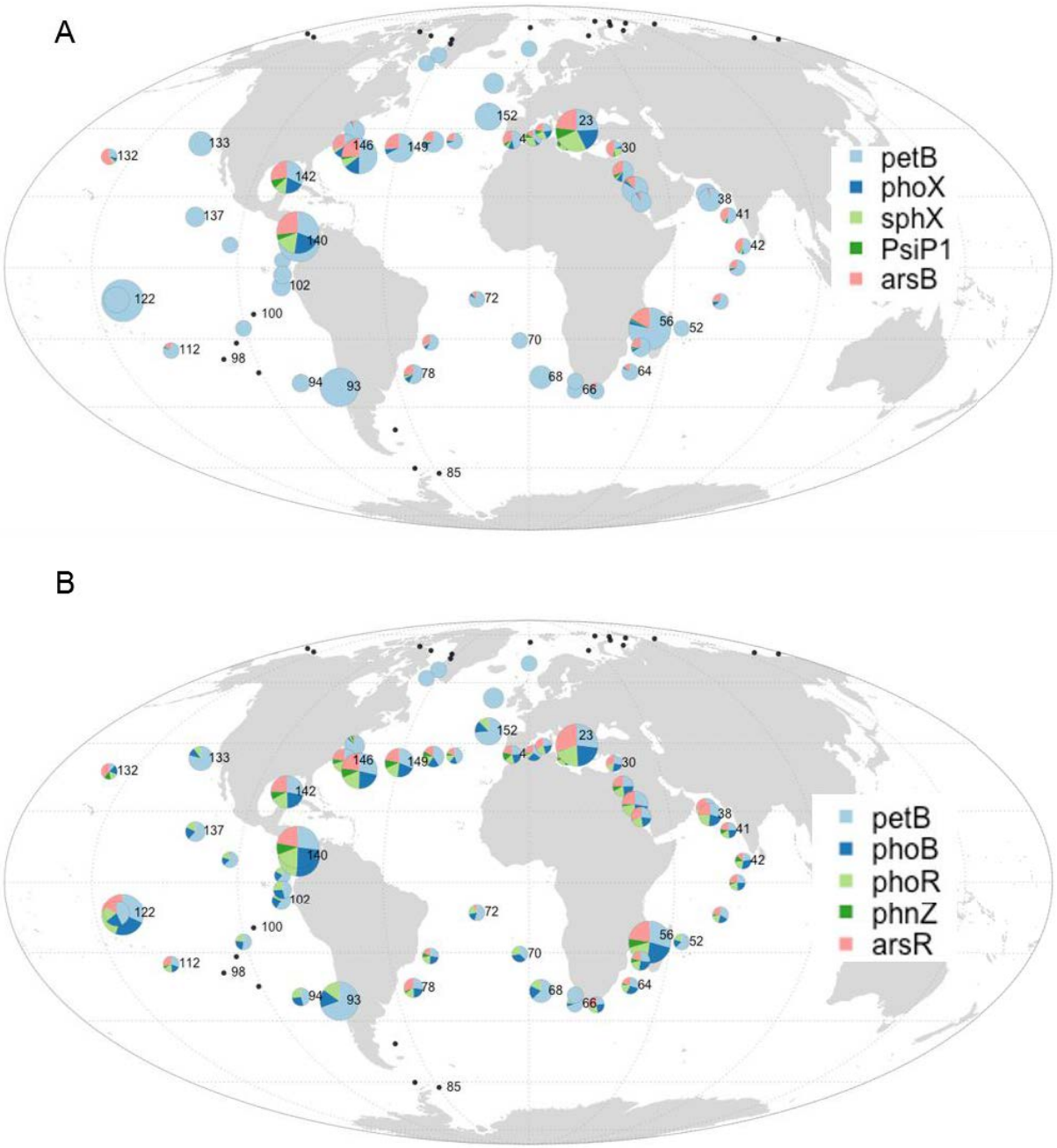


Figure S14 : Global distribution map of *Synechococcus* P-related genes. The size of the circle is proportional to relative abundance of *Synechococcus* as estimated based on the *petB* gene and this gene was also used to estimate the relative abundance of other genes in the *Synechococcus* population. **A.** Salmon module, **B.** Purple module.

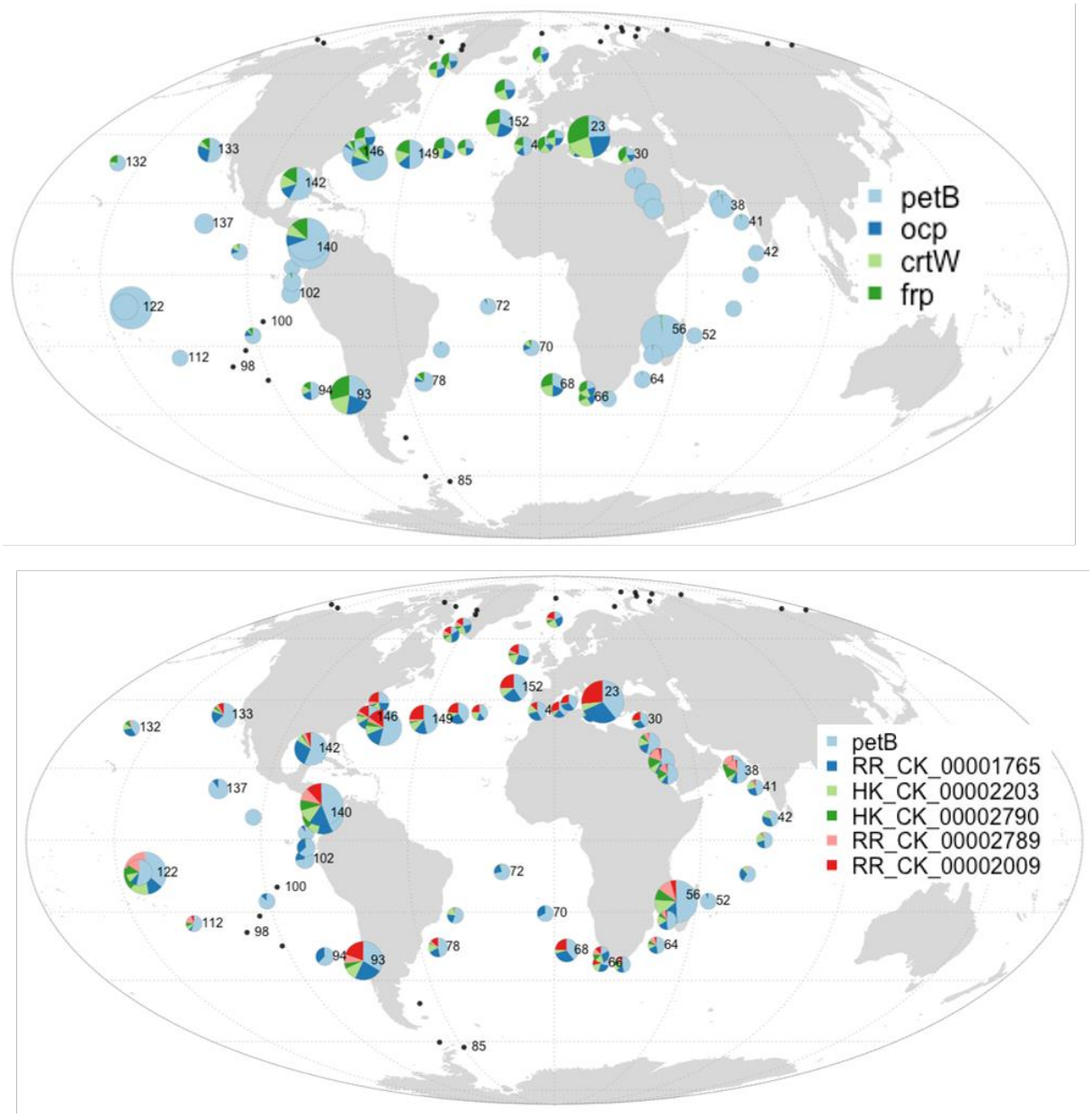


Figure S15 : Global distribution map of key genes of the *Synechococcus* tan module, representative of temperate and cold, chlorophyll-rich waters. The size of the circle is proportional to relative abundance of *Synechococcus* as estimated based on the *petB* gene and this gene was also used to estimate the relative abundance of other genes in the *Synechococcus* population. **A.** *ocp* operon*,* **B.** Two-components systems.

Supplementary Tables

Dataset 1: Composition of *Prochlorococcus* and *Synechococcus* gene modules. 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 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 *Prochlorococcus* gene modules. Same as Dataset 1, but for annotated genes. Genes were kept only if some information about its function was available.

Dataset 3: Same as dataset 2 but for *Synechococcus*

III.2 Distribution globale des gènes de désaturases, impliqués dans la thermoregulation de membranes, chez les picocyanobactéries marines

L'analyse des données méta-omiques de l'expédition *Tara* Oceans m'ont non seulement permis d'identifier des gènes de niches d'un point de vue global, comme décrit dans la partie précédente, mais également d'analyser la distribution d'un ou plusieurs gènes spécifiques comme j'ai pu le faire dans le cadre de l'étude de (Breton et al., 2019 ; en annexe 1) visant à mieux comprendre les bases physiologiques de la spécialisation des différents thermotypes (ou écotypes thermiques) chez *Synechococcus*.

En effet, bien que la température constitue l'un des paramètres les plus discriminants pour la distribution latitudinale des ESTUs de *Synechococcus* (Farrant *et al*., 2016), les mécanismes responsables de ces adaptations restent peu connus. Jusqu'à présent, seules quelques études ont mis en évidence une différence dans le maintien de l'activité photosynthétique des différents thermotypes en réponse aux stress thermiques (Pittera *et al.*, 2014) et également une thermostabilité différentielle des protéines constituant les phycobilisomes (Pittera *et al.*, 2017). On sait par ailleurs que la température a un effet direct sur la fluidité membranaire qui elle-même influence la conformation et donc l'activité des protéines membranaires (Mikami and Murata, 2003) et notamment des photosystèmes I et II. C'est pourquoi la capacité des cellules à moduler la fluidité de ces membranes est indispensable pour maintenir la fitness des organismes photosynthétique dans une niche thermique donnée (Mackey *et al*., 2013; Varkey *et al.*, 2016). Dans ce contexte, une étude récente du laboratoire a montré qu'en réponse au froid, *Synechococcus* sp. WH7803 est capable d'induire un raccourcissement des chaînes d'acide gras composant des lipides membranaires (conduisant à un amincissement de la membrane) et d'activer la désaturation des acides gras à des sites spécifiques des trois glycolipides présents chez cet organisme (Pittera et al., 2018). Des insaturations sont insérées dans les chaînes d'acides gras par des enzymes de la famille des désaturases permettant de maintenir la fluidité de la membrane (Chi et al., 2008; Varkey et al., 2016). La comparaison des génomes disponibles de *Synechococcus* a également permis d'identifier quatre membres principaux de cette famille, codant pour deux Δ9-désaturases (*desC3* et *desC4*) et deux Δ12-désaturases (*desA2* et *desA3*) dont le pattern phylétique diffère entre les thermotypes chauds des clades II et III d'une part et les thermotypes froids des clades I et IV d'autre part (Tableau 1). Ces résultats suggèrent que les capacités de désaturation pourraient avoir joué un rôle clé dans la spécialisation des lignées de *Synechococcus* à différentes niches thermiques (Pittera et al., 2018). La comparaison des lipidomes de souches représentatives des clades I à IV (1 souche par clade), acclimatées à différentes température, a permis de montrer que pour les désaturases à mono-insaturation, DesC3 serait une désaturase commune à toutes les souches, faiblement régio-spécifique mais avec une préférence pour la position sn-1 des chaines en C16 (chaînes carbonées contenant 16 atomes de carbones) alors que DesC4, qui est spécifique des thermotypes froids, permettrait d'accroître la capacité de monoinsaturation des lipides membranaires

de ces souches. En ce qui concerne les désaturases responsables des doubles insaturations, DesA2, présent spécifiquement chez les thermotypes chauds, permettrait d'obtenir de plus hauts niveaux de C16:2 (chaînes carbonées composées de 16 atomes de carbone et présentant 2 insaturations) et serait fortement régio-spécifique en agissant sur la chaine sn-1 des galactolipides. Enfin DesA3, qui semble avoir été spécifiquement perdu par les membres du clade II, serait également fortement régio-spécifique et permettrait d'accroître la quantité de C16:2 chez les membres du clade III alors qu'elle est présente mais semble inactive chez les thermotypes froids. Dans ce contexte, ma contribution à cet article a consisté à analyser la distribution globale des gènes codant pour les 4 désaturases dans les stations *Tara* Oceans dominées par un ou plusieurs des quatre ESTUs majeurs (IA, IIA, IIIA et IVA) en fonction de la température de prélèvement de l'échantillon (Figure 16). Le gène *desC3* s'avère présenter une distribution constante quelle que soit la température, ce qui est cohérent avec le fait qu'elle est commune à (et semble active chez) toutes les souches de *Synechococcus* examinées. Le gène *desC4* est retrouvé uniquement dans les zones où la température est inférieure 20°C, en accord avec son rôle supposé d'accroissement du niveau de monoinsaturation des membranes des clades I et IV. En ce qui concerne les désaturases responsables des doubles insaturations, *desA2* présente une distribution inverse de celle de *desC4*, ce qui semble confirmer que cette désaturase serait limitée aux environnements chauds, colonisés par les clades II et III tandis que *desA3* présente une distribution similaire à *desC4*, à l'exception de la plupart des stations dominées par le clade III dans lesquelles ce gène est présent entre 17 et 25°C. Cette distribution suggère que la coexistence de *desA2* et *desA3* chez les membres du clade III pourrait leur conférer un avantage adaptatif dans des environnements soumis à des variations saisonnières importantes de la température, telles que celles observées en Mer Méditerranée (dominée par le clade III), par rapport aux membres du clade II, qui ne possèdent colonisent des environnements beaucoup plus stables et qui ne possèdent que *desA2*. Globalement, le fait que la distribution de ces désaturases *in situ* soit en accord avec les patrons phylétiques mis en évidence chez les souches représentatives des écotypes tend à confirmer que la capacité à moduler la fluidité membranaire pourrait avoir joué un rôle critique dans la colonisation des différentes niches thermiques par les *Synechococcus* marins.

Tableau 1 - Criblage du génome pour les gènes présumés de désaturase dans 53 génomes marins de *Synechococcus* et *Cyanobium*, classés par subclusters et les clades phylogénétiques issu de Pittera et al., 2018.

			Δ9 desaturases			Δ12 desaturases			Other desaturases					
Sub-cluster[1]	Clade[2]	Representative sequenced strains	*desC3*	*desC4*	*desC6*	*desA2*	*desA3*	*desA4*	*desC*	*desC*	*des*	*des*	*des*	Number of *des* genes
5.1	I	CC9311, MVIR-18-1, PROS-9-1, WH8016, ROS8604	■	■			■							3
		SYN20	■	■			■					■		4
	II	A15-62, CC9605, M16.1, RS9902, RS9907, TAK9802	■			■								2
		KORDI-52	■											1
		A15-44	■			■				■				3
		WH8109, PROS-U-1	■			■	■							3
	III	WH8102, WH8103, A15-24, A18-46.1, BOUM118, RS9915, A15-28, A18-40	■			■	■							3
	IV	BL107, CC9902	■	■			■							3
	V	WH7803, BMK-MC-1	■	■		■	■							4
	VI	WH7805, MEDNS5	■	■		■	■							4
		PROS-7-1	■	■		■	■		■		■			6
	VII	A15-60, A18-25c	■	■			■							3
		NOUM97013	■	■		■	■							4
	VIII	RS9909, RS9917	■	■			■							3
		WH8101	■	■										2
	IX	RS9916	■	■										2
	CRD1	MITS9220, BIOS-E4-1	■	■			■							3
		BIOS-U3-1	■	■		■	■							4
	WPC1	A15-127, KORDI-49	■			■	■							3
	XX	CC9616	■				■							2
	UC-A	KORDI-100	■											1
5.2		NS01, WH5701	■	■	■		■	■				■		6
		PCC6307	■	■	■		■	■					■	7
		CB0101	■	■		■	■	■		■				6
		CB0205	■	■		■	■	■						5
		PCC7001	■	■	■			■						4
5.3		RCC307, MINOS11	■	■			■							3

[1] *sensu* Herdman et al. (2001) ; [2] see Mazard et al. (2012) and Choi & Noh (2009).

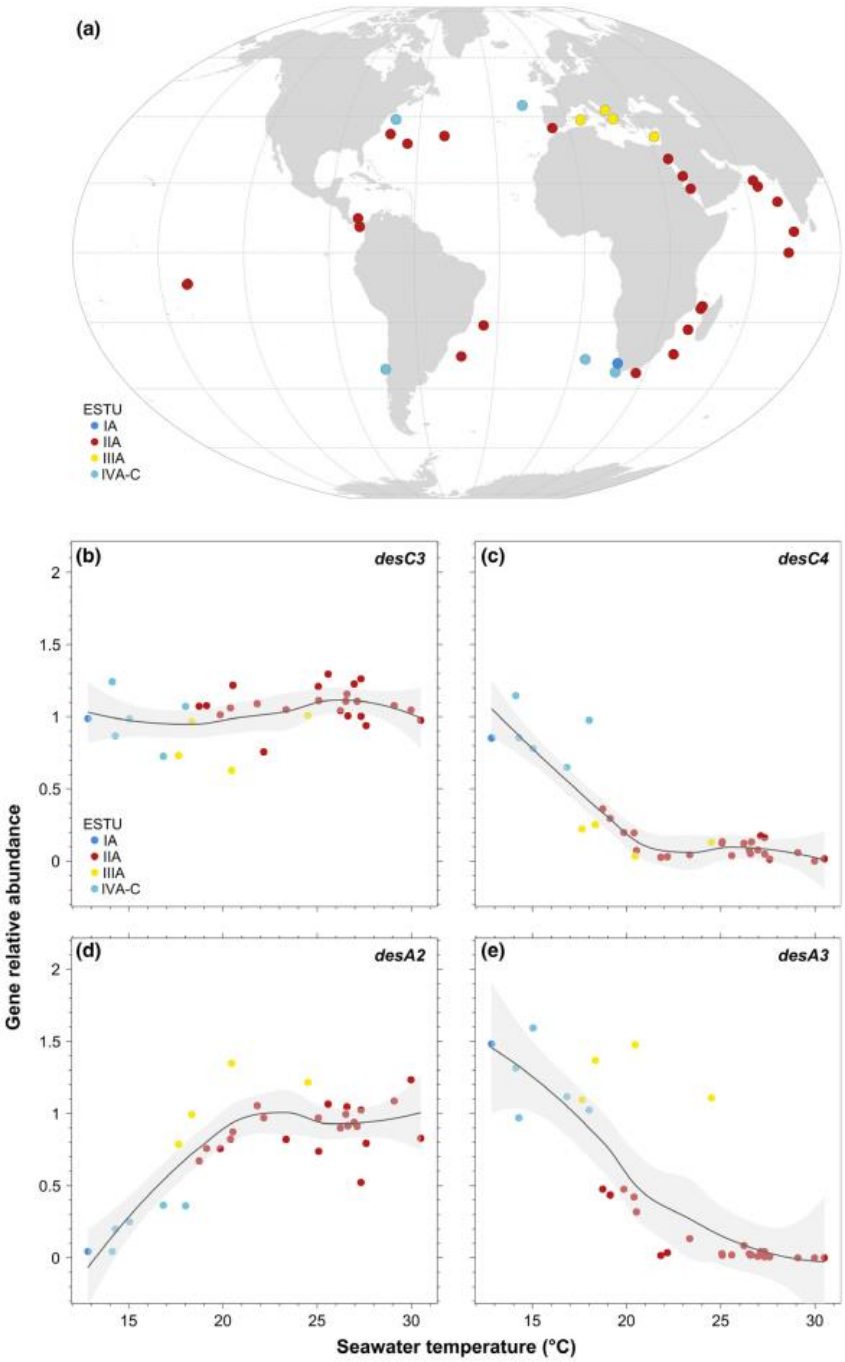


Figure 16 - Localisation des stations de l'expédition *Tara* Océans utilisées dans cette étude (a), et abondances relatives des gènes de désaturases lipidiques Δ9 *desC3* (b) et *desC4* (c) et de désaturases lipidiques Δ12 *desA2* (d) et *desA3* (e), en fonction de la température de surface aux différentes stations. Ces données ont été générées en utisant les métagénomes des stations dominées par les ESTUs ('ecologically significant taxonomic units', *sensu* Farrant et al., 2016) majeurs des clades I-IV de *Synechococcus*. A chaque station, l'abondance relative de chaque gène de désaturase a été normalisée à l'abondance relative du gène commun et simple copie *petB*. La courbe de 'fitting' a été réalisée en utilisant le modèle de régression LOESS (span α = 0.8) en utilisant le package R ggplot2. La zone grisée correspond à l'intervalle de confiance à 95% (Breton et al., 2019).

III.3 Physiologie comparative et acclimatation aux facteurs environnementaux des principaux écotypes de *Synechococcus*

Contexte du travail

L'analyse de la réponse au stress de la souche modèle *Synechococcus* sp. WH7803 a permis de mettre en évidence un certain nombre de caractéristiques communes de la réponse à divers stress environnementaux et d'autres plus spécifiques à un stress particulier (cf. Chapitre II). Cependant, cette souche appartient à un clade peu abondant dans le milieu naturel et dont la niche écologique reste assez mal connue, même si certaines études récentes semblent indiquer que le clade V serait plutôt présent dans les zones chaudes, côtières et dessalées, telles que certaines régions de la mer Méditerranée (Doré et al., en prep ; Marsan, 2016). De façon à vérifier dans quelle mesure la réponse de WH7803 est représentative du genre *Synechococcus*, des analyses physiologiques similaires à celles présentés dans le Chapitre II ont été réalisées sur quatre souches représentatives des clades dominants de *Synechococcus* (clade I-IV) mais en limitant les conditions testées aux fortes lumières et basses températures. En complément de ces analyses, j'ai également initié l'analyse des metatranscriptomes de *Tara* Oceans afin d'essayer de mettre en évidence directement à partir de ces échantillons environnementaux les capacités d'acclimatation des différents écotypes de *Synechococcus* en lien avec les niches écologiques qu'ils colonisent.

Matériel et méthodes

Choix des souches

Afin de mettre en évidence de potentielles différences dans la plasticité phénotypique entre les lignées, qui pourraient expliquer la distribution des populations naturelles de *Synechococcus*, nous avons sélectionné quatre souches répondant aux critères suivants : i) représentatives de quatre des cinq clades majeurs de *Synechococcus* dans le milieu naturel (clades I, II, III et IV), ii) de type pigmentaire 3 afin de limiter la variabilité en réponse à la lumière et enfin iii) ayant un génome séquencé, afin de faciliter les analyses de transcriptomique et d'utiliser les données de génomique pour interpréter la réponse physiologique de ces organismes. Les souches sélectionnées sont répertoriées dans le Tableau 2, incluant la souche WH7803 pour comparaison.

Tableau 2 - Caractéristiques des souches de *Synechococcus* utilisées dans cette étude

Souche	*MVIR-18-1*	*A15-62*	*WH8102*	*BL107*	*WH7803*
Clade	I	II	III	IV	V
ESTU	IA	IIA	IIIA	IVA	VA
Type pigmentaire	3a	3dB	3c	3dA	3a
Numéro d'accession	-	-	BX548020	AATZ00000000	CT971583
Axenie	Non	Non	Oui	Non	Oui
Site d'isolement	Mer de Norvège Sud	Océan Atlantique	Mer des Caraïbes	Mer Méditerranée	Mer des Sargasses
Latitude d'isolement	61°00' N	17°37'N	22°29'N	41°43'N	33°45'N
Longitude d'isolement	1°59' E	20°57'W	65°36'W	3°33'E	67°30'W
Date d'isolement	23/07/2007	04/10/2004	15/03/1981	15/03/2001	01/07/1978
Prof d'isolement (m)	25	15	-	1800	25
Température* d'isolement	13.98°C	26.02°C*	25.78°C*	13.89°C*	25.85°C*

* D'après (Pittera *et al.*, 2014)

Conditions expérimentales

Les conditions de culture étaient les mêmes que celle utilisées pour la souche WH7803 (cf. Chapitre II) mais les traitements réalisés se sont limités à des transferts à forte lumière (250 µE.m^{-2}.s^{-1}, HL pour 'high light') ou à basse température (13°C, LT pour 'low temperature') à partir de cultures acclimatées à faible lumière (20 µmol photon.m^{-2}.s^{-1}, LL) et à 22°C. De même que pour WH7803, les paramètres suivants ont été mesurés à différents temps de stress : le taux de croissance des cellules par cytométrie en flux, les spectres d'émission de fluorescence entre 545 et 750 nm (exc: 530 nm) par spectrofluorimétrie afin de calculer les rapports PE/PC et PC/TA, permettant de déterminer le couplage des phycobiliprotéines et le

potentiel démantèlement des bras des PBS et enfin le rendement quantique du PSII (F_V/F_M) et le taux de réparation du PSII par fluorimétrie PAM. Par ailleurs, nous avons également mesuré lors de ces expériences le contenu pigmentaire par HPLC, et le contenu en lipides, qui ont fait l'objet d'une publication indépendante (Breton et al., 2019). Enfin, 108 transcriptomes ont été générés à différents temps de stress (Tableau 3) dans les mêmes conditions que pour la souche WH7803.

Tableau 3 – Description du plan expérimental

Condition d'acclimatation	Condition de stress	Temps d'échantillonnage (heure)	Nombre de réplicats biologiques
LL	LLHL	0, 0.3, 1, 3, 6, (7, 24)	3
LL	LLLT	0, 12, 24, 36, 48	3

Analyse des métatranscriptomes de Tara Océans

Les metatranscriptomes de l'expédition *Tara* Océans ont été analysés en utilisant la même approche que celle utilisée pour les métagénomes (cf. Chapitre III, partie II). En bref, les lectures recrutées par BLASTN (v2.2.28+) sur les 97 génomes de picocyanobactéries de Cyanorak v2.1 (http://sb-roscoff.fr/Phyto/cyanorak) ont tout d'abord été taxonomiquement assignés à *Prochlorococcus* ou *Synechococcus* en fonction de leur meilleur hit blast, puis fonctionnellement assignés à un CLOG en fonction de la position de ce hit sur le génome le plus proche, cette position correspondant à un gène particulier. Pour chacun des deux genres, le cumul des lectures par CLOG a permis d'obtenir un tableau de contingence répertoriant l'abondance relative de chaque transcrit dans chacune des stations *Tara*. Les profils d'abondance relative obtenus ont été utilisés pour regrouper les stations en utilisant une similarité de Bray-Curtis et pour effectuer des analyses de réseaux de cooccurrence.

Résultats et discussion

Réponse physiologique au stress

Les variations de rendement quantique du PSII montrent que les cinq souches sont différentiellement affectées par les stress lumineux et thermiques (Figure 17). Le PSII de MVIR-18-1 (clade I) et A15-62 (clade II) semble peu affecté par les fortes lumières (baisse de 10 % de leur F_V:F_M en 6h), par rapport à celui des souches BL107 (clade IV) et WH8102 (clade III), qui présentent une forte diminution de leur F_V:F_M après 6h de stress (40% et 60%, respectivement). Ce comportement semble être en partie lié à la capacité de réparation du PSII des souches puisque MVIR-18-1 et A15-62 sont les deux souches qui présentent le plus fort taux de réparation en HL et la plus forte induction de ce taux entre les conditions LL et HL (Figure 18B). En revanche, BL107 et WH8102 présentent un taux de réparation du PSII initial significativement plus élevé dans les cultures acclimatées en LL et une plus faible induction de cette réparation en réponse au stress HL. En comparaison, la souche modèle WH7803 (clade V) montre une sensibilité intermédiaire du PSII et une capacité de réparation plus faible à HL bien qu'elle soit capable de maintenir un F_V:F_M à 80 % de son niveau initial après 6h de stress lumineux. Ceci suggère que cette souche présenterait une plus faible 'cross section' de photoinactivation (σ_i) en HL que les autres souches (Six et al., 2007b). Dans tous les cas, les souches semblent capables de récupérer de ce stress lumineux et de rétablir un F_V:F_M proche du T0, après avoir été replacées dans les conditions initiales de lumière, montrant que ce stress n'est pas létal pour ces organismes.

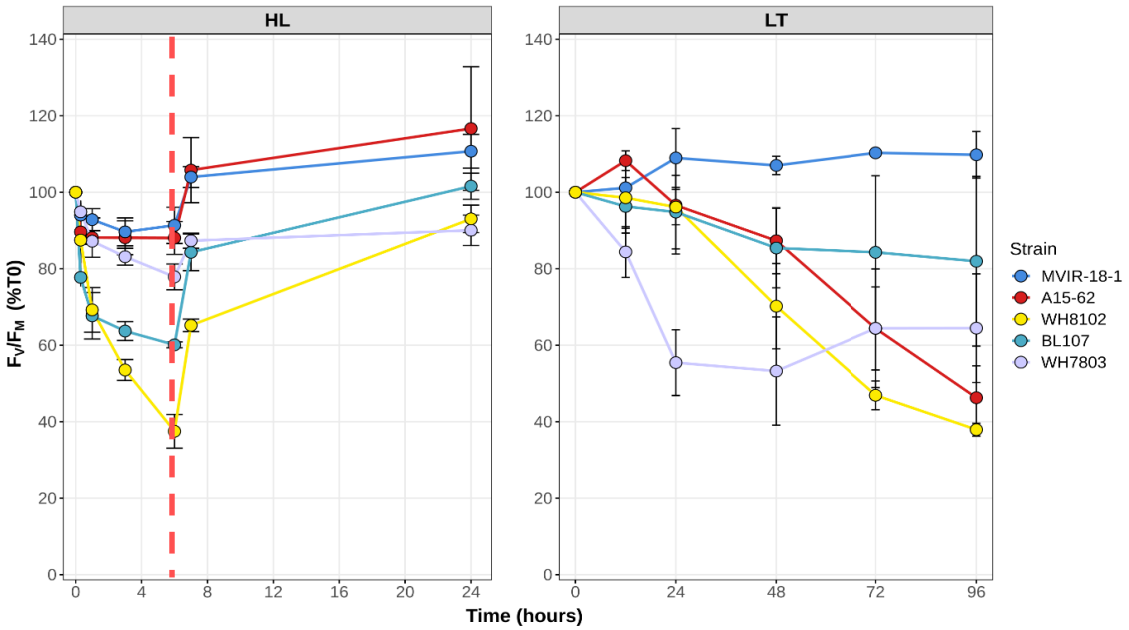


Figure 17 - Variation du rendement quantique du PSII (FV/FM) en réponse au stress dû aux fortes lumières (HL, 250 µE.m-2.s-1) et aux basses températures (LT, 13°C) pour différentes souches de Synechococcus acclimatées en faible lumière (20 µE.m-2.s-1) et à 22°C. Ces données sont basées sur au moins trois expériences indépendantes.

En ce qui concerne le stress thermique, les deux thermotypes froids, MVIR-18-1 (clade I) et BL107 (clade IV), semblent pas ou peu affectés par ce stress alors que les trois thermotypes chauds présentent une diminution de leur F_V:F_M de 40 à 60 % mais avec une dynamique qui diffère selon les souches. En effet, alors que la souche WH7803 présente déjà un maximum d'inhibition au bout d'un jour de stress, cette inhibition est plus progressive sur l'ensemble de l'expérience pour les deux autres souches.

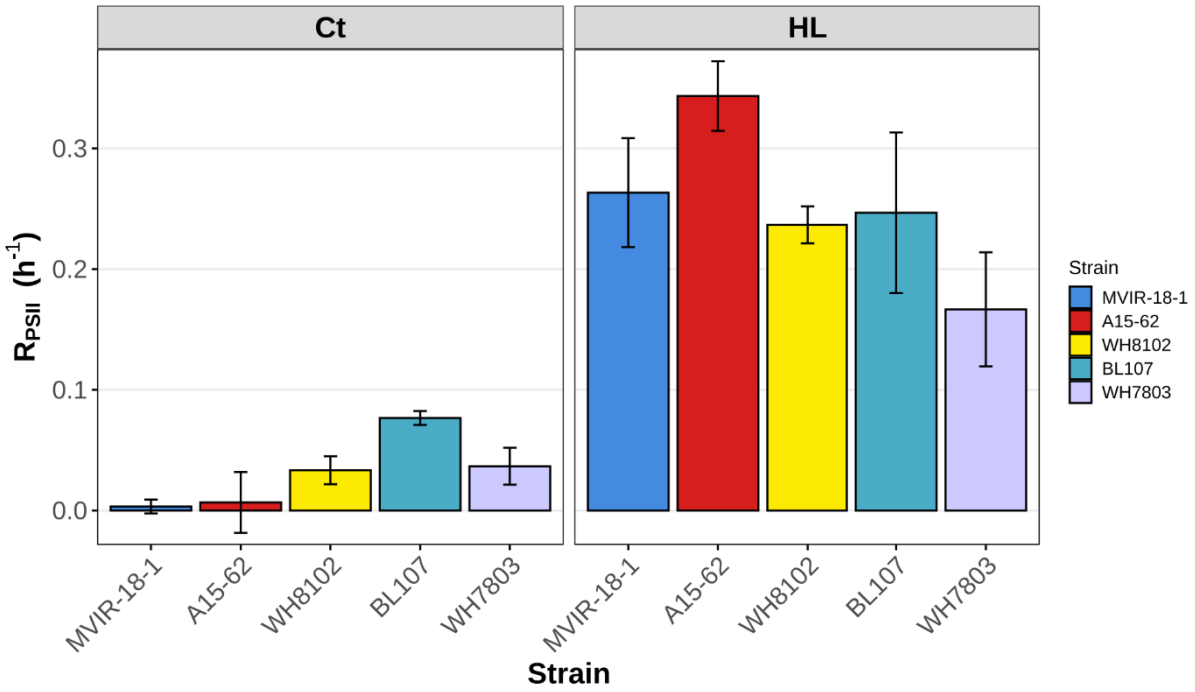


Figure 18 - Taux de réparation cumulatif du PSII pendant les 6 heures de l'expérience pour les cultures contrôle (Ct, panneau de gauche) ou soumises à un stress lumineux (HL, panneau de droite). Ces données sont basées sur au moins trois expériences indépendantes.

L'évolution du rapport PE:PC des PBS montre que seules deux souches présentent des variations significatives de ce rapport en réponse au stress HL, avec une augmentation pour WH7803 et une diminution pour BL107 (Fig. 19). Ces variations de fluorescence sont probablement dues à une déconnexion des hexamères PEII distaux chez WH7803, comme précédemment observé chez WH8102 en réponse aux UV (Six et al., 2007b), et à une déconnexion entre PC et APC pour BL107 à la base des bras des PBS. De façon intéressante, il semble que la dégradation des PBS soit indépendante des dommages générés au niveau du PSII, puisque la souche la plus affectée au niveau du PSII (WH8102) n'a montré aucune variation du rapport PE:PC. En réponse au stress LT, toutes les souches à l'exception de MVIR-18-1, subissent une augmentation du rapport PE:PC (Figure 19), indiquant une altération du couplage de la phycobiliprotéine conduisant probablement à la déconnexion du PE des bras des PBS, comme précédemment observé (Pittera et al., 2014; Six et al., 2007c). Cette déconnexion pourrait permettre de limiter le stress oxydant des cellules d'une part en diminuant la quantité d'énergie atteignant le centre

réactionnel du PSII et d'autre part en dissipant l'excès d'énergie sous forme de fluorescence de la PE libre (Pittera et al., 2014).

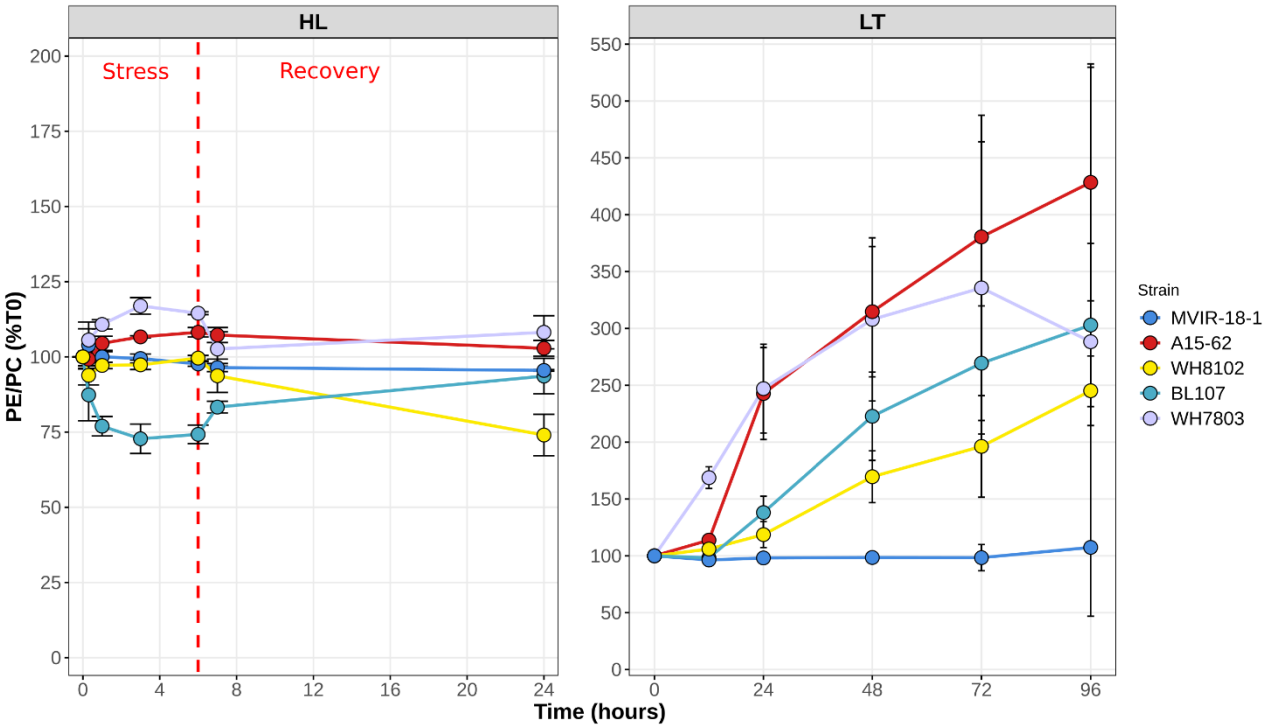


Figure 19 - Variations du rapport d'émission de fluorescence de phycoérythrine (PE) à phycocyanine (PC) chez différentes souches de *Synechococcus* acclimatées à faible lumière (LL, 25 µE.m^{-2}.s^{-1}). La ligne pointillée indique le temps (6h) auquel les cultures soumises à des contraintes légères ont été replacés dans leurs conditions lumineuses initiales pour étudier la récupération. Ces données sont basées sur au moins trois expériences indépendantes.

Globalement, ces résultats montrent qu'alors que la réponse au stress lumineux n'est pas thermotype-dépendante, les souches représentatives des thermotypes froids résistent mieux au stress LT que les trois autres souches, et la faible photoinactivation du PSII telle qu'indiquée par la faible diminution de leur F_V:F_M, suggère que ces souches sont probablement capables de mieux gérer le déséquilibre entre l'absorption d'énergie lumineuse et son utilisation dans les réactions intervenant en aval dans l'appareil photosynthétique (Latifi et al., 2009).

Analyse préliminaire de la réponse transcriptomique au stress

Une analyse globale de l'expression des gènes pendant les stress HL et LT révèle une réponse assez similaire entre les deux types de stress. En effet, une majorité de gènes différentiellement exprimés dans le stress HL montrent les mêmes patrons d'expression que dans le stress LT (Figure 20). Cependant, alors qu'en réponse au stress HL, l'intensité des profils d'expression est assez similaire pour l'ensemble des souches étudiées, les deux thermotypes froids, et en particulier la souche MVIR-18-1 isolée à plus haute

latitude, présentent un niveau d'expression différentielle globalement plus faible que celui des autres souches en réponse au stress LT, en accord avec les données de physiologie qui montrent que ces souches sont moins affectées par ce traitement (Figures 18 et 20). En revanche, pour WH7803 (clade V) et dans une moindre mesure pour A15-62 (clade II), les niveaux d'expression différentielle se sont avérés plus élevés en réponse au stress LT que HL.

Le clustering des gènes par profils d'expression a permis de mettre en évidence une réponse commune à l'ensemble des souches et des stress, qui se manifeste par un ensemble de gènes fortement sous-exprimés (cluster 1 et dans une moindre mesure cluster 3) et surexprimés (cluster 2 et dans une moindre mesure clusters 6 et 7), et ce dans les deux conditions testées (Figure 20A). Parmi les gènes les plus sous-exprimés dans toutes les souches, on retrouve dans le cluster 1 de nombreux gènes photosynthétiques notamment ceux codant pour les sous-unités du PSI (*psaA-F, I-M*), la plupart des gènes impliqués dans la synthèse des PBS (*apcA-B, C, E, F, cpcA-B, G1-G2, cpeA, C, E, S, T, mpeA, B, D* et *E*) ainsi que *uspG*, codant pour une protéine de stress universelle et un gène codant pour une protéine HLIP. Le cluster 3 regroupe également des gènes photosynthétiques parmi lesquels de nombreux gènes de biosynthèse du PSII (*psbB, C, I, K, M, O, Q, T, U, X, Y, Z, 30, 32*), du cytochrome b_6/f (*petD, L, M*), les gènes codant pour les autres sous-unités du PBS (*apcD, cpcM, cpeR, U, Y, Z, mpeY, rpcT)* ou impliqués dans la synthèse de phycobilines (*pebA, B*), dans la fixation du CO_2 (RuBisCO, carboxysomes, *ndhD4* et *F4, cupB*) ainsi que tous les gènes impliqués dans la synthèse de l'ATPase et quelques gènes impliqués dans la division cellulaire ou l'horloge circadienne (*fstQ, kaiA, B, zipN, pex, sepF*). Bien que la plupart de ces gènes soient comme ceux du cluster 1 sous-exprimés dans toutes les conditions, les gènes de biosynthèse de l'ATP synthase et ceux impliqués dans la fixation du CO_2 sont en revanche majoritairement surexprimés en réponse au stress HL spécifiquement chez les thermotypes chauds (A15-62, WH8102 et WH7803). Ainsi, il semble que les stress lumineux et thermique provoquent une forte inhibition des phases claires de la photosynthèse chez les souches représentatives des 4 écotypes tandis que la phase sombre est différentiellement affectée entre les thermotypes. Par ailleurs, on retrouve également au sein de ces clusters des réponses souches-spécifiques au niveau du métabolisme des sucres puisque, par exemple, 3 gènes impliqués dans la synthèse du glycogène (*glgA-C*) sont réprimés chez la plupart des souches et dans les deux conditions de stress, alors qu'ils sont faiblement induits chez A15-62, ce qui suggère que cette souche pourrait continuer d'accumuler du glycogène en conditions de stress lumineux et thermique.

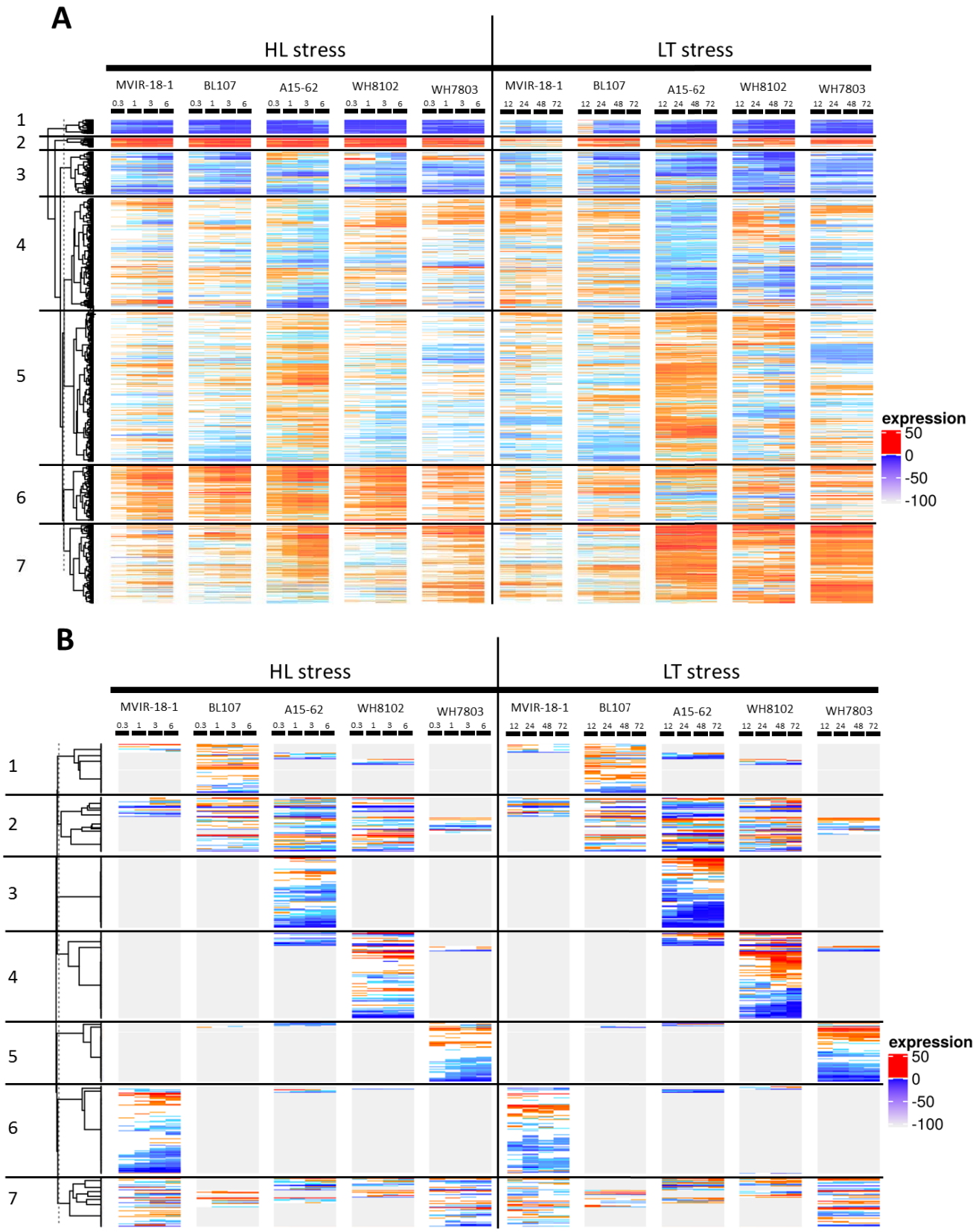


Figure 20 - Clustering des gènes par profils d'expression (log2FC). En rouge, les gènes sur-exprimés et en bleu les gènes sous-exprimés.

A l'inverse des clusters 1 et 3, les clusters 2, 6 et 7, majoritairement constitués de gènes surexprimés chez toutes souches en réponse au stress HL et LT, regroupent des gènes codant pour des enzymes impliquées dans la réplication et/ou réparation de l'ADN (cry2, *dnaA*, *B*, *X*, *holB, mutS2,* nudF, *radA, ruvB-C, umuC,* uvrBC, xseB, phrA, recFR, *xseA, ybaZ, yedK),* des enzymes de biosynthèse des vitamines b1 (*iscS2, rsgA, thiE*), b6, b7 (*bioC, F, H, birA*), b9 (*pabB, queC, E*) et b12 (*cbiG-cobJ, cobA, D, K, L, Q, S, N, cbiD*), des gènes impliqués dans la réponse au stress oxydatif (*ccsC1, evrA, ftrC, msrB, prxQ, trxB*) et dans le système cystéine désulfurase SUF (*sufCD*), responsable de la biosynthèse des centres Fe-S en condition de limitation en Fe ou de stress oxydant chez de nombreuses bactéries (Bolstad and Wood, 2010; Dai and Outten, 2012; Nodop et al., 2008). Les clusters 2,6 et 7 comprennent également de nombreux gènes codant pour des ribonucléases (8 gènes), des HLIPs (5 gènes) ainsi que des gènes codant pour les enzymes impliquées dans la synthèse de β-carotène (*crtB, D, P, Q, R, L-b*). En accord avec les données de photophysiologie, les gènes de cette dernière catégorie ainsi que le pool de gènes codant pour l'isoforme D1:2 de *psbA* se sont également avérés plus faiblement induits chez les thermotypes froids que les thermotypes chauds en réponse au stress LT.

Bien que ces clusters comprennent également de nombreuses protéines heat-shocks ou chaperones (*clpP1-4, clpB1, clpC, dnaK3, dnaJ1, 2, 5, groES, groL1-2, grpE, hsp20, htpG*), certaines d'entre elles ont un comportement souche et/ou stress-spécifique. *dnaK3, groES, groL1 et htpG* sont majoritairement surexprimés en HL et sous-exprimés en LT mais en revanche sous-exprimés chez MVIR-18-1 en HL et surexprimés chez WH7803 en LT. De façon similaire les 4 *clpP* sont assez fortement induits chez les souches des clades I à IV mais ne présentent aucune expression différentielle chez WH7803, chez qui nous avons précédemment montré qu'ils sont spécifiquement induits par le stress UV (cf. Chapitre II). Enfin *clpB1, dnaJ1, 2, groL2, grpE* et *hsp20* sont plutôt sous-exprimés et/ou non différentiellement chez MVIR-18-1 et BL107 en LT alors qu'ils sont fortement surexprimés chez les thermotypes chauds dans cette condition. La réponse des gènes de cette catégorie fonctionnelle, souvent considérés comme étant impliqués dans la réponse au stress en général, semble donc assez variable selon les souches et les conditions de stress.

Parmi les gènes montrant un comportement très différent entre souches, on peut également citer différents gènes impliqués dans la biosynthèse du cytochrome *c* oxidase (*ctaA, B, CI, DI, EI*) qui en acceptant des électrons du cytochrome c_6 contribue à la force proton-motrice et augmente la production d'ATP au dépend du NADPH. De même, 3 gènes de la partie haute de la glycolyse (*zwf, gnd, pgl*, cf. Chapitre II Guyet *et al*., Fig.6) sont particulièrement induits chez les deux thermotypes froids et chez WH8102 par les deux stress alors qu'ils sont réprimés ou non DE chez A15-62 et WH7803, suggérant une régulation différentielle de la glycolyse et de la voie métabolique des pentoses phosphate.

La figure 20B met en évidence les profils d'expression différentielle des gènes accessoires, c-à-d absent dans au moins l'une des souches. Globalement, les profils semblent assez similaires dans les deux conditions de stress. Parmi les gènes présents chez plusieurs souches (clusters 2 et 7), on retrouve notamment les gènes du cluster de la caroteno-protéine orange (*ocp, frp* et *crtW)*, qui réagissent très fortement chez toutes les souches qui le possèdent en réponse à la lumière mais plus ou moins fortement selon les souches en réponse au stress LT. Dans cette même catégorie, on peut noter que les gènes codant pour la delta-9 désaturase DesC4 et dans une moindre mesure la delta-12 désaturase DesA3, impliquées dans la modulation de la fluidité membranaire et présents dans les eaux froides (et tempérées dominées par le clade III pour DesA2, cf. chapitre III-partie YY-Fig. XXX), sont particulièrement induits par le stress LT chez WH7803 alors qu'ils sont assez faiblement induits chez le thermotype froid BL107, voir sous-exprimé chez la souche la plus septentrionale (MVIR-18-1), montrant une gradation dans la réponse au stress LT selon l'intensité ressentie de ce stress pour les différentes souches. Ces deux désaturases semblent donc jouer un rôle majeur, non seulement dans l'adaptation des thermotypes froids à leur niche thermale spécifique (Breton *et al*., 2019), mais également dans les environnements subissant de fortes variations saisonnières de la température, comme c'est le cas des environnements dominés par le clade III en Mer Méditerranée ou dans les zones côtières tempérées, telles que celles colonisées par les membres du clade V (Doré *et al*., en préparation).

Contrairement aux clusters 2 et 7, les autres clusters regroupent des gènes pour la plupart spécifique d'une souche donnée, le cluster 3 regroupant les gènes spécifiques d'A15-62, le cluster 4 de WH8102, le cluster 5 de WH7803 et le cluster 6 de MVIR-18-1 (Figure 20B). La plupart des gènes de ces clusters souches-spécifiques n'ont pas de fonction connue mais certains sont très fortement différentiellement exprimés et pourrait constituer des cibles intéressantes pour une caractérisation fonctionnelle par mutagénèse, notamment les gènes fortement surexprimés en HL chez A15-62, une souche qui ne possède pas de cluster d'OCP permettant de dissiper l'excès d'énergie lumineuse, et qui pourrait utiliser une stratégie alternative de photoprotection. Par ailleurs, en accord avec les comparaisons génomiques réalisées dans le chapitre 1 de ma thèse, aucun cluster de gènes DE spécifique d'un thermotype donné n'a pu être identifié. Par exemple, parmi les 18 gènes identifiés dans l'étude Doré *et al*. (Chapitre I - Tableau S4) comme étant spécifiques des thermotypes froids (clades I et IV), seuls deux sont surexprimés en réponse au stress LT chez les deux thermotypes froids (CK_00003324, CK_00002286) dont l'un, codant pour une pterin-4-alpha-carbinolamine dehydratase, est spécifiquement induit en réponse à ce stress. Ces résultats suggèrent que l'acclimatation aux niches thermiques repose plutôt sur une régulation différentielle de gènes communs aux différentes souches et/ou de gènes spécifiques d'un clade ou d'une souche.

CONCLUSIONS
ET PERSPECTIVES

I. Mécanismes évolutifs des génomes de picocyanobactéries et rôle dans l'adaptation à la niche écologique

Malgré une bonne connaissance de la diversité génétique des picocyanobactéries marines et de leurs préférences environnementales, la base génétique de la plupart des caractères phénotypiques et les processus sous-jacents qui ont régi la diversification de ces picocyanobactéries, commencent tout juste à être élucidés. Dans le chapitre I de cette thèse, à l'aide de 81 génomes non redondants de picocyanobactéries, dont 34 génomes de *Synechococcus* non publiés, nous avons évalué la diversité génomique de ces deux genres et son influence sur la capacité d'adaptation des différentes lignées à diverses niches écologiques. L'identité protéique (AAI pour "Average Amino acid Identity") et la comparaison de l'ARN ribosomal 16S entre paires de génomes nous ont permis de distinguer clairement d'un point de vue génomique les 3 lignées profondes (ou sous-clusters) de *Synechococcus*/*Cyanobium* et tous les clades phylogénétiques majeurs au sein de cette radiation. De même, au sein du genre *Prochlorococcus*, ces analyses ont mis en évidence un fort niveau de divergence au sein du clade LLII-III, qui suggère qu'il serait nécessaire de re-scinder ce clade en deux clades distincts, comme suggéré par Yan et al. (2018) et plus généralement, de définir de nouveaux clades au sein du phototype polyphylétique LL, au vu des analyses d'échantillons environnementaux par metabarcode (Martiny et al., 2009) et plus récemment par SAGs et MAGs (Berube et al., 2018). En termes de contenu génétique, alors que les études précédentes soutenaient l'idée que l'adaptation à une niche écologique reposait principalement sur le contenu différentiel des cellules en gènes (Rocap et al., 2003; Palenik et al., 2006; Dufresne et al., 2008; Kettler et al., 2007; Scanlan et al., 2009), l'étalonnage temporel d'un arbre phylogénétique des protéines communes à ces génomes a montré que les gains/pertes de gènes et la fixation de ces événements au cours de l'évolution se produisaient à un taux étonnamment bas (par exemple, 5,6 gènes gagnés My pour la lignée principale de *Synechococcus* dont seulement 0,71/My sont fixés à long terme). Pourtant, un certain nombre de gènes se sont avérés spécifiques d'un l'écotype ou d'un clade, et donc probablement liés à la colonisation de niches écologiques particulières, comme le confirme l'analyse des métagénomes de *Tara* Oceans (cf. ci-dessous). Il est intéressant de noter que le nombre de gènes varie en fonction de la niche considérée, avec un nombre assez important impliqué dans la carence en fer ou en phosphore, tandis que peu de gènes se sont révélés communs aux clades colonisant les niches froides (I, IV) ou chaudes (II, III). Ces observations remettent donc en question les rôles respectifs des gains/pertes de gènes et des variations de séquence (substitutions) dans la diversification des picocyanobactéries marines. Afin d'élucider les bases génétiques de l'adaptation à la température, nous avons exploré le rôle adaptatif des substitutions d'acides aminés et mis en évidence un certain nombre de gènes contenant des substitutions spécifiques des thermotypes froids (clades I et IV). Parmi les exemples les plus frappants, on peut citer les gènes impliqués dans la biosynthèse des caroténoïdes, ce qui est en accord avec des travaux antérieurs de

physiologie comparative entre les différents thermotypes de *Synechococcus* (Pittera et al., 2014) qui ont montré que les membres des clades I et IV étaient capables de maintenir, voire d'augmenter leur rapport β-carotène:chlorophylle *a* en réponse au stress du froid, alors que ce rapport diminuait dans les souches représentatives de thermotypes chauds. Bien que cela reste à vérifier par mutagénèse dirigée, il est donc possible que ces substitutions permettent aux cellules des clades I et IV de maintenir la synthèse du β-carotène dans les eaux froides, ce qui pourrait également entraîner une réduction du stress oxydatif induit par le froid chez ces organismes. Un autre aspect original de cette étude a concerné la délimitation semi-automatique des îlots génomiques de chaque génome de picocyanobactéries et la comparaison de leur contenu en gènes entre toutes les souches en utilisant également une approche de construction de réseau. Ce réseau a notamment permis de mettre en évidence des transferts latéraux potentiels entre souches aussi bien proches qu'éloignées phylogénétiquement, et notamment certains transferts spécifiques des LLI, LLII-III ou LLIV, dont l'investigation plus approfondie devrait permettre de mieux comprendre les spécificités génomiques de ces différents clades. Globalement, les résultats du chapitre I m'ont permis d'affiner considérablement la compréhension des processus évolutifs impliqués dans la diversification de ces membres clés du phytoplancton marin. Ils soulignent la nécessité de prendre en compte les échelles de temps de l'évolution du génome pour interpréter les rôles adaptatifs respectifs des variations du contenu génétique et celles des séquences nucléotidiques et leurs liens avec une niche particulière. Bien qu'un certain nombre de ces variations aient été mises en évidence en comparant 81 génomes complets de picocyanobactéries, cela reste un nombre limité de génomes compte tenu de l'importante diversité génétique existant au sein de ces deux genres, et la comparaison d'un bien plus grand nombre de génomes proches d'un point de vue phylogénétique permettrait sans doute de faciliter l'extraction du signal du bruit de fond et de révéler probablement de nombreux autres variants.

Au cours de ma thèse, une solution alternative a consisté à analyser directement les populations naturelles via l'analyse de données de métagénomique disponibles, qui permettent d'avoir un aperçu des processus d'adaptation mis en place par les organismes dans une niche écologique donnée. L'analyse de la distribution des gènes des picocyanobactéries dans l'océan mondial par WGCNA a en effet révélé qu'il existe un répertoire de gènes spécifique de chaque communauté, comme le montre la concordance entre les groupes de stations définis sur la base de la composition de la communauté (assemblage d'ESTUs), telle qu'analysée en utilisant le gène marqueur *petB,* et ceux définis à partir des profils d'abondance des gènes flexibles recrutés dans les métagénomes de *Tara* Océans, en utilisant comme références les génomes présents dans la base de données Cyanorak v2.1 (voir chapitre I). La distribution des modules de gènes définis par WGCNA est non seulement étroitement corrélée à la distribution des ESTUs, mais également aux différents paramètres physico-chimiques acquis lors de l'expédition *Tara* Oceans. On retrouve en effet aussi bien pour *Prochlorococcus* que pour *Synechococcus*, des modules de gènes qui correspondent

grossièrement à 4 niches majeures de l'océan, à savoir les zones chaudes carencées en azote, en phosphate ou en fer et enfin les zones tempérées et froides trouvées au niveau des upwelling ou à haute latitude. Dans le cas de *Prochlorococcus*, le module '*blue*', qui est dominé par l'ESTU HLIIA et qui est fortement anti-corrélé à la disponibilité en N inorganique comprend de nombreux gènes impliqués dans le transport et l'assimilation de l'azote. Le fait que certains de ces gènes, notamment des réductases du nitrite et du nitrate et les cofacteurs associés ou des gènes de transport et d'assimilation du cyanate, aient déjà été mis en évidence dans des études précédentes (Berube et al., 2014, 2019; Kent et al., 2016; Garcia et al., 2020), constitue une preuve de concept de l'efficacité de cette approche pour identifier de gènes de niche à partir de données métagénomiques. Comparativement à ces travaux, l'une des forces majeures de notre étude est qu'elle a permis, en intégrant l'information des génomes et celles des métagénomes, d'identifier non seulement des gènes individuels mais également, au sein des îlots génomiques définis dans le chapitre 1, des clusters de gènes de taille plus réduite partagés par tout ou partie d'une population colonisant une même niche. Ainsi, l'examen du module *blue* a, par exemple, permis d'identifier un cluster de 7 gènes, potentiellement impliqué dans le transport et l'assimilation d'acides aminés tels que l'arginine et/ou de dérivés d'acides aminés tels que la taurine, ce dernier composé étant une source majeure de carbone et d'énergie pour les procaryotes dans l'Océan Atlantique Nord (Clifford et al. 2019). De même, comme décrit précédemment (Malmstrom et al., 2013; Rusch et al., 2010), le module *brown* qui est anti-corrélé à la disponibilité en fer, contient des gènes d'acquisition et de stockage du fer, tandis que les gènes codant pour des protéines contenant du fer sont fortement anti-corrélés à ce module. Un résultat original venant de l'analyse de ce dernier a été la mise en évidence de la présence d'un cluster de 6 gènes codant notamment une cytochrome oxydase alternative (ARTO), qui pourrait être impliquée dans la réduction du Fe(III) en Fe(II) avant son transport à travers la membrane par le transporteur de Fe(II) FeoB (Kranzler et al. 2014).

Il est important de noter que de nombreux gènes des modules *blue* et *brown* sont regroupés en clusters de gènes adjacents dans les génomes qui semblent issus soit de transferts latéraux très anciens, soit de rétention sélective par certaines lignées à partir d'ancêtres cyanobactériens, comme le suggère leur présence chez des cyanobactéries d'eau douce, telle que *Synechocystis* sp. PCC 6803, chez qui les gènes sont souvent situés dans le même ordre au sein des clusters. Au contraire, la plupart des gènes du module *red*, qui est anti-corrélé à la disponibilité en phosphate, et dont beaucoup sont potentiellement impliqués dans le transport et le métabolisme du phosphore, sont dispersés entre différentes régions variables du génome, tels que l'ISL5 définie originellement par Coleman et al. (2006) et retrouvée ensuite par analyse de réseau (voir Chapitre I), et la composition en gènes de ces régions ainsi que leur ordre sur le génome est souvent beaucoup plus erratique que dans le cas des modules précédents. Cette forte variabilité laisse à penser que ces gènes ont été acquis du milieu environnant par une série de transferts

latéraux beaucoup plus récents, impliquant potentiellement des phages, et qui se traduit par une grande variabilité des stratégies adoptées par les populations colonisant ces niches pour faire face à la carence en phosphore. A noter cependant que le cas du transport du phosphonate, nous avons identifié deux clusters de gènes pour qui, comme ceux des autres modules, la composition et l'ordre des gènes semble bien conservée entre les souches qui les possèdent. De façon intéressante, l'un deux semble présent chez tous les membres des populations de *Prochlorococcus* des zones pauvres en phosphore, comme la Méditerranée, alors que l'autre, qui, en plus des gènes de transport inclut également des gènes du métabolisme des phosphonates (*phnY-Z*), n'est présent que chez moins de 10% des cellules de *Prochlorococcus* peuplant ces zones.

La même approche de réseau appliquée à *Synechococcus* a également permis de récupérer de nombreux gènes potentiellement liés à l'adaptation à la niche, mais comparativement moins de grands clusters de gènes que pour *Prochlorococcus*. Ceci est peut-être dû au fait que *Synechococcus* possède un génome beaucoup plus grand et diversifié que la plupart des clades de *Prochlorococcus* et ont par conséquent un répertoire de gènes plus étendu, mais il est également possible que le plus faible nombre de clusters de gènes *complets* retrouvés chez *Synechococcus* pourrait être dû à l'utilisation d'un seuil trop stringent pour définir les gènes flexibles dans l'environnement et il serait donc intéressant de tester d'autres valeurs de seuils. Comme c'est le cas pour les modules *brown* et *blue* de *Prochlorococcus*, on retrouve pour *Synechococcus* deux modules dont les gènes présentent des corrélations opposées et donc des distributions complémentaires dans l'environnement. En effet, le module *salmon* regroupe de nombreux gènes qui codent pour des protéines riches en Fer (ferredoxines, succinate déshydrogénase, PTOX, phéophorbide *a* oxygenase, etc.) alors que le module *yellow* rassemble des gènes impliqués dans le stockage (ferritine), ou la synthèse de protéines permettant de limiter l'utilisation du fer (flavodoxine, IsiA). A noter également la présence au sein du module *yellow* de nombreuses copies des petits gènes de type *hli* et *nif11*. Si le premier semble notamment impliqué dans la photoprotection, le rôle des seconds qui sont très largement amplifiés dans les génomes de CRD1C reste à élucider. Nous avons également pu mettre en évidence un certain nombre de gènes et clusters communs entre *Synechococcus* et *Prochlorococcus* et retrouvés dans des niches similaires, notamment l'opéron *sdhABC* impliqué dans la respiration et le cluster *asnB-pydC-pyrB2* impliqué dans le métabolisme des acides aminés, qui sont spécifiquement absents dans les zones pauvres en fer chez les deux organismes, alors que d'autres gènes/clusters sont spécifiques de *Synechococcus* tel que la glycolate oxidase, impliquée dans la photorespiration dans les zones non limitées en fer.

Outre les gènes possédant une annotation fonctionnelle, la méthode de recrutement que j'ai utilisée a permis de mettre en évidence la présence spécifique dans certaines niches de nombreux gènes de fonction inconnue ou mal définie qui pourraient également être impliqués dans l'adaptation à ces niches.

L'attribution d'une fonction putative à ceux qui sont inclus dans des clusters de gènes devrait être facilitée par le fait que d'autres gènes de ces îlots sont homologues de gènes caractérisés chez d'autres organismes et connus pour être impliqués dans une voie métabolique particulière. Cependant, il serait important dans les années à venir de caractériser directement chez une ou plusieurs souches de picocyanobactérie(s) marine(s) la fonction de ces gènes et des clusters dans leur ensemble par des approches de mutagenèse et de biochimie afin de démontrer leur rôle dans la survie de ces organismes dans une niche particulière du milieu marin.

Par ailleurs, bien que cette analyse ait permis d'identifier des gènes impliqués dans la colonisation de la quasi-totalité des niches de la zone euphotique des océans, nous n'avons, de façon similaire aux analyses de génomique comparative (Chapitre I), mis en évidence que peu de gènes potentiellement impliqués dans l'adaptation à la température, alors que ce paramètre a clairement joué un rôle prépondérant dans la diversification des différentes lignées de *Prochlorococcus* et *Synechococcus*. Une piste que j'ai commencé à explorer vers la fin de ma thèse, et qui mériterait d'être poursuivie dans l'avenir, est l'analyse des variants de séquences (SNPs) dans les données *Tara* Océans, afin de faire ressortir des mutations qui pourraient permettre une adaptation des cellules aux variations de température. Cette analyse est délicate du fait de l'énorme diversité génomique des picocyanobactéries, même au sein de clades donnés, comme démontré par l'analyse de centaines de SAGs issus d'une même population de *Prochlorococcus* HLII (Kashtan et al. 2014), une diversité qui induit un bruit de fond conséquent qui gêne l'identification de substitutions spécifiques d'écotypes donnés. Néanmoins, comme cela a été réalisé sur les génomes dans le chapitre I, la comparaison de séquences des membres des clades I et IV qui colonisent les mêmes niches thermiques mais qui sont éloignés phylogénétiquement, pourraient sans doute permettre d'identifier des substitutions potentiellement impliquées dans l'adaptation à la température.

II. Caractéristiques générales et spécificités de la réponse transcriptomique de *Synechococcus* aux variations des paramètres environnementaux

Le chapitre II de cette thèse a été consacrée à l'étude de la réponse physiologique de la souche modèle *Synechococcus* sp. WH7803 à diverses conditions de lumière et température. Cette étude, qui a notamment impliqué l'analyse de 154 transcriptomes, est la plus exhaustive jamais réalisée sur des picocyanobactéries marines à ce jour. Elle présente l'intérêt, comme celle de Blot et al (2011), de prendre en compte l'histoire lumineuse des cellules (pré-acclimation à forte ou à basse lumière) et également d'étudier l'effet des variations nycthémérales déjà en partie abordé par PCR quantitative par Mella-Flores et al. (2012). Cette étude, qui a bénéficié de la riche annotation fonctionnelle de la base Cyanorak v2.1, a permis de comprendre quelles voies métaboliques étaient les plus affectées par les différentes conditions testées. D'un point de vue global, nous avons pu observer un effet synergique entre l'exposition préalable à forte intensité lumineuse (HL) et le stress basse température (LT), comme précédemment observé par Blot et al. (2011) entre l'intensité lumineuse et le stress oxydant. Dans cette étude, ce comportement avait été attribué à un effet plus important sur les souches acclimatées à forte intensité lumineuse des espèces réactives de l'oxygène (ROS), induisant à la fois des dommages directs aux centres réactionnels et une inhibition de leurs cycles de réparation. Par analogie, il est possible que l'effet délétère des ROS puisse être également amplifié par le ralentissement du métabolisme, et donc du turnover de la protéine D1 du photosystème II, en réponse au stress LT. En revanche, les cellules acclimatées en HL semblent être mieux préparées à supporter le stress UV que les cellules acclimatées en LL, avec un nombre de gènes différentiellement exprimés beaucoup plus faible à tous les temps de stress UV pour les cultures acclimatées en HL qu'en LL et une baisse beaucoup plus modérée du F_V/F_M dans les conditions HL, en partie attribuable à une forte augmentation de la capacité de réparation de la protéine D1. Ces résultats suggèrent donc que le turnover de cette protéine et sa sensibilité au stress oxydant joue un rôle central dans la réponse de ces organismes à toutes les variations des conditions environnementales affectant le métabolisme des cellules.

La comparaison des réponses transcriptomiques de cette souche à différents stress a également permis de mettre en évidence des réponses communes ou au contraire spécifiques des diverses variations de conditions environnementales testées. Dans ce contexte, on peut noter que toutes les conditions de stress semblent favoriser la dégradation du glycogène, puis du glucose en pyruvate principalement par la voie Entner-Doudoroff, avec la production concomitante d'ATP, de NAD(P)H et de précurseurs biosynthétiques pour les acides aminés, les nucléotides et les acides gras. En revanche, les UV induisent de façon spécifique une surexpression des gènes codant pour des protéases et des protéines chaperonnes, du gène *dprA* impliqué dans la réparation de l'ADN simple brin, de certaines Hlip (« High-light inducible protein »), et également d'un certain nombre de gènes du PSII qui semblent impliqués dans la néo-

synthèse de sous-unités du PSII et pas seulement le réassemblage de sous-unités préexistantes après remplacement D1, comme cela est le cas pour les autres stress testés. Ces analyses ont également permis de mettre en évidence des différences importantes avec le modèle de cyanobactéries d'eau douce le plus étudié, à savoir *Synechocystis* sp. PCC 6803, et justifie de ce fait l'intérêt d'utiliser des modèles marins, même si la souche WH7803 n'appartient pas à un des clades les plus abondants du milieu marin.

Pour cette raison, nous avons également commencé à comparer les réponses transcriptomiques de souches représentatives de quatre des cinq clades dominants *in situ* (clades I-IV) pour un nombre plus limité de stress, afin d'évaluer la variabilité écotypique de la réponse au stress chez *Synechococcus.* Les analyses préliminaires de ces données ont permis de mettre en évidence des réponses communes entre souches, en particulier une sous-expression d'une grande partie des gènes impliqués dans la photosynthèse et une surexpression des gènes impliqués dans la réponse au stress oxydatif et dans la réplication et la réparation de l'ADN, alors que les stress semblent induire un certain nombre de réponses spécifiques de souches, notamment pour les gènes impliqués dans la synthèse du glycogène. Il est important de noter que nous n'avons comparé qu'une souche représentative de chaque écotype dominant *in situ* et qu'il serait nécessaire d'avoir plus de représentants de chaque écotype afin d'estimer la variabilité intra-clade de la réponse au stress. De plus, les analyses transcriptomiques ont été réalisées afin de pouvoir détecter l'expression de transcrits d'ARN non-codants antisens, dont l'importance dans la régulation de l'expression des gènes chez les picocyanobactéries a été récemment démontrée (Lambrecht et al., 2020; Muro-Pastor and Hess, 2020; Steglich et al., 2008) et à ont même pu être identifié de données environnementales chez *Prochlorococcus* (Lott et al., 2020), mais leur identification ainsi que l'étude de leur expression reste à réaliser. Par ailleurs, je n'ai pas eu l'occasion pendant ma thèse de travailler sur l'effet de la limitation en nutriments sur ces organismes, qui a été étudié par plusieurs autres groupes mais sur un nombre limité de souches (Dupont et al., 2012; Ostrowski et al., 2010; Su et al., 2006; Tetu et al., 2009), ni sur le cinquième écotype dominant dans le milieu marin, à savoir le clade CRD1 qui est prédominant dans les zones pauvres en fer. Dans ce contexte, des analyses sont actuellement en cours au sein de l'équipe afin de mieux comprendre la réponse à la carence en fer de souches représentatives des 3 ESTUs définis par Farrant et al. (2016) au sein de ce clade et qui colonisent des niches thermiques distinctes. Enfin, bien que l'étude de la réponse transcriptomique permette d'avoir un aperçu global de l'état de stress des cellules, elle ne reflète pas forcément la réponse protéomique des organismes aux stress. Il serait donc nécessaire pour avoir une vision plus holistique de la réponse au stress de ces organismes de compléter ces résultats par des analyses de protéomique et de quantification des métabolites, ce qui faciliterait la construction de réseaux métaboliques réalistes pour ces organismes dans différentes conditions environnementales.

Malgré ces inconvénients, le vaste jeu de données transcriptomiques que j'ai eu l'occasion d'analyser, et en particulier les résultats concernant la variabilité circadienne et écotypique de l'expression des gènes dans différentes conditions environnementales, devrait être particulièrement utile afin de mieux interpréter les données de métatranscriptomique toujours plus nombreuses pour le milieu marin, générées à partir d'échantillons prélevés dans une vaste gamme de niches écologiques et à différents moments de la journée.

L'intégration de l'ensemble de ces données d'expression ont également permis d'identifier un certain nombre de gènes cibles qui restent à caractériser d'un point de vue biochimique et fonctionnel afin de mieux comprendre les mécanismes mis en place dans la réponse au stress par les différents écotypes. On peut citer par exemple *crtL-b,* qui code pour une enzyme permettant de transformer le lycopène en bêta-carotène, et qui est le gène accumule le plus de substitutions spécifiques aux clades I et IV et aussi le plus fortement différentiellement exprimé en réponse au stress basse température. Le rôle éventuel de ces substitutions dans l'adaptation à la niche thermique froide pourrait ainsi être testé en construisant des mutants dans lesquels les acides aminés concernés seraient remplacés par leur équivalent en milieu chaud.

III. Importance du maintien et de l'amélioration des bases de données génomiques de référence dans un contexte de croissance exponentielle des données omiques marines

Le génome est clairement au centre de la plupart des analyses présentées au cours de cette thèse. En effet, les études concernant les réponses physiologiques et métaboliques des picocyanobactéries à divers stress environnementaux, des mécanismes évolutifs impliqués dans la diversification fonctionnelle des écotypes ainsi que de la distribution biogéographique des gènes, nécessitent toutes de disposer d'un base fiable et la plus complète possible de génomes de référence. Dans le cadre de cette thèse, le système d'information Cyanorak v2.1 a donc constitué un atout précieux afin i) d'identifier les gènes communs ou spécifiques au sein de clades/écotypes, ii) de réaliser des analyses phylogénomiques, iii) d'analyser les réponses transcriptomiques de ces organismes et enfin, iv) d'extraire les gènes de picocyanobactéries de l'environnement et de d'identifier les répertoires de gènes spécifiques des différentes niches colonisées par ces organismes. Cependant, bien que les 97 génomes intégrés dans Cyanorak v2.1 couvrent déjà assez bien la diversité génétique des souches cultivées des deux genres de picocyanobactéries, il est important de noter que les pan-génomes de *Prochlorococcus* et *Synechococcus* restent ouverts et que donc seule la distribution dans l'environnement des gènes présents dans notre base de données a pu être examinée. En particulier, certains clades tels que EnvA et EnvB, qui sont parfois prépondérants dans les zones pauvres en fer, restent à ce jour non cultivés et il en est de même pour certaines lignées de *Prochlorococcus* au sein du phototype LL.

Afin de combler ces manques pour de futures analyses méta-omiques, nous avons récemment entrepris de mettre en place un workflow d'analyse permettant d'utiliser l'ensemble des génomes (WGS pour 'Whole Genome Sequences', SAGs pour 'Single Amplified Genomes' ou MAGs pour 'Metagenome Assembled Genomes') nouvellement disponibles dans les bases de données publiques. Une recherche dans la base du NCBI à la date du 18 juin 2019 a permis d'extraire 782 nouveaux génomes (40 WGS, 73 MAG et 669 SAGs), générés avec différentes méthodes d'assemblage, suivant les études (Fig. 21). Ce nombre important de nouveaux génomes séquencés est dû à l'avènement des techniques NGS, récemment appliquées à l'écologie des picocyanobactéries marines et permettant de reconstruire des génomes directement à partir d'échantillons environnementaux et ainsi de s'affranchir de l'étape d'isolement et de culture de souches pour extraire les données génomiques, comme nécessaire pour les WGS (Berube et al., 2018; Cornet et al., 2018; Delmont et al., 2018; Haro-Moreno et al., 2018; Haroon et al., 2016; Kashtan et al., 2014, 2017; Malmstrom et al., 2013; Parks et al., 2017, 2018; Powell et al., 2005; Tully et al., 2017, 2018). Il est important de noter qu'un MAG correspond à un génome consensus, puisqu'il est construit à partir d'un (ou plusieurs) échantillon(s) pouvant potentiellement contenir plusieurs clades

et/ou ecotypes d'un organisme donné, chacune des populations de ces clades/écotypes étant elle-même constituée d'individus génétiquement distincts, et l'assemblage obtenu peut par ailleurs être contaminé par des fragments d'ADN d'organismes co-occurrents. En ce qui concerne les SAGs, ils sont le résultat du séquençage d'une cellule unique et sont, de ce fait, moins sujets à la contamination, mais ces génomes sont très souvent incomplets du fait des amorces utilisées qui n'amplifient pas de façon homogène les différentes parties du génome (Blainey, 2013; Kogawa et al., 2018; Lasken and McLean, 2014; Shapiro et al., 2013). Un avantage des SAGs par rapport aux MAGs est cependant de pouvoir analyser le génome à l'échelle d'un organisme et d'identifier ainsi la variabilité génétique d'une cellule à l'autre au sein d'une population naturelle, ce qui permet de différencier directement *in situ*, les variants génétiques maintenus dans l'ensemble d'une population et ceux qui correspondent à des variations ponctuelles.

Etant donné que la plupart de ces génomes sont incomplets, nous n'envisageons pas de les intégrer à la base de données Cyanorak. Cependant, le workflow que nous avons développé nous a permis de transférer les riches annotations fonctionnelles des génomes intégrés à Cyanorak à toutes ces séquences (Figure 21). Ce workflow est composé de plusieurs étapes: i) une prédiction de gènes est réalisée avec l'outil Prokka (Seemann, 2014) sur les génomes non annotés; ii) un Blast réciproque est réalisé entre les séquences protéiques des nouveaux génomes et les clusters de gènes orthologues (CLOGs) de Cyanorak afin d'assigner dans la mesure du possible les gènes à un CLOG existant et de transférer à ce gène l'annotation fonctionnelle de ce CLOG; iii) les gènes non associés à un CLOG sont soumis à une tentative supplémentaire d'assignation à un CLOG sur la base de profils HMM des CLOGs de Cyanorak en utilisant HMMER (hmmer.org), et enfin iv) les séquences restantes sont clusterisées entre elles avec orthoMCL (Li et al., 2003) afin de générer de nouveaux CLOGs mono- ou multi-séquences.

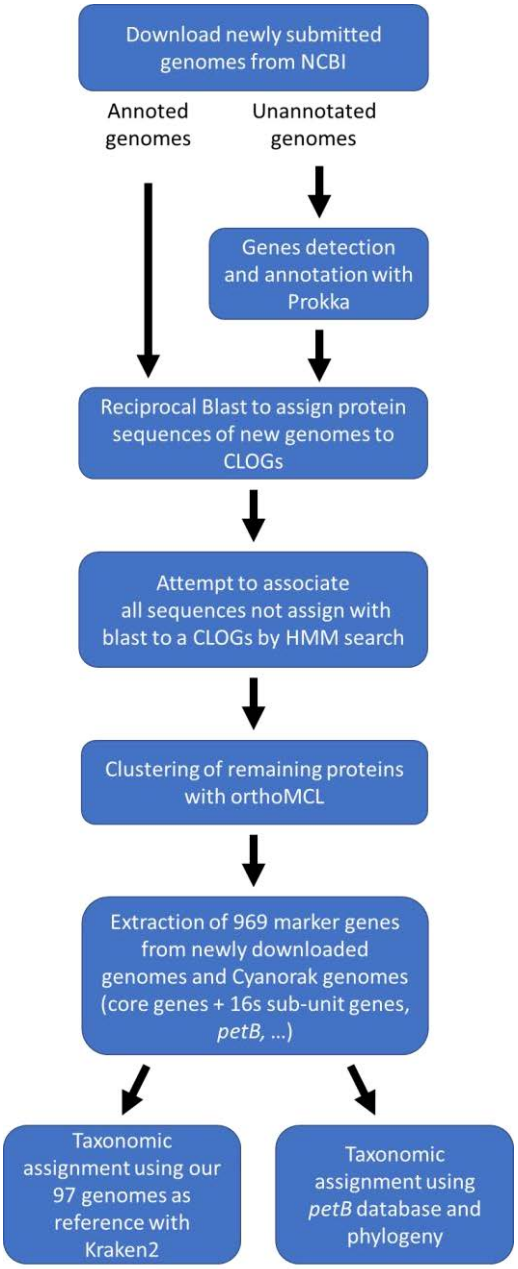


Figure 21- **Description du workflow d'intégration de nouveaux génomes.** Abréviation : CLOG, Cluster of Likely Orthologous Genes.

Afin d'estimer la complétude de ces génomes, j'ai par ailleurs estimé la proportion de gènes communs issus des 81 génomes non-redondants de Cyanorak v2.1, ce qui permettra éventuellement de supprimer un certain nombre de génomes trop incomplets en fonction des analyses réalisées. On peut notamment voir sur la figure 22 que les SAGs sont les génomes les moins complets. Les génomes ont ensuite été assignés taxonomiquement en utilisant plusieurs méthodes.

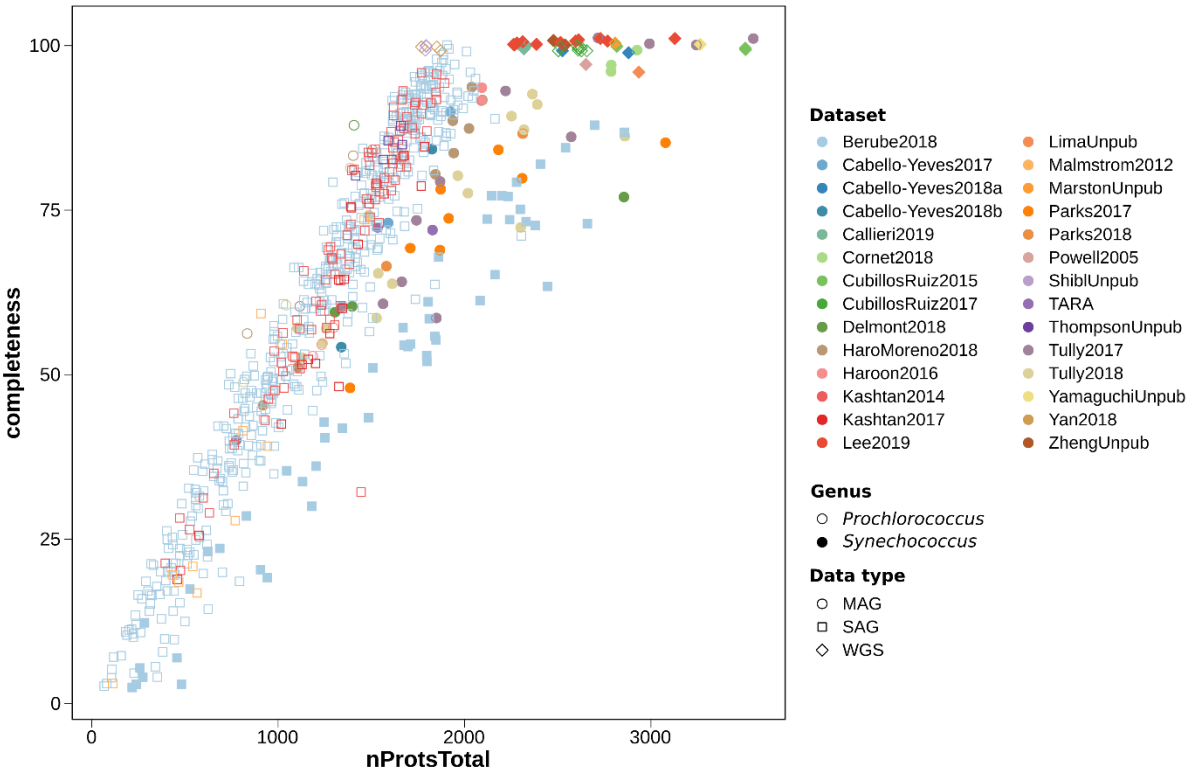


Figure 22 - Représentation de la complétude des génomes en fonction de leur nombre total de protéines communes. La complétude des génomes est estimée par la proportion du nombre de gènes communs retrouvés dans les génomes nouvellement téléchargés par rapport aux 825 gènes communs aux 81 génomes non redondants de la base Cyanorak v2.1 (cf. Doré et al., soumis). Les références complètes des datasets utilisés ici sont listés dans la liste bibliographique située à la fin de cette thèse.

Une première approche a consisté à annoter taxonomiquement ces génomes en utilisant KrakenUniq (Breitweiser et al. 2018), un outil permettant de classifier rapidement les reads en fonction de leur composition en k-mer par rapport à une base de référence, ici constituée des 81 génomes non redondants de Cyanorak v2.1. Cette approche a été utilisée à la fois sur une concaténation des contigs, aboutissant à une annotation unique par génome, soit sur les contigs découpés en fragments de 150 bp, permettant d'obtenir une annotation par fragment et donc de détecter de potentielles chimères au sein de ces contigs. Par ailleurs, une deuxième approche a consisté à extraire différents marqueurs phylogénétiques (ARNr 16S, ITS, *petB*, *rpoC1*), qui ont permis de vérifier l'assignation taxonomique obtenue pour l'ensemble des génomes par krakenUniq, de mettre en évidence de potentielles chimères (lorsque les différents marqueurs aboutissaient à une annotation différente) et également de commencer à réconcilier la taxonomie obtenue par ces différents marqueurs phylogénétiques, pour lesquels la correspondance n'est pas toujours connue (p.ex. le clade XV de *Synechococcus* défini sur la base de l'ITS correspond en fait au sous-clade IIh, en utilisant *petB*). Parmi ces 782 génomes, bien que 27 génomes apparaissent être fortement chimériques, ces génomes ont permis d'augmenter la représentativité de certains clades pour lesquels nous disposions jusqu'à présent d'aucun (*Synechococcus* EnvA, EnvB) ou seulement quelques représentants en culture (p.ex. *Prochlorococcus* LLI, LLIV et *Synechococcus* clades IV, CRD1, WPC1; Figure 23).

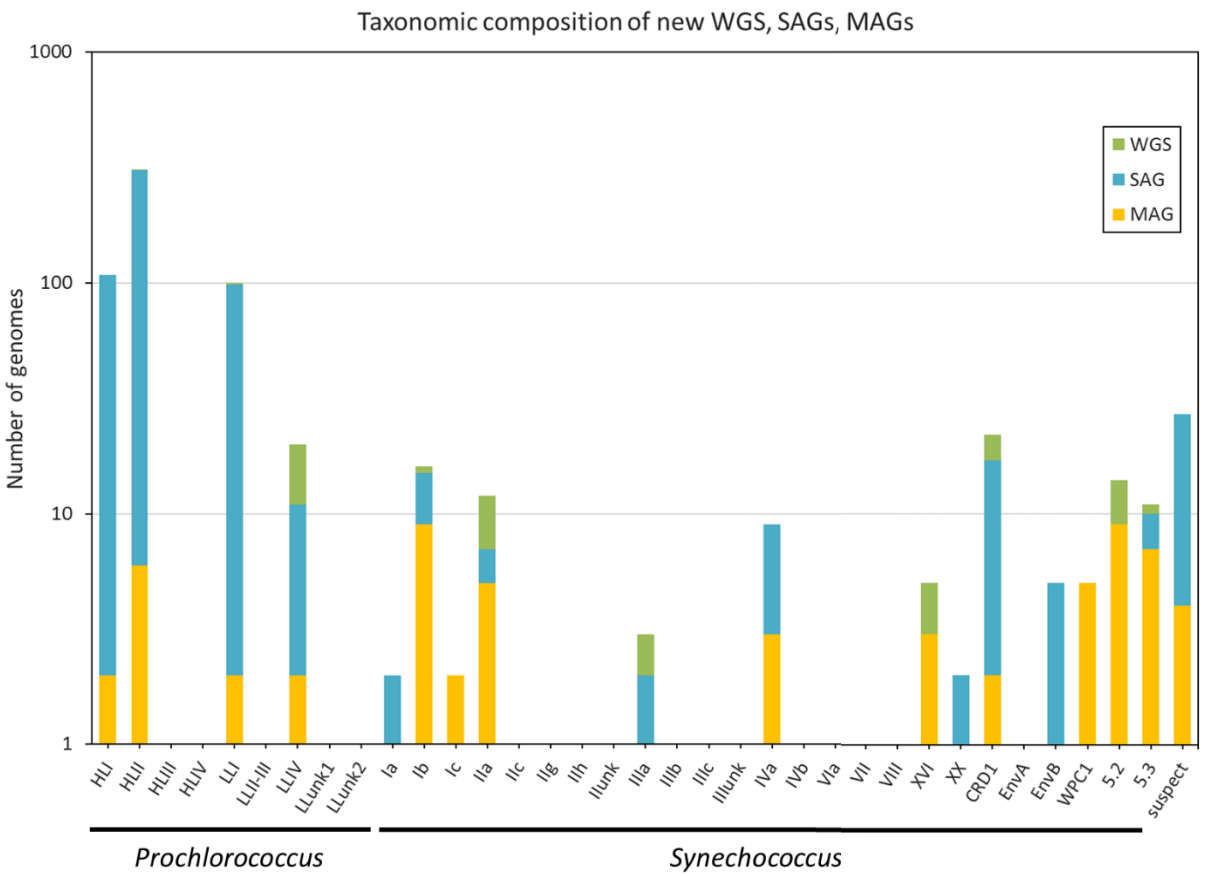


Figure 23 - Représentation du nombre de génomes par clades pour *Prochlorococcus et Synechococcus*. La couleur correspond à la méthode avec laquelle a été générée les génomes telle que précisée dans l'insert. Les génomes indiqués comme "suspect" correspondent aux probables chimères.

En résumé, nous disposons donc désormais d'un set de WGS richement annotés et qui pourront être ajoutés à la prochaine version de la base de données Cyanorak ainsi que d'un grand nombre de SAGs et MAGs qui eux aussi ont bénéficié des annotations de Cyanorak et pourront être utilisés, en complément des génomes de référence, pour de futures analyses de méta-omique. A noter que malgré l'ajout de ces 782 WGS/SAGs/MAGs, multipliant par 8 le nombre de génomes disponibles pour ces deux genres, on est encore bien loin de la saturation du pangenome de *Synechococcus* et celui de *Prochlorococcus* semble plus proche d'atteindre un plateau mais n'est toujours pas saturé malgré presque 700 génomes séquencés pour ce seul genre (Fig. 24).

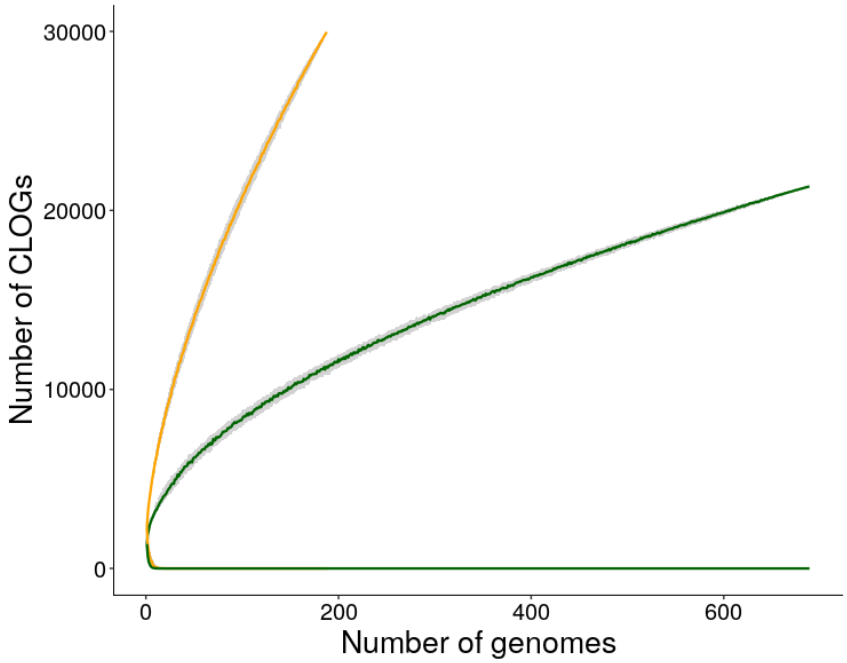


Figure 24 - Analyse des pangénomes de *Prochlorococcus* (vert) et *Synechococcus* (orange) calculés à partir des génomes nouvellement téléchargés et des génomes de Cyanorak. Analyses réalisées avec 188 génomes de *Synechococcus* (dont 54 génomes présents dans Cyanorak) et 688 génomes de *Prochlorococcus* (dont 43 présents dans Cyanorak).

Je n'ai malheureusement pas eu le temps d'analyser plus en détail ce nouveau set de données, mais il a déjà par exemple permis de vérifier que plusieurs génomes de CRD1 et de plusieurs autres clades possédaient, comme MITS9220, une copie du gène *ftrC* mais pas de *ftrV*, ce qui semble confirmer l'hypothèse formulée dans le chapitre III à propos de la présence éventuelle d'une forme tronquée de cette protéine ne fixant pas de cluster [4Fe-4S], et qui serait spécifique des populations de *Synechococcus* peuplant les zones pauvres en fer. Cette nouvelle ressource devrait donc s'avérer très précieuse pour l'équipe, notamment dans le cadre i) de l'analyse des métagénomes prélevés lors de la campagne TONGA en novembre 2019, dont le but est d'explorer l'influence des gradients de concentration en fer d'origine hydrothermale dans le Pacifique sud-ouest et ii) du projet ANR EFFICACY (2020-24) qui prévoit d'analyser les variations saisonnières de la diversité taxonomique et fonctionnelle des communautés de *Synechococcus* à deux stations contrastées d'un point de vue océanographique, l'une située en Manche au large de Roscoff (Astan) et l'autre en Méditerranée au large de Villefranche-sur-Mer (Boussole). Ces analyses des variations saisonnières de la composition des communautés et de leur diversité fonctionnelle seront très complémentaires des études réalisées à large échelle spatiale dans le cadre de ma thèse.

Références

A

Ahlgren, N.A., and 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*, 7193–7204.

Ahlgren, N.A., and 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*, 1–24.

Ahlgren, N.A., Belisle, B.S., and Lee, M.D. (2020). Genomic mosaicism underlies the adaptation of marine *Synechococcus* ecotypes to distinct oceanic iron niches. Environ. Microbiol. *22*, 1801–1815.

Allahverdiyeva, Y., Ermakova, M., Eisenhut, M., Zhang, P., Richaud, P., Hagemann, M., Cournac, L., and Aro, E.-M. (2011). Interplay between Flavodiiron Proteins and Photorespiration in Synechocystis sp. PCC 6803. J. Biol. Chem. *286*, 24007–24014.

Allahverdiyeva, Y., Mustila, H., Ermakova, M., Bersanini, L., Richaud, P., Ajlani, G., Battchikova, N., Cournac, L., and Aro, E.-M. (2013). Flavodiiron proteins Flv1 and Flv3 enable cyanobacterial growth and photosynthesis under fluctuating light. Proc. Natl. Acad. Sci. *110*, 4111–4116.

Allen, J.F. (2003). State transitions–a question of balance. Science *299*, 1530–1532.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Azam, F., Fenchel, T., Field, J., Gray, J., Meyer-Reil, L., and Thingstad, F. (1983). The Ecological Role of Water-Column Microbes in the Sea. Mar. Ecol. Prog. Ser. *10*, 257–263.

B

Bailey, S., Mann, N.H., Robinson, C., and Scanlan, D.J. (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, K.R.M., Cardol, P., Finazzi, G., van Dijken, G., Berg, G.M., Arrigo, K., Shrager, J., and Grossman, A. (2008). Alternative photosynthetic electron flow to oxygen in marine <i>Synechococcus<\i>. Biochim. Biophys. Acta - Bioenerg. *1777*, 269–276.

Baumdicker, F., Hess, W.R., and Pfaffelhuber, P. (2012). The Infinitely Many Genes Model for the Distributed Genome of Bacteria. Genome Biol. Evol. *4*, 443–456.

Behrenfeld, M.J., and Kolber, Z.S. (1999). Widespread iron limitation of phytoplankton in the south Pacific Ocean. Science *283*, 840–843.

Bersanini, L., Battchikova, N., Jokel, M., Rehman, A., Vass, I., Allahverdiyeva, Y., and Aro, E.-M. (2014). Flavodiiron Protein Flv2/Flv4-Related Photoprotective Mechanism Dissipates Excitation Pressure of PSII in Cooperation with Phycobilisomes in Cyanobacteria1[C][W][OPEN]. Plant Physiol. *164*, 805–818.

Bertilsson, S., Berglund, O., Karl, D.M., and Chisholm, S.W. (2003). Elemental composition of marine Prochlorococcus and Synechococcus: Implications for the ecological stoichiometry of the sea. Limnol. Oceanogr. *48*, 1721–1731.

Berube, P.M., Biller, S.J., Kent, A.G., Berta-Thompson, J.W., Roggensack, S.E., Roache-Johnson, K.H., Ackerman, M., Moore, L.R., Meisel, J.D., Sher, D., et al. (2014).

Physiology and evolution of nitrate acquisition in Prochlorococcus. ISME J. *9*, 1195–1207.

Berube, P.M., Biller, S.J., Hackl, T., Hogle, S.L., Satinsky, B.M., Becker, J.W., Braakman, R., Collins, S.B., Kelly, L., Berta-Thompson, J., et al. (2018). Single cell genomes of Prochlorococcus, Synechococcus, and sympatric microbes from diverse marine environments. Sci. Data *5*, 180154.

Berube, P.M., Rasmussen, A., Braakman, R., Stepanauskas, R., and Chisholm, S.W. (2019). Emergence of trait variability through the lens of nitrogen assimilation in Prochlorococcus. ELife *8*, e41043.

Bibby, T.S., Nield, J., and Barber, J. (2001). Iron deficiency induces the formation of an antenna ring around trimeric photosystem I in cyanobacteria. Nature *412*, 743–745.

Bibby, T.S., Mary, I., Nield, J., Partensky, F., and Barber, J. (2003). Low-light-adapted Prochlorococcus species possess specific antennae for each photosystem. Nature *424*, 1051–1054.

Biller, S.J., Schubotz, F., Roggensack, S.E., Thompson, A.W., Summons, R.E., and Chisholm, S.W. (2014). Bacterial vesicles in marine ecosystems. Science *343*, 183–6.

Biller, S.J., Berube, P.M., Lindell, D., and Chisholm, S.W. (2015). *Prochlorococcus*: the structure and function of collective diversity. Nat. Rev. Microbiol. *13*, 13–27.

Blainey, P.C. (2013). The future is now: single-cell genomics of bacteria and archaea. FEMS Microbiol. Rev. *37*, 407–427.

Blot, N., Mella-Flores, D., Six, C., Lecorguille, G., Boutte, C., Peyrat, A., Monnier, A., Ratin, M., Gourvil, P., Campbell, D.A., et al. (2011). Light history influences the response of the marine cyanobacterium Synechococcus sp. WH7803 to oxidative stress. Plant Physiol. *156*, 1934–1954.

Boekema, E.J., Hifney, A., Yakushevska, A.E., Piotrowski, M., Keegstra, W., Berry, S., Michel, K.-P., Pistorius, E.K., and Kruip, J. (2001). A giant chlorophyll–protein complex induced by iron deficiency in cyanobacteria. Nature *412*, 745–748.

Bolstad, H.M., and Wood, M.J. (2010). An In Vivo Method for Characterization of Protein Interactions within Sulfur Trafficking Systems of E. coli. J. Proteome Res. *9*, 6740–6751.

Boulay, C., Abasova, L., Six, C., Vass, I., and Kirilovsky, D. (2008). Occurrence and function of the orange carotenoid protein in photoprotective mechanisms in various cyanobacteria. Biochim. Biophys. Acta BBA - Bioenerg. *1777*, 1344–1354.

Boulay, C., Wilson, A., D'Haene, S., and Kirilovsky, D. (2010). Identification of a protein required for recovery of full antenna capacity in OCP-related photoprotective mechanism in cyanobacteria. Proc. Natl. Acad. Sci. *107*, 11620–11625.

Boyd, P.W., Watson, A.J., Law, C.S., Abraham, E.R., Trull, T., Murdoch, R., Bakker, D.C.E., Bowie, A.R., Buesseler, K.O., Chang, H., et al. (2000). A mesoscale phytoplankton bloom in the polar Southern Ocean stimulated by iron fertilization. Nature *407*, 695–702.

Breitbart, M. (2012). Marine Viruses: Truth or Dare. Annu. Rev. Mar. Sci. *4*, 425–448.

Breton, S., Jouhet, J., Guyet, U., Gros, V., Pittera, J., Demory, D., Partensky, F., Doré, H., Ratin, M., Maréchal, E., et al. (2019). Unveiling membrane thermoregulation strategies in marine picocyanobacteria. New Phytol. nph.16239.

Buchan, A., LeCleir, G.R., Gulvik, C.A., and Gonzalez, J.M. (2014). Master recyclers: features and functions of bacteria associated with phytoplankton blooms. Nat. Rev. Microbiol. *12*, 686–698.

C

Cabello-Yeves, P.J., Haro-Moreno, J.M., Martin-Cuadrado, A.B., Ghai, R., Picazo, A., Camacho, A., and Rodriguez-Valera, F. (2017). Novel Synechococcus genomes reconstructed from freshwater reservoirs. Front. Microbiol. *8*, 1–13.

Cai, H., Wang, K., Huang, S., Jiao, N., and Chen, F. (2010). Distinct Patterns of Picocyanobacterial Communities in Winter and Summer in the Chesapeake Bay. Appl. Environ. Microbiol. *76*, 2955–2960.

Campbell, L., and Vaulot, D. (1993). Photosynthetic picoplankton community structure in the subtropical North Pacific Ocean near Hawaii (station ALOHA). Deep-Sea Res. Part I *40*, 2043–2060.

Casey, J.R., Mardinoglu, A., Nielsen, J., and Karl, D.M. (2016). Adaptive Evolution of Phosphorus Metabolism in Prochlorococcus. MSystems *1*, 1–15.

Chandler, J.W., Lin, Y., Gainer, P.J., Post, A.F., Johnson, Z.I., and Zinser, E.R. (2016). Variable but persistent coexistence of \textlessi\textgreaterProchlorococcus\textless/i\textgreater 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, M.T., and Wommack, K.E. (2006). Diverse and Unique Picocyanobacteria in Chesapeake Bay , Revealed by 16S-23S rRNA Internal Transcribed Spacer Sequences. Appl. Environ. Microbiol. *72*, 2239–2243.

Chi, X., Yang, Q., Zhao, F., Qin, S., Yang, Y., Shen, J., and Lin, H. (2008). Comparative Analysis of Fatty Acid Desaturases in Cyanobacterial Genomes. Comp. Funct. Genomics *2008*.

Chisholm, S.W. (1992). Phytoplankton Size. In Primary Productivity and Biogeochemical Cycles in the Sea, P.G. Falkowski, A.D. Woodhead, and K. Vivirito, eds. (Boston, MA: Springer US), pp. 213–237.

Chisholm, S.W., Olson, R.J., Zettler, E.R., Goericke, R., Waterbury, J.B., and Welschmeyer, N.A. (1988). A novel free-living prochlorophyte abundant in the oceanic euphotic zone. Nature *334*, 340–343.

Chisholm, S.W., Frankel, S.L., Goericke, R., Olson, R.J., Palenik, B., Waterbury, J.B., West-Johnsrud, L., and Zettler, E.R. (1992). Prochlorococcus marinus nov. gen. nov. sp.: an oxyphototrophic marine prokaryote containing divinyl chlorophyll a and b. Arch. Microbiol. *157*, 297–300.

Choi, D.H., and Noh, J.H. (2009). Phylogenetic diversity of Synechococcus strains isolated from the East China Sea and the East Sea. FEMS Microbiol. Ecol. *69*, 439–448.

Clark, L.L., Ingall, E.D., and Benner, R. (1998). Marine phosphorus is selectively remineralized. Nature *393*, 426–426.

Clifford, E.L., Varela, M.M., De Corte, D., Bode, A., Ortiz, V., Herndl, G.J., and Sintes, E. (2019). Taurine Is a Major Carbon and Energy Source for Marine Prokaryotes in the North Atlantic Ocean off the Iberian Peninsula. Microb. Ecol. *78*, 299–312.

Coale, K.H., Johnson, K.S., Fitzwater, S.E., Gordon, R.M., Tanner, S., Chavez, F.P., Ferioli, L., Sakamoto, C., Rogers, P., Millero, F., et al. (1996). A massive phytoplankton bloom induced by an ecosystem-scale iron fertilization experiment in the equatorial Pacific Ocean. Nature *383*, 495–501.

Coleman, M.L., Sullivan, M.B., Martiny, A.C., Steglich, C., Barry, K., DeLong, E.F., and Chisholm, S.W. (2006). Genomic Islands and the Ecology and Evolution of Prochlorococcus. Science *311*, 1768–1770.

Collier, J.L., Brahamsha, B., and 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.

Cooley, J.W., and Vermaas, W.F.J. (2001). Succinate Dehydrogenase and Other Respiratory Pathways in Thylakoid Membranes of Synechocystis sp. Strain PCC 6803: Capacity Comparisons and Physiological Function. J. Bacteriol. *183*, 4251–4258.

Cornet, L., Bertrand, A.R., Hanikenne, M., Javaux, E.J., Wilmotte, A., and Baurain, D. (2018). Metagenomic assembly of new (sub)polar Cyanobacteria and their associated microbiome from non-axenic cultures. Microb. Genomics 15.

D

Dai, Y., and Outten, F.W. (2012). The E. coli SufS–SufE sulfur transfer system is more resistant to oxidative stress than IscS–IscU. FEBS Lett. *586*, 4016–4022.

Dai, S., Johansson, K., Miginiac-Maslow, M., Schürmann, P., and Eklund, H. (2004). Structural Basis of Redox Signaling in Photosynthesis: Structure and Function of Ferredoxin:thioredoxin Reductase and Target Enzymes. Photosynth. Res. *79*, 233–248.

Delmont, T.O., Quince, C., Shaiber, A., Esen, Ö.C., Lee, S.T., Rappé, M.S., McLellan, S.L., Lücker, S., and Eren, A.M. (2018). Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. Nat. Microbiol. *3*, 804–813.

Dufresne, A., Salanoubat, M., Partensky, F., Artiguenave, F., Axmann, I.M., Barbe, V., Duprat, S., Galperin, M.Y., Koonin, E.V., Le Gall, F., et al. (2003). Genome sequence of the cyanobacterium Prochlorococcus marinus SS120, a nearly minimal oxyphototrophic genome. Proc. Natl. Acad. Sci. U. S. A. *100*, 10020–10025.

Dufresne, A., Garczarek, L., and 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, D.J., Garczarek, L., Mazard, S., Palenik, B., Paulsen, I.T., de Marsac, N.T., Wincker, P., Dossat, C., et al. (2008). Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. Genome Biol. *9*, R90.

Dupont, C.L., Johnson, D.A., Phillippy, K., Paulsen, I.T., Brahamsha, B., and Palenik, B. (2012). Genetic Identification of a High-Affinity Ni Transporter and the Transcriptional Response to Ni Deprivation in Synechococcus sp. Strain WH8102. Appl. Environ. Microbiol. *78*, 7822–7832.

E

Eisenhut, M., Ruth, W., Haimovich, M., Bauwe, H., Kaplan, A., and Hagemann, M. (2008). The photorespiratory glycolate metabolism is essential for cyanobacteria and might have been conveyed endosymbiontically to plants. Proc. Natl. Acad. Sci. *105*, 17199–17204.

Emlyn-Jones, D., Ashby, M.K., and Mullineaux, C.W. (1999). A gene required for the regulation of photosynthetic light harvesting in the cyanobacterium Synechocystis 6803. Mol. Microbiol. *33*, 1050–1058.

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

Everroad, C., Six, C., Partensky, F., Thomas, J.-C., Holtzendorff, J., and Wood, A.M. (2006). Biochemical Bases of Type IV Chromatic Adaptation in Marine Synechococcus spp. J. Bacteriol. *188*, 3345–3356.

F

Farrant, G.K., Doré, H., Cornejo-Castillo, F.M., Partensky, F., Ratin, M., Ostrowski, M., Pitt, F.D., Wincker, P., Scanlan, D.J., Iudicone, D., et al. (2016). Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. Proc. Natl. Acad. Sci. *113*, E3365–E3374.

Feingersch, R., Philosof, A., Mejuch, T., Glaser, F., Alalouf, O., Shoham, Y., and Béjà, O. (2012). Potential for phosphite and phosphonate utilization by Prochlorococcus. ISME J. *6*, 827–834.

Ferris, M.J., and Palenik, B. (1998). Niche adaptation in ocean cyanobacteria. Nature *396*, 226–228.

Field, C.B., Behrenfeld, M.J., Randerson, J.T., and Falkowski, P. (1998). Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. Science *281*, 237–240.

Findlay, V.J., Tapiero, H., and Townsend, D.M. (2005). Sulfiredoxin: a potential therapeutic agent? Biomed. Pharmacother. *59*, 374–379.

Flombaum, P., Gallegos, J.L., Gordillo, R. a, Rincón, J., Zabala, L.L., Jiao, N., Karl, D., Li, W., Lomas, M., Veneziano, D., et al. (2013). Present and future global distributions of the marine Cyanobacteria Prochlrococcus and Synechococcus. Pnas *110*, 9824–9829.

Frías, J.E., Herrero, A., and Flores, E. (2003). Open Reading Frame all0601 from Anabaena sp. Strain PCC 7120 Represents a Novel Gene, cnaT, Required for Expression of the Nitrate Assimilation nir Operon. J. Bacteriol. *185*, 5037–5044.

Fu, F., Zhang, Y., Feng, Y., and Hutchins, D.A. (2006). Phosphate and ATP uptake and growth kinetics in axenic cultures of the cyanobacterium Synechococcus CCMP 1334. Eur. J. Phycol. *41*, 15–28.

Fuller, N.J., Marie, D., Partensky, F., Vaulot, D., Post, A.F., and Scanlan, D.J. (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, N.J., West, N.J., Marie, D., Yallop, M., Rivlin, T., Post, A.F., and Scanlan, D.J. (2005). Dynamics of community structure and phosphate status of picocyanobacterial populations in the Gulf of Aqaba, Red Sea. Limnol. Oceanogr. *50*, 363–375.

Fung, I.Y., Meyn, S.K., Tegen, I., Doney, S.C., John, J.G., and Bishop, J.K.B. (2000). Iron supply and demand in the upper ocean. Glob. Biogeochem. Cycles *14*, 281–295.

G

Garcia, C.A., Hagstrom, G.I., Larkin, A.A., Ustick, L.J., Levin, S.A., Lomas, M.W., and Martiny, A.C. (2020). Linking regional shifts in microbial genome adaptation with surface ocean biogeochemistry. Philos. Trans. R. Soc. B Biol. Sci. *375*, 20190254.

Garczarek, L., Dufresne, A., Blot, N., Cockshutt, A.M., Peyrat, A., Campbell, D.A., Joubin, L., and Six, C. (2008). Function and evolution of the psbA gene family in marine Synechococcus : Synechococcus sp. WH7803 as a case study. ISME J. *2*, 937–953.

Geiß, U., Vinnemeier, J., Kunert, A., Lindner, I., Gemmer, B., Lorenz, M., Hagemann, M., and Schoor, A. (2001). Detection of the isiA Gene across Cyanobacterial Strains: Potential for Probing Iron Deficiency. Appl. Environ. Microbiol. *67*, 5247–5253.

Grébert, T., Doré, H., Partensky, F., Farrant, G.K., Boss, E.S., Picheral, M., Guidi, L., Pesant, S., Scanlan, D.J., Wincker, P., et al. (2018). Light color acclimation is a key process in the global ocean distribution of *Synechococcus cyanobacteria*. Proc. Natl. Acad. Sci. *115*, E2010–E2019.

Grotjohann, I., and Fromme, P. (2005). Structure of cyanobacterial Photosystem I. Photosynth. Res. *85*, 51–72.

H

Hagemann, M., Richter, S., and Mikkat, S. (1997). The ggtA gene encodes a subunit of the transport system for the osmoprotective compound glucosylglycerol in Synechocystis sp. strain PCC 6803. J. Bacteriol. *179*, 714–720.

Haro-Moreno, J.M., López-Pérez, M., de la Torre, J.R., Picazo, A., Camacho, A., and Rodriguez-Valera, F. (2018). Fine metagenomic profile of the Mediterranean stratified and mixed water columns revealed by assembly and recruitment. Microbiome *6*, 128.

Haroon, M.F., Thompson, L.R., Parks, D.H., Hugenholtz, P., and Stingl, U. (2016). A catalogue of 136 microbial draft genomes from Red Sea metagenomes. Sci. Data *3*, 160050.

Hassler, C.S., Sinoir, M., Clementson, L.A., and Butler, E.C.V. (2012). Exploring the Link between Micronutrients and Phytoplankton in the Southern Ocean during the 2007 Austral Summer. Front. Microbiol. *3*.

Havaux, M., Guedeney, G., He, Q., and Grossman, A.R. (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, H.C.P., and 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.

Haverkamp, T., Acinas, S.G., Doeleman, M., Stomp, M., Huisman, J., and Stal, L.J. (2008). Diversity and phylogeny of Baltic Sea picocyanobacteria inferred from their ITS and phycobiliprotein operons. Environ. Microbiol. *10*, 174–188.

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

Herdman, M., Castenholz, R., Iteman, I., Waterbury, J., and Rippka, R. (2001). The Archaea and the deeply branching and phototrophic bacteria. Bergey's Man. Syst. Bacteriol. 493–514.

Hoffmann, D., Gutekunst, K., Klissenbauer, M., Schulz-Friedrich, R., and Appel, J. (2006). Mutagenesis of hydrogenase accessory genes of *Synechocystis* sp. PCC 6803: Additional homologues of *hypA* and *hypB* are not active in hydrogenase maturation. FEBS J. *273*, 4516–4527.

Hoiczyk, E., and Hansel, A. (2000). Cyanobacterial Cell Walls: News from an Unusual Prokaryotic Envelope. J. Bacteriol. *182*, 1191–1199.

Holtzendorff, J., Partensky, F., Mella, D., Lennon, J.-F., Hess, W.R., and Garczarek, L. (2008). Genome streamlining results in loss of robustness of the circadian clock in the marine cyanobacterium Prochlorococcus marinus PCC 9511. J. Biol. Rhythms *23*, 187–199.

Huang, S., Wilhelm, S.W., Harvey, H.R., Taylor, K., Jiao, N., and Chen, F. (2012). Novel lineages of Prochlorococcus and Synechococcus in the global oceans. ISME J. *6*, 285–297.

Humily, F., Partensky, F., Six, C., Farrant, G.K., Ratin, M., Marie, D., and Garczarek, L. (2013). A gene island with two possible configurations is involved in chromatic acclimation in marine Synechococcus. PLoS ONE *8*, e84459.

Humily, F., Farrant, G.K., Marie, D., Partensky, F., Mazard, S., Perennou, M.,

Labadie, K., Aury, J.-M., Wincker, P., Segui, A.N., 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*, 231–249.

Hunt, D.E., Lin, Y., Church, M.J., Karl, D.M., Tringe, S.G., Izzo, L.K., and Johnson, I. (2013). Relationship between Abundance and Specific Activity of Bacterioplankton in Open Ocean Surface Waters. Appl. Environ. Microbiol. *79*, 177–184.

Hunter-Cevera, K.R., Post, A.F., Peacock, E.E., and Sosik, H.M. (2016). Diversity of Synechococcus at the Martha's Vineyard Coastal Observatory: Insights from Culture Isolations, Clone Libraries, and Flow Cytometry. Microb. Ecol. *71*, 276–289.

J

Johnson, Z.I., Zinser, E.R., Coe, A., Mcnulty, N.P., Malcolm, E.S., Chisholm, S.W., Woodward, E.M.S., and Chisholm, S.W. (2006). Partitioning Among Prochlorococcus Ecotypes Along Environmental Gradients. Science *311*, 1737–1740.

K

Kashtan, N., Roggensack, S.E., Rodrigue, S., Thompson, J.W., Biller, S.J., Coe, A., Ding, H., Marttinen, P., Malmstrom, R.R., Stocker, R., et al. (2014). Single-cell genomics reveals hundreds of coexisting subpopulations in wild Prochlorococcus. Science *344*, 416–420.

Kashtan, N., Roggensack, S.E., Berta-Thompson, J.W., Grinberg, M., Stepanauskas, R., and Chisholm, S.W. (2017). Fundamental differences in diversity and genomic population structure between Atlantic and Pacific Prochlorococcus. ISME J. 1–15.

Kathuria, S., and Martiny, A.C. (2011). Prevalence of a calcium-based alkaline phosphatase associated with the marine cyanobacterium *Prochlorococcus* and other ocean bacteria: PhoX in *Prochlorococcus*. Environ. Microbiol. *13*, 74–83.

Kent, A.G., Dupont, C.L., Yooseph, S., and Martiny, A.C. (2016). Global biogeography of Prochlorococcus genome diversity in the surface ocean. ISME J. *10*, 1856–1865.

Kent, A.G., Baer, S.E., Mouginot, C., Huang, J.S., Larkin, A.A., Lomas, M.W., and Martiny, A.C. (2019). Parallel phylogeography of Prochlorococcus and Synechococcus. ISME J. *13*, 430–441.

Kerfeld, C.A., Melnicki, M.R., Sutter, M., and Dominguez-martin, M.A. (2017). Structure, function and evolution of the cyanobacterial orange carotenoid protein and its homologs. New Phytol. *215*, 937–951.

Kettler, G.C., Martiny, A.C., Huang, K., Zucker, J., Coleman, M.L., Rodrigue, S., Chen, F., Lapidus, A., Ferriera, S., Johnson, J., et al. (2007). Patterns and implications of gene gain and loss in the evolution of Prochlorococcus. PLoS Genet. *3*, 2515–2528.

Kirilovsky, D., and A. Kerfeld, C. (2013). The Orange Carotenoid Protein : a blue-green light photoactive protein. Photochem. Photobiol. Sci. *12*, 1135–1143.

Kogawa, M., Hosokawa, M., Nishikawa, Y., Mori, K., and Takeyama, H. (2018). Obtaining high-quality draft genomes from uncultured microbes by cleaning and co-assembly of single-cell amplified genomes. Sci. Rep. *8*, 2059.

Kranzler, C., Rudolf, M., Keren, N., and Schleiff, E. (2013). Iron in Cyanobacteria. In Advances in Botanical Research, (Elsevier), pp. 57–105.

Kranzler, C., Lis, H., Finkel, O.M., Schmetterer, G., Shaked, Y., and Keren, N. (2014). Coordinated transporter activity shapes high-affinity iron acquisition in cyanobacteria. ISME J. *8*, 409–417.

Krieger-Liszkay, A., Fufezan, C., and Trebst, A. (2008). Singlet oxygen production in photosystem II and related protection mechanism. Photosynth. Res. *98*, 551–564.

L

Lambrecht, S.J., Steglich, C., and Hess, W.R. (2020). A minimum set of regulators to thrive in the ocean. FEMS Microbiol. Rev. fuaa005.

Landfald, B., and Strøm, A.R. (1986). Choline-glycine betaine pathway confers a high level of osmotic tolerance in Escherichia coli. J. Bacteriol. *165*, 849–855.

Lantoine, F., and Neveux, J. (1997). Spatial and seasonal variations in abundance and spectral characteristics of phycoerithrins in the tropical northeastern Atlantic Ocean. Deep-Sea Res. I *44*, 223–246.

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

Larkin, A.A., Blinebry, S.K., Howes, C., Lin, Y., Loftus, S.E., Schmaus, C.A., Zinser, E.R., and Johnson, Z.I. (2016). Niche partitioning and biogeography of high light adapted Prochlorococcus across taxonomic ranks in the North Pacific. ISME J. 1–13.

Lasken, R.S., and McLean, J.S. (2014). Recent advances in genomic DNA sequencing of microbial species from single cells. Nat. Rev. Genet. *15*, 577–584.

Latifi, A., Ruiz, M., and Zhang, C.C. (2009). Oxidative stress in cyanobacteria. FEMS Microbiol. Rev. *33*, 258–278.

Lavin, P., González, B., Santibáñez, J.F., Scanlan, D.J., and 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.

Lea-Smith, D.J., Ross, N., Zori, M., Bendall, D.S., Dennis, J.S., Scott, S.A., Smith, A.G., and Howe, C.J. (2013). Thylakoid terminal oxidases are essential for the cyanobacterium Synechocystis sp. PCC 6803 to survive rapidly changing light intensities. Plant Physiol. *162*, 484–495.

Lemeille, S., and Rochaix, J.-D. (2010). State transitions at the crossroad of thylakoid signalling pathways. Photosynth. Res. *106*, 33–46.

Li, L., Stoeckert, C.J.J., and Roos, D.S. (2003). OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes – Li et al. 13 (9): 2178 – Genome Research. Genome Res. *13*, 2178–2189.

Lindell, D., and Padan, E. (1998). Regulation of ntcA Expression and Nitrite Uptake in the Marine Synechococcus sp . Strain WH7803. J. Bacteriol. *180*, 1878–1886.

Lindell, D., and Post, A.F. (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., Erdner, D., Marie, D., Prásil, O., Koblíjek, M., Le Gall, F., Rippka, R., Partensky, F., Scanlan, D.J., and Post, A.F. (2002). Nitrogen Stress Response of Prochlorococcus Strain PCC9511 (Oxyphotobacteria) Involves Contrasting Regulation of ntcA and amt1. J. Phycol. *38*, 1113–1124.

Lindell, D., Sullivan, M.B., Johnson, Z.I., Tolonen, A.C., Rohwer, F., and Chisholm, S.W. (2004). Transfer of photosynthesis genes to and from Prochlorococcus viruses. Proc Natl Acad Sci U A *101*, 11013–11018.

Los, D.A., and Murata, N. (1998). Structure and expression of fatty acid desaturases. Biochim. Biophys. Acta BBA - Lipids Lipid Metab. *1394*, 3–15.

Lott, S.C., Voigt, K., Lambrecht, S.J., Hess, W.R., and Steglich, C. (2020). A framework

for the computational prediction and analysis of non-coding RNAs in microbial environmental populations and their experimental validation. ISME J.

Luque, I., Flores, E., and Herrero, A. (1993). Nitrite reductase gene from Synechococcus sp. PCC 7942: homology between cyanobacterial and higher-plant nitrite reductases. Plant Mol. Biol. *21*, 1201–1205.

M

Ma, L., Calfee, B.C., Morris, J.J., Johnson, Z.I., and Zinser, E.R. (2018). Degradation of hydrogen peroxide at the ocean's surface: the influence of the microbial community on the realized thermal niche of Prochlorococcus. ISME J. *12*, 473–484.

Mackey, K.R.M., Paytan, A., Caldeira, K., Grossman, A.R., Moran, D., McIlvin, M., and Saito, M.A. (2013). Effect of Temperature on Photosynthesis and Growth in Marine Synechococcus spp. Plant Physiol. *163*, 815–829.

Mackey, K.R.M., Post, A.F., Mcilvin, M.R., Cutter, G.A., John, S.G., and Saito, M.A. (2015). Divergent responses of Atlantic coastal and oceanic Synechococcus to iron limitation. *112*, 9944–9949.

Malmstrom, R.R., Coe, A., Kettler, G.C., Martiny, A.C., Frias-Lopez, J., Zinser, E.R., and Chisholm, S.W. (2010). Temporal dynamics of Prochlorococcus ecotypes in the Atlantic and Pacific oceans. ISME J. *4*, 1252–1264.

Malmstrom, R.R., Rodrigue, S., Huang, K.H., Kelly, L., Kern, S.E., Thompson, A., Roggensack, S., Berube, P.M., Henn, M.R., and Chisholm, S.W. (2013). Ecology of uncultured Prochlorococcus clades revealed through single-cell genomics and biogeographic analysis - ismej201289a.pdf. ISME J. *7*, 184–198.

Mann, E.L., and Chisholm, S.W. (2000). Iron limits the cell division rate of Prochlorococcus in the eastern equatorial Pacific. Limnol. Oceanogr. *45*, 1067–1076.

Marsan, D.W. (2016). Adaptive mechanisms of an estuarine Synechococcusbased on genomics, transcriptomics, and proteomics. PhD Thesis. University of Maryland, College Park.

Martin, J.H., Coale, K.H., Johnson, K.S., Fitzwater, S.E., Gordon, R.M., Tanner, S.J., Hunter, C.N., Elrod, V.A., Nowicki, J.L., Coley, T.L., et al. (1994). Testing the iron hypothesis in ecosystems of the equatorial Pacific Ocean. Nature *371*, 123–129.

Martinez, A., Tyson, G.W., and DeLong, E.F. (2010). Widespread known and novel phosphonate utilization pathways in marine bacteria revealed by functional screening and metagenomic analyses. Environ. Microbiol. *12*, 222–238.

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

Martiny, A.C., Tai, A.P.K., Veneziano, D., Primeau, F., and Chisholm, S.W. (2009a). Taxonomic resolution, ecotypes and the biogeography of Prochlorococcus. Environ. Microbiol. *11*, 823–832.

Martiny, A.C., Kathuria, S., and Berube, P.M. (2009b). Widespread metabolic potential for nitrite and nitrate assimilation among \emph{Prochlorococcus} ecotypes. Proc. Natl. Acad. Sci. *106*, 10787–10792.

Martiny, A.C., Huang, Y., and Li, W. (2009c). Occurrence of phosphate acquisition genes in Prochlorococcus cells from different ocean regions. Environ. Microbiol. *11*, 1340–1347.

Martiny, J.B.H., Jones, S.E., Lennon, J.T., and Martiny, A.C. (2015). Microbiomes in light of traits: A phylogenetic perspective. Science *350*.

Mazard, S., Ostrowski, M., Partensky, F., and Scanlan, D.J. (2012a). Multi-locus sequence analysis, taxonomic resolution and biogeography of marine Synechococcus. Environ. Microbiol. *14*, 372–386.

Mazard, S., Wilson, W.H., and Scanlan, D.J. (2012b). Dissecting the physiological response to phosphorus stress in marine synechococcus isolates (cyanophyceae). J. Phycol. *48*, 94–105.

McDonald, B.R., and Currie, C.R. (2017). Lateral gene transfer dynamics in the ancient bacterial genus Streptomyces. MBio *8*, e00644–17.

McSorley, F.R., Wyatt, P.B., Martinez, A., DeLong, E.F., Hove-Jensen, B., and Zechel, D.L. (2012). PhnY and PhnZ Comprise a New Oxidative Pathway for Enzymatic Cleavage of a Carbon–Phosphorus Bond. J. Am. Chem. Soc. *134*, 8364–8367.

Mella-Flores, D., Mazard, S., Humily, F., Partensky, F., Mahé, F., Bariat, L., Courties, C., Marie, D., Ras, J., Mauriac, R., 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., Marie, D., Blot, N., Gourvil, P., Kolowrat, C., et al. (2012). Prochlorococcus and Synechococcus have Evolved Different Adaptive Mechanisms to Cope with Light and UV Stress. Front. Microbiol. *3*.

Michel, K.-P., and Pistorius, E.K. (2004). Adaptation of the photosynthetic electron transport chain in cyanobacteria to iron deficiency: The function of IdiA and IsiA. Physiol. Plant. *120*, 36–50.

Mikami, K., and Murata, N. (2003). Membrane fluidity and the perception of environmental signals in cyanobacteria and plants. Prog. Lipid Res. *42*, 527–543.

Moore, L.R., and Chisholm, S.W. (1999). Photophysiology of the marine cyanobacterium Prochlorococcus: Ecotypic differences among cultured isolates. Limnol. Oceanogr. *44*, 628–638.

Moore, J.K., Doney, S.C., and Lindsay, K. (2004). Upper ocean ecosystem dynamics and iron cycling in a global three-dimensional model. Glob. Biogeochem. Cycles *18*.

Moore, L.R., Goericke, R., and Chisholm, S.W. (1995). Comparative physiology of Synechococcus and Prochlorococcus: influence of light and temperature on growth, pigments, fluorescence and absorptive properties. Mar. Ecol. Prog. Ser. 259–275.

Moore, L.R., Rocap, G., and Chisholm, S.W. (1998). Physiology and molecular phylogeny of coexisting Prochlorococcus ecotypes. Nature *393*, 464–467.

Moore, L.R., Ostrowski, M., Scanlan, D.J., Feren, K., and Sweetsir, T. (2005). Ecotypic variation in phosphorus-acquisition mechanisms within marine picocyanobacteria. Aquat. Microb. Ecol. *39*, 257–269.

Morel, A., Ahn, Y.-H., Partensky, F., Vaulot, D., and 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, J.J., Kirkegaard, R., Szul, M.J., Johnson, Z.I., and Zinser, E.R. (2008). Facilitation of Robust Growth of Prochlorococcus Colonies and Dilute Liquid Cultures by “Helper” Heterotrophic Bacteria. Appl. Environ. Microbiol. *74*, 4530–4534.

Morris, J.J., Johnson, Z.I., Szul, M.J., Keller, M., and Zinser, E.R. (2011). Dependence of the Cyanobacterium Prochlorococcus on Hydrogen Peroxide Scavenging Microbes for Growth at the Ocean’s Surface. PLOS ONE *6*, e16805.

Morris, J.J., Lenski, R.E., and Zinser, E.R. (2012). The Black Queen Hypothesis : Evolution of Dependencies through Adaptative Gene Loss. Mbio *3*, 1–7.

Morrissey, J., and Bowler, C. (2012). Iron utilization in marine cyanobacteria and eukaryotic algae. Front. Microbiol. *3*, 1–13.

Mühling, M., Fuller, N.J., Somerfield, P.J., Post, A.F., Wilson, W.H., Scanlan, D.J., Joint, I., and Mann, N.H. (2006). High resolution genetic diversity studies of marine Synechococcus isolates using rpoC1-based restriction fragment length polymorphism. Aquat. Microb. Ecol. *45*, 263–275.

Mullineaux, C.W. (2014). Electron transport and light-harvesting switches in cyanobacteria. Front. Plant Sci. *5*, 1–6.

Muro-Pastor, A.M., and Hess, W.R. (2020). Regulatory RNA at the crossroads of carbon and nitrogen metabolism in photosynthetic cyanobacteria. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1863*, 194477.

N

Niyogi, K.K., and Truong, T.B. (2013). Evolution of flexible non-photochemical quenching mechanisms that regulate light harvesting in oxygenic photosynthesis. Curr. Opin. Plant Biol. *16*, 307–314.

Nodop, A., Pietsch, D., Höcker, R., Becker, A., Pistorius, E.K., Forchhammer, K., and Michel, K.-P. (2008). Transcript Profiling Reveals New Insights into the Acclimation of the Mesophilic Fresh-Water Cyanobacterium Synechococcus elongatus PCC 7942 to Iron Starvation. Plant Physiol. *147*, 747–763.

O

Ohnishi, N., Allakhverdiev, S.I., Takahashi, S., Higashi, S., Watanabe, M., Nishiyama, Y., and Murata, N. (2005). Two-Step Mechanism of Photodamage to Photosystem II: Step 1 Occurs at the Oxygen-Evolving Complex and Step 2 Occurs at the Photochemical Reaction Center. Biochemistry *44*, 8494–8499.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., and Wagner, H. (2015). vegan: Community Ecology Package.

Olson, R., and Chisholm, S. (1990). Pigments, size, and distribution of Synechococcus in the North Atlantic and Pacific Oceans. Limnol. … *35*.

Olson, R.J., Chisholm, S.W., Zettler, E.R., Altabet, M.A., and Dusenberry, J.A. (1990). Spatial and temporal distributions of prochlorophyte picoplankton in the North Atlantic Ocean. Deep Sea Res. Part Oceanogr. Res. Pap. *37*, 1033–1051.

Ostrowski, M., Mazard, S., Tetu, S.G., Phillippy, K., Johnson, A., Palenik, B., Paulsen, I.T., and Scanlan, D.J. (2010). PtrA is required for coordinate regulation of gene expression during phosphate stress in a marine Synechococcus. ISME J. *4*, 908–921.

P

Pablo, F., Stauber, J.L., and Buckney, R.T. (1997). Toxicity of cyanide and cyanide complexes to the marine diatom Nitzschia closterium. Water Res. *31*, 2435–2442.

Palenik, B. (1994). Cyanobacterial community structure as seen from RNA polymerase gene sequence analysis. Appl. Environ. Microbiol. *60*, 3212–3219.

Palenik, B. (2001). Chromatic Adaptation in Marine Synechococcus Strains. Appl. Environ. Microbiol. *67*, 991–994.

Palenik, B., Brahamsha, B., Larimer, F.W., Land, M., Hauser, L., Chain, P., Lamerdin, J., Regala, W., Allen, E.E., McCarren, J., et al. (2003). The genome of a motile marine Synechococcus. Nature *424*, 1037–1042.

Palenik, B., Ren, Q., Dupont, C.L., Myers, G.S., Heidelberg, J.F., Badger, J.H., Madupu, R., Nelson, W.C., Brinkac, L.M., Dodson, R.J., et al. (2006). Genome sequence of Synechococcus CC9311: Insights into adaptation to a coastal environment. Proc. Natl. Acad. Sci. *103*, 13555–13559.

Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B.J., Evans, P.N., Hugenholtz, P., and Tyson, G.W. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat. Microbiol. *2*, 1533–1542.

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., and Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. *36*, 996–1004.

Partensky, F., and Garczarek, L. (2010). Prochlorococcus: Advantages and Limits of Minimalism. Annu. Rev. Mar. Sci. *2*, 305–331.

Partensky, F., Hoepffner, N., Li, WKW., Ulloa, O., and Vaulot, D. (1993). Photoacclimation of Prochlorococcus sp. (Prochlorophyta) Strains Isolated from the North Atlantic and the Mediterranean Sea. Plant Physiol. *101*, 285–296.

Partensky, F., Blanchot, J., and Vaulot, D. (1999a). Differential distribution and ecology of Prochlorococcus and Synechococcus in oceanic waters: a review. Bull.-Inst. Oceanogr. Monaco-Numero Spec. 457–476.

Partensky, F., Hess, W.R., and Vaulot, D. (1999b). Prochlorococcus, a marine photosynthetic prokaryote of global significance. MicrobiolMol BiolRev *63*, 106–127.

Paz-Yepes, J., Brahamsha, B., and 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, P.B., Guisan, A., Broennimann, O., and Randin, C.F. (2008). Niche dynamics in space and time. Trends Ecol. Evol. *23*, 149–158.

Penno, S., Lindell, D., and Post, A.F. (2006). Diversity of Synechococcus and Prochlorococcus populations determined from DNA sequences of the N-regulatory gene ntcA. Environ. Microbiol. *8*, 1200–1211.

Pernil, R., Picossi, S., Mariscal, V., Herrero, A., and Flores, E. (2008). ABC-type amino acid uptake transporters Bgt and N-II of Anabaena sp. strain PCC 7120 share an ATPase subunit and are expressed in vegetative cells and heterocysts. Mol. Microbiol. *67*, 1067–1080.

Pessarakli, M. (2016). Handbook of photosynthesis (CRC press).

Pitt, F.D., Mazard, S., Humphreys, L., and Scanlan, D.J. (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., Humily, F., Thorel, M., Grulois, D., Garczarek, L., and Six, C. (2014). Connecting thermal physiology and latitudinal niche partitioning in marine Synechococcus. ISME J. 1–16.

Pittera, J., Partensky, F., and Six, C. (2017). Adaptive thermostability of light-harvesting complexes in marine picocyanobacteria. ISME J. *11*, 112–124.

Pittera, J., Jouhet, J., Breton, S., Garczarek, L., Partensky, F., Maréchal, É., Nguyen, N.A., Doré, H., Ratin, M., Pitt, F.D., et al. (2018). Thermoacclimation and genome adaptation of the membrane lipidome in marine *Synechococcus*: Membrane

thermoadaptation in marine *Synechococcus*. Environ. Microbiol. *20*, 612–631.

Podar, M., Eads, J.R., and Richardson, T.H. (2005). Evolution of a microbial nitrilase gene family: a comparative and environmental genomics study. BMC Evol. Biol. *5*, 42.

Polz, M.F., Alm, E.J., and Hanage, W.P. (2013). Horizontal gene transfer and the evolution of bacterial and archaeal population structure. Trends Genet. *29*, 170–175.

Pospíšil, P. (2009). Production of reactive oxygen species by photosystem II. Biochim. Biophys. Acta BBA - Bioenerg. *1787*, 1151–1160.

Powell, L., Bowman, J., Skerratt, J., Franzmann, P., and Burton, H. (2005). Ecology of a novel Synechococcus clade occurring in dense populations in saline Antarctic lakes. Mar. Ecol. Prog. Ser. *291*, 65–80.

R

Raven, J.A., Evans, M.C.W., and Korb, R.E. (1999). The role of trace metals in photosynthetic electron transport in O2-evolving organisms. Photosynth. Res. *60*, 111–149.

Rocap, G., Distel, D.L., Waterbury, J.B., and Chisholm, S.W. (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, F.W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N.A., Arellano, A., Coleman, M., Hauser, L., Hess, W.R., et al. (2003). Genome divergence in two Prochlorococcus ecotypes reflects oceanic niche differentiation. Nature *424*, 1042–1047.

Roose, J.L., Wegener, K.M., and Pakrasi, H.B. (2007). The extrinsic proteins of Photosystem II. Photosynth. Res. *92*, 369–387.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., Wu, D., Eisen, J.A., Hoffman, J.M., Remington, K., et al. (2007). The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. PLOS Biol. *5*, e77.

Rusch, D.B., Martiny, A.C., Dupont, C.L., Halpern, A.L., and Venter, J.C. (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., 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.

Saito, M.A., Moffett, J.W., and DiTullio, G.R. (2004). Cobalt and nickel in the Peru upwelling region: A major flux of labile cobalt utilized as a micronutrient. Glob. Biogeochem. Cycles *18*.

Saito, M.A., McIlvin, M.R., Moran, D.M., Goepfert, T.J., DiTullio, G.R., Post, A.F., and Lamborg, C.H. (2014). Multiple nutrient stresses at intersecting Pacific Ocean biomes detected by protein biomarkers. Science *345*, 1173–7.

Sánchez-Riego, A.M., Mata-Cabana, A., Galmozzi, C.V., and Florencio, F.J. (2016). NADPH-Thioredoxin Reductase C Mediates the Response to Oxidative Stress and Thermotolerance in the Cyanobacterium Anabaena sp. PCC7120. Front. Microbiol. *7*.

Sander, J., Nowaczyk, M., Buchta, J., Dau, H., Vass, I., Deák, Z., Dorogi, M., Iwai, M., and Rögner, M. (2010). Functional Characterization and Quantification of the

Alternative PsbA Copies in Thermosynechococcus elongatus and Their Role in Photoprotection. J. Biol. Chem. *285*, 29851–29856.

Sass, L., Spetea, C., Máté, Z., Nagy, F., and Vass, I. (1997). Repair of UV-B induced damage of Photosystem II via de novo synthesis of the D1 and D2 reaction centre subunits in Synechocystis sp. PCC 6803. Photosynth. Res. *54*, 55–62.

Satoh, S., and Tanaka, A. (2006). Identification of Chlorophyllide a Oxygenase in the Prochlorococcus Genome by a Comparative Genomic Approach. Plant Cell Physiol. *47*, 1622–1629.

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

Scanlan, D.J., Ostrowski, M., Mazard, S., Dufresne, a, Garczarek, L., Hess, W.R., Post, a F., Hagemann, M., Paulsen, I., and Partensky, F. (2009). Ecological genomics of marine picocyanobacteria. Microbiol. Mol. Biol. Rev. MMBR *73*, 249–299.

Schlebusch, M., and Forchhammer, K. (2010). Requirement of the Nitrogen Starvation-Induced Protein Sll0783 for Polyhydroxybutyrate Accumulation in Synechocystis sp. Strain PCC 6803. Appl. Environ. Microbiol. *76*, 6101–6107.

Schubert, H., and Forster, R.M. (1997). Sources of variability in the factors used for modelling primary productivity in eutrophic waters. Hydrobiologia *349*, 75–85.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics *30*, 2068–2069.

Shapiro, E., Biezuner, T., and Linnarsson, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. Nat. Rev. Genet. *14*, 618–630.

Shcolnick, S., and Keren, N. (2006). Metal Homeostasis in Cyanobacteria and Chloroplasts. Balancing Benefits and Risks to the Photosynthetic Apparatus. Plant Physiol. *141*, 805–810.

Sherr, E.B., and Sherr, B.F. (2002). Significance of predation by protists in aquatic microbial food webs. Antonie Van Leeuwenhoek *81*, 293–308.

Shih, P.M., Wu, D., Latifi, A., Axen, S.D., Fewer, D.P., Talla, E., Calteau, A., Cai, F., Tandeau de Marsac, N., Rippka, R., et al. (2013). Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. Proc Natl Acad Sci U A *110*, 1053–1058.

Shilova, I.N., Mills, M.M., Robidart, J.C., Turk-Kubo, K.A., Björkman, K.M., Kolber, Z., Rapp, I., van Dijken, G.L., Church, M.J., Arrigo, K.R., et al. (2017). Differential effects of nitrate, ammonium, and urea as N sources for microbial communities in the North Pacific Ocean: N effects on microbial communities. Limnol. Oceanogr. *62*, 2550–2574.

Sidler, W.A. (1994). Phycobilisome and Phycobiliprotein Structures. In The Molecular Biology of Cyanobacteria, D.A. Bryant, ed. (Dordrecht: Springer Netherlands), pp. 139–216.

Sieburth, J.McN., Smetacek, V., and Lenz, J. (1978). Pelagic ecosystem structure: Heterotrophic compartments of the plankton and their relationship to plankton size fractions. Limnol. Oceanogr. *23*, 1256–1263.

Silver, R.P., Prior, K., Nsahlai, C., and Wright, L.F. (2001). ABC transporters and the export of capsular polysaccharides from Gram-negative bacteria. Res. Microbiol. *152*, 357–364.

Six, C., Thomas, J., Brahamsha, B., Lemoine, Y., and Partensky, F. (2004). Photophysiology of the marine cyanobacterium Synechococcus sp.

WH8102, a new model organism. Aquat. Microb. Ecol. *35*, 17–29.

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

Six, C., Finkel, Z.V., Irwin, A.J., and Campbell, D.A. (2007b). Light variability illuminates niche-partitioning among marine picocyanobacteria. PLoS ONE *2*.

Six, C., Joubin, L., Partensky, F., Holtzendorff, J., and Garczarek, L. (2007c). UV-induced phycobilisome dismantling in the marine picocyanobacterium Synechococcus sp. WH8102. Photosynth. Res. *92*, 75–86.

Sohm, J. a, Ahlgren, N. a, Thomson, Z.J., Williams, C., Moffett, J.W., Saito, M. a, Webb, E. a, and Rocap, G. (2015). Co-occurring Synechococcus ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. ISME J. *10*, 1–13.

Sohm, J.A., Ahlgren, N.A., Thomson, Z.J., Williams, C., Moffett, J.W., Saito, M.A., Webb, E.A., and Rocap, G. (2016). Co-occurring Synechococcus ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. ISME J. *10*, 333–345.

Song, Q., Gordon, A. L., & Visbeck, M. (2004). Spreading of the Indonesian throughflow in the Indian Ocean. Journal of Physical Oceanography, 34(4), 772-792.

Sonoike, K. (2011). Photoinhibition of photosystem I. Physiol. Plant. *142*, 56–64.

Steglich, C., Behrenfeld, M., Koblizek, M., Claustre, H., Penno, S., Prasil, O., Partensky, F., and Hess, W.R. (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.E., Lindell, D., Voss, B., Chisholm, S.W., and Hess, W.R. (2008). The challenge of regulation in a minimal photoautotroph: Non-coding RNAs in Prochlorococcus. PLoS Genet. *4*, e1000173.

Stickforth, P., Steiger, S., Hess, W.R., and Sandmann, G. (2003). A novel type of lycopene ε-cyclase in the marine cyanobacterium Prochlorococcus marinus MED4. Arch. Microbiol. *179*, 409–415.

Stuart, R.K., Dupont, C.L., Johnson, D.A., Paulsen, I.T., and 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., Mao, F., Dam, P., Wu, H., Olman, V., Paulsen, I.T., Palenik, B., and Xu, Y. (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., and Xu, Y. (2007). Computational prediction of Pho regulons in cyanobacteria. BMC Genomics *8*, 156.

Sugiura, M., Azami, C., Koyama, K., Rutherford, A.W., Rappaport, F., and Boussac, A. (2014). Modification of the pheophytin redox potential in Thermosynechococcus elongatus Photosystem II with PsbA3 as D1. Biochim. Biophys. Acta BBA - Bioenerg. *1837*, 139–148.

Sunda, W.G., and Huntsman, S.A. (1995). Cobalt and zinc interreplacement in marine phytoplankton: Biological and geochemical implications. Limnol. Oceanogr. *40*, 1404–1417.

Suttle, C.A. (2007). Marine viruses — major players in the global ecosystem. Nat. Rev. Microbiol. *5*, 801–812.

T

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

Takahashi, S., and Badger, M.R. (2011). Photoprotection in plants: a new light on photosystem II damage. Trends Plant Sci. *16*, 53–60.

Tetu, S.G., Brahamsha, B., Johnson, D.A., Tai, V., Phillippy, K., Palenik, B., and Paulsen, I.T. (2009). Microarray analysis of phosphate regulation in the marine cyanobacterium *Synechococcus* sp. WH8102. ISME J. *3*, 835–849.

Thompson, A.W., Huang, K., Saito, M.A., and Chisholm, S.W. (2011). Transcriptome response of high- and low-light-adapted Prochlorococcus strains to changing iron availability. ISME J. *5*, 1580–1594.

Thrash, J.C., Temperton, B., Swan, B.K., Landry, Z.C., Woyke, T., DeLong, E.F., Stepanauskas, R., and Giovannoni, S.J. (2014). Single-cell enabled comparative genomics of a deep ocean SAR11 bathytype. ISME J. *8*, 1440–51.

Tian, L., van Stokkum, I.H.M., Koehorst, R.B.M., Jongerius, A., Kirilovsky, D., and van Amerongen, H. (2011). Site, Rate, and Mechanism of Photoprotective Quenching in Cyanobacteria. J. Am. Chem. Soc. *133*, 18304–18311.

Toledo, G., and 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, A.C., Aach, J., Lindell, D., Johnson, Z.I., Rector, T., Steen, R., Church, G.M., and Chisholm, S.W. (2006). Global gene expression of Prochlorococcus ecotypes in response to changes in nitrogen availability. Mol. Syst. Biol. *2*, 1–15.

Triantaphylidès, C., Krischke, M., Hoeberichts, F.A., Ksas, B., Gresser, G., Havaux, M., Breusegem, F.V., and Mueller, M.J. (2008). Singlet Oxygen Is the Major Reactive Oxygen Species Involved in Photooxidative Damage to Plants. Plant Physiol. *148*, 960–968.

Tully, B.J., Sachdeva, R., Graham, E.D., and Heidelberg, J.F. (2017). 290 metagenome-assembled genomes from the Mediterranean Sea: a resource for marine microbiology. PeerJ *5*, e3558.

Tully, B.J., Graham, E.D., and Heidelberg, J.F. (2018). The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. Sci. Data *5*, 170203.

V

Valladares, A., Montesinos, M.L., Herrero, A., and Flores, E. (2002). An ABC-type, high-affinity urea permease identified in cyanobacteria. Mol. Microbiol. *43*, 703–715.

Van Mooy, B.A.S., Rocap, G., Fredricks, H.F., Evans, C.T., and Devol, A.H. (2006). Sulfolipids dramatically decrease phosphorus demand by picocyanobacteria in oligotrophic marine environments. Proc. Natl. Acad. Sci. U. S. A. *103*, 8607–8612.

Van Mooy, B.A.S., Fredricks, H.F., Pedler, B.E., Dyhrman, S.T., Karl, D.M., Koblizek, M., Lomas, M.W., Mincer, T.J., Moore, L.R., Moutin, T., et al. (2009). Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. Nature *458*, 69–72.

Varkey, D., Mazard, S., Ostrowski, M., Tetu, S.G., Haynes, P., and Paulsen, I.T. (2016). Effects of low temperature on tropical and

temperate isolates of marine Synechococcus. ISME J. *10*, 1252–1263.

Vass, I. (2012). Molecular mechanisms of photodamage in the Photosystem II complex. Biochim. Biophys. Acta BBA - Bioenerg. *1817*, 209–217.

Volk, T., and Hoffert, M.I. (1985). Ocean carbon pumps: Analysis of relative strengths and efficiencies in ocean-driven atmospheric CO2 changes. Carbon Cycle Atmospheric CO2 Nat. Var. Archean Present *32*, 99–110.

W

Waterbury, J., and Rippka, R. (1989). Cyanobacteria. subsection i. order chroococcales. Bergey's Man. Syst. Bacteriol. Krieg NR Holt JB Eds Baltim. MD USA Williams Wilkins *3*, 1728–1746.

Waterbury, J.B., Watson, S.W., Guillard, R.R.L., and Brand, L.E. (1979). Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. Nature *277*, 293–294.

West, N.J., and Scanlan, D.J. (1999). Niche-partitioning of Prochlorococcus populations in a stratified water column in the eastern North Atlantic Ocean. Appl. Environ. Microbiol. *65*, 2585–2591.

West, N.J., Lebaron, P., Strutton, P.G., and Suzuki, M.T. (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.

Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998). Prokaryotes: The unseen majority. Proc. Natl. Acad. Sci. *95*, 6578–6583.

Willey, J.M., and Waterbury, J.B. (1989). Chemotaxis toward Nitrogenous Compounds by Swimming Strains of Marine Synechococcus spp. Appl. Environ. Microbiol. *55*, 1888–1894.

X

Xia, X., Guo, W., Tan, S., and Liu, H. (2017). Synechococcus Assemblages across the Salinity Gradient in a Salt Wedge Estuary. Front. Microbiol. *8*, 1254.

Xia, X., Liu, H., Choi, D., and Noh, J.H. (2018). Variation of Synechococcus Pigment Genetic Diversity Along Two Turbidity Gradients in the China Seas. Microb. Ecol. *75*, 10–21.

Xu, H., Vavilin, D., Funk, C., and Vermaas, W. (2004). Multiple Deletions of Small Cab-like Proteins in the Cyanobacterium Synechocystis sp. PCC 6803 CONSEQUENCES FOR PIGMENT BIOSYNTHESIS AND ACCUMULATION. J. Biol. Chem. *279*, 27971–27979.

Y

Yang, S., Zhang, R., Hu, C., Xie, J., and Zhao, J. (2009). The dynamic behavior of phycobilisome movement during light state transitions in cyanobacterium Synechocystis PCC6803. Photosynth. Res. *99*, 99–106.

Yeremenko, N., Kouřil, R., Ihalainen, J.A., D'Haene, S., van Oosterwijk, N., Andrizhiyevskaya, E.G., Keegstra, W., Dekker, H.L., Hagemann, M., Boekema, E.J., et al. (2004). Supramolecular Organization and Dual Function of the IsiA Chlorophyll-Binding Protein in Cyanobacteria. Biochemistry *43*, 10308–10313.

Z

Zhang, P., Allahverdiyeva, Y., Eisenhut, M., and Aro, E.-M. (2009). Flavodiiron proteins in oxygenic photosynthetic organisms: photoprotection of photosystem II by Flv2 and Flv4 in Synechocystis sp. PCC 6803. PloS One *4*, e5331.

Zhang, P., Eisenhut, M., Brandt, A.-M., Carmel, D., Silén, H.M., Vass, I., Allahverdiyeva, Y., Salminen, T.A., and Aro, E.-M. (2012). Operon flv4-flv2 provides cyanobacterial photosystem II with flexibility of electron transfer. Plant Cell *24*, 1952–1971.

Zinser, E.R., Coe, A., Johnson, Z.I., Martiny, A.C., Fuller, N.J., Scanlan, D.J., and Chisholm, S.W. (2006). Prochlorococcus ecotype abundances in the North Atlantic Ocean as revealed by an improved quantitative PCR method. Appl. Environ. Microbiol. *72*, 723–732.

Zinser, E.R., Johnson, Z.I., Coe, A., Karaca, E., Veneziano, D., and Chisholm, S.W. (2007). Influence of light and temperature on Prochlorococcus ecotype distributions in the Atlantic Ocean. Limnol. Oceanogr. *52*, 2205–2220.

Zwirglmaier, K., Heywood, J.L., Chamberlain, K., Woodward, E.M.S., Zubkov, M.V., and Scanlan, D.J. (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., Not, F., Massana, R., Ulloa, O., and Scanlan, D.J. (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 : Unveiling membrane thermoregulation strategies in marinepicocyanobacteria.

Breton, S., Jouhet, J., Guyet, U., Gros, V., Pittera, J., Demory, D., Partensky, F., Doré, H., Ratin, M., Maréchal, E., et al. (2019). Unveiling membrane thermoregulation strategies in marine picocyanobacteria. New Phytol. nph.16239.

New Phytologist

Research

Unveiling membrane thermoregulation strategies in marine picocyanobacteria

Solène Breton[1], Juliette Jouhet[2], Ulysse Guyet[1], Valérie Gros[2], Justine Pittera[1], David Demory[3], Frédéric Partensky[1], Hugo Doré[1], Morgane Ratin[1], Eric Maréchal[2], Ngoc An Nguyen[1], Laurence Garczarek[1] and Christophe Six[1]

[1]Sorbonne Université, Centre National de la Recherche Scientifique, UMR 7144 Adaptation et Diversité en Milieu Marin (AD2M), Ecology of Marine Plankton (ECOMAP) Team, Station Biologique de Roscoff (SBR), 29680, Roscoff, France; [2]Laboratoire de Physiologie Cellulaire et Végétale, Unité mixe de recherche 5168 CNRS, CEA, INRA, Université Grenoble Alpes, IRIG, CEA Grenoble, 17, rue des Martyrs, 38000, Grenoble, France; [3]School of Biology, Georgia Institute of Technology, Atlanta, GA 30332, USA

Author for correspondence:
Christophe Six
Tel: +33 2 98292534
Email: christophe.six@sb-roscoff.fr

Received: *8 July 2019*
Accepted: *29 September 2019*

New Phytologist (2019)
doi: 10.1111/nph.16239

Key words: adaptation, cyanobacteria, marine *Synechococcus*, membrane lipids, temperature ecotype.

Summary

- The wide latitudinal distribution of marine *Synechococcus* cyanobacteria partly relies on the differentiation of lineages adapted to distinct thermal environments. Membranes are highly thermosensitive cell components, and the ability to modulate their fluidity can be critical for the fitness of an ecotype in a particular thermal niche.
- We compared the thermophysiology of *Synechococcus* strains representative of major temperature ecotypes in the field. We measured growth, photosynthetic capacities and membrane lipidome variations. We carried out a metagenomic analysis of stations of the *Tara* Oceans expedition to describe the latitudinal distribution of the lipid desaturase genes in the oceans.
- All strains maintained efficient photosynthetic capacities over their different temperature growth ranges. Subpolar and cold temperate strains showed enhanced capacities for lipid monodesaturation at low temperature thanks to an additional, poorly regiospecific Δ9-desaturase. By contrast, tropical and warm temperate strains displayed moderate monodesaturation capacities but high proportions of double unsaturations in response to cold, thanks to regiospecific Δ12-desaturases. The desaturase genes displayed specific distributions directly related to latitudinal variations in ocean surface temperature.
- This study highlights the critical importance of membrane fluidity modulation by desaturases in the adaptive strategies of *Synechococcus* cyanobacteria during the colonization of novel thermal niches.

Introduction

Marine *Synechococcus* are among the smallest photoautotrophs on Earth but are thought to be responsible for about 17% of the global marine net primary production (Flombaum *et al.*, 2013). They constitute a much diversified cyanobacterial radiation, most of them being gathered in the phylogenetic subcluster 5.1 *sensu* Herdman *et al.* (2001) that, according to analyses based on the *petB* gene marker, includes 15 clades and 28 subclades (Mazard *et al.*, 2012). A number of phylogeographic studies have shown that five clades (I, II, III, IV and CRD1) predominate in the ocean, occupying different environmental niches. Clades I and IV co-occur in nutrient-rich, temperate and cold regions, whereas clades II and III are usually found in warmer waters, with clades II and III dominating in N- and P-depleted areas, respectively (Zwirglmaier *et al.*, 2008; Sohm *et al.*, 2015; Farrant *et al.*, 2016; Paulsen *et al.*, 2016; Kent *et al.*, 2018). More recently discovered, the fifth major clade, CRD1, appears to be specialized to iron-limited environments, such as the so-called High Chlorophyll Low Nutrient areas of the South Pacific Ocean (Sohm *et al.*, 2015; Farrant *et al.*, 2016; Kent *et al.*, 2018).

The wide latitudinal distribution of marine *Synechococcus*, spanning from the equator to polar circles (Zwirglmaier *et al.*, 2008; Huang *et al.*, 2012), suggests that the different lineages have developed efficient strategies to colonize distinct thermal environments. Several studies have shown that representative strains of specific clades exhibit thermal *preferenda* that are in good agreement with the temperature of their isolation sites (Moore *et al.*, 1995; Mackey *et al.*, 2013; Pittera *et al.*, 2014; Varkey *et al.*, 2016). These genetically distinct lineages, physiologically adapted to distinct niches, thus constitute 'temperature ecotypes' (i.e. 'thermotypes'), a concept previously defined for the other dominant marine cyanobacterium, *Prochlorococcus* (Johnson *et al.*, 2006; Coleman & Chisholm, 2007). The physiological bases that underpin the process of thermal niche partitioning of marine picocyanobacteria remain, however, poorly known.

© 2019 The Authors
New Phytologist © 2019 New Phytologist Trust

Several studies suggest that the ability to adapt the photosynthetic mechanisms to different temperatures could constitute a bottleneck for competitiveness in different thermal environments (Pittera *et al.*, 2014, 2016). In particular, it has been shown that the photosystem II of different *Synechococcus* temperature ecotypes is differentially affected by thermal stress (Pittera *et al.*, 2014), and that their photosynthetic antenna, called the phycobilisome, display differential functional stability consistent with the temperature *preferenda* of the strains (Pittera *et al.*, 2016). Another important component that may directly impact the smooth running of the photosynthetic apparatus at different temperatures is the lipid matrix of thylakoid membranes. Indeed, it is well known that membranes are among the most thermosensitive cell components, because temperature drastically affects their fluidity and permeability and therefore the activity of all the membrane-embedded proteins (Mikami & Murata, 2003). The ability to modulate the fluidity of membranes, notably the thylakoids, can thus be critical for the fitness of photosynthetic organisms in specific thermal niches.

Cyanobacterial membranes are composed of three main glycolipids, mono- and digalactosyldiacylglycerol (MGDG and DGDG) and a charged sulfolipid, sulfoquinivosyldiacylglycerol (SQDG). In addition, smaller amounts of a phospholipid, phosphatidyldiacylglycerol (PG), are usually observed (Murata & Wada, 1995). Whereas a number of studies have dealt with the membrane thermoregulation of freshwater cyanobacteria, the composition of the membranes and the impact of temperature in their marine counterparts remain poorly studied (Merritt *et al.*, 1991; Van Mooy *et al.*, 2009; Varkey *et al.*, 2016). A recent study has shown that the model marine strain *Synechococcus* sp. WH7803 contains membrane lipids with shorter fatty acid chains than freshwater cyanobacteria and is able to adjust its membrane fluidity to different growth temperatures (Pittera *et al.*, 2018). Indeed, in response to cold, the WH7803 strain induces shortening of the fatty acid chains at the *sn*-1 position of the galactolipids, leading to membrane thinning, and activates fatty acid desaturation at specific sites of the three glycolipids (Pittera *et al.*, 2018).

Unsaturations can be dynamically inserted in fatty acid chains by lipid desaturase enzymes. The catalytic site of these enzymes comprises histidine-rich boxes, including a nonheme iron center whose activity requires electrons and oxygen (Los & Murata, 1998). Genomic studies of the fatty acid desaturase gene content in marine *Synechococcus* have shown the presence of four main distinct genes in the genomes, encoding two Δ9- (*desC3* and *desC4*) and two Δ12-desaturases (*desA2* and *desA3*; Chi *et al.*, 2008; Varkey *et al.*, 2016). A closer look at the phyletic profiles of these four genes showed that the marine *Synechococcus* strains representative of the four dominant clades (I–IV) exhibit distinct desaturase gene contents, suggesting that the desaturation capacities could be part of the physiological strategies underlying the specialization of these lineages to different thermal environments (Pittera *et al.*, 2018). However, it is not yet known how these differential gene contents affect the composition and thermoacclimation processes of the membrane lipidome in the different temperature ecotypes.

In this study, we analyzed the composition of the membrane lipidome of a selection of four strains representative of each of the four dominant marine *Synechococcus* clades (I–IV) in natural communities and reveal marked differences in the respective regulation processes adjusting membrane fluidity in response to temperature. Furthermore, we show that the global oceanic distribution of the desaturase genes is fully consistent with their differential distribution among the ecotypes, demonstrating the significance of these thermoadaptation strategies in thermal niche partitioning of *Synechococcus in situ*. We discuss these findings in the context of the important role of marine *Synechococcus* in ocean primary productivity.

Materials and Methods

Growth conditions and experimental design

Four marine *Synechococcus* strains retrieved from the Roscoff Culture Collection (http://roscoff-culture-collection.org/) were used in this study (Table 1): MVIR-18-1 (clade I), A15-62 (clade II), WH8102 (clade III) and BL107 (clade IV), of which only WH8102 was axenic. Strains were grown in PCR-S11 culture medium (Rippka *et al.*, 2000; Moore *et al.*, 2007) supplemented with 1 mM sodium nitrate under low irradiance

Table 1 Characteristics of the four *Synechococcus* strains used in this study.

Strain name	RCC number	Clade	Isolation site	Coordinates	Average seawater temperature (°C)	T_{MIN} (°C)	T_{MAX} (°C)	Lipid-desaturase gene content: Δ9-desaturase *desC3*	Δ9-desaturase *desC4*	Δ12-desaturase *desA2*	Δ12-desaturase *desA3*
MVIR-18-1	2385	I	North Sea	+61°0′, +1°59′	10.4 ± 0.3	1.0	28.8	■	■		■
BL107	515	IV	Mediterranean Sea	+43°43′, +3°33′	17.9 ± 0.3	10.37	28.4	■	■		■
A15-62	2374	II	Offshore Mauritania	+17°37′, −20°27′	23.7 ± 0.4	9.0	30.8	■		■	
WH8102	539	III	Caribbean Sea	+22°29′, −65°3′	27.1 ± 0.1	13.9	32.2	■		■	■

Taxonomic assignments at the clade level and desaturase gene content are derived from Farrant *et al.* (2016) and Pittera *et al.* (2018), respectively. The average seawater temperature is the annual mean temperature at the isolation site of the strain over the period 2000–2010, as derived from the NASA Sea Surface Temperature data (https://oceandata.sci.gsfc.nasa.g). The modeled values of thermal limits of growth were calculated using the phytoplankton growth model defined by Boatman *et al.* (2017). The presence of a desaturase gene in a strain is indicated by shading.

© 2019 The Authors
New Phytologist © 2019 New Phytologist Trust

(20 µmol photons m^{-2} s^{-1}), in order to ensure the presence of several thylakoidal lamellae in the cells (Kana & Glibert, 1987). Continuous light was provided by multicolor LED systems (Alpheus, France). The axenicity of *Synechococcus* sp. WH8102 cultures and minimal contamination levels of the other strains (< 20%) was regularly checked by flow cytometry using SYBR-Green staining (Marie *et al.*, 2001).

Cultures of the four cyanobacteria were acclimated for several weeks to a range of temperatures (10–30°C) in temperature-controlled chambers (Lovibond, Dortmund, Germany), monitored by flow cytometry and sampled during the exponential growth phase to measure photosynthetic parameters and analyze the membrane lipid structure in the acclimated state. To study the dynamics of the temperature-induced remodeling of the membranes, we also carried out temperature shift experiments. Three liters of exponentially growing cultures (at 1–3 × 10^7 cells ml^{-1}) maintained at 22°C was split into subcultures and transferred to either 13°C or 30°C, under identical light conditions. One subculture per condition was harvested every day over 4 d. All experiments were repeated three to four times.

Flow cytometry, *in vivo* fluorimetry and pigment analyses

Aliquots of cultures were preserved using 0.25% (v/v) glutaraldehyde (final concentration; grade II, Sigma Aldrich) and stored at −80°C until analysis. Cell concentrations were determined using a flow cytometer (FACS Canto II, Becton Dickinson, San Jose, CA, USA) as described previously (Pittera *et al.*, 2014). Growth rates were calculated on cultures pre-acclimated to a range of temperatures as the slope of an ln(Nt) vs time plot function, where Nt is the cell concentration at time t. The growth models of Bernard & Rémond (2012) and Boatman *et al.* (2017) were used to model the optimal temperature for growth (T_{OPT}) and the temperature limits for growth (T_{MIN} and T_{MAX}; for more details, see Notes S1).

The photosystem II (PSII) quantum yield (F_V/F_M) was measured using a Pulse Amplitude Modulation fluorometer (PhytoPAM I; Walz, Effeltrich, Germany) in the presence of 100 µM of the PSII blocker 3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU), following a previously described procedure (Pittera *et al.*, 2014). The quantum yield was calculated as:

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

where F_0 is the basal fluorescence level, F_M the maximal fluorescence level and F_V the variable fluorescence (Campbell *et al.*, 1998; Ogawa *et al.*, 2017).

For pigment analyses, 50–100 ml of culture was harvested by centrifugation over 10 min, at 4°C and 20 000 ***g*** (Eppendorf 5804R) in the presence of 0.01% (v/v) pluronic acid (Sigma Aldrich). The cell pellets were covered with 200 µl of cold methanol (HPLC grade; Sigma Aldrich), vortexed until complete resuspension and incubated for 30 min on ice. The pigment extracts were vortexed again and then centrifuged as described above (Eppendorf 5417R). The supernatants were transferred and supplemented with 10% (v/v) distilled water (final volume), in order to avoid peak distortion. Pigments were measured by HPLC using an HPLC 1100 Series System equipped with a photodiode array (Hewlett-Packard) as previously described (Pittera *et al.*, 2014).

Membrane lipidome analyses

The global detailed procedure has been previously published in Jouhet *et al.* (2017) and Pittera *et al.*, (2018). For each condition, 400 ml of culture was harvested by two successive centrifugation steps of 10 min, at 20 000 ***g*** and 4°C (centrifuges Beckman Avanti J-25 and Eppendorf 5804R) and stored at −80°C until analysis. Membrane lipids were extracted in glass hardware following a modified version of the Bligh & Dyer (1959) procedure (Pittera *et al.* 2018).

Fatty acid regiolocalization We first identified the positional distribution of the fatty acids esterified to the four main glycerolipids of the four *Synechococcus* strain samples. Each strain was grown at three different temperatures spanning their respective thermal growth range. The glycerolipids were then extracted and separated by two-dimensional thin layer chromatography (TLC) as previously described (Pittera *et al.*, 2018). Glycerolipids spots were revealed under UV light in the presence of malvidine-3-*O*-galactoside (primuline, 0.2% in pure methanol, Sigma Aldrich) and scraped off the plate. Each separated lipid class were recovered from the silica by extraction in glass hardware and dried under nitrogen gas. The position of the different fatty acid molecular species esterified to the glycerol backbone of the purified glycerolipids was determined by MS/MS analyses as described by Jouhet *et al.* (2017).

Lipid quantification The quantity of fatty acids in the samples was first determined by GC coupled with a flame ionization detector (GC-FID). The lipid samples were first methanolyzed and analyzed following a previously described procedure (Jouhet *et al.*, 2017). The lipid samples were then prepared in order to inject 25 nmol of total fatty acids dissolved in 100 µl of chloroform : methanol (2 : 1, v/v) containing 125 pmol of each internal standard. The membrane lipids were separated by HPLC (Agilent 1200) equipped with a diol column (150 mm × 3 mm × 5 µm; Macherey-Nagel, Düren, Germany) at 40°C and quantified by MS/MS with a triple quadrupole mass spectrometer equipped with a jet stream electrospray ion source (Agilent 6460) (Jouhet *et al.*, 2017; Pittera *et al.*, 2018). Mass spectra were processed with the Agilent MassHunter Workstation software for identification and quantification of lipids. Lipid amounts (pmol) were corrected for response differences between internal standards and endogenous lipids and by comparison with a qualified control (QC). The QC extract corresponded to a known lipid extract from *Arabidopsis* cell culture qualified and quantified by TLC and GC-FID as described by Jouhet *et al.* (2017).

© 2019 The Authors
New Phytologist © 2019 New Phytologist Trust

In situ metagenomic analyses

***Tara* Oceans dataset** We analyzed 33 metagenomes collected from surface waters during the *Tara* Oceans expedition (Sunagawa *et al.*, 2015; http://www.taraoceans-dataportal.org). Metagenomes corresponding to the bacterial size fraction (0.2–1.6 µm for TARA_004 to TARA_052 and 0.2–3 µm for the other stations) were sequenced as 100–108 bp Illumina overlapping paired-end reads. Reads were merged using FLASH v.1.2.7 (Magoč & Salzberg, 2011) and trimmed with CLC QUALITYTRIM v.4.10.86742 (CLC Bio, Aarhus, Denmark), resulting in 100–215 bp fragments. Temperature data for each station were retrieved from PANGAEA (https://doi.pangaea.de/10.1594/PANGAEA.840718).

Functional assignment of metagenomics data Reads were first recruited against a database of 155 protein sequences of the four lipid desaturases (DesC3, DesC4, DesA2, DesA3) extracted from the 53 *Synechococcus*/*Cyanobium* genomes of the information system CYANORAK v.2 (http://sb-roscoff.fr/Phyto/cyanorak; Pittera *et al.*, 2018; Dataset S1) using BLASTX analyses (v.2.7.1+; Altschul *et al.*, 1990) with default parameters, limiting the results to one target sequence (--max_target_seqs 1), and considering only the results with an e-value < 0.001. The recruited reads were then mapped with BLASTN (with the same options as above) against a database including the 155 above mentioned *Synechococcus*/*Cyanobium* desaturases as well as many sequences used as outgroups, including 117 *Prochlorococcus* fatty acid desaturase sequences (Dataset S2), 185 outgroup cyanobacterial genomes from Cyanobase plus 537 genomes from proGenomes (Dataset S3), 6092 additional nonpicocyanobacterial fatty acid desaturase sequences retrieved from NCBI (Dataset S4) as well as all marine picocyanobacterial desaturases from Cyanorak. Reads showing a best hit to outgroup sequences or with < 90% of their sequence aligned were removed and the remaining reads were assigned to *Synechococcus* according to their BLASTN best hit. Reads were then functionally assigned to one of the four desaturase clusters of likely orthologous genes (CLOGs) from the Cyanorak database based on their best hit sequence, and read counts were normalized to the average gene length of each CLOG. To study the presence/absence of lipid desaturases with regard to thermal environments, only stations dominated by major ecologically significant taxonomic units (ESTUs) of *Synechococcus* clades I–IV, namely ESTUs IA, IIA, IIIA or IVA-C *sensu* Farrant *et al.* (2016), were used. Furthermore, to ensure sufficient coverage for the results to be reliable, only stations with a minimum of 2500 distinct CLOGs, corresponding to the average number of genes in *Synechococcus* genomes, were considered. Altogether, 52% (33 out of 63) of the *Tara* Oceans stations were retained for analysis. Reads assigned to genes encoding the four lipid desaturases (*desC3*, *desC4*, *desA2*, *desA3*) were then extracted and their respective abundances were normalized to the corresponding *petB* reads abundance, retrieved from Farrant *et al.* (2016).

Results

Thermal *preferenda* and photosynthetic activity

We acclimated four *Synechococcus* strains isolated from different thermal environments to a global range of temperatures spanning from 10 to 32°C. The strains displayed thermal *preferenda* that differed significantly with regard to their thermal growth limits, especially at low temperatures, and less so by their optimal growth temperature (Fig. 1a). The subpolar clade I strain MVIR-18-1 showed significant growth down to 9–10°C but could not cope with temperatures above 26°C. The cold-temperate clade IV strain BL107 grew between 14 and 28°C, while the warm temperate clade III strain WH8102 and the tropical clade II strain A15-62 grew from 16°C to more than 32°C. Growth vs temperature models were used to estimate the growth parameters (Table 1; Fig. S1). Although some differences in these parameters could be observed depending on the model used (Tables S1, S2), the subpolar and cold-temperate strains displayed lower estimates of T_{MIN} and T_{MAX} than the warm-adapted strains (see Notes S1).

The quantum yield of the PSII reaction center (F_V/F_M) remained generally high over the whole growth temperature ranges (Fig. 1b). However, the values were somewhat lower at low growth temperatures (16°C) for strains WH8102 and A15-62, and at high growth temperatures for the subpolar MVIR-18-1 and cold temperate BL107 strains.

The β-carotene : Chl *a* ratio increased at high growth temperature and this trend was more pronounced for the high temperature strains WH8102 and A15-62 (Fig. 1c). For all strains, the zeaxanthin : Chl *a* mass ratio was high at low growth temperature and lower at high temperature. This ratio was globally higher in the subpolar strain MVIR-18-1 and in the temperate strain BL107 than in the two other strains (Fig. 1d). These variations probably mostly originated from an increase in the zeaxanthin cell content with decreasing temperature, as suggested by the cold-induced increase of the zeaxanthin : β-cryptoxanthin ratio, the direct precursor of zeaxanthin in its biosynthetic pathway (Fig. S2). It is also likely that the Chl *a* cell content decreased in response to cold.

Membrane lipidomic analyses

General composition of the membrane lipidome We analyzed the composition of the total membrane lipidome, which in cyanobacteria includes a very high proportion of thylakoidal lipids. The two-dimensional TLC analyses revealed four major spots corresponding to the lipid classes usually observed in cyanobacteria: MGDG, DGDG, SQDG and PG. In total, there were about 30 different molecular species of complex lipids in the four studied strains of marine *Synechococcus*. The two galactolipids (MGDG and DGDG) were generally the major components of membranes and the fatty acids bound to these lipids were mostly C14, C16 and traces of C18 fatty acid chains. The number of unsaturations on these chains never exceeded 2. The global composition of the fatty acids bound to the four different

© 2019 The Authors
New Phytologist © 2019 New Phytologist Trust

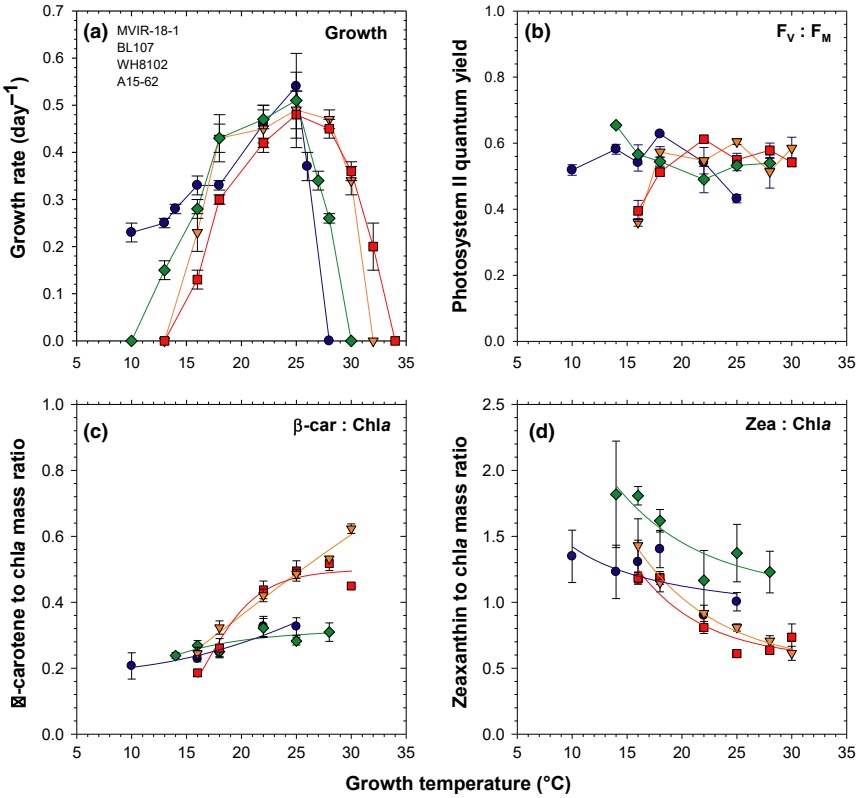


Fig. 1 Variations in growth rates (a), photosystem II quantum yield (F_V/F_M; b) and membrane pigment mass ratios (c, d) in *Synechococcus* sp. MVIR-18-1 (blue circles), BL107 (green diamonds), WH8102 (orange triangles) and A15-62 (red squares), acclimated to a range of temperatures spanning from 10 to 30°C. Zea, zeaxanthin; Chl*a*, chlorophyll *a*; β-car, β-carotene. Measurements are the average of four replicates with error bars ± SD.

lipids in these strains, reported in detail in Tables S3–S6 and Figs S3–S6 (see Notes S2), was overall similar to that of *Synechococcus* sp. WH7803 (Pittera *et al.*, 2018). The *sn*-1 position most often bound C16 chains, whereas the *sn*-2 bound exclusively C14 chains on the galactolipids (Figs S3, S4) and a mix of C14 and C16 chains on SQDG (Fig. S5).

PG is present in heterotrophic bacteria, complicating the interpretation of the PG data of the three nonaxenic strains. In cyanobacteria, this minor lipid is thought to play roles in the stabilization of photosystems (Itoh *et al.*, 2012; Mizusawa and Wada, 2012; Boudière *et al.*, 2014) rather than being directly involved in membrane thermoacclimation processes, as supported by the results for the axenic strains WH8102 and WH7803 (Pittera *et al.*, 2018; Fig. S6). Hereafter we therefore compare the major mechanisms involved in thermoacclimation only for the three major membrane glycolipids present in cyanobacteria.

Membrane thickness variations in response to temperature As palmitic (C16 : 0) and myristic (C14 : 0) fatty acid chains were dominant in all glycolipids of all strains, we used the C14 : C16 molar ratio as a proxy for assessing variations in membrane thickness. In the four strains, such variations may occur at the *sn*-1 position of the two most abundant lipids, MGDG and DGDG, as already observed in *Synechococcus* sp. WH7803 (Pittera *et al.*, 2018), and at the *sn*-2 position of SQDG. The results showed an increase of the C14 : 16 ratio at the *sn*-1 position of MGDG and DGDG in response to low growth temperatures only for the subpolar strain MVIR-18-1 and the tropical strain A15-62 (Fig. 2). This process was more pronounced for MGDG than for DGDG. By contrast, strains BL107 and WH8102 showed a slight increase in fatty acid length at these positions in response to low growth temperature. MVIR-18-1 was additionally capable of decreasing the length of the fatty acid bound to the *sn*-2 position of SQDG in response to growth temperature lower than 14°C (Fig. S5). In cultures grown at 22°C and shifted to either 13 or 30°C, only minor variations of the C14 : C16 ratio were observed (Figs S7–S14).

Mono-desaturation processes on galactolipids MGDG showed the highest levels of mono-unsaturation, with high proportions of 14 : 1 and 16 : 1 fatty acid chains. At the *sn*-1 position, the proportion of C16 : 1 chains increased in response to low growth temperature, especially in the subpolar strain MVIR-18-1 and the cold temperate strain BL107 (Fig. 3a,b). By contrast, the warm-adapted strains WH8102 and A15-62 showed much lower variations in the relative level of C16 : 1.

The proportion of C14 : 1 chains at the MGDG *sn*-1 position also increased at lower temperatures in all strains except for the warm temperate WH8102, for which the C14 species were low at this position. The largest C14 : 1 to C14 : 0 amplitude changes were achieved by the subpolar strain MVIR-18-1. This probably resulted from the replacement of C16 chains by C14 : 1 chains in *de novo* synthetized MGDG molecules, rather than active desaturation of C14 : 0 chains by desaturase enzymes, inducing the above-mentioned chain shortening. This conclusion is supported

© 2019 The Authors
New Phytologist © 2019 New Phytologist Trust

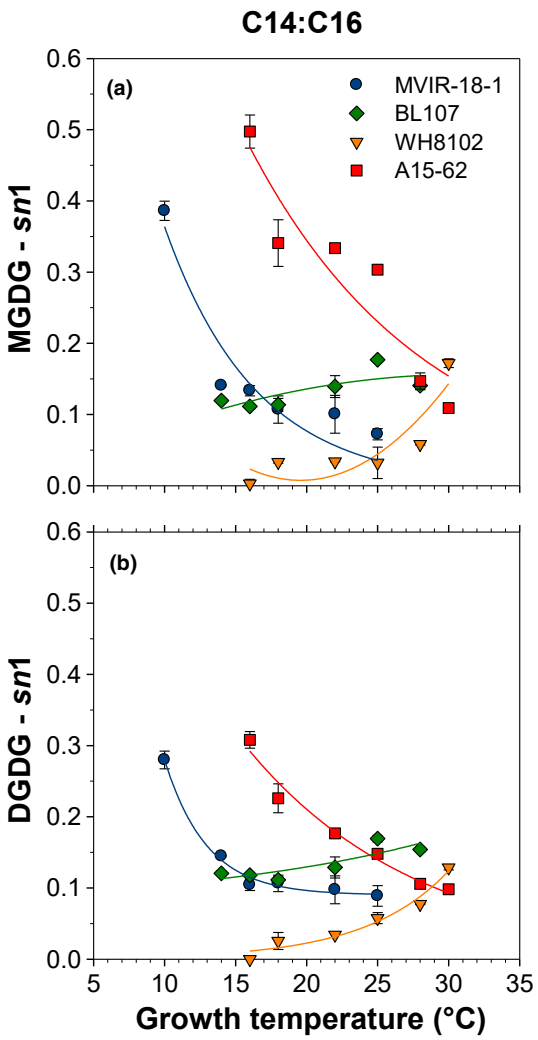


Fig. 2 Variations of the C14 : C16 molar ratio, related to the average length of the fatty acid bound to the *sn*-1 position of the monogalactosyldiacylglycerol (MGDG, a) and the digalactosyldiacylglycerol (DGDG, b) for the four marine *Synechococcus* strains: MVIR-18-1 (blue circles), BL107 (green diamonds), WH8102 (orange triangles) and A15-62 (red squares), acclimated to a range of temperatures spanning from 10 to 30°C. Measurements are the average of three to four replicates with error bars ± SD.

by the fact that no rapid complementary decrease of C14 : 0 was observed in acclimated cultures (Fig. 3) or during thermal shift experiments (Figs S7, S11).

The proportion of C14 : 1 chains increased at the MGDG *sn*-2 position in the subpolar MVIR-18-1 and in both temperate strains in response to low temperature, with WH8102 showing the largest changes (Fig. 4a). Complementary changes between C14 : 0 and C14 : 1 chains were observed during the shift experiments, suggesting a Δ9-desaturation activity at this position (Figs S7, S11).

DGDG showed overall similar variations to MGDG at both *sn*-1 (Fig. 3c,d) and *sn*-2 (Fig. 4b) positions, but global C16 and C14 mono-unsaturation levels were lower than for MGDG. This suggests the occurrence of an active C16 : 0 mono-desaturation process at the *sn*-1 site, especially for the subpolar strain MVIR-18-1, as also observed in the cold (Fig. S8) and warm (Fig. S12) shift experiment results. The insertion of *de novo* synthetized C14 : 1 chains at the DGDG *sn*-1 (Fig. 3d) and *sn*-2 (Fig. 4b) positions in cultures acclimated to low temperature were observable for all strains, but were weaker than for MGDG.

Mono-desaturation processes on SQDG As already observed in *Synechococcus* sp. WH7803, SQDG showed a quite different fatty acid composition from the two galactolipids (Pittera *et al.*, 2018). In the four strains, the *sn*-1 position bound quasi exclusively C16 chains (Fig. S5) and was seemingly the site of an active Δ9-mono-desaturation activity in cultures acclimated to low temperature (Fig. 3e,f), as well as in cultures subjected to temperature shifts (Figs S9, S13). The *sn*-2 position bound C14 and C16 chains, often in comparable proportions (Fig. S5). Active C16 : 0 mono-desaturation was induced in cultures acclimated to low temperature in the four strains, again with greater amplitudes in the subpolar strain MVIR-18-1 and the cold-temperate strain BL107, along with a weak increase in the C14 : 1 proportion (Figs 4d, S9, S13).

Double desaturation processes We used the C16 : 2 to C16 : 1 ratio to study the level of double unsaturation of the fatty chains of the different lipids (Figs 5, S3–S6). Interestingly, double unsaturations occurred exclusively on C16 chains, at the *sn*-1 position of the two galactolipids. The global level of double unsaturation was quite high in the strains adapted to warm environments, WH8102 and A15-62, whereas C16: 2 chains were barely detected in the cold-temperate and subpolar strains BL107 and MVIR-18-1 (Fig. 5). Of note, strain WH8102 responded to low temperature by a large induction of C16 : 2 at the *sn*-1 position of MGDG and, to a lesser extent, DGDG. The capacity for efficient double desaturation activity of strain WH8102 was also observable during the cold shift experiments (Fig. S7).

Metagenomics of lipid-desaturase genes

Previous studies have shown that marine *Synechococcus* temperature ecotypes display different sets of genes encoding lipid-desaturases, including the Δ9-desaturases DesC3 and DesC4 and the Δ12-desaturases DesA2 and DesA3 (Notes S3; Varkey *et al.*, 2016; Pittera *et al.*, 2018). To complement the results obtained on strains representative of cold- (I and IV) and warm-adapted (II and III) clades, we used the *Tara* Oceans dataset to study the global distribution and prevalence of the four desaturase genes along natural thermal gradients (Fig. 6).

The prevalence of the *desC3* gene was close to 1 at all stations, independently of the seawater temperature and ESTU composition (Fig. 6b). This shows that this desaturase gene is present in all *Synechococcus* cells *in situ* in a single copy, just like *petB*. By contrast, the *desC4* gene was virtually absent from waters warmer than 20°C (Fig. 6c). Below this threshold, the prevalence of this gene increased quasilinearly as seawater temperature decreased, reaching values close to 1 in 13°C waters. This gene was most abundant in populations colonizing waters below 17°C, dominated by *Synechococcus* ESTUs IA and IVA-C (Farrant *et al.*, 2016).

© 2019 The Authors
New Phytologist © 2019 New Phytologist Trust

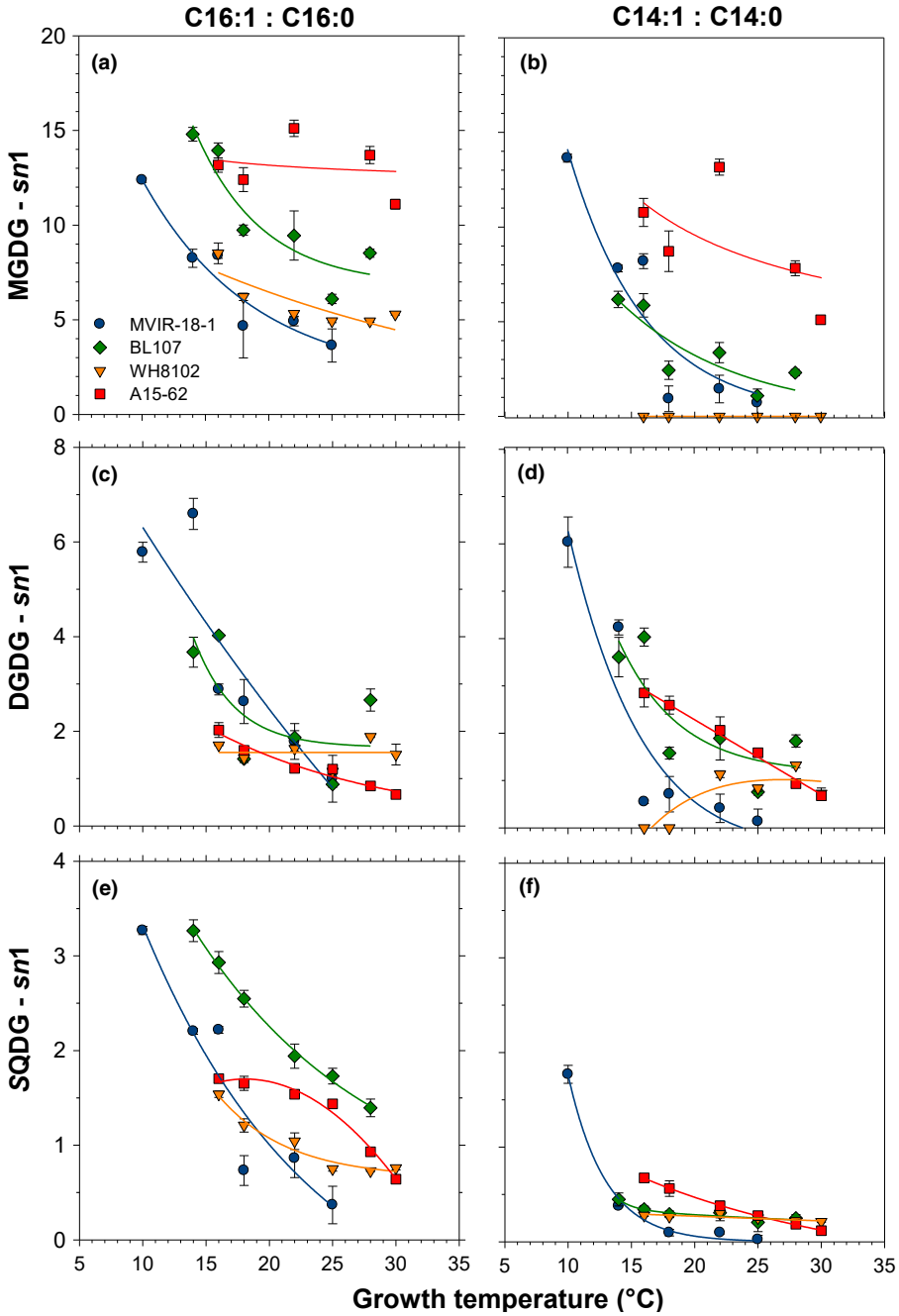


Fig. 3 Variations of the monodesaturation level, expressed as the C16:1 : C16:0 (left) and C14:1 : C14:0 (right) molar ratios, of the fatty acid bound to the *sn*-1 position of the monogalactosyldiacylglycerol (MGDG, a and b), digalactosyldiacylglycerol (DGDG, c and d) and sulfoquinovosyldiacylglycerol (SQDG, e and f) for the four marine *Synechococcus* strains: MVIR-18-1 (blue circles), BL107 (green diamonds), WH8102 (orange triangles) and A15-62 (red squares), acclimated to a range of temperatures spanning from 10 to 30°C. Measurements are the average of three to four replicates with error bars ± SD.

The prevalence of the *desA2* gene, which encodes a Δ12-desaturase inserting a double unsaturation, showed its maximal value of *c.* 1 for temperatures above 20°C (Fig. 6d). Below this temperature, prevalence decreased progressively down to zero for seawater temperatures lower than 14°C. In this dataset, the *desA2* gene is mostly present in populations dominated by *Synechococcus* ESTUs IIA and IIIA.

The other Δ12-desaturase, *desA3*, showed a similar global pattern to the *desC4* gene with a marked shift in the relative abundance of this gene between populations dominated by *Synechococcus* ESTUs IA and IVA-C, colonizing waters below 18°C, and those dominated by ESTU IIA, thriving in waters above this temperature. By contrast, note that in Mediterranean Sea stations dominated by *Synechococcus* ESTU IIIA, prevalence of the *desA3* gene remained high independent of temperature (Fig. 6e), consistently with the fact that *desA3* is present in the genome of all strains representative of clade III, while it is generally absent from most clade II strains (Pittera *et al.*, 2018).

Discussion

Photosynthesis regulation in cold- and warm-adapted *Synechococcus* ecotypes

Previous studies have shown that adaptation to temperature in marine *Synechococcus* relies on the existence of physiologically

© 2019 The Authors
New Phytologist © 2019 New Phytologist Trust

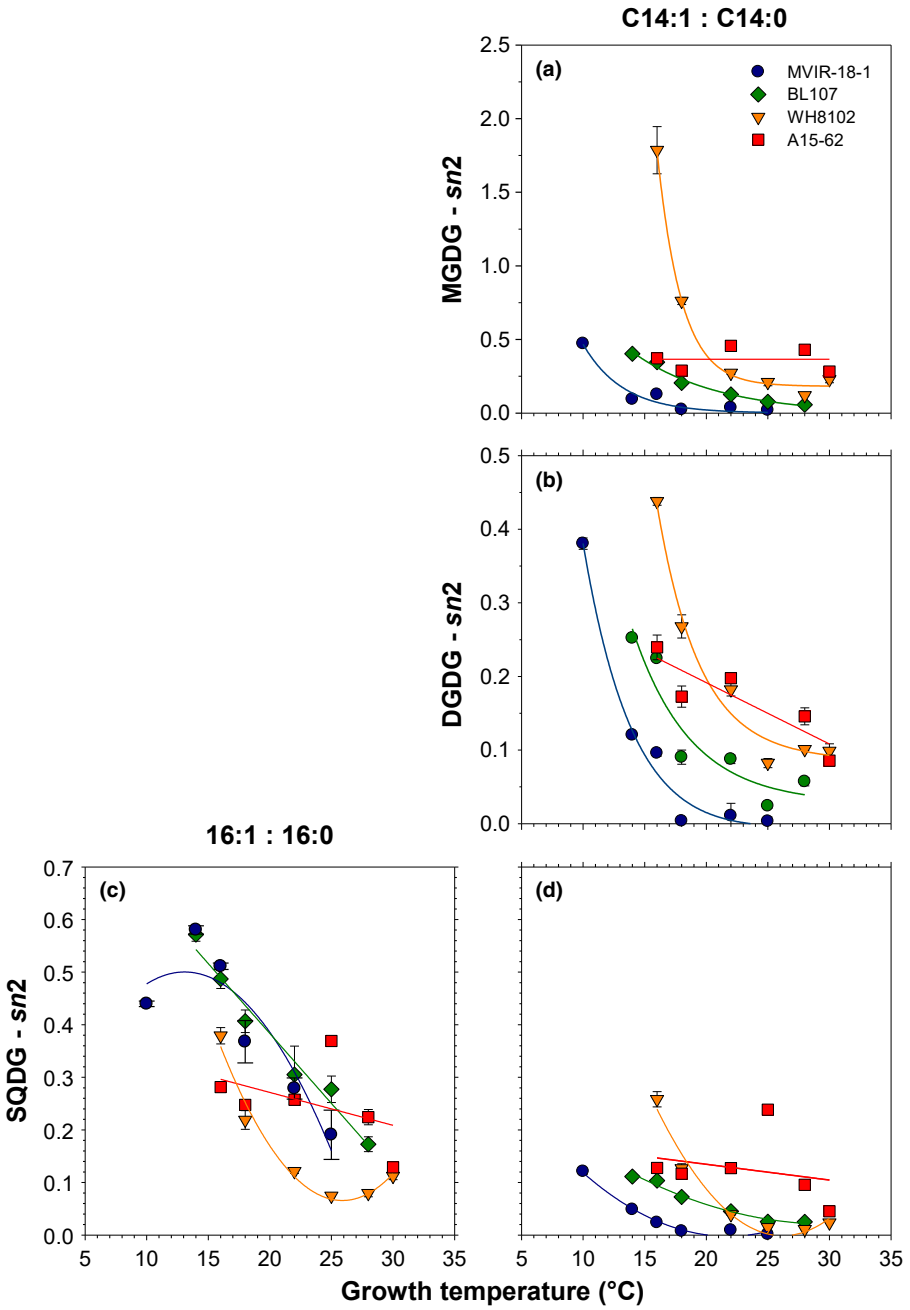


Fig. 4 Variations of the monodesaturation level, expressed as the C16:1 : C16:0 (left) and C14:1 : C14:0 (right) molar ratios, of the C14 fatty acid bound to the *sn*-2 position of the monogalactosyldiacylglycerol (MGDG, a), the digalactosyldiacylglycerol (DGDG, b) and sulfoquinovosyldiacylglycerol (SQDG, c and d) for the four marine *Synechococcus* strains MVIR-18-1 (blue circles), BL107 (green diamonds), WH8102 (orange triangles) and A15-62 (red squares), acclimated to a range of temperatures spanning from 10 to 30°C. Measurements are average of three to four replicates with error bars ± SD.

specialized lineages that have colonized different thermal environments (Pittera *et al.*, 2014; Varkey *et al.*, 2016). The strains used in the present study were isolated in water masses with contrasting temperatures, from tropical to subpolar habitats, and belong to the four major *Synechococcus* clades (Table 1). Study of their thermal *preferenda* showed that the two strains isolated from warm environments, *Synechococcus* spp. A15-62 (clade II) and WH8102 (clade III), were able to grow at temperatures above 30°C. This is consistent with previous studies on the Sargasso Sea strains WH8103 (clade III) and WH7803 (clade V; Moore *et al.*, 1995; Pittera *et al.*, 2014, 2018). By contrast, strains BL107 (clade IV) and MVIR-18-1 (clade I), isolated from temperate and subpolar environments respectively, could not cope with these high temperatures but were able to grow down to 12°C or less (Fig. 1a). These results are in good agreement with previous observations that clades I and IV predominate in relatively cold environments, whereas clades II and III preferentially thrive in warmer waters (Zwirglmaier *et al.*, 2008; Sohm *et al.*, 2015; Farrant *et al.*, 2016; Paulsen *et al.*, 2016; Kent *et al.*, 2018). Moreover, note that these differences in thermal *preferenda* would be likely even more apparent with cultures grown under higher light irradiance, as discussed in previous studies (Pittera *et al.*, 2014, 2018).

All four strains managed to maintain fairly high PSII quantum yield over their whole temperature growth range, showing an efficient acclimation of their photosynthetic apparatus to different

© 2019 The Authors
New Phytologist © 2019 New Phytologist Trust

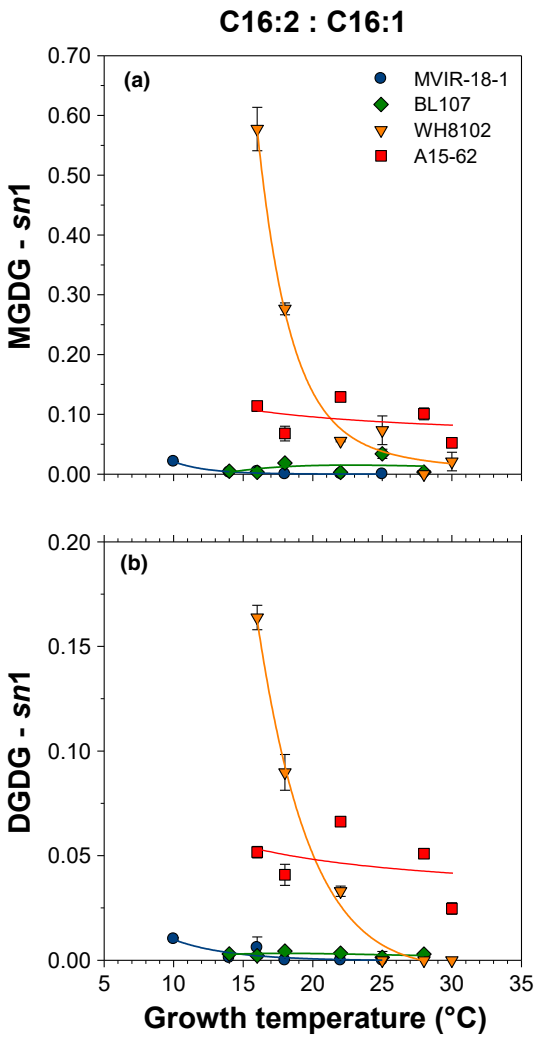


Fig. 5 Variations of the double-desaturation level, expressed as C16:2 : C16:1 molar ratios, of the fatty acid bound to the *sn*-1 position of the monogalactosyldiacylglycerol (MGDG, a) and digalactosyldiacylglycerol (DGDG, b), for the four marine *Synechococcus* strains MVIR-18-1 (blue circles), BL107 (green diamonds), WH8102 (orange triangles) and A15-62 (red squares) acclimated to a range of temperatures spanning from 10 to 30°C. Measurements are the average of three to four replicates with error bars ± SD.

temperatures (Fig. 1b). Pigment analyses suggest that this ability relies partly upon changes in the relative proportions of major pigments in thylakoid membranes, which differed between strains. *Synechococcus* spp. WH8102 and A15-62 showed a marked increase of the β-carotene : Chl *a* ratio in response to high growth temperature, whereas this increase was moderate in cold-adapted strains (Fig. 1c). Such variations might originate from a change in the photosystem ratio because these complexes bind the two pigments in fixed but very different proportions. As the β-carotene : Chl *a* molar ratio is 0.23 in PSI (Xu & Wang, 2017) and 0.31 in PSII monomers (Umena *et al.*, 2011), our results thus suggest an increase of the PSII : PSI ratio in response to high temperatures. This would lead to a decrease the cyclic electron transport around PSI and increase the transthylakoidal pH gradient, then enhancing the ATP production in the cytosol of the warm-adapted ecotypes at high growth temperature. In addition, β-carotene and Chl *a* molecules have recently been found to be bound to High Light Inducible Proteins (HLIPs) in *Synechocystis* sp. PCC 6803, which are stress-inducible proteins allowing dissipation of excess light (Dolganov *et al.*, 1995; Komenda & Sobotka, 2016). The characterized proteins, whose dissipative activity is temperature-dependent, have a 0.33 (HliD) and 0.5 (HliC) β-carotene : Chl *a* ratio (Staleva *et al.*, 2015; Shukla *et al.*, 2018). Although such proteins have never been biochemically characterized in marine *Synechococcus*, their genomes contain many *hli* genes, which probably encode similar pigmented proteins (Bhaya *et al.*, 2002; Scanlan *et al.*, 2009). Therefore, the regulation of such proteins might also be involved in temperature acclimation in marine *Synechococcus*, especially in the warm-adapted strains WH8102 and A15-62.

In all strains, the zeaxanthin : Chl *a* ratio increased in response to low temperature (Fig. 1d). Moreover, this ratio was globally higher in the cold-adapted strains MVIR-18-1 and BL107. Variations of the zeaxanthin : Chl *a* ratio probably originated partly from a decrease of the Chl *a* cell content, a classical cold-induced response, in order to regulate light utilization in slow growth conditions (Inoue *et al.*, 2001). However, an upregulation in zeaxanthin synthesis was also probably involved, as suggested by the increase of the zeaxanthin : β-cryptoxanthin ratio (Fig. S2), the latter pigment being directly converted to zeaxanthin by the β-carotene hydroxylase (CrtR). Although the localization and function of zeaxanthin have never been formally demonstrated in marine *Synechococcus*, it could have a role in photoprotection by dissipating excess light under low temperature conditions, a process seemingly more prevalent in cold-adapted strains (Fig. 1). However, it is well known that the integration of polar xanthophylls into biological and model membranes strongly affects their properties (for a review see Popova & Andreeva, 2013). If zeaxanthin occurs as a free molecule in marine *Synechococcus* thylakoids, it is expected to decrease the membrane fluidity at low temperature, as shown in Antarctic bacteria (Jagannadham *et al.*, 2000). Yet, this potentially deleterious increase in membrane rigidity at low temperature seems to be counteracted by modifications of the membrane composition in order to maintain proper membrane fluidity.

Marine Synechococcus thermoregulate the membrane composition differently

Analysis of all molecular species of membrane lipids in four phylogenetically and physiologically distinct strains, as well as comparisons with our previous data on the model strain WH7803 (Pittera *et al.*, 2018), allowed us for the first time to draw a broad picture of the composition of the membranes in marine *Synechococcus*. These five strains exhibit general features that seem to be specific for this group, including notably variability of the chain bound to the SQDG *sn*-2 position and the relative insensitivity of the composition of PG to temperature (Murata & Wada, 1995; Pittera *et al.*, 2018). Also in sharp contrast to freshwater strains, all five marine *Synechococcus* strains contain very low relative concentrations of C18 chains and are C14-rich. Therefore,

© 2019 The Authors
New Phytologist © 2019 New Phytologist Trust

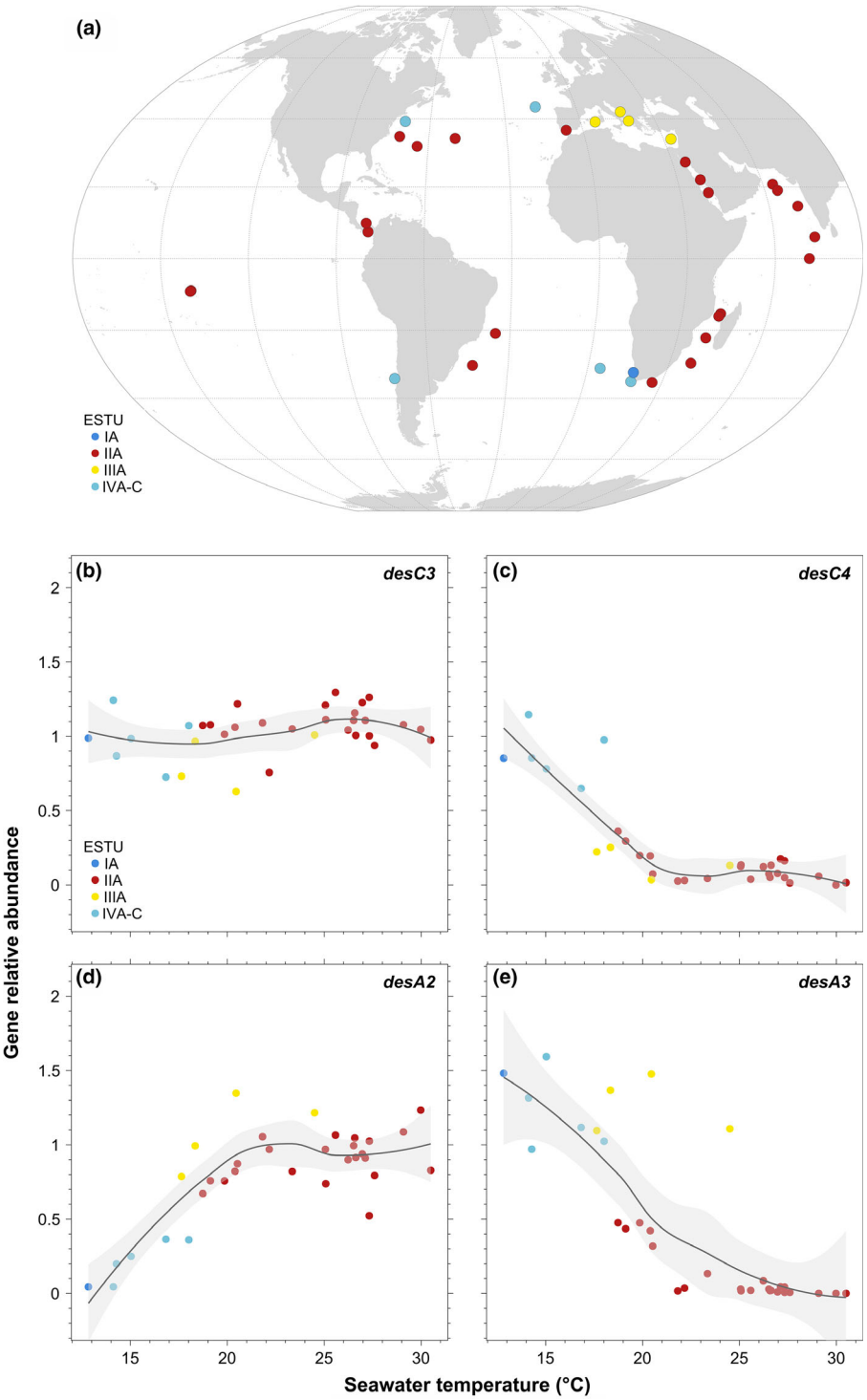


Fig. 6 Localization of the stations of the *Tara* Oceans expedition used in this study (a), and prevalence of the Δ9 lipid-desaturase genes *desC3* (b) and *desC4* (c) and the Δ12 lipid-desaturase genes *desA2* (d) and *desA3* (e), as a function of seawater temperature at these stations. This was determined using metagenomic data from the *Tara* Oceans expedition for stations dominated by major ecologically significant taxonomic units (ESTUs) *sensu* Farrant *et al*. (2016) within *Synechococcus* clades I–IV. At each station, the relative abundance of each desaturase gene was normalized by the relative abundance of the core, single-copy gene *petB*. The LOESS regression model was used as a curve fitting method with a span $\alpha = 0.8$ using the ggplot2 R package (https://cran.r-project.org/web/packages/ggplot2/index.html). The gray shading areas correspond to the 95% confidence interval.

while freshwater strains have C18 and C16 at the *sn*-1 and *sn*-2 positions respectively, marine *Synechococcus* lipids preferentially bind C16 and C14 at these positions, respectively. This fatty acid composition implies that marine *Synechococcus* thylakoids are probably thinner and globally more fluid than those of freshwater strains. This is also in good agreement with the much smaller cell volume of these picocyanobacteria compared to their freshwater counterparts. In addition, it is worth noting that C14 : 0 is usually one of the major fatty acids in samples from surface ocean regions (e.g. Wakeham and Canuel, 1988; Van Mooy and Fredricks, 2010). Thus, the finding that C14 dominates in all marine *Synechococcus* strains investigated so far further confirms that marine picocyanobacteria are a major source of this lipid in the oceans.

However, some strains have the capacity to modify the C16 : C14 ratio, and thus to modulate the membrane thickness.

© 2019 The Authors
New Phytologist © 2019 New Phytologist Trust

Indeed, our results and a previous study show that, whereas it slightly increased in BL107 (clade IV) and WH8102 (clade III) in response to cold temperature, strains MVIR-18-1 (clade I), A15-62 (clade II) and WH7803 (clade V) were able to significantly decrease the membrane thickness by synthetizing *de novo* shorter lipids, therefore increasing its fluidity (Fig. 2; Pittera *et al.*, 2018). As these strains inhabit very different thermal environments, the capacity to carry out this process does not appear to be related to the thermotype differentiation. To better understand these different responses, future research should investigate the influence of other abiotic parameters on a larger number of strains.

In response to temperature variations, the *Synechococcus* strains deployed a panel of different desaturation activities. Our results show that the cold-adapted strains MVIR-18-1 (clade I) and BL107 (clade IV) can carry out extensive desaturation of fatty acid mono-desaturation, on both C16 and C14 chains and at both *sn* positions of the three membrane glycolipids (Figs 3, 4). Of note, the primary Δ9 monounsaturations are expected to have larger membrane fluidifying effects than the secondary Δ12 unsaturations (Hyvonen *et al.*, 1997; Hyvonen & Kovanen, 2005; Ollila *et al.*, 2007). These large Δ9 monodesaturation capacities are thus likely to be critical for tolerance to low temperature. The warm-adapted strains A15-62 (clade II) and WH8102 (clade III) were overall able to carry out lower monodesaturation activities (Figs 3, 4). Nevertheless, the tropical A15-62 strain showed stable and relatively high C16 : 2 content in the galactolipids, and WH8102 grown at low temperature (Figs 5, S7) induced the highest contents of double unsaturations, whereas these were only detected at trace levels in the cold-adapted strains. Thus, in contrast to cold-adapted strains, strains adapted to warm environments seem to display much higher relative contents of double unsaturated fatty acids in their membranes.

Synechococcus thermotypes use different lipid-desaturase enzymes

The five strains of marine *Synechococcus* studied to date in the context of membrane lipid thermoacclimation capacities possess different sets of desaturase genes (Pittera *et al.*, 2018; Table 1). Lipidomic analyses showed that all strains were able to induce Δ9-monodesaturations, which is consistent with the presence of *desC3* in all strains. The lipidic profiles of the A15-62 and WH8102 strains (Figs 3, 4), which only possess DesC3 to synthesize monounsaturated chains, show that this enzyme can incorporate Δ9 double bonds on the three glycolipids, at the two *sn* positions and on both palmitic (C16 : 0) and myristic (C14 : 0) chains. Therefore, it appears to be a poorly regiospecific enzyme, responsible for the basal (but essential) monodesaturation capacity of all marine *Synechococcus*. The cold-adapted strains MVIR-18-1 and BL107, which contain an additional Δ9-desaturase gene (Table 1), *desC4*, showed increased monodesaturation capacities. This gene, thought to have derived from a recent gene duplication event (Pittera *et al.*, 2018), could then provide additional monodesaturation capacity in order to cope with lower temperatures. However, on the lipids of these cold-adapted strains, we found no evidence for sites of monodesaturation that would be specific of the DesC4 enzyme. Therefore, our results suggest that the two DesC enzymes in marine *Synechococcus* are rather generalist monodesaturases of the C16 and C14 chains of glycolipids, that is they display a relatively weak regiospecificity. This contrasts with the DesC1 and DesC2 enzymes of freshwater strains, which show differential site specificities (Chintalapati *et al.*, 2006). Future work should aim to study the possible induction of the *desC4* gene by low temperature, which would provide further evidence of a monodesaturation booster function of the DesC4 enzyme in marine *Synechococcus*.

Double unsaturations were detected mostly in A15-62 and WH8102, which like all other warm-adapted ecotypes sequenced to date, possess a specific Δ12-desaturase, DesA2 (Table 1). In contrast to monounsaturations, the double unsaturations are located specifically on C16 chains at the *sn*-1 of the galactolipids binding a C14 : 1 chain at the *sn*-2 site, indicating a high level of regiospecificity of DesA enzymes in marine *Synechococcus*. Note that the WH8102 strain possesses two Δ12-desaturases, DesA3 and DesA2. This probably allows this strain to induce the strong double desaturation activities that were observed in response to low temperature (Figs 5, S7h,k). The role of DesA3 in cold-adapted strains remains enigmatic because very few double unsaturations were detected in these strains, both in acclimated cultures and in temperature shift experiments. It is possible that, in these strains, this enzyme is activated in response to other environmental factors, such as high irradiance.

The lipid-desaturase genes show differential distributions in the ocean

To provide more insight into the physiological strategies used by *Synechococcus* temperature ecotypes to thermoregulate their membranes, we studied the distribution of the four lipid desaturase genes at 33 widespread stations along the *Tara* Oceans transect (Fig. 6a). As expected, the core gene *desC3* displayed a temperature-independent distribution. By contrast, the Δ9-desaturase gene *desC4*, which is specific of cold-adapted strains (Table 1), was found exclusively in waters colder than 20°C, consistent with its hypothesized role of monodesaturation enhancer enabling growth in cold thermal environments.

The distribution of the *desA2* gene displayed a mirror pattern to *desC4*, confirming that this Δ12-desaturase is specific to warm environments, dominated by either clade II or clade III *Synechococcus* cells. Like *desC4*, the Δ12-desaturase gene *desA3* was abundant in waters colder than 20°C. This gene was also abundant between 17 and 25°C, but specifically in assemblages dominated by clade III strains, consistent with the presence of this gene in all clade III but no clade II strains sequenced thus far (Table 1). Clade III *Synechococcus* are most abundant in oligotrophic, phosphorus-limited areas such as the Mediterranean Sea (Mella-Flores *et al.*, 2011; Farrant *et al.*, 2016), which is subjected to a strong seasonality, being warm in summer (up to

© 2019 The Authors
New Phytologist © 2019 New Phytologist Trust

30°C) but cold in winter, with minimal temperatures as low as 13°C in the northwestern basin. Our results thus suggest that the presence of both *desC2* and *desC3* genes confers fitness to clade III strains, which are subjected to much stronger seasonal variability over the year than their (sub)tropical clade II counterparts.

Conclusion and perspectives

Our study shows that marine *Synechococcus* temperature ecotypes can manage, via different mechanisms, to regulate their photosynthetic apparatus in order to maintain optimal photosynthetic capacity over most of their respective temperature growth range. We also show that marine *Synechococcus* have evolved different physiological strategies based on the presence of specific lipid desaturase gene sets, allowing them to cope with changes in the fluidity of the photosynthetic membrane in different thermal niches. Although more work such as mutant characterization is necessary, our data allow us to propose functions and ecological roles for the four *Synechococcus* desaturases. DesC3 is a constitutive Δ9-desaturase with weak regiospecificity and its activity is probably mostly temperature-independent. *Synechococcus* strains belonging to clades I and IV use an additional enzyme, DesC4, which provides them with enhanced Δ9 monodesaturation capacities and allow them to greatly increase the membrane fluidity in cold environments. The occurrence of the *desC4* gene, probably by duplication of its common ancestor with *desC3* (Pittera *et al.*, 2018), has probably been an important event for the colonization of high-latitude environments by marine *Synechococcus*. In warm environments, there is little need to have strong variations in membrane fluidity and marine *Synechococcus* have seemingly adopted a different strategy. Clade II strains, which live in warm and stable environments, have globally poor desaturation capacities. However, they constitutively contain membrane lipids with more double unsaturation thanks to DesA2, a Δ12-desaturase specific to warm environments. In warm-temperate environments, where temperature is more variable, clade III strains use the combination of two Δ12-desaturases, DesA2 and DesA3. They confer high capacities of double desaturations, which probably help in tolerating the low winter temperatures in this niche. The preferential use of double unsaturations, which have less influence on the membrane fluidity than the monounsaturations, probably allows a finer tuning of this physical property in the thermal environments of clade II and III strains than their cold-adapted counterparts. An alternative hypothesis is that Δ12-desaturases might be most efficient at high temperature, while Δ9-desaturases were selected in colder environments during the evolution of marine *Synechococcus*. In conclusion, this study suggests strongly that the capacity to greatly modulate membrane fluidity has been critical for the colonization of different thermal niches by marine *Synechococcus*, leading to the differentiation in physiologically distinct temperature ecotypes. It furthermore shows that lipid-desaturase genes are interesting markers for monitoring the dynamics of ocean microbial communities in the context of global change.

Acknowledgements

This work was funded by the French programs EC2CO-Microbien (METALIC), ANR CINNAMON (ANR-17-CE2-0014-01) and EMBRC France (INFRA-2010-2.2.5). SB and UG were supported by the Région Bretagne and the French Ministry of Higher Education and Research. JJ and EM were supported by the French National Research Agency (ANR-10-LABEX-04 GRAL Labex, Grenoble Alliance for Integrated Structural Cell Biology; ANR-11-BTBR-0008 Océanomics; ANR-15-IDEX-02 GlycoAlps Cross Disciplinary Project). We are grateful to the Roscoff Culture Collection for maintaining the *Synechococcus* strain used in this study. We thank the ABIMS Platform (Station Biologique de Roscoff) for their help in the metagenomics data computing. 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). We certify that there is no conflict of interest associated with this manuscript.

Author contributions

CS, LG and FP designed the research. SB, JJ, UG, VG, JP, DD, FP, HD, MR, EM, NAN, LG and CS contributed to the performance of this work and the analysis of the data. CS, LG, FP, SB, NN, JJ, EM, UG and DD worked on the writing of the manuscript.

ORCID

David Demory https://orcid.org/0000-0003-4928-8925
Hugo Doré https://orcid.org/0000-0003-4160-3679
Laurence Garczarek https://orcid.org/0000-0002-8191-8395
Valérie Gros https://orcid.org/0000-0002-6082-4613
Juliette Jouhet https://orcid.org/0000-0002-4402-2194
Eric Maréchal https://orcid.org/0000-0002-0060-1696
Frédéric Partensky https://orcid.org/0000-0003-1274-4050
Christophe Six https://orcid.org/0000-0002-8506-1149

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410.

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

Bhaya D, Dufresne A, Vaulot D, Grossman A. 2002. Analysis of the *hli* gene family in marine and freshwater cyanobacteria. *FEMS Microbiology Letters* **215**: 209–219.

Bligh E, Dyer W. 1959. A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology* **37**: 911–917.

© 2019 The Authors
New Phytologist © 2019 New Phytologist Trust

Boatman TG, Lawson T, Geider RJ. 2017. A key marine diazotroph in a changing ocean: the interacting effects of temperature, CO_2 and light on the growth of *Trichodesmium erythraeum* IMS101. *PLoS ONE* **12**: 1–20.

Boudière L, Michaud M, Petroutsos D, Rébeillé F, Falconet D, Bastien O, Roy S, Finazzi G, Rolland N, Jouhet J *et al.* 2014. Glycerolipids in photosynthesis: Composition, synthesis and trafficking. *Biochimica et Biophysica Acta (BBA)–Bioenergetics* **1837**: 470–480.

Campbell DA, Hurry V, Clarke AK, Gustafsson P, Oquist G. 1998. Chlorophyll fluorescence analysis of cyanobacterial photosynthesis and acclimation. *Microbiology and Molecular Biology Reviews* **62**: 667–683.

Chi X, Yang Q, Zhao F, Qin S, Yang Y, Shen J, Lin H. 2008. Comparative analysis of fatty acid desaturases in cyanobacterial genomes. *Comparative and Functional Genomics* **2008**: 1–25.

Chintalapati S, Prakash JSS, Gupta P, Ohtani S, Suzuki I, Sakamoto T, Murata N, Shivaji S. 2006. A novel Delta9 acyl-lipid desaturase, DesC2, from cyanobacteria acts on fatty acids esterified to the sn-2 position of glycerolipids. *Biochemical Journal* **398**: 207–14.

Coleman ML, Chisholm SW. 2007. Code and context: *Prochlorococcus* as a model for cross-scale biology. *Trends in Microbiology* **15**: 398–407.

Dolganov NA, Bhaya D, Grossman AR. 1995. Cyanobacterial protein with similarity to the chlorophyll a/b binding proteins of higher plants: evolution and regulation. *Proceedings of the National Academy of Sciences, USA* **92**: 636–640.

Farrant GK, Doré H, Cornejo-Castillo FM, Partensky F, Ratin M, Ostrowski M, Pitt FD, Wincker P, Scanlan DJ, Iudicone D *et al.* 2016. Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proceedings of the National Academy of Sciences, USA* **113**: E3365–E3374.

Flombaum P, Gallegos JL, Gordillo RA, Rincon J, Zabala LL, Jiao N, Karl DM, Li WKW, Lomas MW, Veneziano D *et al.* 2013. Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proceedings of the National Academy of Sciences, USA* **110**: 9824–9829.

Herdman M, Castenholz RW, Waterbury JB, Rippka R. 2001. Form-genus XIII. *Synechococcus*. In: Boone DR, Castenholz RW, Garrity GM, eds. *Bergey's manual of systematic bacteriology*. Berlin, Germany: Springer, 508–512.

Huang S, Wilhelm SW, Harvey HR, Taylor K, Jiao N, Chen F. 2012. Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *ISME Journal* **6**: 285–297.

Hyvonen MT, Ala-korpela M, Vaara J, Rantala TT, Jokisaari J. 1997. Inequivalence of single CHa and CHb methylene bonds in the interior of a diunsaturated lipid bilayer from a molecular dynamics simulation. *Chemical Physics Letters* **268**: 55–60.

Hyvonen MT, Kovanen PT. 2005. Molecular dynamics simulations of unsaturated lipid bilayers: effects of varying the numbers of double bonds. *European Biophysics Journal* **34**: 294–305.

Inoue N, Taira Y, Emi T, Yamane Y, Kashino Y, Koike H, Satoh K. 2001. Acclimation to the growth temperature and the high-temperature effects on photosystem II and plasma membranes in a mesophilic cyanobacterium, *Synechocystis* sp. PCC 6803. *Plant & Cell Physiology* **42**: 1140–1148.

Itoh S, Kozuki T, Nishida K, Fukushima Y, Yamakawa H, Domonkos I, Laczkó-Dobos H, Kis M, Ughy B, Gombos Z. 2012. Two functional sites of phosphatidylglycerol for regulation of reaction of plastoquinone QBin photosystem II. *Biochimica et Biophysica Acta – Bioenergetics* **1817**: 287–297.

Jagannadham MV, Chattopadhyay MK, Subbalakshmi C, Vairamani M, Narayanan K, Rao CM, Shivaji S. 2000. Carotenoids of an Antarctic psychrotolerant bacterium, *Sphingobacterium antarcticus*, and a mesophilic bacterium, *Sphingobacterium multivorum*. *Archives of Microbiology* **173**: 418–424.

Johnson ZI, Zinser ER, Coe A, Mcnulty NP, Malcolm ES, Chisholm SW, Woodward EMS, Chisholm SW. 2006. Partitioning among *Prochlorococcus* ecotypes along environmental gradients. *Science* **311**: 1737–1740.

Jouhet J, Lupette J, Clerc O, Magneschi L, Bedhomme M, Collin S, Roy S, Maréchal E, Rébeillé F. 2017. LC-MS/MS versus TLC plus GC methods: consistency of glycerolipid and fatty acid profiles in microalgae and higher plant cells and effect of a nitrogen starvation. *PLoS ONE* **12**: 1–21.

Kana TM, Glibert PM. 1987. Effect of irradiances up to 2000 $\mu E\ m^{-2}\ s^{-1}$ on marine *Synechococcus* WH7803—I. Growth, pigmentation, and cell composition. *Deep Sea Research Part A* **34**: 479–495.

Kent AG, Baer SE, Mouginot C, Huang JS, Lomas MW, Martiny AC. 2018. Parallel phylogeography of *Prochlorococcus* and *Synechococcus*. *ISME Journal* **13**: 430–441.

Komenda J, Sobotka R. 2016. Cyanobacterial high-light-inducible proteins — Protectors of chlorophyll – protein synthesis and assembly. *Biochimica et Biophysica Acta – Bioenergetics* **1857**: 288–295.

Los DA, Murata N. 1998. Structure and expression of fatty acid desaturases. *Biochimica et Biophysica Acta* **1394**: 3–15.

Mackey KRM, Paytan A, Caldeira K, Grossman AR, Moran D, McIlvin M, Saito MA. 2013. Effect of temperature on photosynthesis and growth in marine *Synechococcus* spp. *Plant Physiology* **163**: 815–829.

Magoč T, Salzberg SL. 2011. FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**: 2957–2963.

Marie D, Partensky F, Vaulot D, Brussaard C. 2001. Enumeration of phytoplankton, bacteria, and viruses in marine samples. *Current Protocols in Cytometry* **11**: 1–15.

Mazard S, Ostrowski M, Partensky F, Scanlan DJ. 2012. Multi-locus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*. *Environmental Microbiology* **14**: 372–386.

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

Merritt MV, Rosenstein SP, Loh C, Hsui-sui Chou R, Allen MM. 1991. A comparison of the major lipid classes and fatty acid composition of marine unicellular cyanobacteria with freshwater species. *Archives of Microbiology* **155**: 107–113.

Mikami K, Murata N. 2003. Membrane fluidity and the perception of environmental signals in cyanobacteria and plants. *Progress in Lipid Research* **42**: 527–543.

Mizusawa N, Wada H. 2012. The role of lipids in photosystem II. *Biochimica et Biophysica Acta – Bioenergetics* **1817**: 194–208.

Moore LR, Coe A, Zinser ER, Saito MA, Sullivan MB, Lindell D, Frois-Moniz K, Waterbury J, Chisholm SW. 2007. Culturing the marine cyanobacterium *Prochlorococcus*. *Limnology and Oceanography: Methods* **5**: 353–362.

Moore L, Goericke R, Chisholm S. 1995. Comparative physiology of *Synechococcus* and *Prochlorococcus*: influence of light and temperature on growth, pigments, fluorescence and absorptive properties. *Marine Ecology Progress Series* **116**: 259–275.

Murata N, Wada H. 1995. Acyl-lipid desaturases and their importance in the tolerance and acclimatization to cold of cyanobacteria. *Biochemical Journal* **308**: 1–8.

Ogawa T, Misumi M, Sonoike K. 2017. Estimation of photosynthesis in cyanobacteria by pulse-amplitude modulation chlorophyll fluorescence: problems and solutions. *Photosynthesis Research* **133**: 63–73.

Ollila S, Hyvonen MT, Vattulainen I. 2007. Polyunsaturation in lipid membranes: dynamic properties and lateral pressure profiles. *Journal of Physical Chemistry B* **111**: 3139–3150.

Paulsen ML, Doré H, Garczarek L, Seuthe L, Müller O, Sandaa R, Bratbak G, Larsen A. 2016. *Synechococcus* in the Atlantic gateway to the Arctic Ocean. *Frontiers in Marine Science* **3**: 1–14.

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

Pittera J, Jouhet J, Breton S, Garczarek L, Partensky F, Maréchal É, Nguyen NA, Doré H, Ratin M, Pitt FD *et al.* 2018. Thermoacclimation and genome adaptation of the membrane lipidome in marine *Synechococcus*. *Environmental Microbiology* **20**: 612–631.

Pittera J, Partensky F, Six C. 2016. Adaptive thermostability of light-harvesting complexes in marine picocyanobacteria. *ISME Journal* **11**: 1–13.

Popova AV, Andreeva AS. 2013. *Carotenoid–lipid interactions*. Amsterdam, the Netherlands: Elsevier.

Rippka R, Coursin T, Hess W, Lichtlé C, Canlan DJ, Palinska KA, Iteman I, Partensky F, Houmard J, HerdmanM *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*). *International Journal of Systematic and Evolutionary Microbiology* **50**: 1833–1847.

Scanlan DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR, Post AF, Hagemann M, Paulsen I, Partensky F. 2009. Ecological genomics of

© 2019 The Authors
New Phytologist © 2019 New Phytologist Trust

marine picocyanobacteria. *Microbiology and Molecular Biology Reviews: MMBR* **73**: 249–299.

Shukla MK, Llansola MJ, Martin P, Andrew T, Bruno AP, Sobotka R. 2018. Binding of pigments to the cyanobacterial high-light-inducible protein HliC. *Photosynthesis Research* **137**: 29–39.

Sohm JA, Ahlgren NA, Thomson ZJ, Williams C, Moffett JW, Saito MA, Webb EA, Rocap G. 2015. Co-occurring *Synechococcus* ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. *ISME Journal* **10**: 1–13.

Staleva H, Komenda J, Shukla MK, Kan R, Sobotka R. 2015. Cyanobacterial ancestor of plant antenna proteins. *Nature Chemical Biology* **11**: 287–291.

Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A *et al.* 2015. Structure and function of the global ocean microbiome. *Science* **348**: 1261359–1261359.

Umena Y, Kawakami K, Shen J, Kamiya N. 2011. Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å. *Nature* **473**: 55–60.

Van Mooy BAS, Fredricks HF. 2010. Bacterial and eukaryotic intact polar lipids in the eastern subtropical South Pacific: water-column distribution, planktonic sources, and fatty acid composition. *Geochimica et Cosmochimica Acta* **74**: 6499–6516.

Van Mooy BAS, Fredricks HF, Pedler BE, Dyhrman ST, Karl DM, Koblížek M, Lomas MW, Mincer TJ, Moore LR, Moutin T *et al.* 2009. Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. *Nature* **458**: 69–72.

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 Journal* **10**: 1252–1263.

Wakeham SG, Canuel EA. 1988. Organic geochemistry of particulate matter in the eastern tropical North Pacific Ocean: implications for particle dynamics. *Journal of Marine Research* **46**: 183–213.

Xu W, Wang Y. 2017. Function and structure of cyanobacterial photosystem I. In: Hou H, Najafpour M, Moore G, Allakhverdiev S, eds. *Photosynthesis: structures, mechanisms, and applications*. Cham, Germany: Springer, pp. 111–168.

Zwirglmaier K, Jardillier L, Ostrowski M, Mazard S, Garczarek L, Vaulot D, Not F, Massana R, Ulloa O, Scanlan DJ. 2008. Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environmental Microbiology* **10**: 147–161.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Dataset S1 Desaturase sequences of marine *Synechococcus*/*Cyanobium* used to recruit desaturase reads from *Tara* Oceans metagenomes.

Dataset S2 Desaturase sequences of *Prochlorococcus* used as outgroups to exclude non-*Synechococcus*/*Cyanobium* fatty acid desaturase reads recruited from *Tara* Oceans metagenomes.

Dataset S3 Prokaryotic genomes used as outgroups to exclude non-*Synechococcus*/*Cyanobium* desaturase reads recruited from *Tara* Oceans metagenomes.

Dataset S4 Desaturase sequences used as outgroups to exclude non-*Synechococcus*/*Cyanobium* desaturase reads recruited from *Tara* Oceans metagenomes.

Fig. S1 Growth rate vs temperature modelled curves with the BR and Boatman models.

Fig. S1 Variations of the zeaxanthin : β-cryptoxanthin mass ratio function of growth temperature.

Fig. S2 Variations in acyl chains esterified on the MGDG for the four *Synechococcus* strains grown at different temperatures.

Fig. S3 Variations in acyl chains esterified on the DGDG for the four *Synechococcus* strains grown at different temperatures.

Fig. S4 Variations in acyl chains esterified on the SQDG for the four *Synechococcus* strains grown at different temperatures.

Fig. S5 Variations in acyl chains esterified on the PG for the four *Synechococcus* strains grown at different temperatures.

Fig. S6 Variations in acyl chains esterified on the MGDG for the four *Synechococcus* strains in response to a temperature shift from 22 to 13°C.

Fig. S7 Variations in acyl chains esterified on the DGDG for the four *Synechococcus* strains in response to a temperature shift from 22 to 13°C.

Fig. S8 Variations in acyl chains esterified on the SQDG for the four *Synechococcus* strains in response to a temperature shift from 22 to 13°C.

Fig. S9 Variations in acyl chains esterified on the PG for the four *Synechococcus* strains grown in response to a temperature shift from 22 to 13°C.

Fig. S10 Variations in the acyl chains esterified on the MGDG for the four *Synechococcus* strains in response to a temperature shift from 22 to 30°C.

Fig. S11 Variations in the acyl chains esterified on the DGDG for the four *Synechococcus* strains in response to a temperature shift from 22 to 30°C.

Fig. S12 Variations in the acyl chains esterified on the SQDG for the four *Synechococcus* strains in response to a temperature shift from 22 to 30°C.

Fig. S13 Variations in the acyl chains esterified on the PG for the four *Synechococcus* strains in response to a temperature shift from 22 to 30°C.

Notes S1 Growth rate modeling and pigment analyses.

Notes S2 Membrane lipidomics.

Notes S3 Fatty acid desaturase metagenomics.

Table S1 Parameters calculated with the BR model.

Table S2 Parameters calculated with the Boatman model.

© 2019 The Authors
New Phytologist © 2019 New Phytologist Trust

Table S3 Variations of the acyl chains esterified on the MGDG, DGDG, SQDG and PG of *Synechococcus* sp. MVIR-18-1 grown at different temperatures.

Table S4 Variations of the acyl chains esterified on the MGDG, DGDG, SQDG and PG of *Synechococcus* sp. BL107 grown at different temperatures.

Table S5 Variations of the acyl chains esterified on the MGDG, DGDG, SQDG and PG of *Synechococcus* sp. WH8102 grown at different temperatures.

Table S6 Variations of the acyl chains esterified on the MGDG, DGDG, SQDG and PG of *Synechococcus* sp. A15-62 grown at different temperatures.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

About *New Phytologist*

- *New Phytologist* is an electronic (online-only) journal owned by the New Phytologist Trust, a **not-for-profit organization** dedicated to the promotion of plant science, facilitating projects from symposia to free access for our Tansley reviews and Tansley insights.
- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as ready' via *Early View* – our average time to decision is <26 days. There are **no page or colour charges** and a PDF version will be provided for each article.
- The journal is available online at Wiley Online Library. Visit **www.newphytologist.com** to search the articles and register for table of contents email alerts.
- If you have any questions, do get in touch with Central Office (np-centraloffice@lancaster.ac.uk) or, if it is more convenient, our USA Office (np-usaoffice@lancaster.ac.uk)
- For submission instructions, subscription and all the latest information visit **www.newphytologist.com**

© 2019 The Authors
New Phytologist © 2019 New Phytologist Trust

New Phytologist

Unveiling membrane thermoregulation strategies in marine picocyanobacteria

Solène Breton, Juliette Jouhet, Ulysse Guyet, Valérie Gros, Justine Pittera, David Demory, Frédéric Partensky, Hugo Doré, Morgane Ratin, Eric Maréchal, Ngoc An Nguyen, Laurence Garczarek & Christophe Six.

Acceptance date: 29 September 2019

SUPPORTING INFORMATION

NOTE S1: GROWTH RATE MODELLING AND PIGMENT ANALYSES

S1 A: Modelling growth function of temperature

To estimate the cardinal growth parameters for each strain, we used the cardinal temperature model with inflection (BR model; Bernard and Rémond, 2012). This model helps describing the growth response of acclimated phytoplankton strains to temperature using four parameters: the optimal temperature for growth (T_{OPT}) at which occurs the optimal growth rate (μ_{OPT}), and the minimal and maximal temperatures of growth (T_{MIN} and T_{MAX}) at which $\mu = 0$. The function $\mu(T)$ is described by the following equations:

$$\mu(T) = \begin{cases} 0 \text{ for } T < T_{MIN} \\ \mu_{OPT}\phi(T) \text{ for } T_{MIN} < T < T_{MAX} \\ 0 \text{ for } T > T_{MAX} \end{cases}$$

$$\phi(T) = \frac{(T - T_{MAX})(T - T_{MIN})^{2}}{(T - T_{MIN})[(T_{OPT} - T_{MIN})(T - T_{OPT}) - (T_{OPT} - T_{MAX})(T_{OPT} + T_{MIN} - 2T)]}$$

It order to compare the results of the BR model with another modelling approach, we also used the Boatman model (Boatman *et al.*, 2017) that is based on the following equation:

$$\mu(T) = \mu_{MAX}\left[\sin\left(\pi\frac{T - T_{MIN}}{T_{MAX} - T_{MIN}}\right)^{a}\right]^{b}$$

Table S1: Cardinal parameters calculated with the BR model (Bernard and Rémond, 2012). Values in parentheses are the lower and upper ranges of the confidence interval (95%).

Strain	MVIR-18-1	A15-62	WH8102	BL107
T_{MIN} (°C)	-10.63 (-19.59 to -2.62)	7.14 (-12.54 to 16.64)	11.03 (3.01 to 17.04)	5.83 (-0.35 to 9.89)
T_{OPT} (°C)	26.33 (19.39 to 29.50)	25.80 (24.11 to 30.26)	25.32 (23.76 to 27.56)	23.36 (22.37 to 25.10)
T_{MAX} (°C)	28.29 (26.86 to 29.43)	32.08 (22.97 to 35.43)	33.58 (30.76 to 35.23)	29.18 (27.31 to 30.03)
μ_{OPT} (day^{-1})	0.54 (0.20 to 0.65)	0.51 (0.48 to 0.59)	0.49 (0.47 to 0.53)	0.52 (0.48 to 0.57)

Table S2: Parameter estimations of the Boatman model (Boatman *et al.*, 2017).

Strain	MVIR-18-1	A15-62	WH8102	BL107
μ_{max} (day^{-1})	0.54	0.56	0.48	0.51
T_{min} (°C)	-0.92	9.0487	13.93	10.37
T_{max} (°C)	28.76	30.80	32.20	28.40
a	3.88	2.22	1.44	1.85
b	0.43	0.5	0.5	0.43

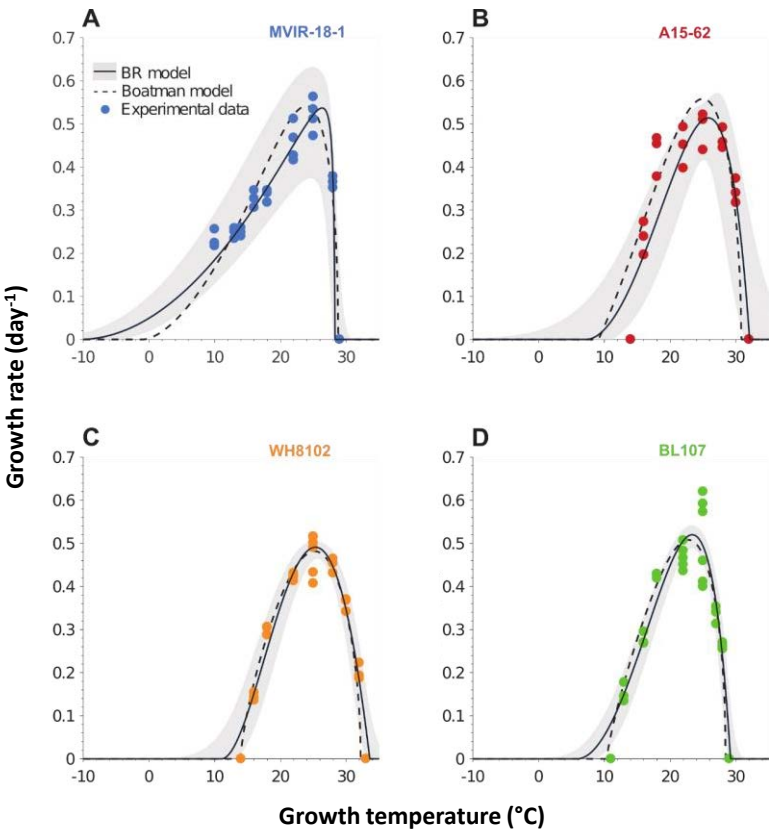


Figure S1: Growth rate *vs.* temperature modelled curves for *Synechococcus* spp. A) MVIR-18-1 (Clade I), B) A15-62 (Clade II), C) WH8102 (Clade III) and D) BL107 (Clade IV), acclimated at temperatures spanning from 10°C to 30°C. The plain, black curve represents the BR model fit with its standard confidence interval (gray shaded area). The dashed curve represents the Boatman model. Dots represent experimentally measured growth rates.

S1 B: Pigment analyses

In the carotenoid pathway, β-cryptoxanthin is synthesized from the β-carotene molecule through mono hydroxylation. Zeaxanthin is then synthesized from β-cryptoxanthin by a second hydroxylation catalyzed by the β-carotene hydroxylase enzyme (CrtR).

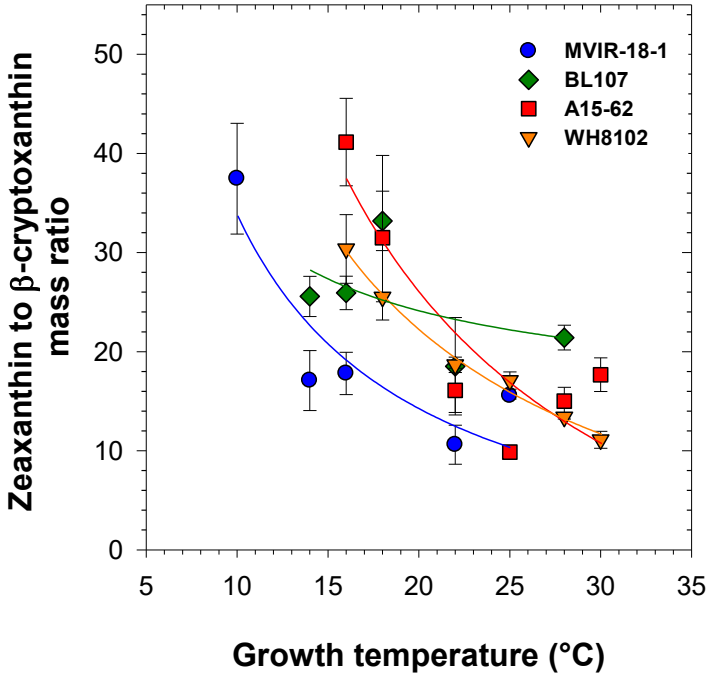


Figure S2: Variations of the zeaxanthin to β-cryptoxanthin mass ratio in *Synechococcus* spp. MVIR-18-1 (blue circles), BL107 (green diamonds), WH8102 (orange triangles) and A15-62 (red squares) acclimated to temperatures ranging from 10°C to 30°C. Each data point is an average of four replicates with error bars ± SD.

NOTE S2: MEMBRANE LIPIDOMICS

S2 A: Cultures acclimated to a range of growth temperatures

Table S3: Relative variations (%) of the acyl chains esterified at the two glycerol positions, *sn*-1 and *sn*-2, of the monogalactosyldiacylglycerol (MGDG), digalacosyldiacylglycerol (DGDG), sulfoquinovosyldiacylglycerol (SQDG) and phosphatidylglycerol (PG), in ***Synechococcus* sp. MVIR-18-1** acclimated to temperature ranging from 10 to 25°C. Each data point is an average of three to four replicates (mean ± SD).

			Growth temperature (°C)							
			10	**14**	**16**	**18**	**22**	**25**	**28**	**30**
MGDG	***sn*-1**	**14:0**	1.9 ± 0.0	1.4 ± 0.0	1.3 ± 0.0	8.2 ± 7.6	4.15 ± 2.1	4.6 ± 2.1		
		14:1	25.9 ± 0.6	10.9 ± 0.1	10.5 ± 0.5	3.91 ± 2.6	4.9 ± 0.1	2.1 ± 1.7		
		16:0	5.2 ± 0.0	9.4 ± 0.4	9.4 ± 0.1	16.8 ± 5.5	15.4 ± 0.4	20.6 ± 3.6		
		16:1	65.4 ± 0.6	77.9 ± 0.4	78.6 ± 0.6	71.1 ± 10.7	75.5 ± 2.2	72.7 ± 4.1		
		16:2	1.3 ± 0.0	0.2 ± 0.0	0.3 ± 0.03	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0		
	***sn*-2**	**14:0**	67.9 ± 0.0	91.5 ± 0.3	88.7 ± 0.2	97.7 ± 1.5	96.4 ± 0.3	98.2 ± 2.8		
		14:1	32.1 ± 0.0	8.5 ± 0.3	11.3 ± 0.2	2.3 ± 1.5	3.6 ± 0.3	1.8 ± 2.8		
DGDG	***sn*-1**	**14:0**	3.1 ± 0.1	2.4 ± 0.1	6.0 ± 0.4	5.8 ± 1.2	6.4 ± 1.4	7.5 ± 2.2		
		14:1	18.7 ± 0.9	10.2 ± 0.1	3.3 ± 0.3	3.8 ± 1.6	2.4 ± 1.6	0.6 ± 1.2		
		16:0	11.4 ± 0.5	11.5 ± 0.5	23.2 ± 0.5	25.2 ± 3.4	33.1 ± 4.1	47.8 ± 10.1		
		16:1	66.0 ± 0.3	75.8 ± 0.5	66.9 ± 1.3	65.1 ± 3.3	58.0 ± 4.5	44.0 ± 11.2		
		16:2	0.6 ± 0.0	0.1 ± 0.1	0.4 ± 0.3	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0		
	***sn*-2**	**14:0**	72.4 ± 0.4	89.2 ± 0.1	91.2 ± 0.3	99.6 ± 0.4	98.9 ± 1.6	99.7 ± 0.6		
		14:1	27.5 ± 0.4	10.7 ± 0.1	8.7 ± 0.3	0.3 ± 0.4	1.0 ± 1.7	0.3 ± 0.6		
SQDG	***sn*-1**	**14:0**	1.1 ± 0.1	2.8 ± 0.1	2.5 ± 0.1	2.9 ± 0.3	3.0 ± 0.3	2.1 ± 0.3		
		14:1	1.9 ± 0.2	1.1 ± 0.1	0.9 ± 0.1	0.3 ± 0.1	0.3 ± 0.0	0.04 ± 0.1		
		16:0	22.5 ± 0.2	29.6 ± 0.3	29.5 ± 0.2	47.6 ± 1.7	48.1 ± 3.2	64.8 ± 5.2		
		16:1	73.6 ± 0.4	65.2 ± 0.1	65.4 ± 0.6	34.6 ± 6.2	40.9 ± 6.7	23.1 ± 10.4		
		18:1	0.5 ± 0.1	0.6 ± 0.1	0.8 ± 0.0	1.7 ± 0.3	1.7 ± 0.1	1.4 ± 0.2		
		18:2	0.3 ± 0.1	0.6 ± 0.1	0.8 ± 0.3	12.7 ± 4.1	5.8 ± 3.1	8.3 ± 5.4		
	***sn*-2**	**14:0**	66.3 ± 0.4	54.6 ± 0.74	65.7 ± 0.8	54.3 ± 1.71	53.4 ± 0.8	53.5 ± 0.4		
		14:1	8.01 ± 0.3	2.6 ± 0.2	1.5 ± 0.0	0.4 ± 0.1	0.5 ± 0.1	0.1 ± 0.13		
		16:0	17.8 ± 0.2	27.1 ± 0.4	21.7 ± 0.5	33.2 ± 0.5	36.0 ± 0.3	39.0 ± 1.1		
		16:1	7.8 ± 0.1	15.7 ± 0.2	11.1 ± 0.4	12.2 ± 1.4	10.0 ± 0.7	7.4 ± 1.6		
PG	***sn*-1**	**16:0**	1.2 ± 0.9	1.2 ± 0.8	1.1 ± 0.6	1.1 ± 0.6	1.7 ± 0.0	2.7 ± 2.4		
		16:1	69.8 ± 17.5	68.1 ± 18.2	65.7 ± 21.8	71.9 ± 15.8	73.9 ± 0.0	76.0 ± 7.1		
		16:2	1.0 ± 1.0	0.9 ± 0.9	0.8 ± 1.0	0.9 ± 0.9	0.3 ± 0.0	0.4 ± 0.6		
		18:0	2.2 ± 2.2	2.1 ± 1.9	2.3 ± 1.8	1.8 ± 1.5	2.4 ± 0.0	2.1 ± 1.2		
		18:1	24.0 ± 15.6	25.8 ± 16.5	27.9 ± 19.4	22.5 ± 14.1	21.4 ± 0.0	17.7 ± 8.2		
		18:2	1.5 ± 1.3	1.7 ± 1.2	2.0 ± 1.9	1.6 ± 1.3	0.1 ± 0.0	1.0 ± 1.4		
	***sn*-2**	**16:0**	12.9 ± 0.2	17.9 ± 0.1	22.3 ± 1.9	31.9 ± 4.7	35.6 ± 3.9	50.1 ± 16.2		
		16:1	80.7 ± 0.7	67.3 ± 1.5	56.3 ± 4.6	63.4 ± 4.6	52.6 ± 6.4	46.6 ± 13.3		
		18:0	1.6 ± 0.2	4.6 ± 0.3	6.4 ± 0.7	1.4 ± 1.2	5.9 ± 1.9	1.6 ± 1.9		
		18:1	4.8 ± 0.4	10.2 ± 1.2	14.9 ± 2.3	3.2 ± 1.3	5.8 ± 0.9	1.6 ± 1.9		

Table S4: Relative variations (%) of the acyl chains esterified at the two glycerol positions, *sn*-1 and *sn*-2, of the monogalactosyldiacylglycerol (MGDG), digalacosyldiacylglycerol (DGDG), sulfoquinovosyldiacylglycerol (SQDG) and phosphatidylglycerol, in ***Synechococcus* sp. BL107** acclimated to temperatures ranging from 14 to 28°C. Each data point is an average of three to four replicates (mean ± SD).

			Growth temperature (°C)							
			10	**14**	**16**	**18**	**22**	**25**	**28**	**30**
MGDG	***sn*-1**	**14:0**		1.5 ± 0.1	1.5 ± 0.2	3.0 ± 0.6	2.8 ± 0.1	7.2 ± 0.3	3.7 ± 0.3	
		14:1		9.2 ± 0.1	8.5 ± 0.1	7.2 ± 0.2	9.4 ± 1.2	7.8 ± 0.3	8.6 ± 0.3	
		16:0		5.6 ± 0.1	6.0 ± 0.1	8.2 ± 0.2	8.5 ± 1.2	1.6 ± 0.4	9.2 ± 0.2	
		16:1		83.3 ± 0.3	83.7 ± 0.4	80.1 ± 0.8	79.1 ± 0.3	70.9 ± 0.5	78.2 ± 0.6	
		16:2		0.4 ± 0.0	0.3 ± 0.0	1.5 ± 0.0	0.2 ± 0.0	2.4 ± 0.5	0.3 ± 0.2	
	***sn*-2**	**14:0**		71.3 ± 0.2	74.3 ± 0.7	83.0 ± 0.6	88.8 ± 1.5	92.9 ± 0.8	94.6 ± 0.1	
		14:1		28.7 ± 0.2	25.7 ± 0.7	17.0 ± 0.6	11.2 ± 1.5	7.1 ± 0.8	5.4 ± 0.1	
DGDG	***sn*-1**	**14:0**		2.3 ± 0.2	2.1 ± 0.1	3.9 ± 0.3	3.9 ± 0.3	8.2 ± 0.3	4.7 ± 0.2	
		14:1		8.4 ± 0.2	8.4 ± 0.1	6.1 ± 0.2	7.4 ± 1.4	6.2 ± 0.4	8.6 ± 0.3	
		16:0		19.1 ± 1.2	17.8 ± 0.1	37.1 ± 1.3	30.9 ± 2.0	45.6 ± 1.5	23.7 ± 1.6	
		16:1		69.9 ± 1.3	71.5 ± 0.1	52.6 ± 1.1	57.5 ± 0.9	39.9 ± 1.4	62.7 ± 1.6	
		16:2		0.2 ± 0.0	0.1 ± 0.03	0.2 ± 0.0	0.2 ± 0.0	0.1 ± 0.1	0.1 ± 0.0	
	***sn*-2**	**14:0**		78.9 ± 0.2	81.7 ± 0.2	91.7 ± 0.8	91.9 ± 0.5	97.7 ± 0.3	94.6 ± 0.5	
		14:1		20.1 ± 0.2	18.3 ± 0.2	8.3 ± 0.8	2.3 ± 0.5	2.3 ± 0.3	5.4 ± 0.5	
SQDG	***sn*-1**	**14:0**		0.8 ± 0.2	1.1 ± 0.1	1.3 ± 0.1	1.5 ± 0.3	1.8 ± 0.3	1.6 ± 0.2	
		14:1		0.4 ± 0.0	0.4 ± 0.0	0.4 ± 0.1	0.4 ± 0.1	0.3 ± 0.1	0.4 ± 0.1	
		16:0		22.9 ± 0.6	24.8 ± 0.7	26.7 ± 0.6	32.8 ± 1.5	32.5 ± 0.5	40.3 ± 1.6	
		16:1		74.8 ± 0.8	72.5 ± 0.9	68.0 ± 0.8	63.4 ± 1.4	56.2 ± 2.5	56.1 ± 1.5	
		18:1		0.7 ± 0.0	0.6 ± 0.0	0.9 ± 0.0	1.1 ± 0.3	1.4 ± 0.1	0.9 ± 0.0	
		18:2		0.6 ± 0.1	0.6 ± 0.1	2.6 ± 0.2	0.7 ± 0.3	7.6 ± 2.2	0.6 ± 0.2	
	***sn*-2**	**14:0**		50.9 ± 0.8	51.8 ± 0.6	53.1 ± 1.1	53.7 ± 3.1	56.0 ± 1.3	60.8 ± 0.8	
		14:1		5.7 ± 0.1	5.4 ± 0.3	3.9 ± 0.2	2.4 ± 0.1	1.4 ± 0.1	1.5 ± 0.1	
		16:0		27.6 ± 0.4	28.7 ± 0.8	30.6 ± 1.0	33.6 ± 0.9	33.3 ± 0.6	32.6 ± 1.1	
		16:1		15.8 ± 0.5	14.0 ± 0.2	12.44 ± 0.5	10.3 ± 2.1	9.2 ± 0.9	5.6 ± 0.3	
PG	***sn*-1**	**16:0**		1.1 ± 0.7	1.0 ± 0.5	1.0 ± 0.6	1.1 ± 0.4	1.9 ± 0.6	1.0 ± 1.0	
		16:1		84.6 ± 5.3	84.5 ± 5.9	86.1 ± 4.4	84.2 ± 2.1	75.6 ± 1.7	79.6 ± 4.3	
		16:2		1.0 ± 0.1	1.0 ± 0.1	0.8 ± 0.2	0.9 ± 0.3	0.5 ± 0.1	0.9 ± 0.1	
		18:0		1.0 ± 0.8	0.9 ± 0.8	0.8 ± 0.5	0.9 ± 0.5	1.1 ± 0.1	1.0 ± 0.2	
		18:1		11.6 ± 3.9	11.61 ± 4.4	10.6 ± 3.5	12.1 ± 3.0	19.3 ± 1.1	16.5 ± 2.7	
		18:2		0.7 ± 0.2	0.0 ± 0.1	0.7 ± 0.2	0.8 ± 0.4	1.5 ± 0.2	1.0 ± 0.5	
	***sn*-2**	**16:0**		15.4 ± 0.2	20.2 ± 0.9	16.9 ± 0.4	20.3 ± 1.9	27.7 ± 1.6	27.3 ± 2.3	
		16:1		80.6 ± 0.9	73.2 ± 1.0	79.9 ± 0.7	73.4 ± 2.4	69.4 ± 1.9	67.6 ± 2.5	
		18:0		0.6 ± 0.1	1.1 ± 0.0	0.5 ± 0.1	1.2 ± 0.09	0.8 ± 0.0	1.1 ± 0.1	
		18:1		3.4 ± 0.3	5.5 ± 0.2	2.7 ± 0.5	5.1 ± 0.5	2.1 ± 0.3	3.9 ± 0.2	

Table S5: Relative variations (%) of the acyl chains esterified at the two glycerol positions, *sn*-1 and *sn*-2, of the monogalactosyldiacylglycerol (MGDG), digalacosyldiacylglycerol (DGDG), sulfoquinovosyldiacylglycerol (SQDG) and phosphatidylglycerol, in ***Synechococcus* sp. WH8102** acclimated to temperatures ranging from 16 to 30°C. Each data point is an average of three to four replicates (mean ± SD).

			Growth temperature (°C)							
			10	**14**	**16**	**18**	**22**	**25**	**28**	**30**
MGDG	***sn*-1**	**14:0**			0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	1.3 ± 0.9
		14:1			0.3 ± 0.7	3.2 ± 0.2	7.0 ± 0.6	1.5 ± 2.1	5.5 ± 0.1	12.9 ± 0.2
		16:0			6.9 ± 0.5	10.8 ± 0.3	14.1 ± 0.4	15.6 ± 0.3	16.0 ± 0.2	13.42 ± 0.2
		16:1			58.8 ± 0.7	67.3 ± 0.6	74.8 ± 0.3	77.9 ± 2.4	78.5 ± 0.2	70.9 ± 1.0
		16:2			33.9 ± 1.8	18.6 ± 0.5	4.1 ± 0.1	5.6 ± 1.7	0.0 ± 0.0	1.4 ± 1.1
	***sn*-2**	**14:0**			36.0 ± 2.2	56.8 ± 0.8	78.6 ± 0.6	82.6 ± 1.6	89.3 ± 0.2	81.4 ± 1. 6
		14:1			64.0 ± 2.2	43.2 ± 0.8	21.4 ± 0.6	17.4 ± 1.7	10.7 ± 0.2	18.6 ± 1.6
DGDG	***sn*-1**	**14:0**			0.0 ± 0.0	0.4 ± 0.6	3.1 ± 0.2	3.0 ± 0.3	3.1 ± 0.1	6.6 ± 0.3
		14:1			0.0 ± 0.0	0.8 ± 1.0	3.5 ± 0.1	2.5 ± 0.4	4.1 ± 0.1	4.8 ± 0.6
		16:0			33.5 ± 0.1	38.6 ± 1.7	34.7 ± 0.8	44.0 ± 1.4	32.2 ± 0.2	35.5 ± 3.3
		16:1			57.1 ± 0.2	55.2 ± 0.5	56.7 ± 0.7	50.6 ± 1.2	60.6 ± 0.1	53.1 ± 3.0
		16:2			9.4 ± 0.3	5.0 ± 0.5	1.9 ± 0.1	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	***sn*-2**	**14:0**			69.5 ± 0.3	78.9 ± 1.0	84.6 ± 0.6	92.4 ± 0.5	90.8 ± 0.1	91.0 ± 0.8
		14:1			30.4 ± 0.3	21.1 ± 1.0	15.4 ± 0.6	9.29 ± 0.5	9.2 ± 0.1	8.9 ± 0.8
SQDG	***sn*-1**	**14:0**			0.3 ± 0.0	0.5 ± 0.0	1.0 ± 0.1	0.8 ± 0.0	1.1 ± 0.2	3.2 ± 0.4
		14:1			1.0 ± 0.0	0.1 ± 0.0	0.3 ± 0.1	0.2 ± 0.0	0.2 ± 0.0	0.7 ± 0.1
		16:0			35.5 ± 0.8	43.6 ± 1.3	47.9 ± 2.2	55.7 ± 0.8	56.6 ± 0.3	54.1 ± 0.7
		16:1			54.6 ± 0.2	52.6 ± 1.46	49.8 ± 2.1	41.7 ± 0.8	41.2 ± 0.2	41.1 ± 0.6
		18:1			3.3 ± 0.2	1.5 ± 0.1	0.5 ± 0.1	0.7 ± 0.0	0.6 ± 0.0	0.6 ± 0.1
		18:2			6.0 ± 0.6	1.6 ± 0.0	0.4 ± 0.05	0.7 ± 0.1	0.3 ± 0.0	0.3 ± 0.06
	***sn*-2**	**14:0**			29.3 ± 1.7	45.1 ± 0.3	63.0 ± 1.9	57.4 ± 0.9	65.6 ± 0.9	65.8 ± 1.5
		14:1			7.± 0.10	5.72 ± 0.4	2.5 ± 0.1	0.8 ± 0.1	0.7 ± 0.0	1. 6 ± 0.1
		16:0			45.7 ± 0.8	40.3 ± 1.0	30.8 ± 1.7	38.8 ± 0.9	31.2 ± 0.8	19.3 ± 1.2
		16:1			17.3 ± 1.0	8.8 ± 0.5	3.7 ± 0.2	2.9 ± 0.1	2.5 ± 0.1	3.3 ± 0.3
PG	***sn*-1**	**16:0**			0.2 ± 0.2	0.1 ± 0.2	0.6 ± 0.5	0.6 ± 0.0	1.3 ± 0.1	1.7 ± 0.1
		16:1			86.1 ± 1.2	95.2 ± 0.1	96.2 ± 1.0	95.9 ± 0.6	94.9 ± 0.4	92.8 ± 0.4
		16:2			9.58 ± 1.0	4.7 ± 0.3	3.2 ± 0.6	4.1 ± 0.6	3.6 ± 0.1	5.1 ± 0.2
		18:1			4.1 ± 1.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.2	0.3 ± 0.2
	***sn*-2**	**16:0**			22.1 ± 0.3	22.1 ± 0.3	21.7 ± 0.4	20.0 ± 0.2	21.4 ± 0.3	22.1 ± 0.2
		16:1			77.9 ± 1.1	77.9 ± 0.3	78.3 ± 0.4	80.0 ± 0.2	78.6 ± 0.3	78.0 ± 0.2

Table S6: Relative variations (%) of the acyl chains esterified at the two glycerol positions, *sn*-1 and *sn*-2, of the monogalactosyldiacylglycerol (MGDG), digalacosyldiacylglycerol (DGDG), sulfoquinovosyldiacylglycerol (SQDG) and phosphatidylglycerol, in ***Synechococcus* sp. A15-62** acclimated to temperatures ranging from 16 to 30°C. Each data point is an average of three to four replicates (mean ± SD).

			Growth temperature (°C)							
			10	**14**	**16**	**18**	**22**	**25**	**28**	**30**
MGDG	***sn*-1**	**14:0**			2.8 ± 0.1	2.6 ± 0.1	1.8 ± 0.1	1.7 ± 0.0	1.5 ± 0.1	1.6 ± 0.1
		14:1			30.4 ± 1.1	22.8 ± 1.9	23.3 ± 0.1	21.5 ± 0.3	11.3 ± 0.8	8.2 ± 02
		16:0			4.3 ± 0.2	5.3 ± 0.4	4.1 ± 0.1	3.2 ± 0.1	5.4 ± 0.2	7.1 ± 0.1
		16:1			56.1 ± 0.9	65.0 ± 2.0	62.7 ± 0.2	59.0 ± 0.2	74.3 ± 1.2	78.9 ± 0.4
		16:2			6.4 ± 0.1	4.4 ± 0.6	8.1 ± 0.3	14.4 ± 0.2	7.5 ± 0.6	4.1 ± 0.1
	***sn*-2**	**14:0**			72.9 ± 0.9	77.7 ± 1.7	68.7 ± 0.7	55.5 ± 0.4	69.9 ± 1.5	78.0 ± 0.3
		14:1			27.1 ± 0.9	22.3 ± 1.7	31.3 ± 0.7	44.5 ± 0.4	30.1 ± 1.5	22.0 ± 0.3
DGDG	***sn*-1**	**14:0**			6.1 ± 0.3	5.1 ± 0.1	4.9 ± 0.1	5.0 ± 0.0	4.9 ± 0.2	5.3 ± 0.2
		14:1			17.4 ± 0.9	13.2 ± 1.2	10.1 ± 0.3	7.9 ± 0.2	4.6 ± 0.3	3.6 ± 0.2
		16:0			24.4 ± 1.4	30.6 ± 0.9	36.9 ± 0.7	37.3 ± 0.4	47.9 ± 1.6	54.1 ± 0.9
		16:1			49.4 ± 0.7	48.9 ± 0.6	45.0 ± 0.5	44.8 ± 0.2	40.4 ± 1.2	36.0 ± 0.5
		16:2			2.6 ± 0.2	2.0 ± 0.2	2.9 ± 0.0	5.0 ± 0.1	2.06 ± 0.1	0.89 ± 0.1
	***sn*-2**	**14:0**			80.7 ± 1.1	85.3 ± 1.0	83.5 ± 0.3	78.1 ± 0.3	87.3 ± 0.9	92.1 ± 0.4
		14:1			19.4 ± 1.1	14.7 ± 1.0	16.5 ± 0.3	21.9 ± 0.3	12.7 ± 0.9	7.9 ± 0.4
SQDG	***sn*-1**	**14:0**			3.3 ± 0.2	2.6 ± 0.2	2.7 ± 0.3	3.1 ± 0.1	2.7 ± 0.2	2.4 ± 0.1
		14:1			2.3 ± 0.2	1.4 ± 0.1	1.0 ± 0.0	0.8 ± 0.1	0.5 ± 0.0	0.3 ± 0.0
		16:0			34.3 ± 0.4	35.7 ± 0.9	37.0 ± 0.3	38.4 ± 0.6	49.0 ± 1.0	58.6 ± 0.4
		16:1			58.3 ± 0.3	59.0 ± 1.1	57.0 ± 0.7	55.2 ± 0.7	45.6 ± 0.9	37.5 ± 0.5
		18:1			1.0 ± 0.1	0.6 ± 0.1	1.2 ± 0.3	1.1 ± 0.1	1.1 ± 0.0	0.6 ± 0.0
		18:2			0.7 ± 0.0	0.5 ± 0.0	1.0 ± 0.1	1.3 ± 0.1	1.0 ± 0.1	0.5 ± 0.0
	***sn*-2**	**14:0**			50.7 ± 1.0	48.7 ± 1.6	39.6 ± 0.5	37.0 ± 0.8	50.2 ± 1.2	61.9 ± 0.5
		14:1			6.5 ± 0.3	5.7 ± 0.3	5.0 ± 0.2	8.8 ± 0.3	4.8 ± 0.2	2.8 ± 0.1
		16:0			33.4 ± 1.2	36.6 ± 1.7	44.0 ± 0.8	39.6 ± 0.9	36.7 ± 0.7	31.3 ± 0.7
		16:1			9.4 ± 0.1	9.0 ± 0.1	11.3 ± 0.2	14.6 ± 0.3	8.2 ± 0.6	4.0 ± 0.1
PG	***sn*-1**	**16:0**			0.8 ± 0.3	0.9 ± 0.1	0.8 ± 0.1	1.0 ± 0.2	1.6 ± 0.5	1.0 ± 0.2
		16:1			74.4 ± 12.6	74.1 ± 7.1	75.8 ± 6.0	75.9 ± 4.9	74.3 ± 4.8	73.7 ± 2.8
		16:2			8.1 ± 5.6	7.5 ± 4.5	7.2 ± 2.4	7.1 ± 0.7	7.8 ± 1.7	8.0 ± 1.6
		18:0			1.1 ± 0.8	1.1 ± 0.5	1.0 ± 0.5	1.1 ± 0.0	0.7 ± 0.1	1.1 ± 0.1
		18:1			14.2 ± 15.2	15.1 ± 9.0	13.9 ± 6.9	13.8 ± 3.6	14.4 ± 4.7	15.1 ± 3.8
		18:2			1.3 ± 1.8	1.4 ± 1.2	1.2 ± 1.0	1.1 ± 0.4	1.2 ± 0.7	1.1 ± 0.7
	***sn*-2**	**16:0**			14.9 ± 0.5	13.2 ± 0.6	15.6 ± 0.3	16.7 ± 0.7	25.0 ± 3.4	26.5 ± 0.7
		16:1			65.1 ± 3.8	73.9 ± 1.0	78.6 ± 0.2	78.0 ± 1.1	65.9 ± 4.0	65.3 ± 0.9
		18:0			3.2 ± 0.6	2.1 ± 0.2	2.1 ± 0.1	1.2 ± 0.1	1.9 ± 0.2	1.8 ± 0.1
		18:1			16.7 ± 2.8	10.8 ± 1.2	4.7 ± 0.4	4.2 ± 0.4	7.1 ± 0.7	6.4 ± 0.4

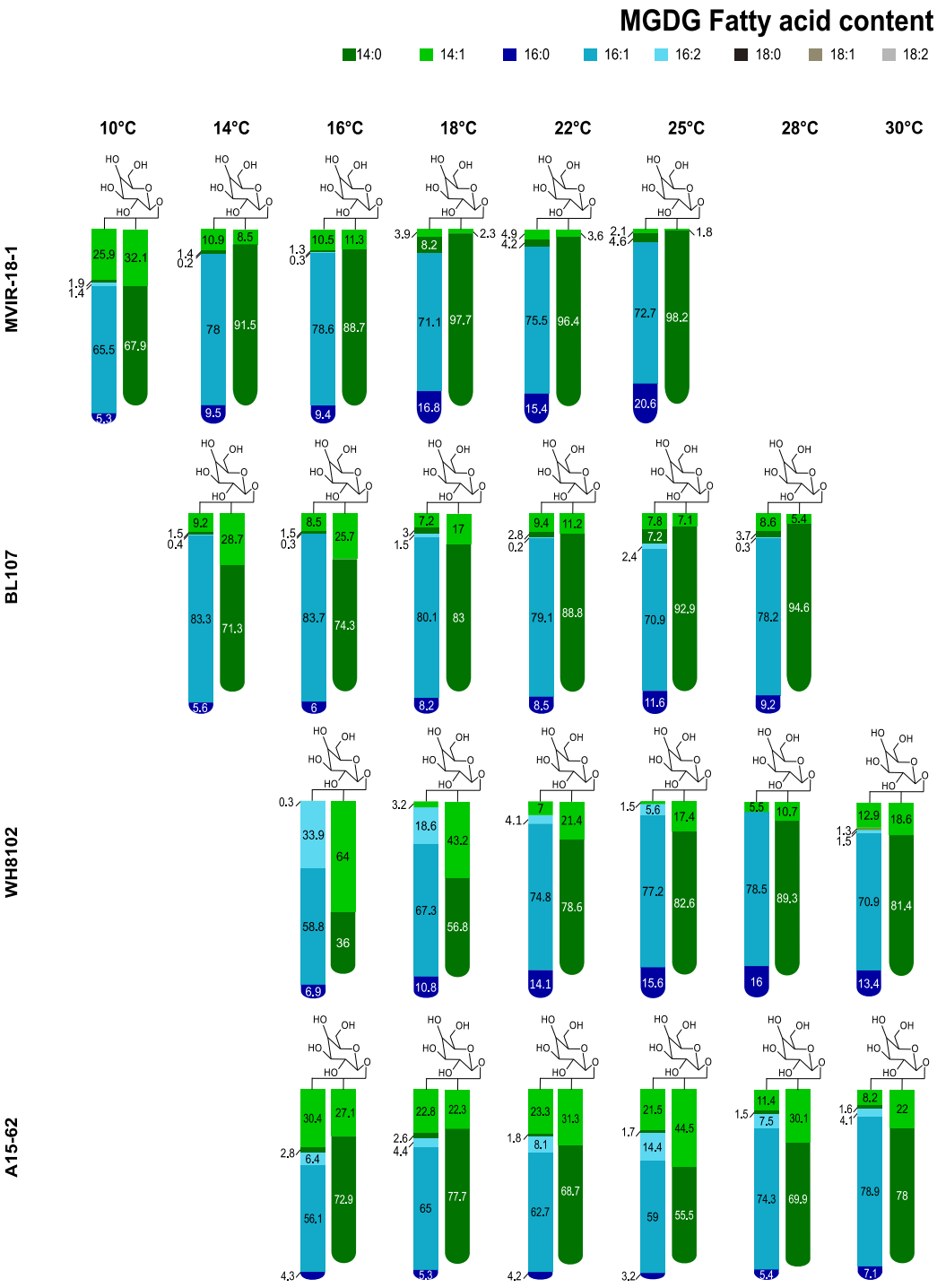


Figure S3: Variations of the relative content (%) in acyl chains esterified at the two glycerol positions, *sn*-1 and *sn*-2, of the **monogalactosyldiacylglycerol (MGDG)** for *Synechococcus* spp. MVIR-18-1, BL107, WH8102, and A15-62, acclimated to a range of temperatures. Each value is an average of three replicates.

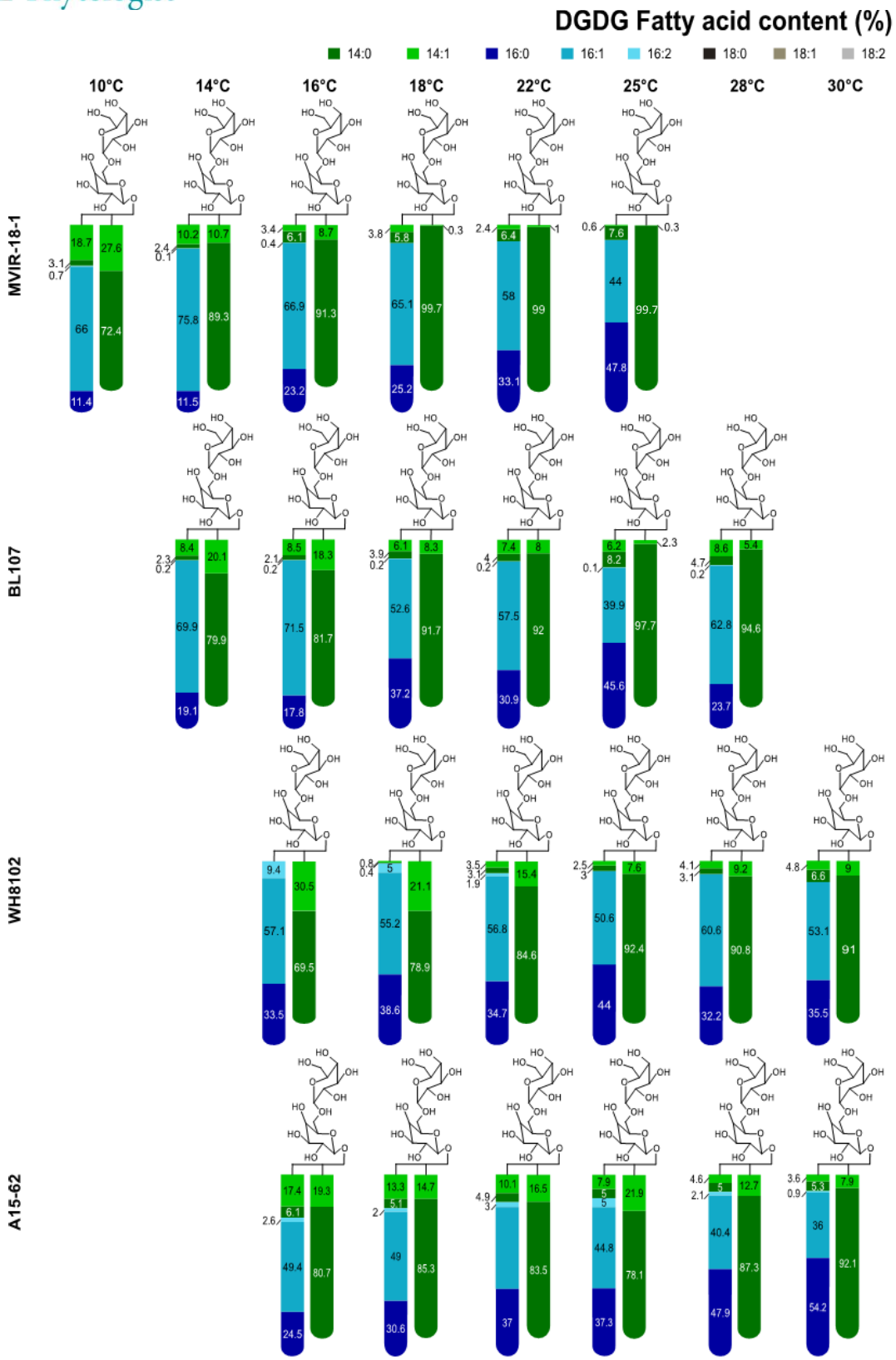


Figure S4: Variations of the relative content (%) in acyl chains esterified at the two glycerol positions, *sn*-1 and *sn*-2, of the **digalactosyldiacylglycerol (DGDG)** for *Synechococcus* spp. MVIR-18-1, BL107, WH8102, and A15-62, acclimated to a range of temperatures. Each value is an average of three to four replicates.

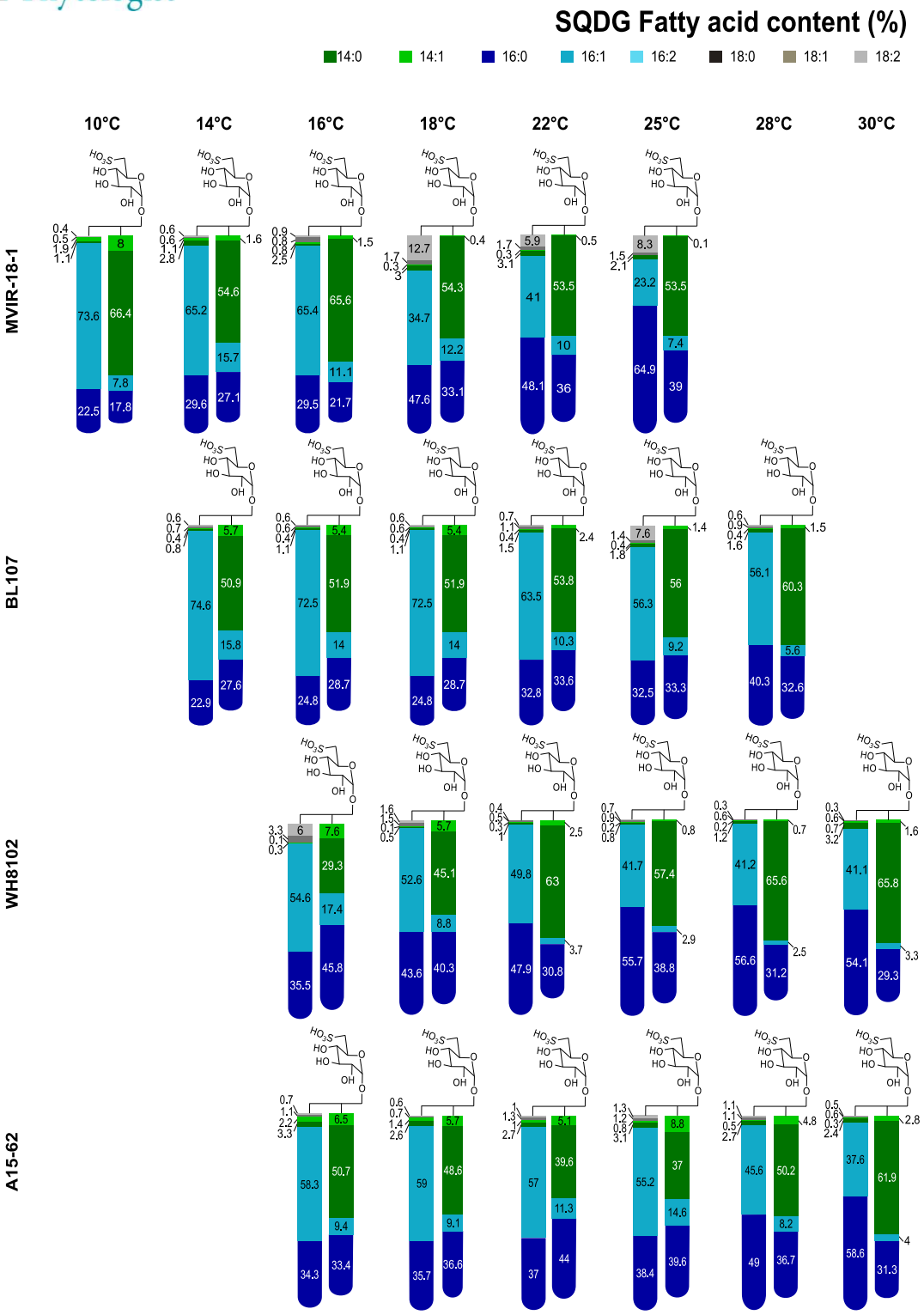


Figure S5: Variations of the relative content (%) in acyl chains esterified at the two glycerol positions, *sn*-1 and *sn*-2, of the **sulfoquinovosyldiacylglycerol (SQDG)** for *Synechococcus* spp. MVIR-18-1, BL107, WH8102, and A15-62, acclimated to a range of temperatures. Each value is an average of three to four replicates.

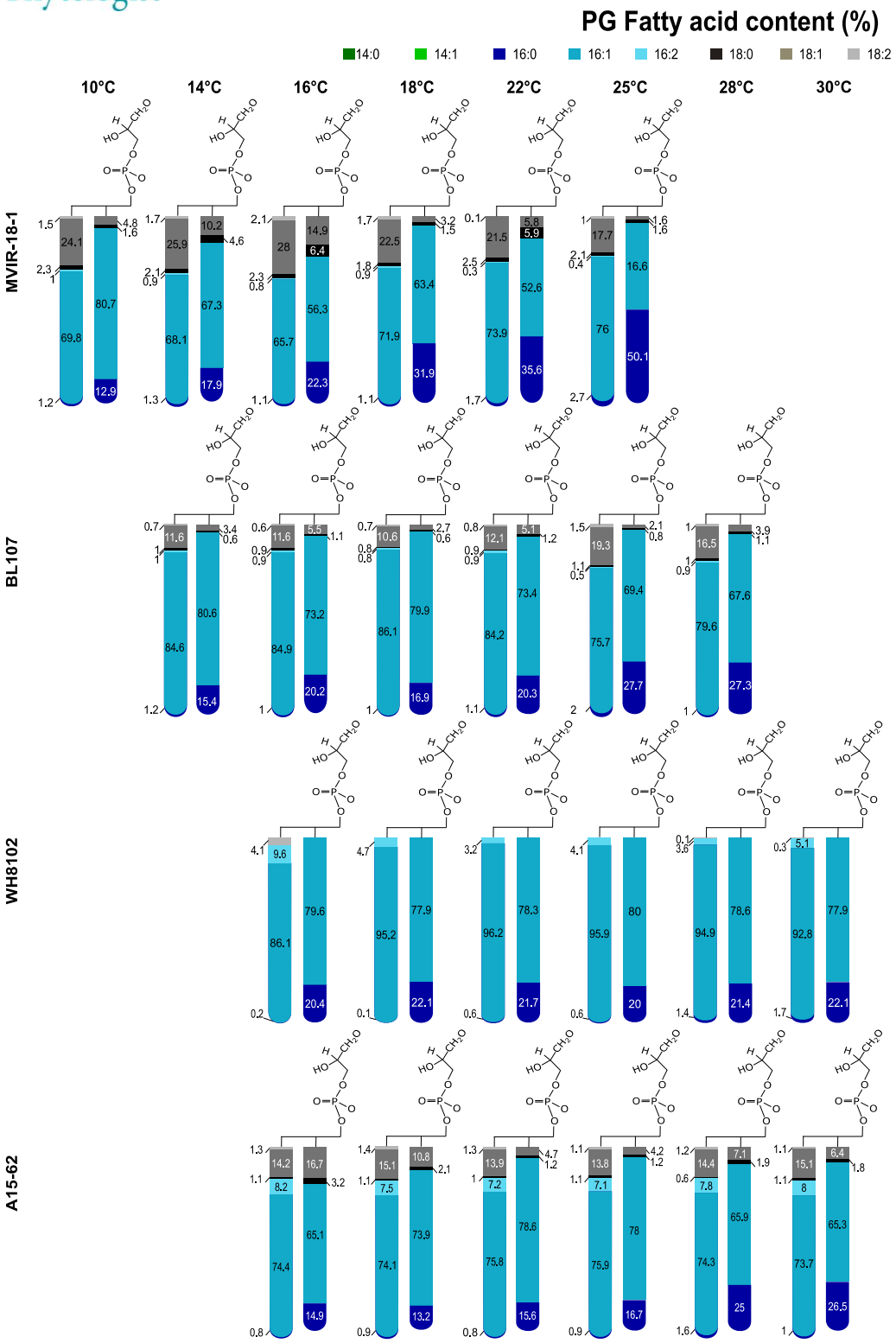


Figure S6: Comparison of the relative variations (%) in the acyl chains esterified at the two glycerol positions of the **phosphatidylglycerol (PG)** for the four marine *Synechococcus* strains MVIR-18-1, BL107, WH8102, and A15-62 acclimated to a range of temperatures (see also Supporting Information Table S2). Each value is an average of four replicates.

II.B Cultures acclimated to 22°C and shifted to 13°C

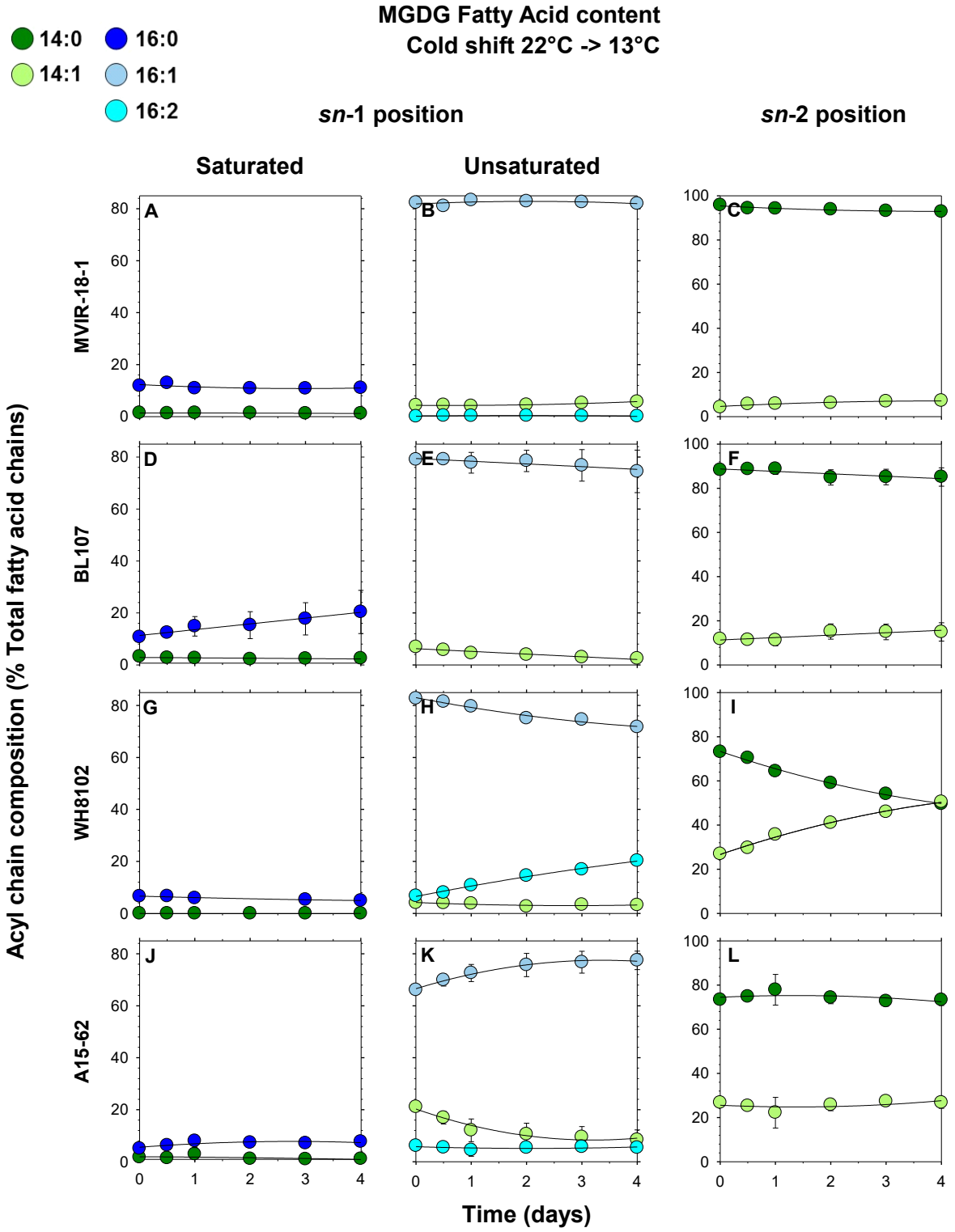


Figure S7: Relative variations (%) in the acyl chains esterified at the two glycerol positions, *sn*-1 (left panels) and *sn*-2 (right panel), of the **monogalactosyldiacylglycerol (MGDG)** for *Synechococcus* spp. MVIR-18-1 (A, B, C), BL107 (D, E, F), WH8102 (G, H, I) and A15-62 (J, K, L), in response to a temperature shift from 22°C to 13°C. The results are expressed as percentages of total acyl chains esterified at the specified stereospecific position of the glycerolipid. Each value is an average of three replicates with error bars ± SD.

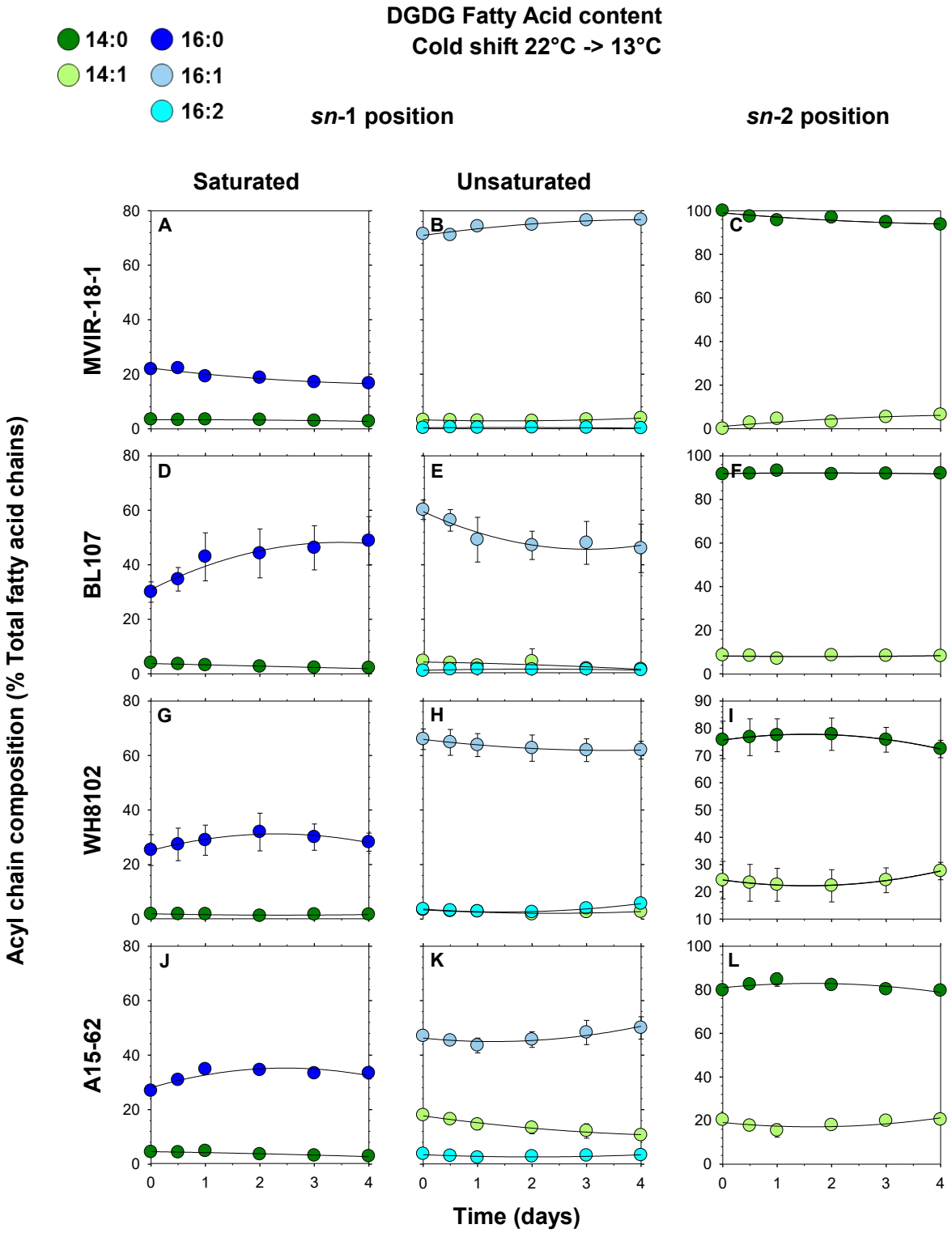


Figure S8: Relative variations (%) in the acyl chains esterified at the two glycerol positions, *sn*-1 (left panels) and *sn*-2 (right panel), of the **digalactosyldiacylglycerol (DGDG)** for *Synechococcus* spp. MVIR-18-1 (A, B, C), BL107 (D, E, F), WH8102 (G, H, I) and A15-62 (J, K, L), in response to a temperature shift from 22°C to 13°C. The results are expressed as percentages of total acyl chains esterified at the specified stereospecific position of the glycerolipid. Each value is an average of three replicates with error bars ± SD.

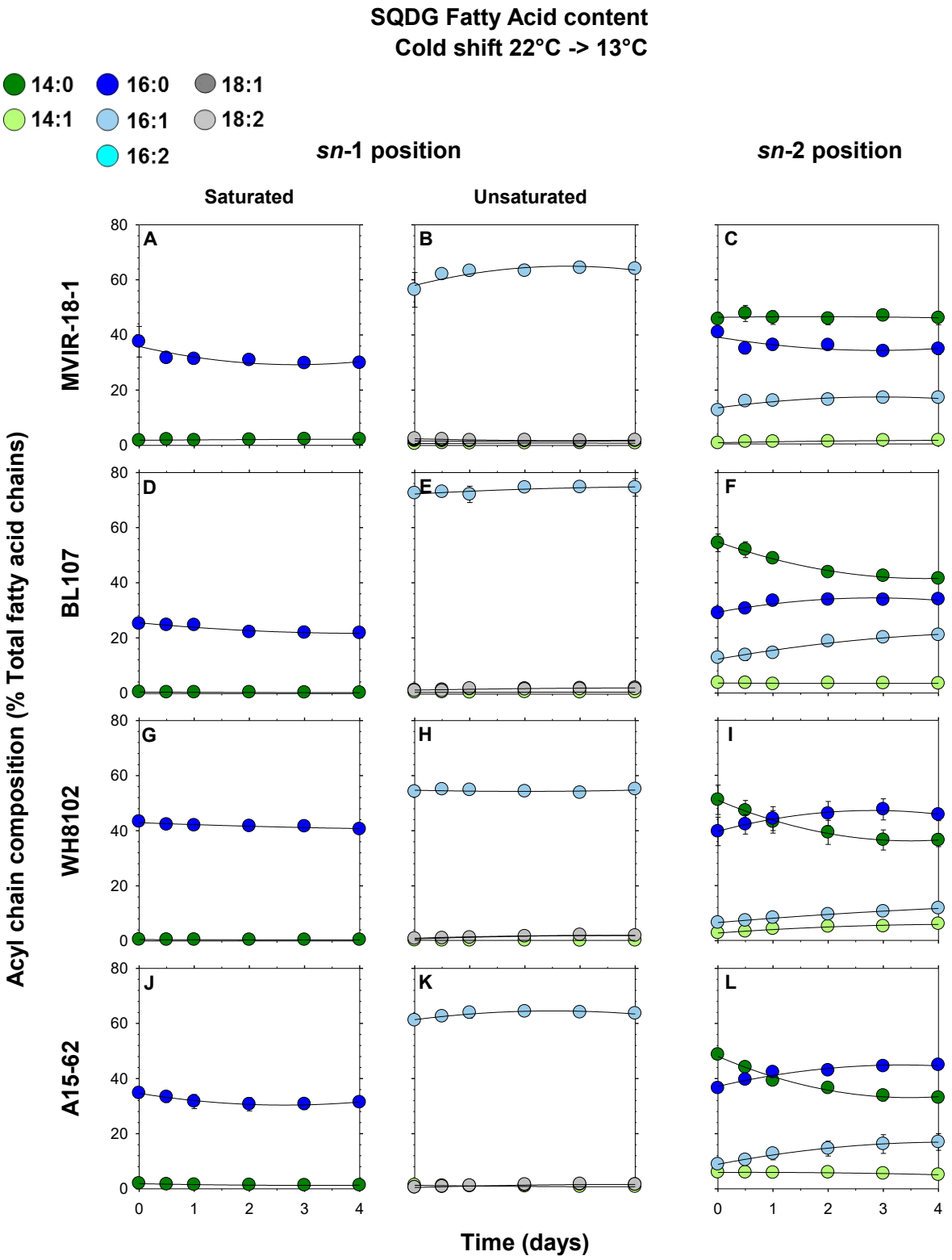


Figure S9: Relative variations (%) in the acyl chains esterified at the two glycerol positions, *sn*-1 (left panels) and *sn*-2 (right panel), of the **sulfoquinovosyldiacylglycerol (SQDG)** for *Synechococcus* spp. MVIR-18-1 (A, B, C), BL107 (D, E, F), WH8102 (G, H, I) and A15-62 (J, K, L), in response to a temperature shift from 22°C to 13°C. The results are expressed as percentages of total acyl chains esterified at the specified stereospecific position of the glycerolipid. Each value is an average of three replicates with error bars ± SD.

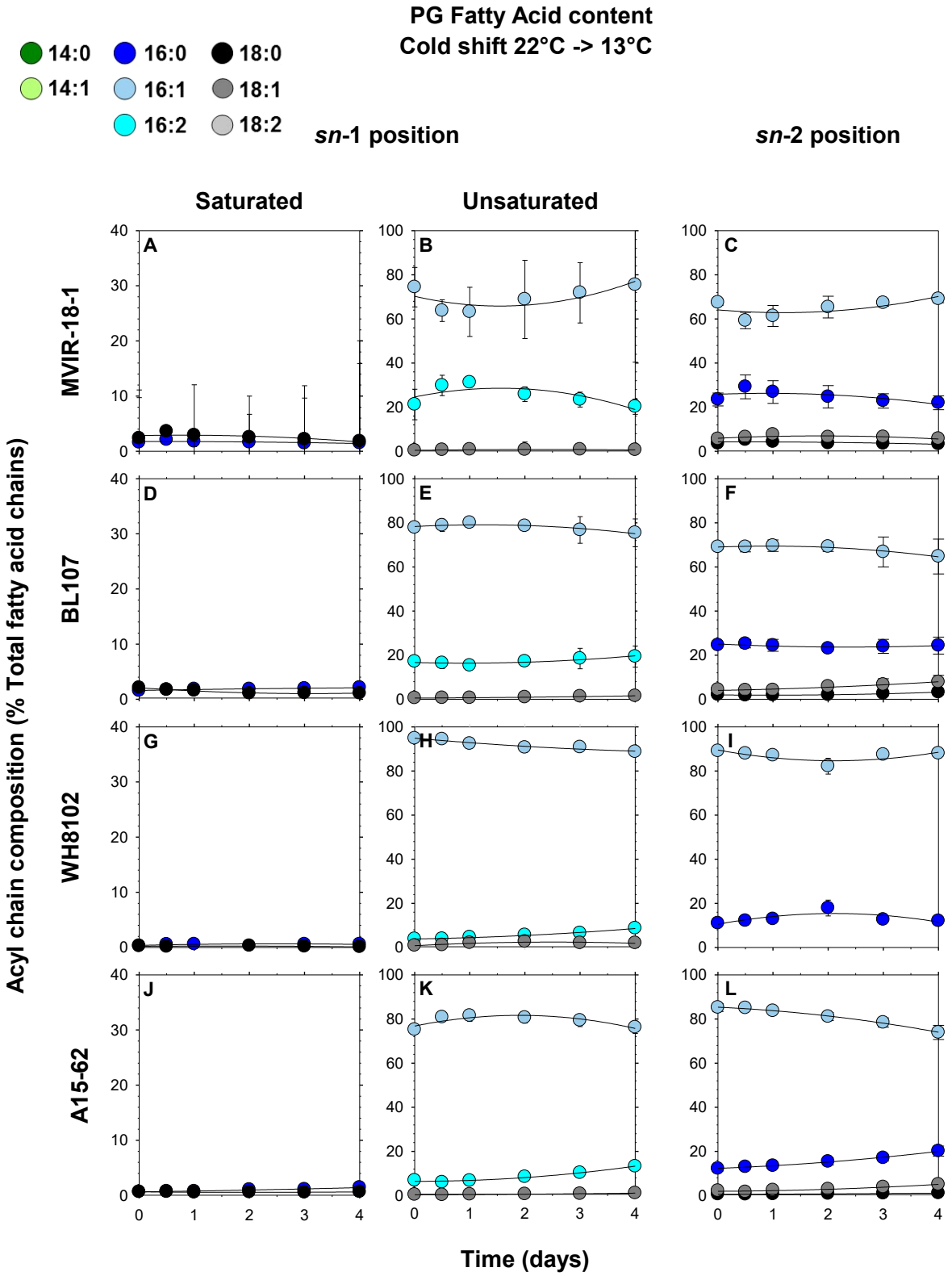


Figure S10: Relative variations (%) in the acyl chains esterified at the two glycerol positions, *sn*-1 (left panels) and *sn*-2 (right panel), of the **phosphatidylglycerol (PG)** for *Synechococcus* spp. MVIR-18-1 (A, B, C), BL107 (D, E, F), WH8102 (G, H, I) and A15-62 (J, K, L), in response to a temperature shift from 22°C to 13°C. WH8102 was the only axenic strain. The results are expressed as percentages of total acyl chains esterified at the specified stereospecific position of the glycerolipid. Each value is an average of three replicates with error bars ± SD.

S2 B: Cultures acclimated to 22°C and shifted to 30°C

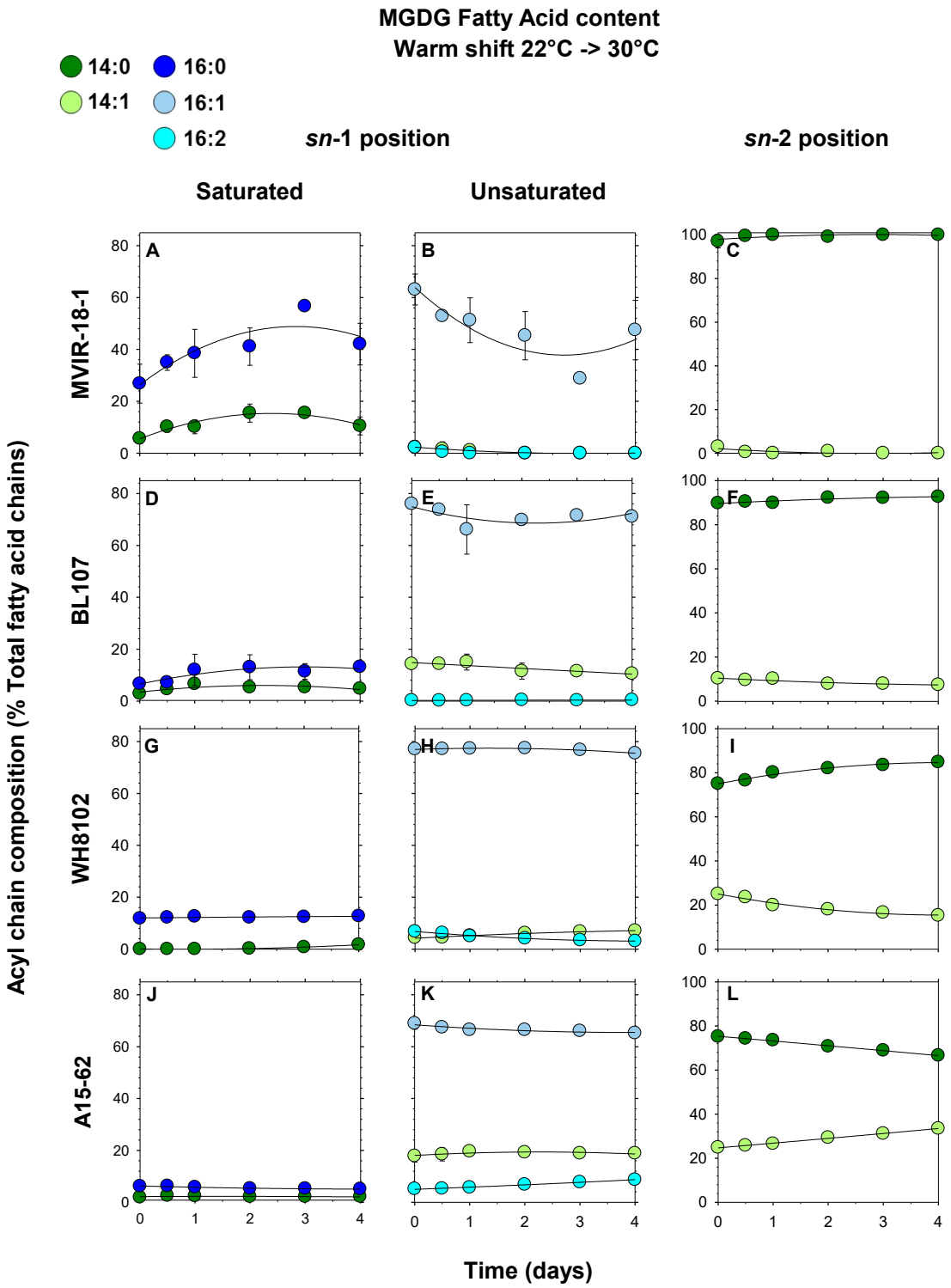


Figure S11: Relative variations (%) in the acyl chains esterified at the two glycerol positions, *sn*-1 (left panels) and *sn*-2 (right panel), of the **monogalactosyldiacylglycerol (MGDG)** for *Synechococcus* spp. MVIR-18-1 (A, B, C), BL107 (D,E,F), WH8102 (G, H, I) and A15-62 (J, K, L), in response to a temperature shift from 22°C to 30°C. The results are expressed as percentages of total acyl chains esterified at the specified stereospecific position of the glycerolipid. Each value is an average of three replicates with error bars ± SD.

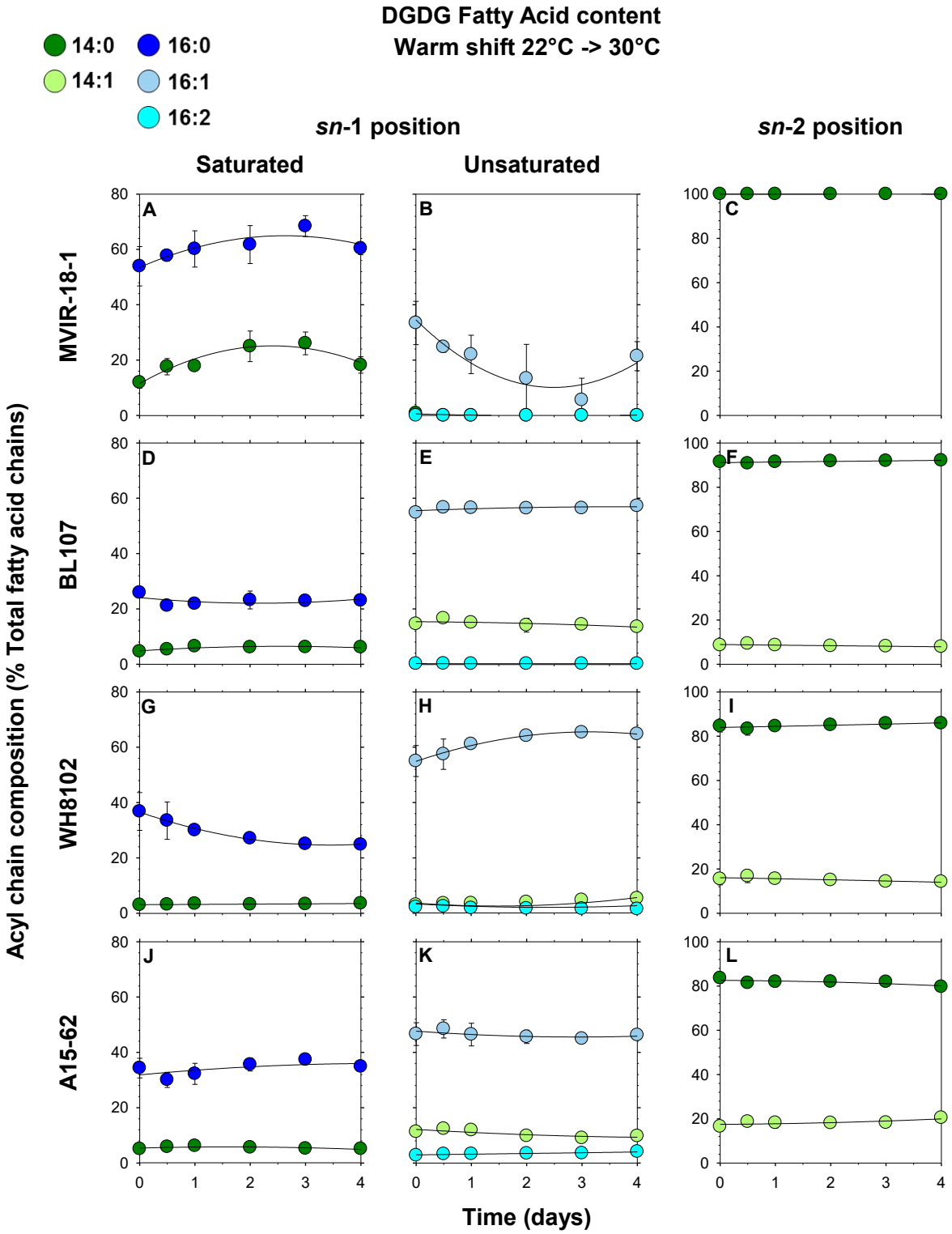


Figure S12: Relative variations (%) in the acyl chains esterified at the two glycerol positions, *sn*-1 (left panels) and *sn*-2 (right panel), of the **digalactosyldiacylglycerol (DGDG)** for *Synechococcus* spp. MVIR-18-1 (A, B, C), BL107 (D ,E ,F), WH8102 (G, H, I) and A15-62 (J, K, L), in response to a temperature shift from 22°C to 30°C. The results are expressed as percentages of total acyl chains esterified at the specified stereospecific position of the glycerolipid. Each value is an average of three replicates with error bars ± SD.

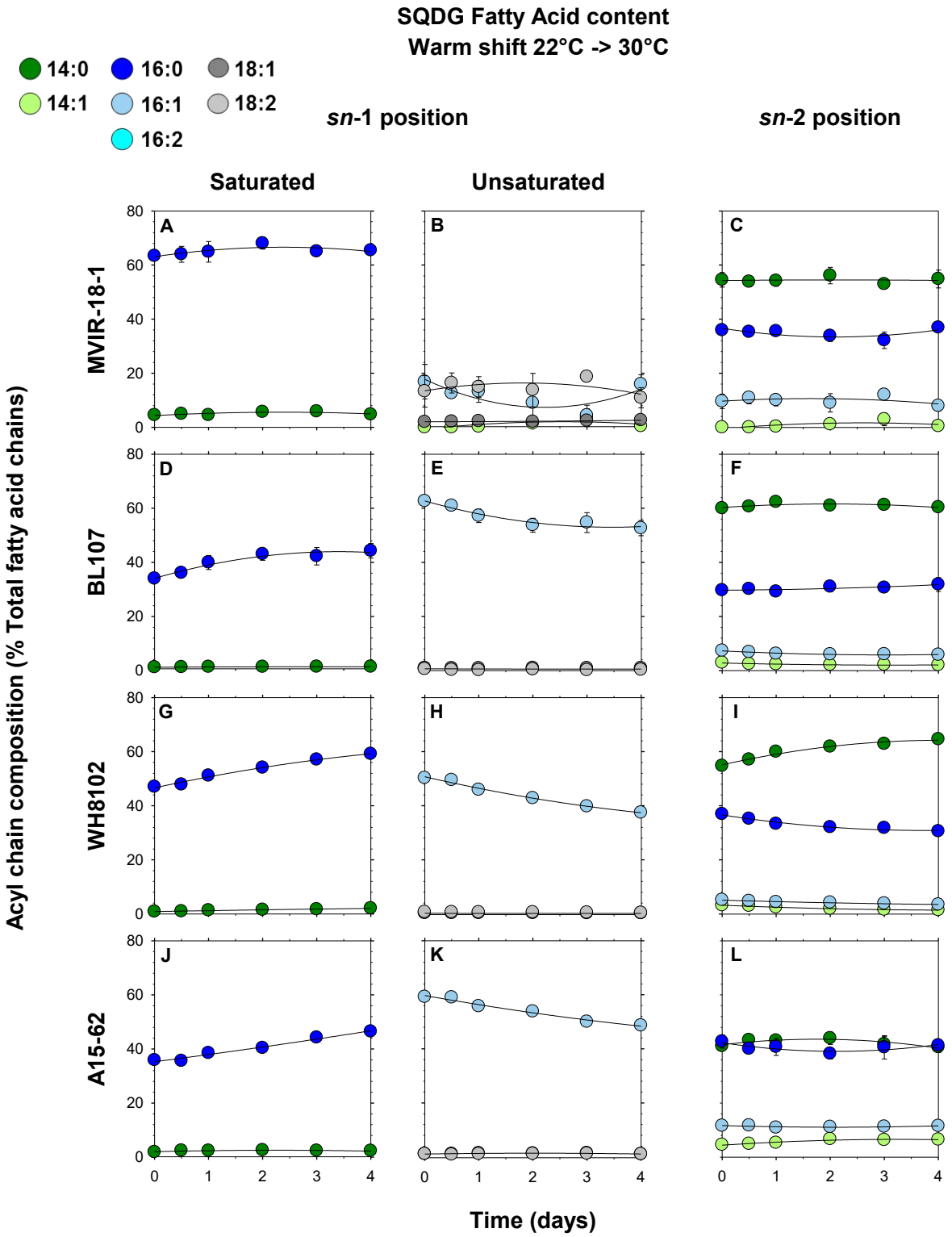


Figure S13: Relative variations (%) in the acyl chains esterified at the two glycerol positions, *sn*-1 (left panels) and *sn*-2 (right panel), of the **sulfoquinovosyldiacylglycerol (SQDG)** for *Synechococcus* spp. MVIR-18-1 (A, B, C), BL107 (D ,E ,F), WH8102 (G, H, I) and A15-62 (J, K, L), in response to a temperature shift from 22°C to 30°C. The results are expressed as percentages of total acyl chains esterified at the specified stereospecific position of the glycerolipid. Each value is an average of three replicates with error bars ± SD.

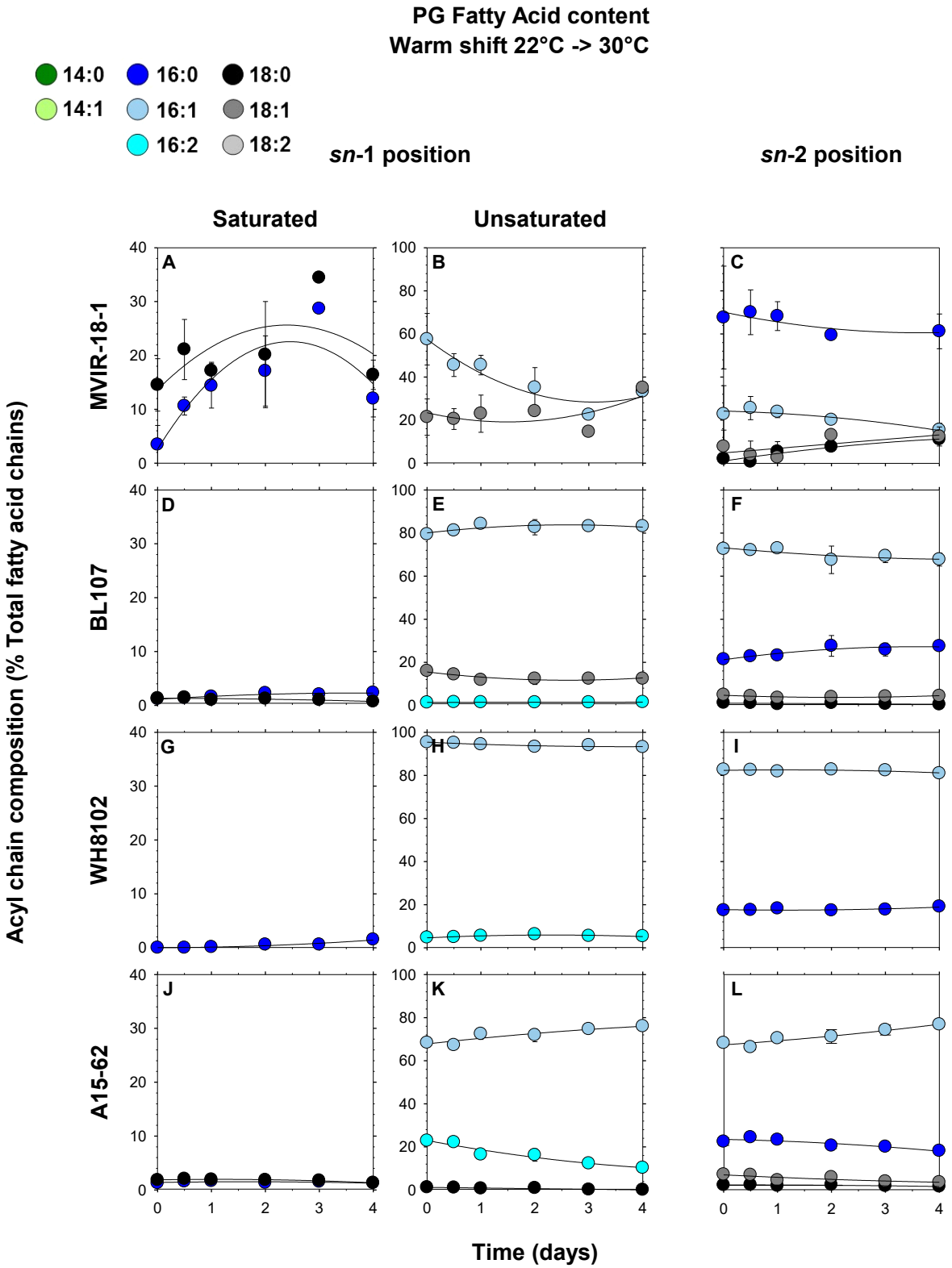


Figure S14: Relative variations (%) in the acyl chains esterified at the two glycerol positions, *sn*-1 (left panels) and *sn*-2 (right panel), of the **phosphatidylglycerol (PG)** for *Synechococcus* spp. MVIR-18-1 (A, B, C), BL107 (D, E, F), WH8102 (G, H, I) and A15-62 (J, K, L), in response to a temperature shift from 22°C to 30°C. WH8102 was the only axenic strain. The results are expressed as percentages of total acyl chains esterified at the specified stereospecific position of the glycerolipid. Each value is an average of three replicates with error bars ± SD.

NOTE S3: FATTY ACID DESATURASE METAGENOMICS

The database including the 155 *Synechococcus*/*Cyanobium* desaturases as well as many sequences used as outgroups including 117 *Prochlorococcus* fatty acid desaturase sequences, 185 outgroup cyanobacterial genomes from Cyanobase plus 537 genomes from proGenomes, 6,092 additional non-picocyanobacterial fatty acid desaturase sequences retrieved from NCBI as well as all marine picocyanobacterial desaturases from Cyanorak, is available in a supplemental file as spreadsheet tables (Datasets S1 - S4).

References of the supplementary information

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

Boatman TG, Lawson T, Geider RJ. **2017**. A key marine diazotroph in a changing ocean: The interacting effects of temperature, C02 and light on the growth of *Trichodesmium erythraeum* IMS101. *PLoS ONE* **12**: 1–20.

RESUMÉ

Les deux picocyanobactéries marines *Prochlorococcus* et *Synechococcus* sont les organismes photosynthétiques les plus abondants de la planète, leur vaste distribution étant sans doute liée à leur importante diversité génomique. La première partie de mon travail a consisté à étudier les bases génétiques de l'adaptation de ces organismes à des niches écologiques distinctes et a révélé, en comparant 81 génomes de ces deux genres, le rôle des gains/pertes de gènes et des variations des séquences nucléotidiques dans la diversification de ces organismes. Une deuxième partie a consisté en l'analyse des réponses physiologiques d'une souche modèle de *Synechococcus* (WH7803) et de quatre autres souches représentatives des écotypes dominants *in situ* (clades I à IV) à divers stress environnementaux, afin d'identifier les gènes impliqués dans les réponses spécifiques ou communes à ces différents stress et/ou écotypes. Enfin, la dernière partie de ma thèse a visé à identifier la distribution de l'ensemble des gènes de picocyanobactéries dans l'océan mondial et à la relier aux paramètres environnementaux et à la distribution des écotypes, ce qui a permis de mettre en évidence les répertoires de gènes spécifiques de niches et/ou d'écotypes. L'intégration de ces résultats m'a permis de mieux comprendre les mécanismes d'adaptation et d'acclimatation, qui ont permis aux picocyanobactéries de coloniser la quasi-totalité des niches écologiques éclairées des océans.

ABSTRACT

The two marine picocyanobacteria Prochlorococcus and Synechococcus are the most abundant photosynthetic organisms on the planet, their wide distribution being probably linked to their large genomic diversity. The first part of my work has consisted in studying the genetic bases of the adaptation of these organisms to distinct ecological niches and revealed, by comparing 81 genomes of these two genera, the role of gene gains/losses and variations in nucleotide sequences in the diversification of these organisms. A second part has consisted in the analysis of the physiological responses of a Synechococcus model strain (WH7803) and of four other strains, representative of the dominant ecotypes in the field (clades I to IV), to various environmental factors in order to identify the genes that are involved in specific or common responses to these different stresses and/or ecotypes. Finally, the last part of my thesis aimed at identifying the distribution of all picocyanobacterial genes in the world ocean and to link it to environmental parameters and to the distribution of ecotypes, which made it possible to highlight the gene repertoires specific of niches and/or ecotypes. The integration of these results led me to get a better understanding of the adaptation and acclimation mechanisms, which allowed marine picocyanobacteria to colonize virtually all lit ecological niches of the oceans.