



**HAL**  
open science

# Pigment diversity in marine *Synechococcus* sp. : molecular basis, evolution and ecological role

Théophile Grébert

► **To cite this version:**

Théophile Grébert. Pigment diversity in marine *Synechococcus* sp. : molecular basis, evolution and ecological role. Ecology, environment. Université Pierre et Marie Curie - Paris VI, 2017. English. NNT : 2017PA066614 . tel-02422222

**HAL Id: tel-02422222**

**<https://theses.hal.science/tel-02422222>**

Submitted on 21 Dec 2019

**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  
L'UNIVERSITÉ PIERRE ET MARIE CURIE**

Spécialité Microbiologie Marine  
Ecole Doctorale « Sciences de la Nature et de l'Homme : évolution et écologie » (ED 227)

**Présentée par**

**Théophile GRÉBERT**

En vue de l'obtention du grade de

DOCTEUR de L'UNIVERSITÉ PIERRE ET MARIE CURIE

**Pigment diversity in marine *Synechococcus* sp.:  
molecular basis, evolution and ecological role**

Soutenue le 19 décembre 2017, devant le jury composé de :

- Dr. Purificación LÓPEZ-GARCÍA..... Rapporteur  
*CNRS/Université Paris-Sud XI*
- Prof. Dr. Nicole FRANKENBERG-DINKEL..... Rapporteur  
*Technische Universität Kaiserslautern (Allemagne)*
- Dr. Alexis DUFRESNE..... Examineur  
*CNRS/Université Rennes 1*
- Dr. Angela FALCIATORE..... Examineur  
*CNRS/Université Pierre et Marie Curie*
- Dr. Mirjam CZJZEK..... Examineur  
*Station Biologique de Roscoff, CNRS/Université Pierre et Marie Curie*
- Prof. David KEHOE..... Co-directeur de thèse  
*Indiana University (Etats-Unis)*
- Dr. Frédéric PARTENSKY..... Directeur de thèse  
*Station Biologique de Roscoff, CNRS/Université Pierre et Marie Curie*



*À Mamine et Papet,  
pour m'avoir transmis votre amour de la nature*

*À Mie Babey et Grand Pa Pierre,  
pour avoir tourné cette passion vers la mer*



## **Pigment diversity in marine *Synechococcus* sp.: molecular basis, evolution and ecological role**

Marine *Synechococcus* are the second most abundant photosynthetic organisms on the planet. These picocyanobacteria present a wide diversity of pigmentation, which comes from differences in the composition of their light-harvesting antenna, called phycobilisome, allowing them to efficiently exploit a wide range of spectral niches. Yet, the evolution, ecology and molecular bases of the different *Synechococcus* pigment types are not well understood. By comparing the genomic region gathering most genes involved in the synthesis of phycobilisome rods from 54 sequenced isolates spanning all cultured pigment types and from natural *Synechococcus* populations, I proposed a scenario for the evolution of the different pigment types, and showed that the pigment diversity of marine *Synechococcus* can be traced back to before the diversification of this genus. Then, I developed a bioinformatic pipeline for reliably quantifying all known *Synechococcus* pigment types from metagenomic data. Applying it to the Tara Oceans dataset allowed me to describe for the first time their distribution in the global ocean, and revealed that type IV chromatic acclimation, a process by which cells can match their absorption properties to the ambient light colour, is widespread and constitutes the dominant pigmentation in *Synechococcus* populations. It also showed that natural chromatic acclimation mutants prevail in wide oligotrophic areas of the southern Pacific Ocean. Finally, I genetically characterized two members of an enzyme family binding chromophores to phycoerythrin-II, a major component of phycobilisomes. This provided new insights into the molecular bases of the chromatic acclimation process, and revealed the importance of allelic variation for the diversity of pigment types.

Keywords: cyanobacteria, marine microbiology, phycobiliprotein, chromatic acclimation, metagenomics, functional genomics, microbial ecology, evolution

---

## **Diversité pigmentaire des *Synechococcus* marins : bases moléculaires, évolution et importance écologique**

Les *Synechococcus* marins sont les seconds organismes photosynthétiques les plus abondants sur la planète. Ces picocyanobactéries marines présentent une grande diversité pigmentaire du fait de différences dans la composition de leur antenne collectrice de lumière appelée phycobilisome, ce qui leur permet d'utiliser efficacement une grande partie du spectre lumineux. Cependant, l'évolution, l'écologie et les bases moléculaires de cette diversité restent mal comprises. La comparaison de la région génomique regroupant la plupart des gènes impliqués dans la synthèse des bras de phycobilisomes provenant de 54 souches séquencées ainsi que de populations naturelles m'a permis de proposer un scénario évolutif pour l'apparition des différents types pigmentaires, et de montrer que cette diversité pigmentaire précède la diversification des *Synechococcus* marins. Par la suite, j'ai développé une procédure bioinformatique permettant de quantifier de façon fiable l'abondance relative de tous les types pigmentaires connus à partir de données de métagénomique. L'utilisation de cette méthode sur l'ensemble des métagénomomes de Tara Oceans m'a permis de décrire pour la première fois leur répartition à l'échelle mondiale, et a révélé que l'acclimatation chromatique de type IV, qui permet aux cellules de modifier leur spectre d'absorption en fonction de la couleur de la lumière, est très répandue et domine les populations naturelles de *Synechococcus*. Cela a aussi montré que des mutants naturels de l'acclimatation chromatique prédominent dans les larges étendues oligotrophes de l'océan Pacifique sud. Enfin, j'ai caractérisé génétiquement deux membres d'une famille d'enzymes liant les chromophores à la phycoérythrine II, un constituant majeur des phycobilisomes. Ces résultats apportent de nouvelles perspectives quant aux bases moléculaires de l'acclimatation chromatique, et ont révélé l'importance des variations alléliques dans la diversité des types pigmentaires.

Mots-clés : cyanobactéries, microbiologie marine, phycobiliprotéine, acclimatation chromatique, métagénomique, génomique fonctionnelle, écologie microbienne, évolution







ACKNOWLEDGEMENTS

&

REMERCIEMENTS

---

## **Acknowledgments et remerciements !**

I would first like to acknowledge the members of my PhD committee, and to express my sincere thanks and gratitude to Drs Nicole Frankenberg-Dinkel and Purificación López-García for accepting to extensively review this manuscript, as well to Drs Mirjam Czjzek, Alexis Dufresne and Angela Falciatore for accepting to evaluate this work.

I would also like to thank Drs Mirjam Czjzek and Laurence Garczarek, as well as Prof. Christophe Destombe for accepting to evaluate my work during yearly PhD committees. Many thanks for your enthusiasm, constructive criticism and new ideas.

Le travail effectué lors de cette thèse n'aurait pas été possible sans soutien financier. A titre personnel, mon travail a été financé par une bourse doctorale du ministère de l'enseignement supérieur, de la recherche et de l'innovation, versée par l'Université Pierre et Marie Curie – Paris 6. Ce financement m'a été accordé sur le contingent des contrats doctoraux spécifiques normaliens de l'École Normale Supérieure de Lyon. A titre professionnel, ce travail a été rendu possible grâce à différents financements de l'Agence Nationale de la Recherche. Je tiens à remercier ces différentes institutions pour m'avoir permis de mener à bien ces recherches.

Au cours de cette thèse, j'ai eu le privilège de pouvoir effectuer un séjour de plusieurs mois aux Etats-Unis. Ce séjour a été rendu possible grâce à une bourse Fulbright. J'adresse ici mes sincères remerciements à la Commission Franco-Américaine Fulbright et au J. William Fulbright Foreign Scholarship Board pour m'avoir accordé cette chance et cet honneur. Un grand merci également à Emily Resnier, Séverine Peyrichou, Madeleine Bouvier d'Yvoire et Jennifer Skiba pour leur aide et leur efficacité.

De même, j'ai pu assister à différents congrès internationaux grâce à des bourses de la Société Phycologique de France et de la British Microbiology Society. Merci de soutenir de la sorte les jeunes docteurs.

A brief word to the Schluchter Lab: I really enjoyed working with you on the characterization of phycobilin lyases, and I would particularly like to thank Adam Ahn Nguyen, Suman Pokhrel and Dr. Wendy Schluchter for your commitment to this project.

To Dr. David Kehoe, thank you for accepting to be my co-advisor, but most importantly I am deeply grateful to you for your warm welcome in your lab. I really enjoyed my stay, and appreciated how you shared not only your knowledge of cyanobacteria and science, but also of so much more. Thank you for your encouragement and commitment, and for your everyday enthusiasm and motivation. If I am not sure to totally get the meaning of IU's motto, at least my visit entirely met my expectations. Many thanks to the Kehoe family as well for kindly welcoming me on several festive occasions, and sharing these good moments. Allissa, Bo, Alan, thanks a lot to you for welcoming me as a new member of the Kehoe lab. You contributed greatly in making me appreciate my short time in Bloomington. Special thanks to Bo for your expert counseling, and your training in the genetic manipulation of *Synechococcus*. I enjoyed working with you on the FrankenSyn project!

Frédéric “Fred” Partensky, merci d’avoir accepté de m’encadrer durant cette thèse. Ces *Synechococcus* ont vraiment des couleurs incroyables ! J’espère que les nombreux mystères restants sur les types pigmentaires et l’acclimatation chromatique seront bientôt résolus par tes soins. Merci pour tes connaissances, ta culture, ton humour subtil ... Mais aussi ta bonne humeur, ta passion, ton émerveillement et ta candeur : ce sera toujours une surprise pour moi de te voir sauter sur et de ta chaise à la suite d’une nouvelle idée ! Merci aussi pour ton investissement, ta disponibilité et ta réactivité dans les moments difficiles (mais pas que), j’ai probablement des progrès à faire de ce côté là... Difficile d’évoquer Fred sans évoquer son double féminin, merci à toi Laurence pour tes conseils, ta patience (!), et pour les franches discussions. Merci à tous les deux pour vos idées, votre dynamisme et votre optimisme ... parfois peut-être un peu trop poussé (« ça prend deux jours ») ! Même si j’aurais parfois préféré avoir à faire à l’un *ou* à l’autre, et non à l’un *et* à l’autre, force est de constater que vous formez un duo de choc. Cette thèse n’aurait pas été aussi aboutie sans l’énergie considérable que vous y avez mise. Je suis fier du travail accompli, et j’espère que vous aussi !

Quelle chance incroyable que de faire une thèse à Roscoff ! Une petite pépite (bien) cachée au fin fond du nord Finistère, parfois ingrate (Je croyais que la saison des pulls c’était l’hiver ?), mais on y est tellement bien ... Et si on y est aussi bien, c’est beaucoup grâce aux personnes avec qui l’on travaille. Merci à toutes ces personnes qui, si elles n’ont pas directement « les mains dans le cambouis » et dont le nom n’apparaît pas sur nos publications scientifiques, contribuent néanmoins à faire tourner la Station et à nous permettre de faire ce que l’on aime dans de bonnes conditions. Mention spéciale pour Céline, d’une efficacité redoutable lorsqu’il s’agit d’affronter les sandwiches administratifs : merci pour ton aide et ta bonne humeur !

Un grand merci à l’équipe MaPP pour la super ambiance de travail ! Christophe, merci pour les échanges que nous avons pu avoir, ainsi que pour tes conseils experts sur la fluorescence et l’extraction de phycobilisomes. Momo ... J’espère que toutes ces manip de mutagénèse vont *enfin* marcher ! C’était vraiment chouette de travailler avec toi et de papoter en manipant. Merci pour ton aide, ta rigueur, ton énergie (mais tu bois quoi le matin ?), tes délires (« J’veux pas retourner à l’hôpital ! »), ton rire communicatif et ton amitié ! Et si jamais tu as besoin de mains supplémentaires pour faire ton potager surélevé, n’hésite pas !

Aux différents membres du groupe plancton, merci pour votre bonne humeur quotidienne, votre dynamisme, et pour vos discussions enrichissantes. Mention spéciale aux bureaux du premier étage : c’est toujours un moment agréable d’aller manger un sandwich au soleil, prendre un café à Kerjefic ou un thé dans le bureau avec vous. Merci Fabienne, Domi, Pris, Flo, Momo encore, pour prendre bien soin de nous avec vos petites attentions et vos encouragements.

Aux anciens du bureau 331 et aux nouveaux du bureau 307, Domi B, Klervi, Marie, Delphine D, Margot (mention spéciale : 3 ans à me supporter sur le même bureau, j’espère que les séquelles ne seront pas trop lourdes ;)), Delphine S, Ruibo, Wei-ting, Pym merci pour les parties de rigolade, les discussions sérieuses et moins sérieuses ...

Merci à tous les jeunes, nouveaux et anciens, Juju, Doriane, Tristan, Laura, Hugo, Ambre, Marine, Alexis, Valérian, David, Laure, Peche, Eloise, Camille, Ulysse, Martin, Laura, Charles pour le rendez-vous hebdomadaire à l’institution roscovite qu’est Ty Pierre, toujours agréable !

Merci aux amis d'ici et d'ailleurs, l'équipe lyonnaise, Hugo, Marc, Félix, Thibault, Camille, Michel, Mathilde, Vincent : ça fait longtemps qu'on a pas fait de rando non ? Petite mention à Thuvarakan, François, Théophile, Florent : il serait temps aussi de remettre ça ...

Tristan, Doriane, Laura, Hugo, Juju, la team malt fermenté -avec ou sans distillation- ... Que de bons moments passés ensemble lors de ces soirées pâtes carbo-jeux ! Vivement nos prochaines retrouvailles, ici ou là-bas, au nouvel an ou en été, à six ou à sept, ou plus :o (non je rigole, calme-toi Momo) !

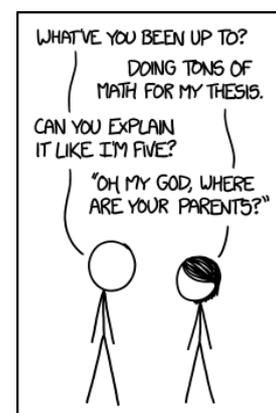
Pierre et PG « de Porz an Park », le duo de choc, merci pour votre amitié et les bons moments passés ensemble autour d'une bière. Merci d'avoir accepté de me laisser un petit lopin de terre pour cultiver quelques racines, une superbe échappatoire pour se rafraîchir les idées et retrouver les siennes, de racines.

Hugo, ça fait "quelques" années qu'on se suit maintenant ! Merci pour tous ces échanges d'idées, ta rigueur scientifique et ton amitié indéfectible. Nul homme n'étant parfait, il va par contre falloir envisager de faire quelques progrès en Tétris de meubles Ikea ;) et puis ... désolé pour tous ces assassinats sauvages à Complots, c'est pour me venger de ta suprématie à Sobek !

Juju ... Merci d'avoir enduré mes monologues sur la vie secrète des phycobiline lyases et autres types pigmentaires, les avoir parfois transformés en dialogue (!), mais aussi pour avoir su régulièrement me rappeler de décrocher. Merci pour m'avoir encouragé et dissipé mes doutes dans les moments difficiles, râlé avec moi quand j'avais raison, et avoir su me faire comprendre quand j'avais tort. Merci pour ton soutien précieux, ta patience, tes surprises, ton rire et ta simple présence à mes côtés qui suffit à rendre la vie plus belle.

Enfin, merci à ma famille, qui a accepté et supporté mes longues absences et la distance imposées ces dernières années, pour votre soutien et vos encouragements, vous être intéressés à ce que je fais, pour tout ce que vous m'avez donné.

2013, 2017 : Anne, la prochaine c'est pour toi ;) !



Randall 'xkcd' Munroe, "Like I'm five", 2014

License Creative Commons by-nc

<https://xkcd.com/1364/>



Roscoff, 11 avril 2015





**Table of contents**

|   |      |
|---|------|
| Acknowledgments et remerciements ! .....                  | i    |
| Table of contents .....                                   | viii |
| List of figures .....                                     | xii  |
| List of tables .....                                      | xv   |
| Introduction .....  | 1    |
| I. Cyanobacteria .....                                    | 2    |
| 1. Evolution and biogeochemical importance.....           | 2    |
| 2. The (oxygenic) photosynthetic apparatus .....          | 4    |
| 3. Light harvesting complex: the phycobilisome .....      | 7    |
| a. Overview.....  | 7    |
| b. Components .....                                       | 10   |
| i. Phycobiliproteins.....                                 | 10   |
| ii. Linker proteins .....                                 | 13   |
| iii. Bilins .....   | 15   |
| c. Energy transfer: structure-function relationship ..... | 19   |
| d. Phycobilin lyases .....                                | 22   |
| i. T clan.....  | 23   |
| ii. S/U clan.....   | 24   |
| iii. E/F clan .....                                       | 25   |
| e. Regulation of phycobilisome activity .....             | 26   |
| i. Nutrients.....   | 26   |

---

|   |     |
|---|-----|
| ii. Regulation of light harvesting and utilization.....   | 27  |
| iii. Chromatic acclimation.....   | 30  |
| f. Biotechnological applications.....   | 33  |
| II. The picocyanobacteria, key members of the phytoplankton.....  | 34  |
| 1. Phytoplankton.....   | 34  |
| a. Taxonomic, morphological and ecological diversity.....   | 34  |
| b. Role in the biological pump.....   | 37  |
| 2. Marine picocyanobacteria.....  | 40  |
| a. Discovery and physiology of marine picocyanobacteria.....  | 40  |
| b. Distribution and ecological role.....  | 42  |
| c. Evolution and diversity of marine picocyanobacteria.....   | 44  |
| 3. Adaptive microdiversity in picocyanobacteria.....  | 48  |
| a. Environmental variability in the ocean.....  | 48  |
| b. <i>Prochlorococcus</i> genetic diversity is shaped by light and temperature.....   | 51  |
| c. <i>Synechococcus</i> diversity.....  | 54  |
| i. Genetic (micro)diversity.....  | 54  |
| ii. <i>Synechococcus</i> pigment types.....   | 58  |
| Objectives of this work.....  | 65  |
| <br>  |     |
| Chapter I - Genomics and evolution of <i>Synechococcus</i> pigment types.....   | 67  |
| I. Context of the work.....   | 68  |
| II. Genomics and evolution of <i>Synechococcus</i> pigment types.....   | 69  |
| 1. Diversity and evolution of the phycobilisome gene region in relation to light niches in marine <i>Synechococcus</i> cyanobacteria..... | 69  |
| 2. Future work.....   | 104 |
| III. Metagenomes suggest adaptation of phycobiliproteins to temperature.....  | 105 |
| 1. Background.....  | 105 |
| 2. Materials and methods.....   | 105 |
| 3. Results and discussion.....  | 106 |

|   |     |
|---|-----|
| Chapter II - Distribution of <i>Synechococcus</i> pigment types .....   | 111 |
| I. Context of the work .....  | 112 |
| II. Light color acclimation: a key process in the global ocean distribution of <i>Synechococcus</i> cyanobacteria .....   | 113 |
| <br>  |     |
| Chapter III - Genetic approaches to <i>Synechococcus</i> pigment types .....  | 135 |
| I. Context of the work .....  | 136 |
| II. Characterization of two enzymes involved in phycoerythrin-II chromophorylation ...  | 139 |
| 1. Functional characterization of the MpeWYZ phycobilin lyase family provides key insights into color niche acclimation and adaptation in marine <i>Synechococcus</i> ..... | 139 |
| 2. Future work.....   | 158 |
| 3. Identification of motifs involved in the lyase-isomerase activity .....  | 158 |
| III. The “ <i>FrankenSynechococcus</i> ” experiment: how easy is it to gain CA4? .....  | 166 |
| 1. Introduction.....  | 166 |
| 2. Results.....   | 167 |
| 3. Discussion.....  | 168 |
| 4. Materials and methods .....  | 170 |
| 5. Supplemental material .....  | 170 |
| <br>  |     |
| Conclusion & Perspectives.....  | 173 |
| I. Considerations on the phycobilin lyases function .....   | 174 |
| II. Type IV Chromatic Acclimation .....   | 179 |
| III. Adaptive value of the different pigment types .....  | 183 |
| IV. Evolution of pigmentation and evolution of <i>Synechococcus</i> .....   | 185 |
| <br>  |     |
| References .....  | 189 |



**List of figures**

|   |    |
|---|----|
| Figure 1: Phylogeny and ancestral state reconstruction of different representative cyanobacteria, based on the small ribosomal (16S rDNA) and RNA polymerase beta prime (RpoC) subunits. .... | 3  |
| Figure 2: Macromolecular complexes involved in oxygenic photosynthesis.....   | 5  |
| Figure 3: Schematic representation of the electron transport chain of oxygenic photosynthesis....   | 6  |
| Figure 4: First direct observation of phycobilisomes.....   | 8  |
| Figure 5: Model of phycobilisome associated to the thylakoid membrane.....  | 9  |
| Figure 6: Assembly of phycobiliprotein into hexamers.. ....   | 11 |
| Figure 7: Scenario proposed by Apt and co-workers for the evolution of phycobiliproteins.....   | 12 |
| Figure 8: Linker proteins. (A) .....  | 14 |
| Figure 9: Phycobilins present in cyanobacteria and red algae, with their thioether bound to phycobiliproteins.....  | 16 |
| Figure 10: Chromophorylation of cyanobacterial phycobiliproteins.. ....   | 17 |
| Figure 11: Biosynthetic pathways of phycobilins.. ....  | 18 |
| Figure 12: Energy transfer in phycobilisomes.....   | 21 |
| Figure 13: Structural model of a phycobilisome associated with both photosystem II (PSII) and photosystem I (PSI) in a megacomplex. ....  | 22 |
| Figure 14: 3D structure of a $\phi$ CpeT dimer from the P-HM1 cyanophage (PDB 5HI8).....  | 23 |
| Figure 15: 3D structure of two members from the S/U clan.. ....   | 25 |
| Figure 16: Effect of light intensity on <i>Synechococcus</i> sp. WH7803 pigmentation.....   | 28 |
| Figure 17: Diverse marine planktonic organisms drawn by Ernst Haeckel in <i>Kunstformen der Natur</i> , 1904. ....  | 35 |
| Figure 18: the different size classes of plankton. ....   | 36 |
| Figure 19: taxonomic and morphological diversity of single-celled eukaryotic marine plankton.   | 36 |
| Figure 20: Marine primary production in August 2017. ....   | 38 |

---

|   |     |
|---|-----|
| Figure 21: The oceanic biological pump.....   | 39  |
| Figure 22: Electron micrographs of <i>Synechococcus</i> (A, C) and <i>Prochlorococcus</i> (B, D).....   | 40  |
| Figure 23: Light-harvesting complexes associated to PSII in marine <i>Synechococcus</i> and <i>Prochlorococcus</i> . .....  | 41  |
| Figure 24: Global latitudinal distribution of <i>Prochlorococcus</i> and <i>Synechococcus</i> .. .....  | 42  |
| Figure 25: Vertical distribution of <i>Prochlorococcus</i> and <i>Synechococcus</i> in different oceans and seas.....   | 43  |
| Figure 26: Phylogenetic relationships amongst marine picocyanobacteria. ....  | 45  |
| Figure 27: Genome size (bar) and G+C content (plot) of marine picocyanobacteria compared to other cyanobacteria. ....   | 47  |
| Figure 28: Overview of the marine environment experienced by marine picocyanobacteria.....  | 49  |
| Figure 29: Gradients of light colour in different aquatic environments.....   | 50  |
| Figure 30: Light ecotypes in <i>Prochlorococcus</i> . ....  | 52  |
| Figure 31: Copy number (#) and phyletic pattern of genes involved in photosynthesis and nutrient uptake in various <i>Prochlorococcus</i> isolates.....   | 53  |
| Figure 32: Temperature ecotypes within the <i>Prochlorococcus</i> HL ecotype. ....  | 54  |
| Figure 33: Major clades of the marine <i>Synechococcus</i> radiation.....   | 55  |
| Figure 34: Distribution of the different <i>Prochlorococcus</i> and <i>Synechococcus</i> clades in oceanic surface waters, assessed with the high-resolution phylogenetic marker <i>petB</i> . .... | 56  |
| Figure 35: thermal range of different <i>Synechococcus</i> strains isolated along a latitudinal gradient.. .....  | 58  |
| Figure 36: Picture of different cultures from strains with different pigment types. ....  | 59  |
| Figure 37: Phycobilisome composition and spectral properties of different pigment types. ....   | 60  |
| Figure 38: Organization of PBS rod genomic region for different pigment types.....  | 62  |
| Figure 39: Amino acids observed at CpcA43 (top) and CpcB42 (bottom) in phycocyanin alpha- and beta-subunit sequences retrieved from the Tara Oceans dataset. ....                                   | 108 |
| Figure 40: Sequence comparison of lyases and lyase-isomerases of the MpeWYZ family. ....  | 160 |

Figure 41: Whole cell fluorescence excitation spectra with emission at 585 nm for two different “*FrankenSynechococcus*” transformants acclimated under green (A, C) or blue light (B,D).. ..... 167

## **List of tables**

|  |     |
|--|-----|
| Table S1: Strains and sequence accession number used in this study.....  | 161 |
| Table S2: Residues sets differing between lyases and isomerases from the MpeWYZ family and<br>homolog residues from CpeY.....  | 164 |
| Table S3: Strains used in the “ <i>FrankenSynechococcus</i> ” study.....   | 171 |
| Table S4: Primers used in the “ <i>FrankenSynechococcus</i> ” study.....   | 171 |
| Table S5: Plasmids used in the “ <i>FrankenSynechococcus</i> ” study.....  | 171 |
| Table 6: Marine <i>Synechococcus</i> phycobiliprotein chromophorylation sites and cognate phycobilin<br>lyases (or lyases-isomerases).....   | 176 |
| Table 7 <i>Prochlorococcus</i> phycobiliprotein chromophorylation sites and cognate phycobilin<br>lyases (isomerases).....   | 177 |
| Table 8: Genes identified by (Sanfilippo <i>et al.</i> , 2016) as presenting different transcript levels<br>under blue- and green-light in the wild-type or between wild-type and <i>fciA</i> - or <i>fciB</i> -<br>mutants..... | 181 |







# INTRODUCTION

---

« Since my chief object is to point out the bearing of my subject on botany,  
I shall enter as little as possible into mere physics and chemistry.  
I must, however, to some extent; or else the evidence on  
which my conclusions are founded could not be understood. »

H.C. Sorby, 1875

## I. Cyanobacteria

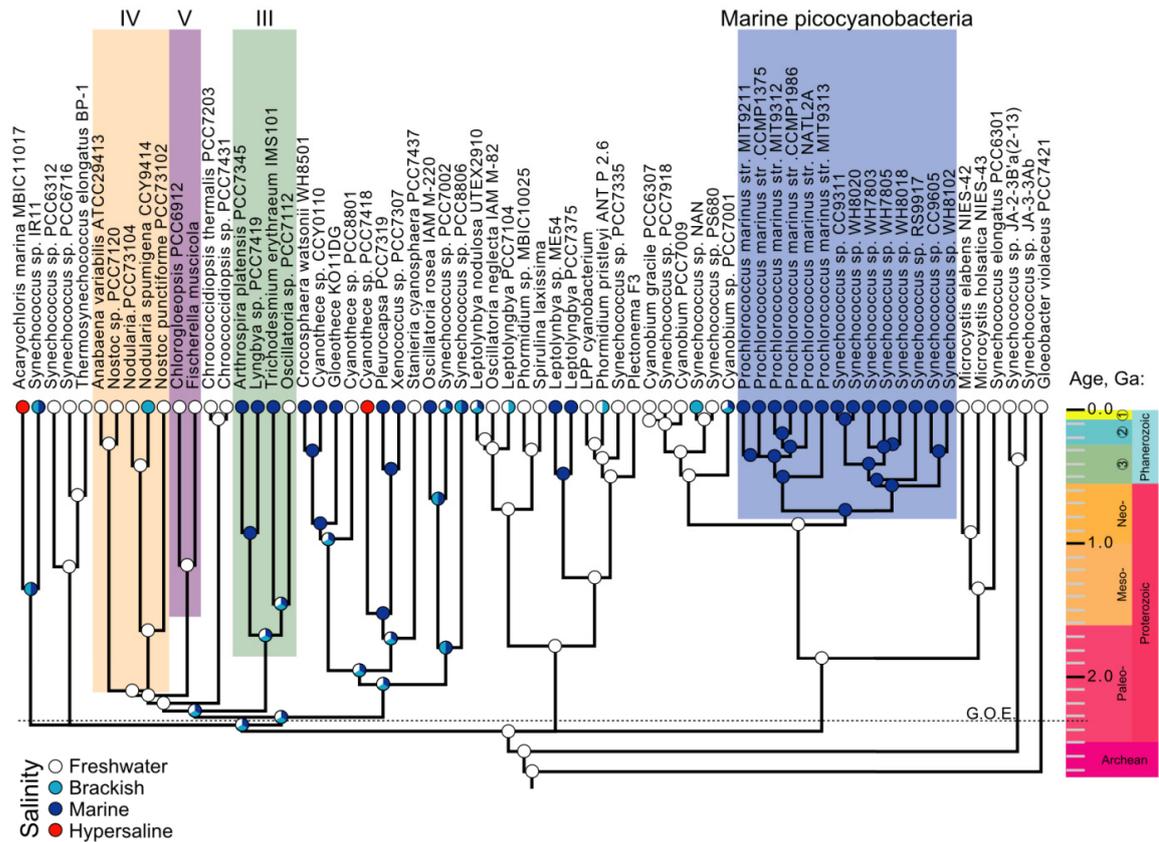
### 1. Evolution and biogeochemical importance

Cyanobacteria form the monophyletic group of Gram-negative bacteria that are able to perform oxygenic photosynthesis (Fox *et al.*, 1980; Woese, 1987; Stanier and Cohen-Bazire, 1977). The similarity between cyanobacteria and chloroplasts of higher plants and red algae has been recognized early, and led Lynn Sagan to formulate the modern theory of endosymbiosis 50 years ago (Sagan, 1967).

Some authors suggested that the appearance of cyanobacteria dates back as early as 3.5 billion years ago (Ga), but the precise dating is still debated. Indeed, all geology-based (fossilized microstructures, carbon isotope fractionation, lipid biomarkers, and sedimentary macrostructures such as stromatolites) and biologically-based (molecular clock analysis) methods individually suffer from potential pitfalls and drawbacks, the major being poor sample preservation for the former and the need for accurate calibration points for the latter (Blank, 2004; Schirrmeister *et al.*, 2016). However, the major geological event known as ‘Great Oxidation Event’ (GOE), during which O<sub>2</sub> levels in the atmosphere permanently raised above 10<sup>-5</sup> of present-day level, occurred between 2.48 and 2.32 Ga (see review from Knoll and Nowak, 2017), and oxygenic photosynthesis is recognized as necessary to explain such an important environmental change. As this metabolic pathway first appeared in cyanobacteria and was then transferred to eukaryotes through an endosymbiosis event, the GOE provides a lower bound for the apparition of cyanobacteria (**figure 1**; Knoll and Nowak, 2017).

Oxygenic photosynthesis permanently modified the redox state of our planet and strongly affected both key biogeochemical cycles (carbon, iron, sulfur, oxygen, nitrogen, phosphorus) and the composition of the atmosphere. The increase of atmospheric O<sub>2</sub> levels ultimately led to the rise of life forms based on aerobic respiration, including plants and animals, to the detriment of the previously predominant anaerobic microbes, which were constrained to inhabit the few remaining O<sub>2</sub>-free niches (Kopp *et al.*, 2005; Knoll and Nowak, 2017; Sánchez-Baracaldo *et al.*, 2014).

Cyanobacteria also have a key role in the evolutionary history of algae and plants as the ancestor of all chloroplasts, an organelle that seemingly arose from a unique endosymbiosis event (Ochoa de Alda *et al.*, 2014; Deusch *et al.*, 2008; Shih *et al.*, 2013). Phylogenetic methods and ancestral state reconstruction suggest a small, unicellular and freshwater-inhabiting cyanobacterial ancestor.



**Figure 1: Phylogeny and ancestral state reconstruction of different representative cyanobacteria, based on the small ribosomal (16S rDNA) and RNA polymerase beta prime (RpoC) subunits.** Clades corresponding to sections III, IV, V, as defined by (Rippka *et al.*, 1979) are highlighted, as is the marine picocyanobacteria clade. Ancestral habitat salinity reconstructed using maximum parsimony. G.O.E, Great Oxidation Event; ①, Cenozoic; ②, Mesozoic; ③, Paleozoic; Ga, Billion years. Figure modified from (Blank and Sánchez-Baracaldo, 2010).

Cyanobacteria display a wide variety of shapes and lifestyles and are found in a vast range of ecosystems. Indeed, they can be single-celled, free-living picocyanobacteria in the oligotrophic ocean, filamentous forms thriving in fresh water, aggregates living in soils, hot springs, desert crusts, or glacier, and even symbionts of lichens, sponges or amoeba (Whitton, 2012; De Los Ríos *et al.*, 2007; Deusch *et al.*, 2008; Hoffmann, 1989). An ecologically important feature of cyanobacteria is the ability to fix atmospheric nitrogen. Although not restricted to and not found in all cyanobacteria, this process is particularly important in nitrogen-poor environments that encompass more than half of all oligotrophic areas of the world ocean, and in which cyanobacteria constitute the major N<sub>2</sub>-fixers, either as free-living organisms or in association with other planktonic organisms, including microalgae (Zehr, 2011).

Taxonomically, cyanobacteria have been historically classified into five sections based on their morphology and reproduction modes (Rippka *et al.*, 1979). Sections I (Chroococcales)

and II (Pleurocapsales) correspond to unicellular or colonial forms, with section I reproducing by binary fission or budding, and section II also reproducing by multiple fission. Sections III to V correspond to filamentous forms. Section III (Oscillatoriales) filaments are always solely made of vegetative cells, whereas section IV (Nostocales) and V (Stigonematales) can produce differentiated cells specialized in N<sub>2</sub> fixation (heterocysts). Finally, section IV divides in only one plane, leading to linear trichomes, whereas section V divide in more than one plane and have ramified filaments (Rippka *et al.*, 1979). Molecular phylogenies revealed that only sections IV and V correspond to monophyletic groups (Tomitani *et al.*, 2006; Schirrmeyer *et al.*, 2015; Shih *et al.*, 2013).

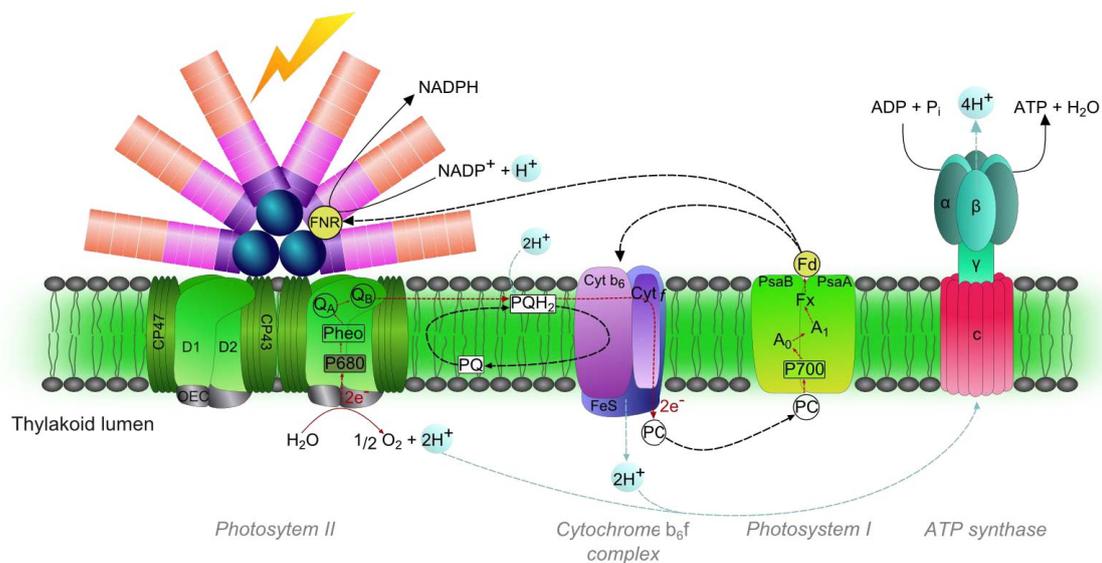
Interestingly, recent phylogenomic studies suggest that the first cyanobacteria occurred first in terrestrial or freshwater environments and that the marine environment was colonized only later, in at least three independent events. One of these events led to the marine picocyanobacteria lineage, which comprises *Prochlorococcus*, *Cyanobium* and marine *Synechococcus* (**figure 1**; Blank and Sánchez-Baracaldo, 2010). The *Synechococcus* genus is polyphyletic, and gathers both freshwater and marine isolates. However, *Cyanobium*, *Prochlorococcus* and marine *Synechococcus* form a monophyletic clade. It has recently been suggested to rename marine *Synechococcus* into *Parasynechococcus* (Coutinho *et al.*, 2016), but the former genus name is more widely recognized and will be used in the rest of this thesis manuscript.

## 2. The (oxygenic) photosynthetic apparatus

Cyanobacteria are photoautotrophs, using light as energy source and a mineral electron donor. More precisely, cyanobacteria perform oxygenic photosynthesis, in which the electron donor is the water molecule. Oxygenic photosynthesis proceeds in two steps that are somewhat decoupled. During the light phase, solar energy is converted into chemical energy through photosynthetic electron transport. This chemical energy is then used during the light-independent of dark phase to fix atmospheric CO<sub>2</sub>, i.e. incorporate it into sugars.

The light reactions of oxygenic photosynthesis take place in the thylakoid membrane, and involve four different transmembrane macromolecular complexes, the two photosystems PSII and PSI, the *b<sub>6</sub>f* cytochrome and the ATP synthase (**figure 2**; Ke, 2001). Light energy absorbed by the extrinsic light harvesting antenna associated to the PSII, called phycobilisome, is transformed into excitation energy (molecules that reach higher energy levels). This excitation is ultimately transferred to a special pair of chlorophyll *a* (Chl *a*) molecules (P680) in the reaction centre of PSII, which is formed by the protein dimer D1/D2. These Chl *a* molecules enter the P680\* state, and then quickly dissipate their excitation energy by emitting an electron and pass

into the oxidized  $P680^+$  state.  $P680^+$  is a very strong oxidant and captures electrons from the special tyrosine Z of the D1 protein. These electrons are in turn collected from water molecules by the Oxygen Evolving Complex (OEC). This complex, also known as the water-splitting complex, sits on the lumen side of the thylakoid membrane. It comprises four manganese and one calcium atom, which are key to its catalytic activity. The OEC catalyzes four successive oxidations of two  $H_2O$  molecules, resulting in the formation of one  $O_2$  molecule and the liberation of four electrons and four protons (**figure 3**, see also **figure 13**; Ke, 2001).

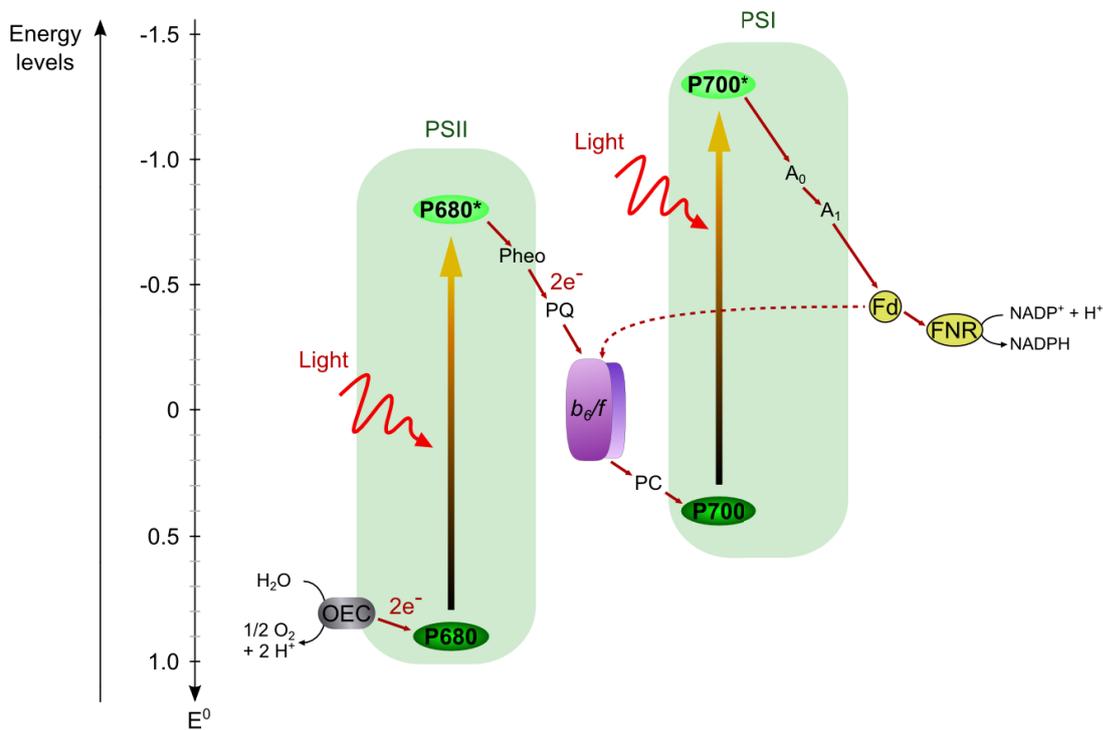


**Figure 2: Macromolecular complexes involved in oxygenic photosynthesis.** Photon energy absorbed by the light-harvesting antenna (phycobilisome) is transferred to the photosystem II and triggers an electron transport chain, which results in the formation of a transmembrane electrochemical gradient of protons, the splitting of water and production of  $O_2$  and NADPH. The electrochemical gradient is then dissipated through the ATP synthase to generate ATP. Note that the phycobilisome is about 1/5 of its real scale (see **figure 13**). Figure modified from (Pittera, 2015).

Electrons are then transferred through successive redox reactions (implying various cofactors) to the  $b_6/f$  cytochrome and then to the PSI (**figure 3**). The  $b_6/f$  cytochrome uses a part of the electron energy to pump protons across the thylakoid membrane, resulting in the formation of an electrochemical gradient.

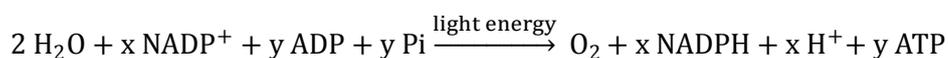
The PSI functions similarly to the PSII, except that its association with a phycobilisome is much less understood (see 3.c and 3.e.ii). Instead, light energy could-be primarily collected by the PSI-associated pigments (essentially Chl  $a$  and  $\beta$ -carotene molecules). As for PSI, light energy is transferred to a special pair of Chl  $a$  (P700). The excited  $P700^*$  emit electrons, which

are transferred to the ferredoxin-NADPH oxidoreductase (FNR) through diverse cofactors, constituting the non-cyclic electron transport. The FNR, which could be physically associated to the phycobilisome, uses the electrons to reduce  $\text{NADP}^+$  into NADPH, which serves as reducing agent in numerous anabolic reactions (**figure 3**; Thomas *et al.*, 2006; Arteni *et al.*, 2009; Korn *et al.*, 2009). Alternatively, the electron can be transferred back to the  $b_6/f$  cytochrome, which pumps additional protons through the membrane, and are then returned to the P700. This cyclic electron transport is used to balance the production of NADPH and the generation of electrochemical gradient. Cyclic electron transport is controlled by short-term (second timescale) state transitions and other longer timescale acclimation processes (see I.3.e.ii).



**Figure 3: Schematic representation of the electron transport chain of oxygenic photosynthesis.** The absorption of light by the PSII triggers the charge separation. Electrons are transferred to the PSI through the  $b_6/f$  cytochrome. The absorption of light by PSI allows the transfer to the terminal acceptor  $\text{NADP}^+$ . Oxidized PSII chlorophylls are reduced with electrons from water, through the action of the Oxygen Evolving Complex (OEC). Pheo, pheophytine; PQ, plastoquinone; PC, plastocyanin;  $A_0/A_1$ , electron carriers; Fd, ferredoxin; FNR, ferredoxin- $\text{NADP}^+$  oxidoreductase.

The transmembrane electrochemical gradient is then dissipated through the ATP synthase, which transforms this potential energy into chemical energy by phosphorylating ADP into ATP. The net reaction for the light phase of photosynthesis is of the form:



with  $x$  and  $y$  depending on the balance between cyclic and non-cyclic electron transport. With only non-cyclic electron transport,  $x=2$  and  $y=3$ .

During the light-independent phase, ATP and NADPH are used to produce ribulose-1, 5-biphosphate (RuBP), which comprises five carbon atoms. The ribulose-1, 5-biphosphate carboxylase/oxygenase (RuBisCO) then combines this substrate with  $\text{CO}_2$  to generate two molecules of 3-phosphoglycerate, which comprise three carbon atoms each. For every three RuBP used, six 3-phosphoglycerate molecules are generated. These are converted to glyceraldehyde-3-phosphate (G3P) by using six ATP and six NADPH molecules. Five G3P are then used to regenerate three RuBP molecules by using an extra three ATP, forming the Calvin cycle.

The sixth glyceraldehyde-3-phosphate is the net product of the cycle, and can be used for the production of six-carbon sugars or other metabolic pathways. The net reaction for one Calvin cycle is



### 3. Light harvesting complex: the phycobilisome

#### a. Overview

In 1836, Esenbeck extracted a “very nice sky blue”<sup>1</sup> water-soluble, photo-labile pigment from the freshwater cyanobacterium *Conferva* (possibly corresponding to the current genus *Oscillatoria*), described its susceptibility to light and various chemicals and attempted to relate it with the different colours observed for different cyanobacteria (von Esenbeck, 1836). Some years later, Kützing found in different cyanobacteria the same water-soluble blue pigment, which he renamed “phykokyan”, and isolated from red algae two water-soluble pigments exhibiting red (“phykoerythrin”) and red-brown (“phykohämatin”) colours and similar biochemical properties to “phykokyan” (Kützing, 1843). Two years after its description of fluorescence based on spectral properties of minerals, Stokes reported the strong fluorescence of biological pigments, including chlorophyll and phycoerythrin (Stokes, 1854). Sorby extended this description, and using thermal denaturation provided the first argument for the composite, multi-component nature of phycoerythrin<sup>2</sup> (Sorby, 1875). A few years later, Engelmann developed the key concept of action spectrum of photosynthesis and provided the first (indirect) evidence that phycoerythrin and phycocyanins were light-harvesting pigments used for photosynthesis

<sup>1</sup> The description is a very interesting reading (in German). A (tentative) translation of the first sentence is: “As it so often leads us to an interesting observation, we owe to chance once again the study described here.” Esenbeck wrote four times on the same page “striking [or intense], very nice colour”, and was mesmerized by it...

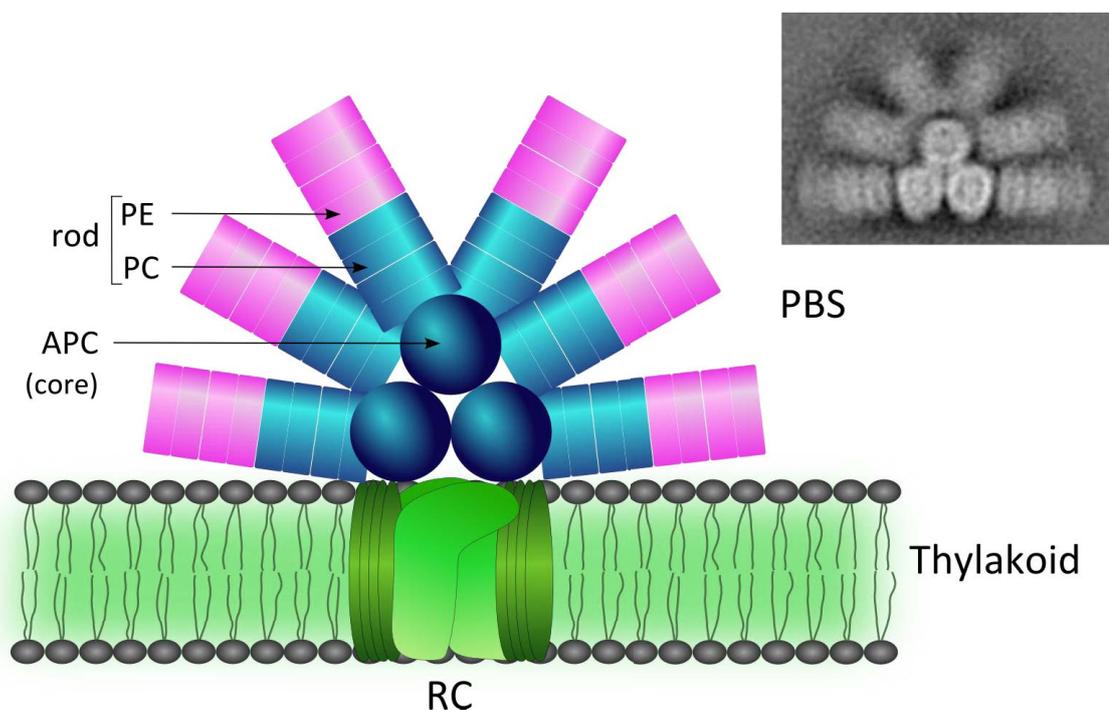
<sup>2</sup> In fact, he probably observed the heat-induced dismantling of phycobilisomes/phycobilisome rods.

(Engelmann, 1882; Drews, 2005; Tandeau de Marsac, 2003). In the 1900's, Kylin deduced from the properties of phycoerythrin and phycocyanin crystals that phycobiliproteins had a globin-like structure and were made of two components, one pigment (chromophore) and one apoprotein (Kylin, 1910, 1912), that Lemberg later managed to separate using acid or hot chloroform/ether treatments (Lemberg, 1933). Spectral characterization (in particular zinc-enhanced fluorescence) led them to suggest that these chromophores are related to bilicyanin, mesobiliviolin and urobilin, which are oxidation products of (meso)bilirubin pigments found in animal bile. They proposed to name the algal chromophores “phycocyanobilin” and “phycoerythrobilin”, and showed that they were essentially the same compound with different oxidation degrees. In the late 1950s, the advent of electronic microscopy combined with theoretical advances on energy transfer between fluorescent molecules, which suggested close proximity between light-harvesting antennae and reaction centres, led Gantt and Conti to the discovery of thylakoid-associated granules twice the size of ribosomes (32-40 nm) in chloroplasts of the red algae *Porphyridium cruentum* (**figure 4**; Gantt and Conti, 1965; Arnold and Oppenheimer, 1950). The subsequent purification of these granules by the same authors allowed them to show that they contained phycocyanin and phycoerythrin; the granules were named phycobilisomes (Gantt and Conti, 1966; see Tandeau de Marsac, 2003 for a more complete historical perspective on the research on phycobilisomes).



**Figure 4: First direct observation of phycobilisomes.** Electron micrograph of a cell from *Porphyridium cruentum*, magnification 43,000s. Note the relative size of ribosomes, dense granules attached on the endoplasmic reticulum (ER), and phycobilisomes, highly ordered and densely packed opaque granules located on thylakoid lamellae in the chloroplast (C). G, Golgi body; S, Starch. Reproduced from (Gantt and Conti, 1965).

Since their first description in red algae, phycobilisomes have been found in most cyanobacteria as well as in cryptophytes and glaucocystophytes, two groups of eukaryotic protists (Toole and Allnutt, 2003). Phycobilisomes differ from Light Harvesting Complexes (LHC) found embedded in the membrane of higher plants or other phytoplanktonic organisms (diatoms, but also the marine cyanobacteria *Prochlorococcus*, see II. 2.a) by its extramembranous localization (**figure 5**). These huge macromolecular complexes (3,000 to 16,800 kDa) can make up to 50% of the soluble protein of the cell, and are essentially made of chromophorylated phycobiliproteins assembled thanks to linker proteins (Sidler, 1994; Adir, 2005; Zhang *et al.*, 2017). Phycobilisomes can present different ultrastructures, but the most commonly found in cyanobacteria is hemidiscoidal<sup>3</sup> and comprises a core made of two, three or five cylinders of allophycocyanin functionally linked with the thylakoid membrane, from which six to eight rods radiate (**figure 5**; Watanabe and Ikeuchi, 2013; Watanabe *et al.*, 2014; Adir *et al.*, 2006; Arteni *et al.*, 2009; Guglielmi *et al.*, 1981). The rods can be made of phycocyanin only, or in combination with phycoerythrocyanin or phycoerythrin (Sidler, 1994).



**Figure 5: Model of phycobilisome associated to the thylakoid membrane.** The tricylindrical core is made of allophycocyanin (APC), the proximal part of rod of phycocyanin (PC) and the distal part of phycoerythrin (PE). The inset is an electron micrograph of isolated phycobilisome, reproduced from (Arteni *et al.*, 2009). Note that the phycobilisome is about 1/5 of its real scale (see **figure 13**). Figure reproduced from (Pittera, 2015).

<sup>3</sup> Fan-shaped

## **b. Components**

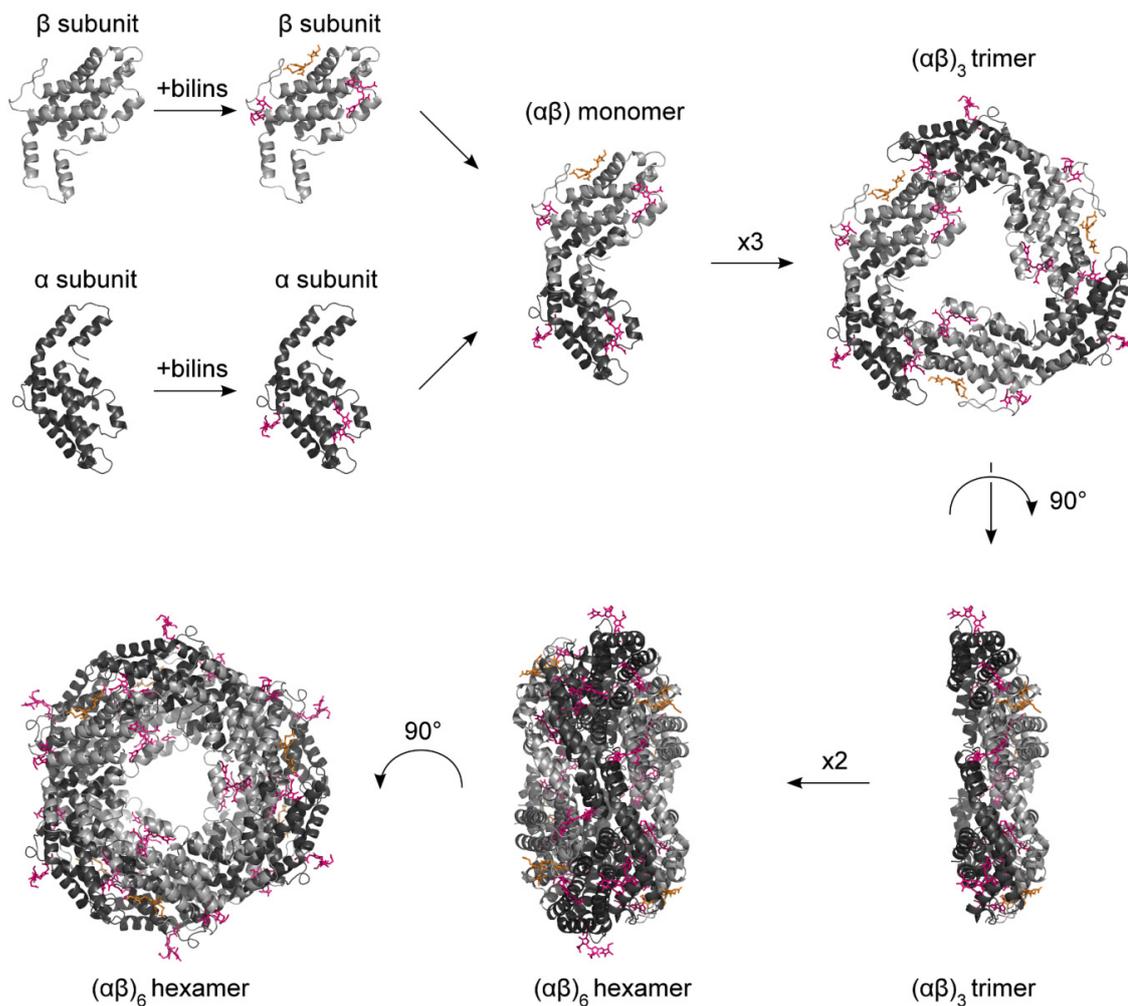
### *i. Phycobiliproteins*

Phycobiliproteins are usually made of two subunits termed  $\alpha$  and  $\beta$ , each about 160-184 amino acid in length and binding between one and three chromophore(s) on conserved cysteine residues (Sidler, 1994; MacColl, 1998). These subunits have a globin-like, boomerang-shaped structure made of eight  $\alpha$ -helices (X, Y, A, B, E, F, G, H), with the X and Y helices from one subunit interacting with the A and B helices of the other subunit (**figure 6**). Bilin attachment onto phycobiliprotein subunits is necessary for the assembly of ( $\alpha\beta$ ) monomers (Adir *et al.*, 2006). The assembly of subunits into monomers is spontaneous and highly stable, with mostly hydrophobic interactions and some residues “locking” the assembly through polar interactions. Recently, Pittera and co-workers suggested that substitutions of two residues located at the interface of  $\alpha$  and  $\beta$  subunits, near to these polar locks, could stabilize the monomer and represent adaptations to temperature in PC of marine *Synechococcus* (Pittera *et al.*, 2016; see Chapter 2). The ( $\alpha\beta$ ) monomer spontaneously assembles into ( $\alpha\beta$ )<sub>3</sub> trimers, forming a *tore*<sup>4</sup> of 11-12 nm in diameter, 3 nm thick and with a hole diameter of about 3 nm, the  $\alpha$  and  $\beta$  subunits forming the two different faces. Trimers then assemble into hexamers (about 6 nm tick) by face-to-face interactions, the  $\alpha$ -subunits being buried in the hexamer and the  $\beta$ -subunits being exposed. These hexamers are then stacked into rods with the help of linker proteins, which protrude in the central hole of hexamers (**figure 6**). While phycobiliprotein trimers are stable in solution at low ionic strength, this is not the case for phycobiliprotein hexamers, suggesting that the interaction with the linker protein is necessary for higher-order assembly (Adir *et al.*, 2006).

Four different rod phycobiliproteins have been described in cyanobacteria: phycocyanin (PC), phycoerythrocyanin (PEC) and phycoerythrin (PE)  $\beta$ - and  $\alpha$ -subunits are encoded by the *cpcBA*, *pecBA* and *cpeBA* operons respectively. PE-II, which is only found in marine *Synechococcus*, is encoded by the *mpeBA* operon. For allophycocyanin (APC), which make the core of the PBS, the operon *apcAB* codes for  $\alpha$ - and  $\beta$ -subunits. In addition, *apcD* and *apcF* code for two variants of APC,  $\alpha^{\text{APB}}$  and  $\beta^{18}$  (sometimes referred to as A-PB and  $\beta^{18.5}$ ) respectively. Both have a PCB attached, which is red-shifted compared to other PCB bound to APC  $\alpha$ - and  $\beta$ -subunits.  $\alpha^{\text{APB}}$  and  $\beta^{18}$  are involved in mediating the formation of core cylinders, the interactions between the phycobilisome and PSII as well as transferring excitation energy from the PBS to the PSII ( $\beta^{18}$ ) or PSI ( $\alpha^{\text{APB}}$ ). Finally, the linker protein ApcE or L<sub>CM</sub>, which is involved in anchoring the PBS to the membrane (see next paragraph), possesses a phycobiliprotein-like domain (**figure 7**, **figure 8**) and is also involved in the energy transfer to the PSII (Arteni *et al.*, 2009; Chang *et al.*, 2015; Dong *et al.*, 2009).

---

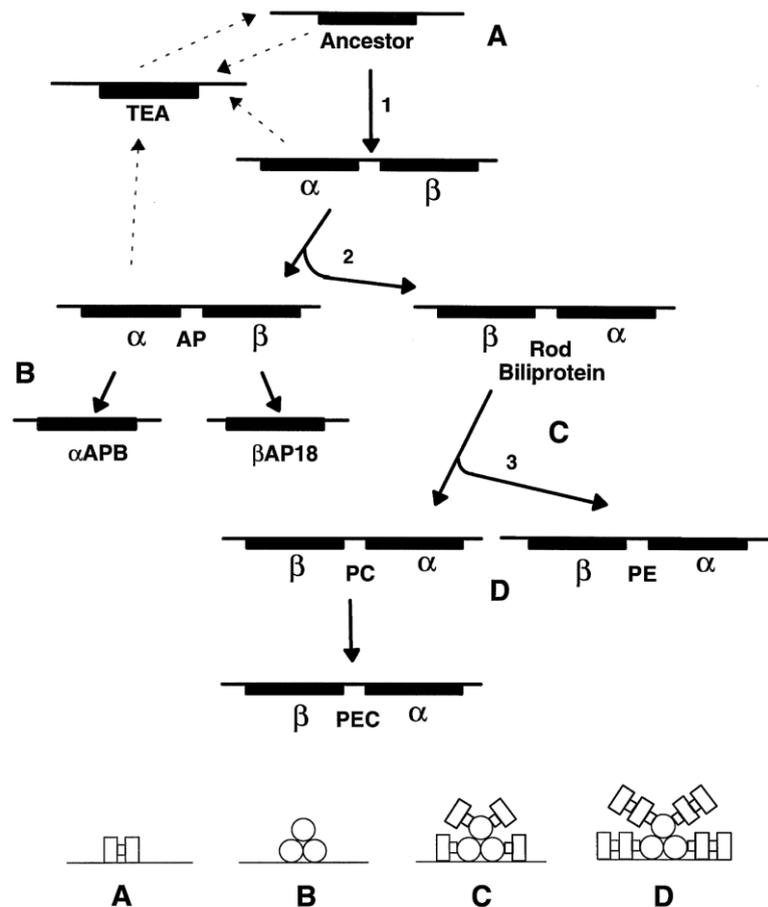
<sup>4</sup> Mathematical term for the shape of donuts



**Figure 6: Assembly of phycobiliprotein into hexamers.** Protein backbone of  $\alpha$ -subunit is displayed as dark gray ribbons, and in light gray for  $\beta$ -subunit. Phycoerythrobilins (located at  $\alpha 82$ ,  $\alpha 139$ ,  $\beta 82$  and  $\beta 158$ ) are displayed as pink sticks and phycourobilins (located at  $\beta 50$ ,  $61$ ) as orange sticks. The assembly into monomers and trimers is spontaneous and stable, but the hexamers are only stable when associated with linker protein (not shown here). The structure used corresponds to the phycoerythrin of the red algae *Griffithsia monilis* (PDB 1B8D).

Phycobiliproteins are found in very different domains of life (cyanobacteria, red algae, and cryptophytes), and have been early recognized as being closely related<sup>5</sup> (Berns, 1967). Further work confirmed the common evolutionary origin of phycobiliproteins, and showed that residues located in the vicinity of bilin-binding cysteines were more conserved, as were residues from the X and Y helices, confirming their functionally important role (Sidler *et al.*, 1990; Ducret *et al.*, 1994; Apt *et al.*, 1995; Zhao and Qin, 2006).

<sup>5</sup> Interestingly, this observation was among those used by Lynn Sagan when she formulated the modern theory of endosymbiosis exactly 50 years ago (Sagan, 1967).



**Figure 7: Scenario proposed by Apt and co-workers for the evolution of phycobiliproteins.** TEA, Terminal Energy Acceptor phycobiliprotein domain of the ApcE; AP, allophycocyanin; PC, phycocyanin; PE, phycoerythrin; PEC, phycoerythrocyanin. Figure reproduced from (Apt *et al.*, 1995).

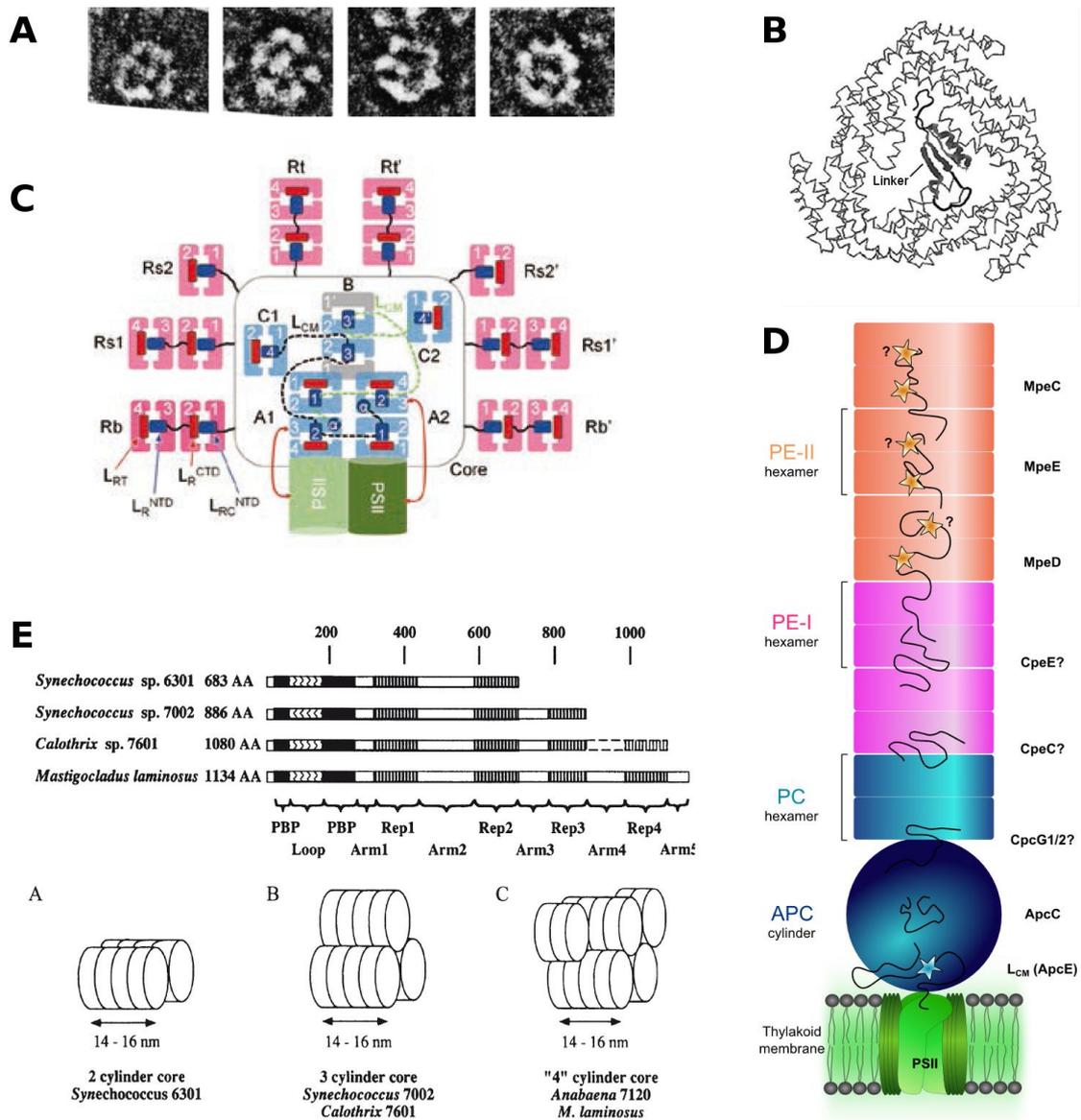
The scenario proposed by Apt and co-workers for phycobiliproteins involves the expansion of this family by several gene duplication events and coevolution (Apt *et al.*, 1995). They proposed a common ancestor possessing a similar globin-fold<sup>6</sup> to contemporary phycobiliproteins, binding a single bilin, able to multimerize, to interact with linker proteins and with the thylakoid membrane, from which a first gene duplication event led to the apparition of a tandem of gene, which coevolved into  $\alpha$  and  $\beta$  subunits (1 in **figure 7**). The co-localization and probable co-transcription of these ancestral subunits probably had a key role in their co-evolution. From these ancestral subunits, a second duplication (2 in **figure 7**) led to the advent of the ancestors of the PBS core subunits (APC) as well as of the rod subunits (PC). The APC ancestor eventually led to  $\alpha$ - and  $\beta$ -APC and to  $\alpha^{\text{APB}}$  (*apcD*) and  $\beta^{18.5}$  (*apcF*) through a duplication event (B in **figure 7**). Finally, the rod ancestor evolved into PC and PE  $\alpha$ - and  $\beta$ -subunits through a fourth duplication. In addition to the overall similarity of all these proteins,

<sup>6</sup> It is interesting to note that phycobiliproteins and globins share a similar structure and bind related pigments, bilins being derived from heme.

the conservation of the chromophore-binding  $\beta$ -82 (based on which organism/paralog is considered, the actual position varies between  $\beta$ -81 and  $\beta$ -84) in all of them is another key observation supporting this scenario. The number of chromophores per subunit increased from one up to three during the subsequent evolution of PC and PE, thus increasing the light absorption cross-section of phycobilisome rods at a relatively low cost. Similarly, the chromophores bound to phycobiliproteins diversified from PCB to PVB, PEB or PUB, allowing organisms to collect shorter wavelength. In particular, Everroad and Road suggested that in marine *Synechococcus*, positive selection drove the evolution of phycobiliproteins towards higher PUB content, and thus toward the absorption of blue photons (Everroad and Wood, 2012).

### ii. Linker proteins

Often colourless, linker proteins (linkers) represent about 15% of the phycobilisome proteins (Tandeau de Marsac and Cohen-bazire, 1977). These proteins are more variable than their phycobiliprotein counterparts, both within and between species. They act as an internal skeleton for the stacking of phycobiliprotein hexamers into rods by fitting inside the hole of phycobiliprotein hexamers, and are necessary for proper PBS assembly, thus constituting key determinants of the cohesion of core APC cylinders and phycobilisome rods and of the length of phycobilisome rods (**figure 8A,B,E**; Six *et al.*, 2007c, 2007b; Schluchter *et al.*, 2010). As an example of their key structural role, the linker *apcE*, which is involved in anchoring the PBS core to the thylakoid membrane and interacting with the photosystem II (PSII), comprises a variable number of a conserved Rep domain (Sidler, 1994; Arteni *et al.*, 2009; Chang *et al.*, 2015). This repeat number controls if there are two, three or five APC cylinders in the PBS core (**figure 8E**). Different linker proteins have been characterized, and classified depending on their position in the PBS: the linker core-membrane ( $L_{CM}$ , encoded by *apcE*), core linkers ( $L_C$ , involved in maintaining APC cylinders in the core, encoded by *apcC*), rod-core linkers ( $L_{RC}$ , involved in the attachment of phycobilisome rods onto the APC core, encoded by different *cpcG* variants, from *cpcG1* to *cpcG4*) and rod linkers ( $L_R$ , for assembling the PC and PE hexamers into rods, encoded by *cpcC*, *cpcD*, for PC-associated linkers, *cpeC*, *cpeD* and *mpeD* for PEI-associated linkers, and the *mpeC*, *mpeE*, *mpeF*, *mpeG* and *mpeH* genes found only in marine *Synechococcus* for PEII-associated linkers; **figure 8C-D**; Sidler, 1994; Six *et al.*, 2007c; Liu *et al.*, 2005; Watanabe and Ikeuchi, 2013). The structural role of linker proteins is made possible by their slightly positive charge (basic) combined with the slightly negative charge (acidic) of phycobiliproteins (Liu *et al.*, 2005).



**Figure 8: Linker proteins.** (A) Electron micrographs of cyanobacterial phycocyanin, showing the PC ring-shaped hexamer filled with a central object. Reproduced from (MacColl, 1998). (B) Structure of the complex APC-L<sub>C</sub><sup>7.8</sup> (ApcC) from *Mastigocladus laminosus* (*Fischerella* sp.), showing how the linker protein (in ribbons) is buried in the phycobiliprotein trimer/hexamer (in sticks). Note the three β-strands assembled into a sheet and the random coil segments. Reproduced from (Liu *et al.*, 2005). (C) Model of PBS structure showing the localization of linker proteins in the phycobilisome rods and core cylinders of *Nostoc* sp. PCC7120. Reproduced from (Chang *et al.*, 2015). (D) Putative model for the localization of linker proteins in the phycobilisome rod of *Synechococcus* sp. strain WH8102. Unsure localization of linker protein is indicated by a "?". ApcE bears a PCB chromophore, and MpeD, MpeE and MpeC each bear one or two PUB chromophore. Modified after (Six *et al.*, 2005; Pittera, 2015). (E) Domain structure of the core-membrane linker ApcE (L<sub>CM</sub>). The number of Rep domains in ApcE determines the number of APC cylinders in the PBS core. Reproduced from (Sidler, 1994).

Linker 3D structure and organization in the PBS have been historically more difficult to resolve than the phycobiliprotein trimers/hexamers due to their more disordered structure compared to phycobiliproteins (**figure 8B**), but recent advances in cryo-electromicroscopy<sup>7</sup> allowed the resolution of entire phycobilisomes structures (Liu *et al.*, 2005; Chang *et al.*, 2015; Zhang *et al.*, 2017). It is interesting to note that this linker possesses some  $\beta$ -strands whereas all phycobiliproteins are solely made of  $\alpha$ -helices (Liu *et al.*, 2005; Tal *et al.*, 2014). Most recently, Chang and co-workers described the structure of intact phycobilisomes with the localization of all conserved linker domains in a PBS, allowing them to propose a global structural model of PBS interacting with the PSII (**figure 8C**; Chang *et al.*, 2015).

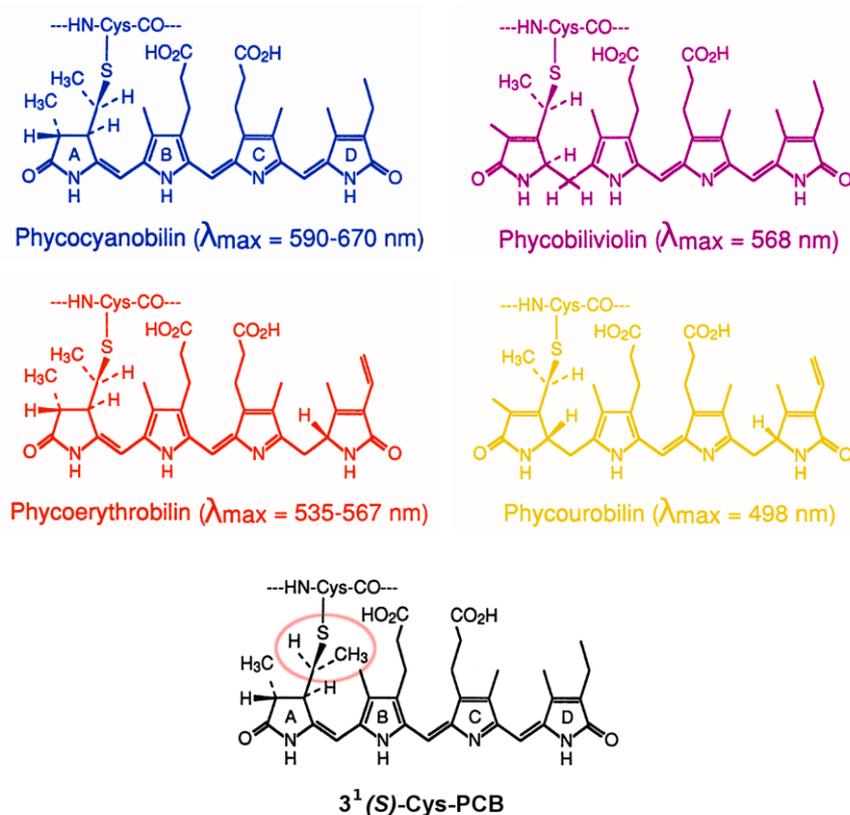
In addition to their structural role, linkers also participate indirectly to the light-harvesting function of phycobilisomes by modulating the spectral properties of the phycobilins in phycobiliprotein hexamers, thus ensuring the overall efficiency and unidirectionality of excitation energy transfer in the PBS (Pizarro and Sauer, 2001; Liu *et al.*, 2005; Watanabe and Ikeuchi, 2013). Also, some linkers are chromophorylated and might participate in light-harvesting:  $L_{CM}$  bears a red-shifted PCB ( $A_{max} = 668-676$  nm) that has the key role of transferring light energy initially received by any chromophore of the PBS to the PSII, constituting the terminal acceptor of the PBS for this transfer (Lundell *et al.*, 1981; Chang *et al.*, 2015; Zhao *et al.*, 2005a). In marine *Synechococcus*, the rod linkers ( $L_R$ ) MpeC, MpeD and MpeE all bear at least one and possibly two PUB chromophores and could be involved in light-harvesting (Six *et al.*, 2005, 2007c).

### iii. Bilins

Four different phycobilins have been found in cyanobacteria and red algae: phycocyanobilin (PCB,  $A_{max} \sim 620$  nm), phycoerythrobilin (PEB,  $A_{max} \sim 550$  nm), phycourobilin (PUB,  $A_{max} \sim 500$  nm) and phycoviolobilin (PVB,  $A_{max} \sim 590$  nm), of which only PCB, PEB and PUB have been observed in marine picocyanobacteria (**figure 9**).

Each cyanobacterial phycobiliprotein subunit can bind from one to three (four for a linker protein found in red algae) phycobilins on conserved cysteine residues, the number of bilins linked increasing from APC to PE-I and to PE-II in marine *Synechococcus*. In the latter organisms, this results in up to 14 different cysteine residues capable of binding phycobilins per  $\alpha\beta$  monomer: three on PC ( $\alpha$ -84,  $\beta$ -84,  $\beta$ -155), five on PE-I ( $\alpha$ -84,  $\alpha$ -143,  $\beta$ -84,  $\beta$ -155 and the doubly linked  $\beta$ 50, 61) and six on PE-II (same as on PE-I plus  $\alpha$ -75; **figure 10**).

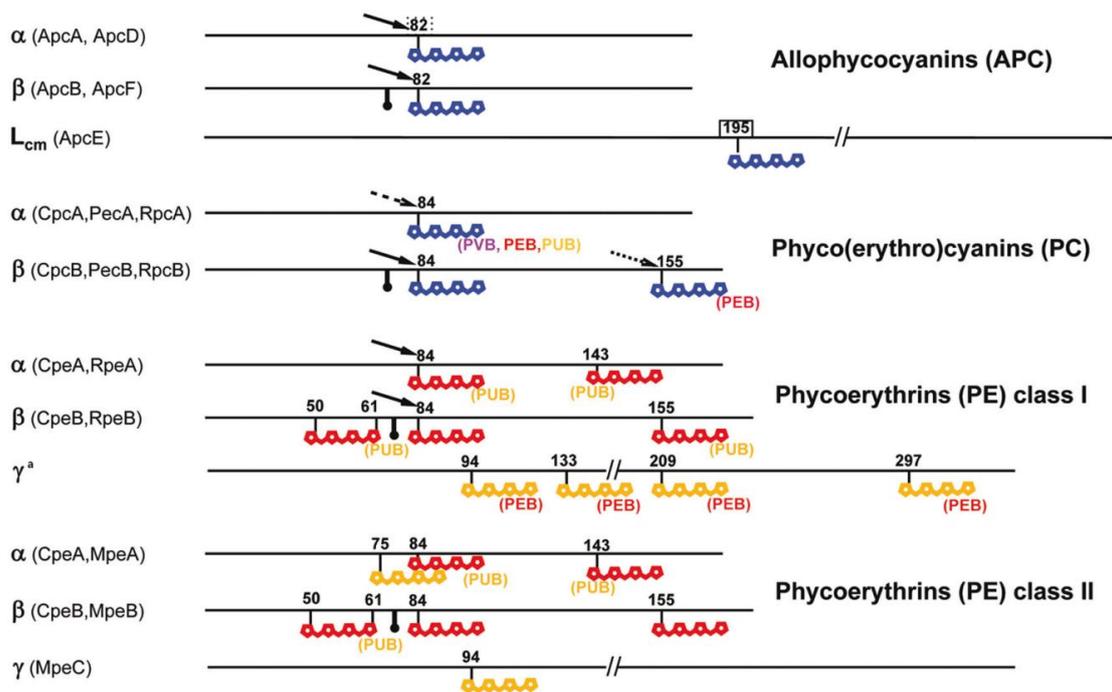
<sup>7</sup> Dubochet, Frank and Henderson received this year the Nobel Prize in Chemistry for their contribution to the development of this method.



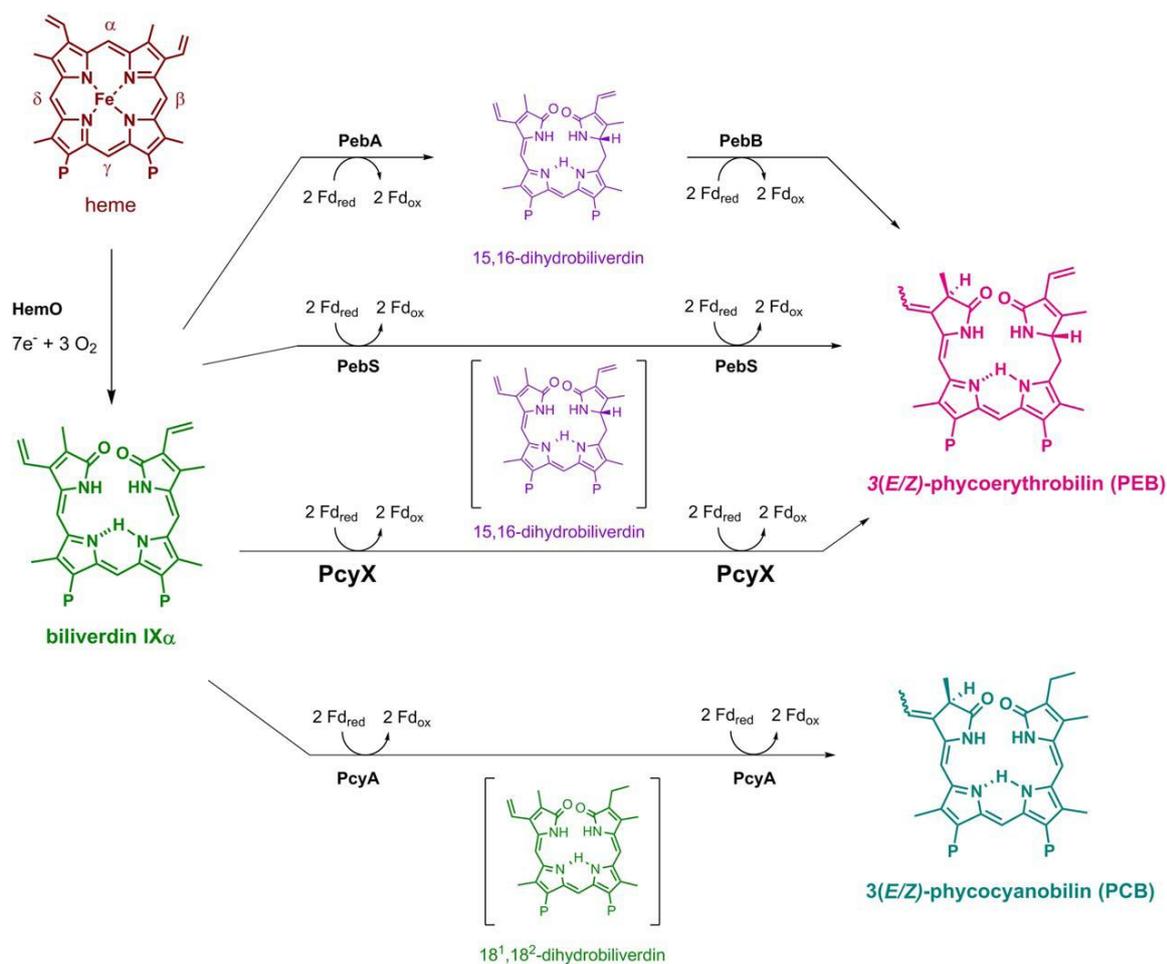
**Figure 9: Phycobilins present in cyanobacteria and red algae, with their thioether bound to phycobiliproteins.** Note that PCB, PVB, PEB and PUB are all isomers, with PCB and PVB differing from PEB and PUB by their A ring, and PCB and PEB differing from PVB and PUB by their D ring. Also note the two possible stereoisomers for the attachment to phycobiliprotein: the top four are  $3^1(R)$  stereoisomers, the bottom one is a  $3^1(S)$  stereoisomer. Modified from (Grossman *et al.*, 1993).

Phycobilins are linear tetrapyrroles derived from heme. Their synthesis first requires the opening of the heme ring in an oxygen-dependent step, which is catalyzed by a heme oxygenase encoded by *hemO* (sometimes designated by *ho1*), resulting in biliverdin IX $\alpha$  (**figure 11**; Frankenberg-Dinkel, 2004). In cyanobacteria, this precursor is then reduced by different ferredoxin-dependent bilin reductases (FDBR) to PCB or PEB (Frankenberg *et al.*, 2001). While PCB synthesis is catalyzed by one enzyme (PcyA) in a two-step reaction, biliverdin IX $\alpha$  reduction to PEB requires first PebA to reduce it to 15, 16-dihydrobiliverdin, which is further reduced into PEB by PebB. Interestingly, homologs of genes implied in this metabolic pathway (both heme oxygenases and FDBR) have been found in viruses infecting marine picocyanobacteria (Dammeyer *et al.*, 2008; Ledermann *et al.*, 2016). These investigations also revealed a new bi-functional enzyme (PebS) homolog to PebA and PebB but sufficient to

catalyze both steps of the reduction of biliverdin IX $\alpha$  into PEB (Dammeyer *et al.*, 2008). More recently, a distant homolog to PcyA named PcyX has been found in viral metagenomes and shown to catalyze the same reaction as PebS (**figure 11**; Ledermann *et al.*, 2016). In contrast to PCB and PEB, PVB and PUB are not directly synthesized in cyanobacteria. Instead, they are produced during the covalent attachment to phycobiliprotein by bi-functional enzymes called phycobilin lyase-isomerases that concomitantly attach and isomerise bilins. In some freshwater cyanobacteria, PecE/F isomerizes PCB bound to phycoerythrocyanin into PVB, whereas in marine *Synechococcus*, PUB is being produced from PEB by different enzymes, of which RpcG and MpeZ have been biochemically characterized (Zhao *et al.*, 2000, 2005b; Blot *et al.*, 2009; Shukla *et al.*, 2012; see also Chapter III).



**Figure 10: Chromophorylation of cyanobacterial phycobiliproteins.** Sites are numbered according to consensus, and bilins are indicated by different colours: PCB in blue, PVB in purple, PEB in red and PUB in orange. Some homologous sites always bind the same bilin whatever the species/strains, whereas others can accommodate different bilins in different species/strains. The most variable site is Cys- $\alpha$ 84, to which the attachment of all four chromophores has been reported. Reproduced from (Scheer and Zhao, 2008).



**Figure 11: Biosynthetic pathways of phycobilins.** Phycobilins are derived from biliverdin IX $\alpha$ , an oxygenation product of heme. The reduction to PCB is catalyzed by PcyA. The reduction to PEB implies either two successive enzymes (PebA and PebB), or the action of the viral-encoded FDBR PebS or PcyX. Reproduced from (Ledermann *et al.*, 2016).

Phycobilins are attached to conserved residues on apoproteins by thioether covalent bonds on their A ring (**figure 9**; Glazer, 1989). Some bilins are doubly attached by both A (C3) and D (C18) rings, as observed for PEB and PUB bound to Cys 50 and Cys 61 of the  $\beta$ -subunit of PE in marine *Synechococcus*. Two possible stereoisomers exist for the carbon of the thioether bond, the (*R*) one being the most commonly observed but the (*S*) also being observed (e.g. on Cys-153  $\beta$ -PC; Shen *et al.*, 2006; Schluchter *et al.*, 2010). PVB and PCB on one hand and PEB and PUB on the other hand only differ by the localization of C:C double bond, located either on the A ring or between A and B rings (**figure 9**). This subtle modification shortens the  $\pi$ -conjugated system in PVB and PUB, resulting in blue-shifted absorption maxima ( $A_{\max}$ ) relative to PCB and PEB respectively. Similarly, the (*E*) configuration of double bonds and high torsion angles between successive pyrrole rings both reduce the effective length of the  $\pi$ -conjugated

system compared to (*Z*)<sup>8</sup> and planar configurations of bilins, and result in blue-shifted light absorption and emission peaks (Falk, 1989; Rossi *et al.*, 1989). The protein environment also influences the absorption/emission spectrum of bilins, for example through excitonic coupling with nearby chromophores or by changing the protonation state of pyrrole rings, a property exploited by some cyanobacteriochromes (Hirose *et al.*, 2013; Tang *et al.*, 2015; see Falk, 1989; Zhao *et al.*, 2012 for extensive reviews of bilin properties). The extended conformation strongly enhances bilins visible light absorption properties compared to free bilins which adopt the almost circular configuration<sup>9</sup> that can be seen (**figure 11**), and also enhances their excitation lifetime by up to four orders of magnitude (Zhao *et al.*, 2012). Interestingly, the geometry of PC-attached bilins depends on the interaction of the nitrogen atom of pyrrole rings B and C with the oxygen atom of conserved aspartate residues, a role similar to iron in heme (Schirmer *et al.*, 1987). Finally, Wang and Moerner recently hypothesized that the environment and other nearby bilins could also result in different light absorption *probabilities* between otherwise chemically identical bilins (Wang and Moerner, 2015).

### ***c. Energy transfer: structure-function relationship***

Light-harvesting complexes are highly optimized structures and represent elegant solutions to the challenge of both absorbing light energy and transferring this energy before pigments dissipate it through radiative (fluorescence) or non-radiative (photoisomerisation, heat) decay. To this end, light-harvesting complexes extend the lifetime of excited pigments by forcing them in a particular conformation (see previous paragraph), and densely yet minutely arrange pigment molecules in space so that when one pigment absorbs a photon, a nearby pigment is able to receive the resulting excitation through excitonic coupling or Förster resonance energy transfer at a much higher rate (usually in the picosecond<sup>-1</sup> range for PBS) than the one of excitation decay (usually in the nanosecond<sup>-1</sup> range; see for example Doust *et al.*, 2004; Acuña *et al.*, 2017; Sauer and Scheer, 1988). This results in an energetically downstream chain of pigments that eventually leads to a final acceptor (**figure 12C**). This final acceptor subsequently transfers the excitation energy to a reaction centre, where the excitation energy will be converted to chemical energy through charge separation (for a review on the diversity of bacterial light harvesting strategies, see Saer and Blankenship, 2017 and references therein).

In this regard, the phycobiliproteins are particularly well adapted to their function. Phycobiliproteins maintain phycobilins in a proper extended and rigid conformation, in order to give them their functional properties. The protein environment also generates slightly different

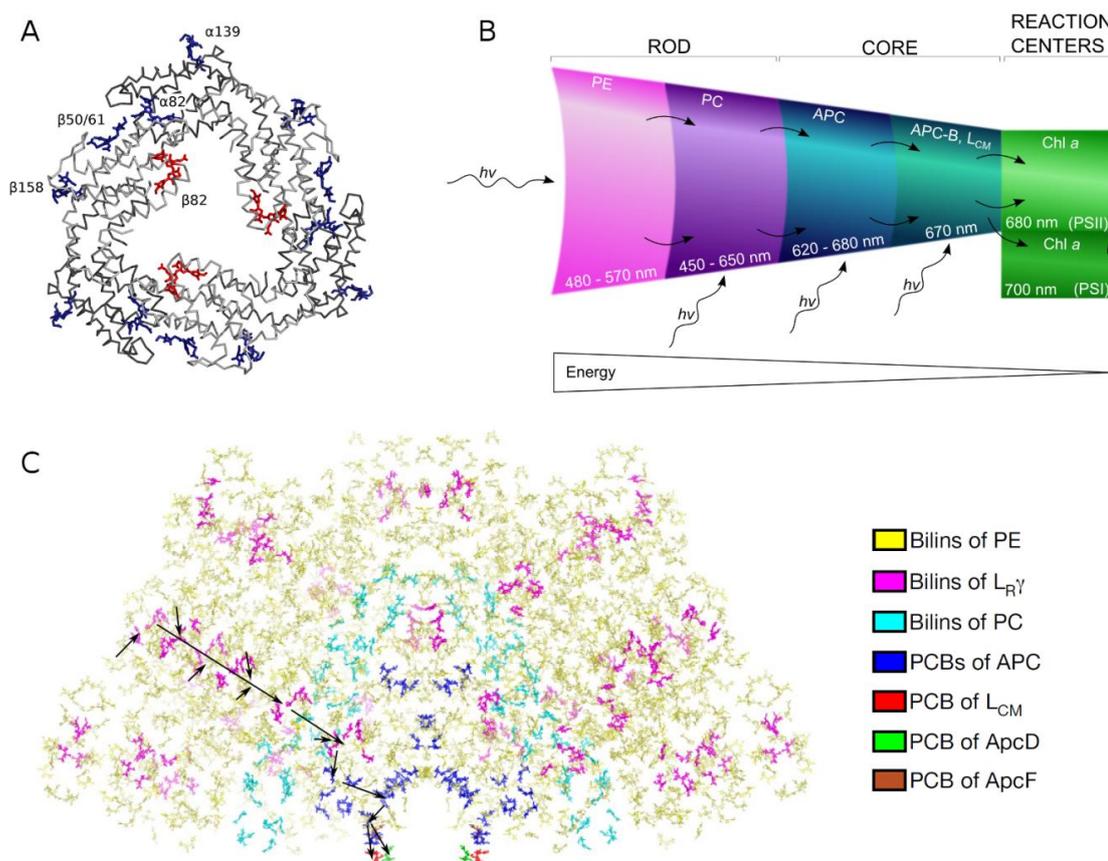
---

<sup>8</sup> The (*E*) configuration is hindered by steric hindrance of pyrrole ring substituents, resulting in high torsion angles of successive rings, which are the true cause for shortened conjugated  $\pi$  systems (Zhao *et al.*, 2012).

<sup>9</sup> The 3D conformation is more helical, due to the steric hindrance of C=O links in rings A and D (Zhao *et al.*, 2012).

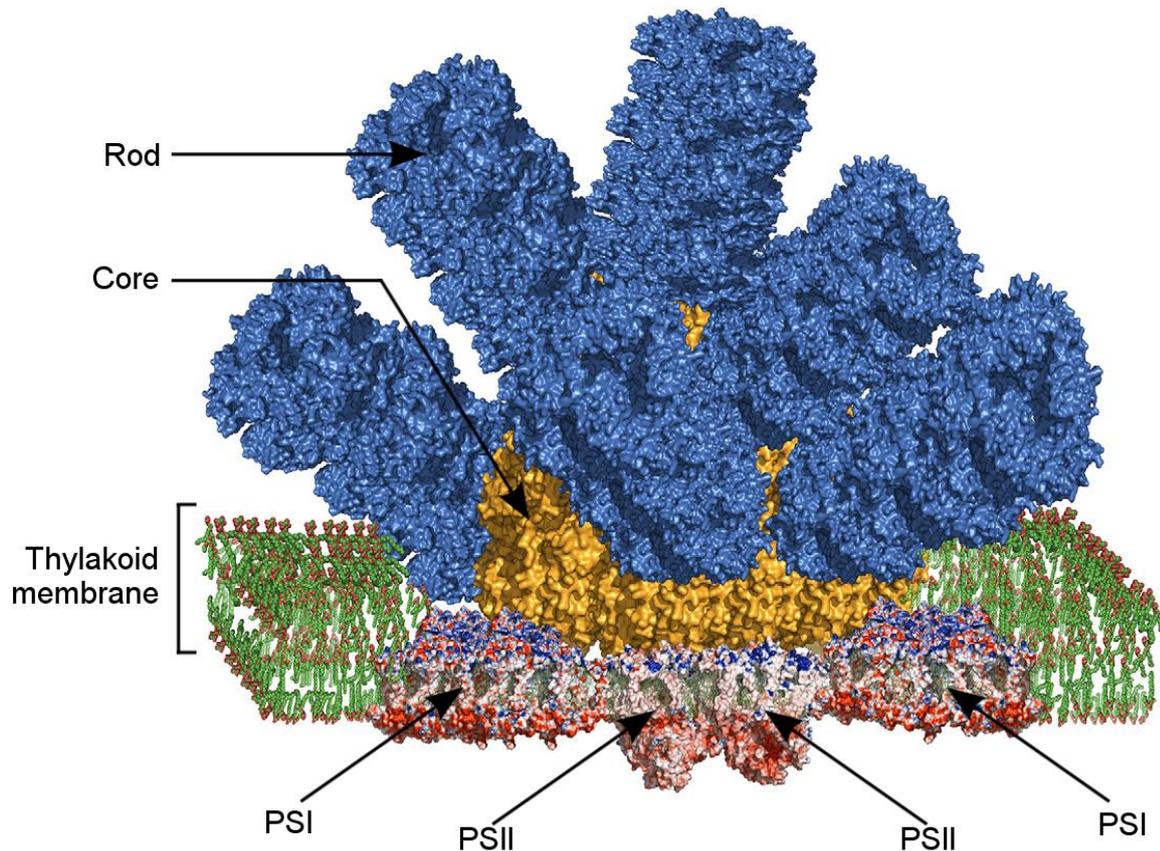
spectral properties for otherwise chemically identical pigment molecules, a crucial condition for proper directional energy transfer. A nice example of such protein-determined spectral properties is the terminal acceptor  $L_{CM}$  (ApcE), in which PCB exhibits an extreme yet functionally important red-shift ( $A_{max} = 665$  nm,  $F_{max} = 676$  nm) when compared to free PCB (Lundell *et al.*, 1981; Tang *et al.*, 2015; Zhao *et al.*, 2005a). Similarly, the multimerization of phycobiliprotein subunits can also change spectral properties compared to the monomeric state. Upon trimerization, the absorption peak of APC is red-shifted of about 30 nm (~650 nm) compared to the monomeric state (~620 nm). This shift has been attributed to the PCB located on  $\alpha$ -APC and related with site-specific trimerization-induced structural changes (Peng *et al.*, 2014; Wang and Moerner, 2015). The different phycobilins have decreasing energy levels both between different phycobiliproteins (e.g. between phycoerythrin and phycocyanin), but also between bilins bound to the same trimer. In particular, spectroscopic studies revealed that the conserved  $\beta$ -84 bilin, which is located near the central hole in phycobiliprotein hexamers, functions as an acceptor for the energy collected by the more peripheral  $\alpha$ -84 and  $\beta$ -155, and the different bilins have been classified as *f* (fluorescent, the  $\beta$ -84 bilin) of *s* (sensitive,  $\alpha$ -84 and  $\beta$ -155; **figure 12A**; Teale and Dale, 1970).

The complex structure of the phycobilisome, with different parts of varying bilin and protein composition (a core of APC and rods made of one proximal PC and several PC or PE, see also II. 3.c.ii), allows collecting different light colours, with the outer part of the phycobilisome (i.e. distal end of the rods) absorbing shorter wavelength compared to the inner part (phycobilisome core), as well as an efficient unidirectional transfer of energy from the distal to the proximal end of the rods, then to the core and to one of the two potential the final acceptors,  $L_{CM}$  (for transfer to PSII) or AP-B (for transfer to PSI; **figure 12B**; Acuña *et al.*, 2017). Glazer estimated that the efficiency for the whole transfer was above 90% and suggested the idea of “light guide” for the phycobilisome (Glazer, 1989). More recently, the advances in ultrafast spectroscopy (“time-resolved spectroscopy”) and single-molecule analyses allowed investigating how the energy is transferred both in isolated PBS components (core only, hexamers only) or in whole PBS. This revealed that the energy transfer both within and between rod hexamers is fast, and that the limiting steps are the transfer between PBS rods and the core, and between the different cylinders of the core (van Stokkum *et al.*, 2017; Squires and Moerner, 2017; Wang and Moerner, 2015).



**Figure 12: Energy transfer in phycobilisomes. (A)** Sensitive (in blue) and fluorescent (red) phycobilins in a phycoerythrin trimer (PDB 1B8D), according to Teale and Dale classification (Teale and Dale, 1970). **(B)** Excitation energy transfer from the distal to the proximal end of the rod, then to the core and to the photosystem II or I. Reproduced from (Pittera, 2015). **(C)** Positioning of chromophores in the phycobilisome of the red algae *Griffithsia pacifica*, with the (putative) energy transfer represented. Note how the two potential terminal acceptors ApcD (green) and  $L_{CM}$  (red) are found near to each other and to the thylakoid membrane (not shown, under the PBS). Modified from (Zhang *et al.*, 2017).

Finally, Liu and co-workers recently demonstrated using a combination of protein cross-linking, mass spectrometry and spectroscopy that the phycobilisome could be associated in vivo to both PSII and PSI in a “megacomplex” in *Synechocystis* PCC6803, thus explaining previous results showing that there are two potential terminal acceptors in the PBS (**figure 13**; Liu *et al.*, 2013).



**Figure 13: Structural model of a phycobilisome associated with both photosystem II (PSII) and photosystem I (PSI) in a megacomplex.** The complex is made of one phycobilisome attached to a dimer of PSII. Two trimeric PSI surround the PSII dimer. Note the PSII downward extensions in the lumen, which correspond to the (accessory) oxygen evolving complex, and the stacking of phycobiliprotein hexamers into rods. Modified from (Blankenship, 2015).

#### ***d. Phycobilin lyases***

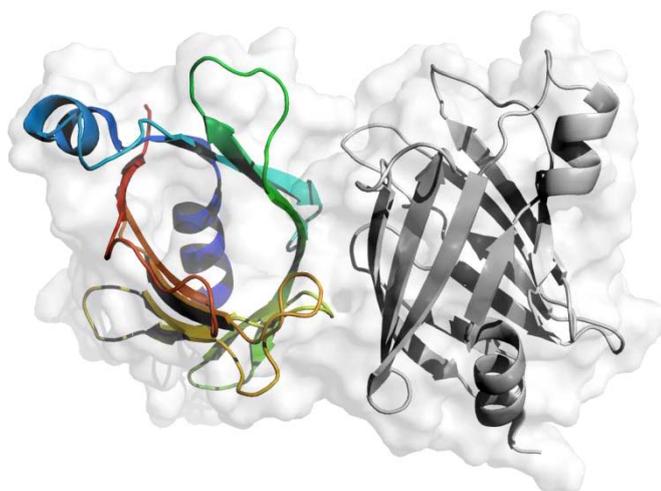
The attachment of phycobilins to phycobiliproteins can occur spontaneously in heterologous expression or *in vitro* experiments, but these spontaneous reactions rarely ensure the correct stereochemistry, and the resulting chromophorylated proteins are not found *in vivo* (Scheer and Zhao, 2008). The only known exception to this general rule is the  $L_{CM}$ , for which the autocatalytic attachment has been demonstrated by (Liu *et al.*, 2005). Six and co-workers suggested that the chromophorylation of PE-II associated linkers found in marine *Synechococcus* could also be autocatalytic, but this has not been demonstrated yet (Six *et al.*, 2007a). For every other bilin, specific enzymes called phycobilin lyases ensure the proper attachment of the correct bilin on the correct phycobiliprotein subunit and in the correct stereochemistry, suggesting a chaperone-like mechanism (Schluchter *et al.*, 2010). Moreover, Zhao and co-workers suggested that the ligation of bilins to phycobiliproteins has to occur in a specific order based on the observation that  $\beta$ -84 chromophorylation inhibited chromophore binding at  $\beta$ -155 but not the opposite (Zhao *et al.*, 2007b). Phycobilin-lyases are very diverse, and have been classified into

three distinct structural clans (i.e. enzymes inferred to share a similar 3D structure), the T clan, the S/U clan and the E/F clan, grouping about 30 different families (i.e. groups of orthologs; Bretaudeau *et al.*, 2013); for more complete reviews on phycobilin lyases see (Scheer and Zhao, 2008; Schluchter *et al.*, 2010). The naming of phycobilin lyases generally corresponds to the name on the phycobiliprotein on which they act, with Cpc/Rpc enzymes acting on phycoerythrin, Cpe on phycoerythrin/phycoerythrin-I, Mpe on phycoerythrin-II and Pec on phycoerythrocyanin. Phycobilin-lyases can be active as homo- or heterodimers (see the CyanoLyase web site <http://cyanolyase.genouest.org/>, described in Bretaudeau *et al.*, 2013, for a complete listing of phycobilin lyase families and their predicted specificity).

#### *i. T clan*

Enzymes of the T clan have not been reported to form heterodimers. CpcT has been characterized in *Synechococcus* sp. PCC7002 (which PBS rods only contain PC) and attaches PCB at PC  $\beta$ -153 (Shen *et al.*, 2006). In the PC- and PEC-containing *Nostoc* sp. PCC7120, CpcT also attaches PCB at PEC  $\beta$ -155, the binding site structurally homologous to PCC 7002 PC  $\beta$ -153 (Zhao *et al.*, 2007b). *Nostoc* PCC7120 CpcT has been crystallized alone or with PCB and its structure resolved, confirming that it forms a homodimer, and showing that it adopts a calyx shaped  $\beta$ -barrel (**figure 14**; Zhou *et al.*, 2014). A reaction mechanism accounting for the region- and stereospecificity has also been proposed by the same authors.

RpcT is only found in marine *Synechococcus*. Its function has not been formally biochemically assessed, but its similarity with CpcT, its phyletic pattern (i.e. distribution among *Synechococcus* isolates) and the characterization of the chromophores attached to PC in strains possessing RpcT all concordantly suggest that it has a PC  $\beta$ -153 PEB lyase function (Ong and Glazer, 1987; Six *et al.*, 2007c; Blot *et al.*, 2009; Bretaudeau *et al.*, 2013).



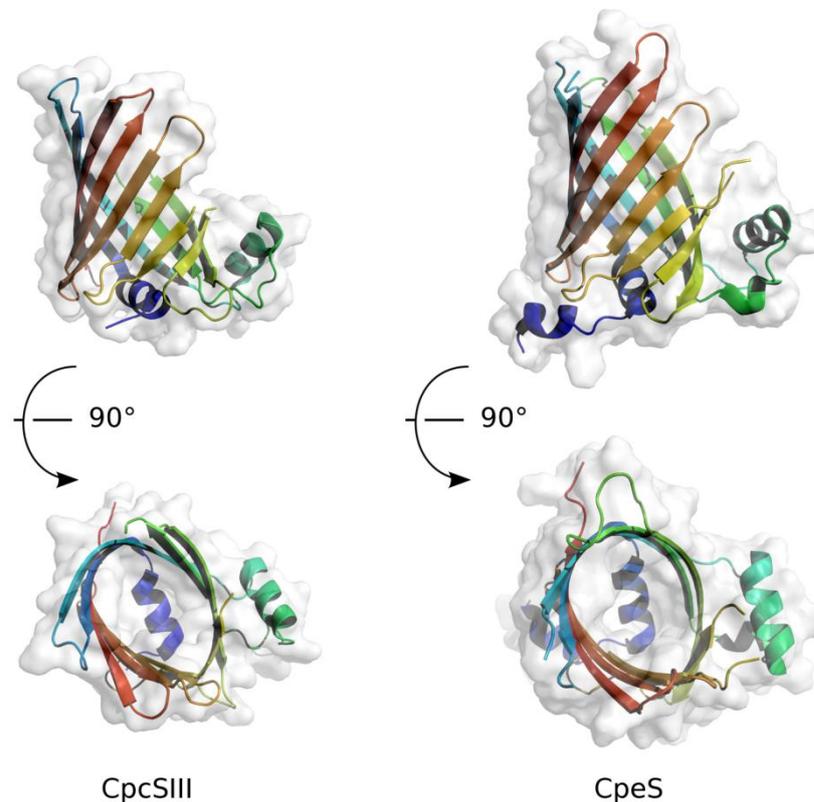
**Figure 14:** 3D structure of a  $\phi$ CpeT dimer from the P-HM1 cyanophage (PDB 5HI8).

A CpeT homolog found in the P-HM1 cyanophage, i.e. a virus infecting cyanobacteria (in this case *Prochlorococcus marinus* strain MED4), has recently been biochemically characterized and crystallized (Gasper *et al.*, 2017). This revealed that  $\phi$ CpeT adopts a similar yet tighter/smaller form than CpcT, a result that the authors related to the phage origin of this enzyme. Discarding crystallographic and biochemical results did not allow them to determine whether  $\phi$ CpeT is active as a monomer or a homodimer. Surprisingly, while  $\phi$ CpeT can stably bind PEB, Gasper, Schwach and co-workers did not observe any direct lyase activity for this enzyme. However, *Prochlorococcus* MED4 (and more generally all high-light-adapted *Prochlorococcus* isolates, see II. 3.b) contains only one phycobiliprotein subunit, which is a divergent form of the  $\beta$ -subunit of phycoerythrin (PpeB) and only has one chromophore binding site ( $\beta$ -82; Hess *et al.*, 1996). As other characterized T clan lyases are specific of the  $\beta$ -153 consensus binding site,  $\phi$ CpeT could have a  $\beta$ -155 lyase function in hosts having a more conventional phycoerythrin, i.e. a low-light *Prochlorococcus* or a *Synechococcus* strain. Alternatively, Gasper, Schwach and co-workers suggested that  $\phi$ CpeT could assist the *Prochlorococcus* host lyase CpeS in binding PEB to  $\beta$ -82 (Gasper *et al.*, 2017).

#### ii. S/U clan

Schluchter and co-workers recently suggested differentiating homologs of CpcS and CpcU, and renamed those CpcS-I to CpcS-III and CpcU-I or CpcU-II (Schluchter *et al.*, 2010). CpcS can be active either as a homodimer (CpcS-III) or as heterodimer (CpcS-I/CpcU-I and probably CpcS-II/CpcU-II, in which case both proteins are necessary for the catalytic function; Schluchter *et al.*, 2010). CpcS-III and CpcS-I/CpcU-I have been shown to catalyze the attachment of PCB on the cysteine 82 of a broad range of phycobiliprotein substrates:  $\beta$ -PC,  $\beta$ -APC,  $\beta$ -PEC, ApcF ( $\beta^{18.5}$ ), but also  $\alpha$ -APC and ApcD ( $\alpha^{ABP}$ ; Zhao *et al.*, 2006, 2007b, 2007a; Saunée *et al.*, 2008; Shen *et al.*, 2008b; Biswas *et al.*, 2010). CpcS-II/CpcU-II is only present in marine *Synechococcus* and has not been formally characterized yet, but likely exhibit the same substrate (both bilin and apoprotein) specificity (Bretaudeau *et al.*, 2013).

Biswas and co-workers demonstrated that CpeS is a lyase attaching PEB at  $\beta$ 82-PEI in the freshwater cyanobacteria *Fremyella diplosiphon* and Wiethaus and co-workers showed a similar role for *Prochlorococcus* CpeS, which attaches PEB on cysteine 82 of PpeB (Biswas *et al.*, 2011; Wiethaus *et al.*, 2010b, 2010a). CpeS from marine *Synechococcus* have not been characterized yet, but probably have the same function, possibly with the help of CpeU, a CpcU paralog co-occurring with CpeU in these organisms (Bretaudeau *et al.*, 2013 and F. Partensky, personal communication).



**Figure 15: 3D structure of two members from the S/U clan.** Left, CpcSIII from *Thermosynechococcus elongatus* (PDB 3BDR); right, CpeS from *Guillardia theta* (PDB 4TQ2).

Interestingly, some putative homologs of CpcS do not have a lyase activity (Saunée 2008, Shen 2008, Zhao 2006). The structure of CpcS-III from *Thermosynechococcus elongatus* strain BP1 exhibits a  $\beta$ -barrel structure with an additional  $\alpha$ -helix and belongs to the lipocalin structural family, some members of which being known to bind bilins (**figure 15**; Kuzin *et al.*; Schluchter *et al.*, 2010).

### iii. E/F clan

In contrast to members of the S/U clan, and especially CpcS, phycobilin lyases of the E/F clan are highly specific of one particular attachment site on one phycobiliprotein subunit, and of one particular bilin. No 3D structure has been obtained yet for this clan, but domain search and homology modelling suggest that members of this clan contains Armadillo repeats, which are made of pairs of alpha-helices forming a hairpin structure. They possess HEAT repeats, which are motifs found in protein/protein interactions.

CpcE and CpcF are both necessary to form a heterodimer that is a phycobilin lyase adding PCB at  $\alpha$ -84 PC in *Synechococcus* sp. PCC 7002 (Fairchild *et al.*, 1992; Zhou *et al.*, 1992). In *Fischerella* sp. PCC 7603, the CpcE/F homologs PecE and PecF form a phycobilin lyase-isomerase, which both catalyzes the attachment of PCB at  $\alpha$ -84 of phycoerythrocyanin

(PEC) and its isomerisation into PVB (Zhao *et al.*, 2000). A fusion protein with a N-terminus showing similarity to PecE and a C-terminus to PecF, called RpcG, is present in some marine *Synechococcus* and has been demonstrated to be a phycobilin lyase-isomerase, also acting on Cys-84 of  $\alpha$ -PC but adding PEB and converting it to PUB (Blot *et al.*, 2009). Finally, based on their similarity to CpcE/F and phyletic pattern, RpcE and RpcF have been predicted to be involved in PEB attachment at Cys-82 of  $\alpha$ -PEI (Six *et al.*, 2007c), but this putative function still has to be confirmed experimentally.

Similarly, CpeY and CpeZ form a heterodimer that attaches PEB at  $\alpha$ -82 PEI in *F. diplosiphon* (Biswas *et al.*, 2011). Contrary to CpcE/F, only CpeY is strictly required for the lyase activity, albeit with a lower yield that when CpeZ is also present. MpeZ, which is only present in some marine *Synechococcus* capable of chromatic acclimation, has recently been shown to be a phycobilin lyase-isomerase acting on Cys-83  $\alpha$ -PEII (Shukla *et al.*, 2012). Before the present study (see chapter IV), it was the only phycobilin lyase acting on PEII biochemically characterized so far. Recently, the characterization of a *mpeU* interruption mutant, a gene present in all marine *Synechococcus* strains with medium to high-PUB content, has suggested that MpeU is a phycobilin lyase-isomerase (Mahmoud *et al.*, 2017), although it remains to be biochemically characterized.

As for the S/U clan, some homologs of members of the E/F clan do not have a phycobilin lyase activity, such as NblB that is involved in PBS degradation (Dolganov and Grossman, 1999).

At least one other enzyme is necessary for a proper PBS rod assembly and energy transfer in PBS rods. This enzyme, called CpcM, methylates a conserved asparagine residue found at position 72 of phycobiliprotein  $\beta$ -subunits, a post-translational modification that takes places between  $\beta$ -subunits chromophorylation and phycobiliprotein trimer formation (Shen *et al.*, 2008a; Miller *et al.*, 2008). This modification is probably important for modifying the spectral properties of the functionally key  $\beta$ -82 chromophore (Schluchter *et al.*, 2010).

### ***e. Regulation of phycobilisome activity***

#### ***i. Nutrients***

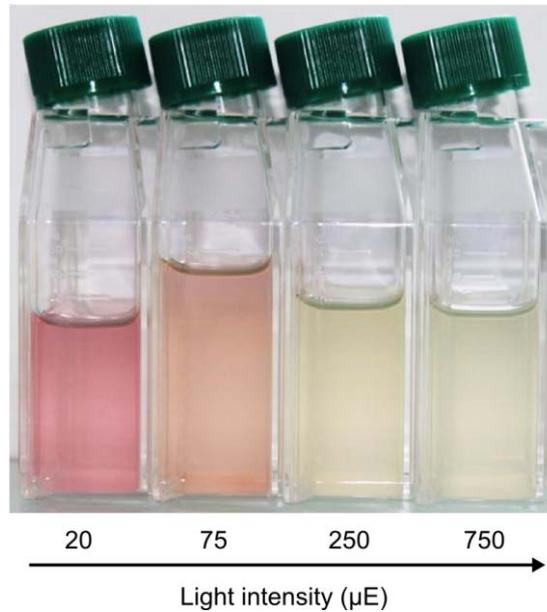
When cyanobacteria are submitted to nutrient deprivation, they exhibit a stress response and a “bleaching” phenotype, i.e. a degradation of pigments that leads to colour loss. In particular, the phycobilisome gets degraded under nitrogen, sulfur or iron deprivation. Two

non-exclusive hypotheses have been proposed to explain this observation (Grossman *et al.*, 1993, 2001; Adir *et al.*, 2006). First, phycobilisomes represent an important pool of cellular proteins, and thus of nutrients such as nitrogen or sulfur. Some authors suggested a nitrogen storage role for phycobiliproteins (Wyman *et al.*, 1985), but this was contradicted by subsequent work (Kana and Glibert, 1987). Alternatively, phycobilisome degradation could be a way for cells to reduce their energy input in a context of nutrient-deprivation induced slowdown of their metabolism, and thus to maintain the energy balance. Indeed, an excess of input energy can lead to the production of harmful reactive oxygen species (Adir *et al.*, 2006). The observation of mutants of the freshwater *Synechocystis* sp. PCC6803 unable to degrade their phycobilisome under nutrient-deprivation stress led to the discovery of proteins involved in PBS degradation. These proteins, called *non bleaching* (*nbl*), include NblA, which may act as an adaptor protein that guides a ClpC-ClpP complex to the phycobiliprotein disks in the rods of phycobilisomes, thereby initiating the degradation process (Karradt *et al.*, 2008), and the abovementioned NblB that is thought to catalyze the removal of chromophores from phycobiliprotein subunits (Grossman *et al.*, 2001; Adir *et al.*, 2006). Interestingly, no marine *Synechococcus* possess any of these two genes, suggesting that have developed different but yet poorly known mechanisms to cope with iron deficiency.

#### *ii. Regulation of light harvesting and utilization*

In marine *Synechococcus*, medium to long-term responses (minute to hour) to high light exposure include the reduction of the number and surface area of thylakoids, and the associated decrease in phycobilisome number and photosynthetic pigments such as chlorophyll *a* (Kana and Glibert, 1987; Six *et al.*, 2004, 2005). Additionally, Six and co-workers observed a progressive decrease in the PE to PC ratios with increasing light levels, indicative of a light-induced shortening of PBS rods (Six *et al.*, 2004). Similarly, a decline in the transcript and protein levels of linkers and phycobiliproteins can be observed under both high-light and UV-induced stress, suggesting that these stress induce both a degradation of phycobilisomes and a reduction of their synthesis, leading to a “bleached” phenotype (**figure 16**; Six *et al.*, 2007b; Mackey *et al.*, 2017). These responses are commonly designated as “photoacclimation”, and include other metabolic changes such as the synthesis of high-light induced proteins (HLIP), heat-shock proteins (HSP) or flavodiiron proteins (Flv), some of which interact with the phycobilisome or the PSII, as well as an increased turnover of the D1 protein of photosystem II, a process that allows repairing photoinactivated reaction (Sato *et al.*, 2010; Palenik *et al.*, 2006; Rocap *et al.*, 2003; He *et al.*, 2001; Edelman and Mattoo, 2008; Mackey *et al.*, 2017; Muramatsu and Hihara, 2012; Bersanini *et al.*, 2014); see also the review (Mullineaux, 2014). All these mechanisms tend to either to reduce the number and effective cross-section of photosystems, thus reducing the light-harvesting capacity, or to alleviate high-light effects (e.g. by alternative electron acceptors from the PSII to

avoid the deleterious saturation of the electron transport chain). Recently, the ambient light colour has also been shown to influence the energy distribution between PSII and PSI by impacting levels of the linker protein CpcG2, which is involved in the formation of PSI-associated “rod-type” phycobilisomes, consisting of rods directly linked to PSI without the APC core (Hirose *et al.*, 2013).



**Figure 16: Effect of light intensity on *Synechococcus* sp. WH7803 pigmentation.** Increasing light intensity leads to lower phycobiliprotein content per cell, which can be observed by the decrease in colour of cells grown under high irradiances (“bleached” phenotype). The change of colour between 20 and 75  $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$  is due to the decrease in the PE:PC ratio. Photograph courtesy of Hugo Doré.

In addition to these medium to long-term responses, changes in irradiance induce short-term responses (seconds to minutes) such as state transitions that consist in changes in the photosynthetic machinery to balance the energy input between the two photosystems. This balance is critical for controlling the relative productions of NADPH and ATP, and most importantly the redox state of the photosynthetic electron transport chain between PSII and PSI. Indeed, when PSII receives too much energy compared to PSI, the PSI activity cannot re-oxidize the plastoquinone/plastocyanin pool as fast as the PSII reduces it. This pool gets progressively reduced and thus cannot accept any new electron from the PSII. Without an electron acceptor, the excitation energy is trapped in the PSII, leading to its destruction (e.g. inactivation of the D1 protein) or the production of harmful radical species. State transitions change the relative electron transport rate in the photosynthetic membrane, and consist in a redistribution of the excitation energy between the two photosystems, with states 1 and 2 respectively corresponding to preferential distribution of excitation energy to PSII and PSI (Biggins and Bruce, 1989;

Mullineaux, 2014). State transitions have been linked to a transfer of energy from the PSII to the PSI (“spillover” hypothesis), as well as the physical relocation of phycobilisomes on the thylakoid membrane (“mobility” hypothesis; Biggins and Bruce, 1989; Mullineaux *et al.*, 1997; Joshua and Mullineaux, 2004; Joshua *et al.*, 2005), and are likely triggered by the redox state of the electron transport chain (Mullineaux, 2014). The recently described CpcG2 “rod-type” phycobilisomes might also be involved in state transitions, as could the PBS/PSII/PSI “megacomplex” discovered by Liu and co-workers (Ueno *et al.*, 2016; Chukhutsina *et al.*, 2015; Kondo *et al.*, 2007, 2009; Watanabe and Ikeuchi, 2013; Watanabe *et al.*, 2014).

High irradiances also induce the dissipation of excitation energy before its transformation into chemical energy (i.e. the photochemical quenching of excitation energy). This non-photochemical quenching (NPQ) can either consist in the emission of a photon (fluorescence) or in the transformation of excitation energy into vibrational energy (heat). The orange carotenoid protein (OCP) is a key actor of the latter (for complete reviews on the OCP, see Kirilovsky and Kerfeld, 2016 and Kerfeld *et al.*, 2017). This wide protein family found in cyanobacteria binds a carotenoid that gives it its orange colour (Melnicki *et al.*, 2016; Bao *et al.*, 2017). Upon absorption of blue light, the carotenoid undergoes structural changes (rotations of the  $\beta$  rings relative to the polyene chain, possibly due to a transient keto-enol tautomerization), which disrupt some interactions with the OCP apoprotein (Bandara *et al.*, 2017; Leverenz *et al.*, 2015). This results in the translocation of the carotenoid from the N-terminal domain (NTD) to the C-terminal domain (CTD) of the OCP, leading to the separation of the two domains and the conversion into the active “red” form<sup>10</sup> of the OCP (OCP<sup>R</sup>). This allows the NTD to bind to the PBS APC core. The precise site and interaction mechanism with the PBS core is not precisely known, but likely candidates include the L<sub>CM</sub> ApcE and ApcC (Stadnichuk *et al.*, 2015; Harris *et al.*, 2016; Kerfeld *et al.*, 2017). Strikingly, one single OCP<sup>R</sup> is sufficient to efficiently quench a whole PBS (Stadnichuk *et al.*, 2015; Kerfeld *et al.*, 2017). The quenching mechanism likely involves Förster resonance energy transfer between a phycobilin and the carotenoid, but the interaction between the NTD and the PBS core could also disrupt the excitation energy transfer in the latter by increasing the distances between successive bilins (Harris *et al.*, 2016; Kerfeld *et al.*, 2017). The fluorescence recovery protein (FRP) allows the dissociation of the OCP<sup>R</sup>-PBS complex and the back conversion of OCP<sup>R</sup> into the inactive orange form OCP<sup>O</sup> (Kirilovsky and Kerfeld, 2016; Kerfeld *et al.*, 2017). OCP and OCP homologs could also be involved in the quenching of reactive oxygen species such as <sup>1</sup>O<sub>2</sub> (Kerfeld *et al.*, 2017). Interestingly, the *ocp* gene has not been found in the genome of all marine *Synechococcus*, so these OCP-lacking

---

<sup>10</sup> Similarly to bilins bound to phycobiliproteins, the change in colour is due to changes in the conformation of the carotenoid polyene chain, which is more bended and twisted in the orange form, thus reducing its effective conjugation length compared to the “red” conformation (Leverenz *et al.*, 2015).

strains might be more sensitive to light stress or could possess yet undetected OCP or OCP domains homologs or other uncharacterized quenching mechanisms.

Most recently, very short-term responses of the phycobilisome to high-light levels have been discovered, such as the excitonic and physical uncoupling of the phycobilisome from the reaction centres, probably through the destruction of the most labile components (Tamary *et al.*, 2012). Even more intriguing is the recent suggestion of an excitation quenching role (i.e. a photoprotective role through excitation energy dissipation) for some phycobilins (Eisenberg *et al.*, 2017), such as the PCB of the  $\beta$ -subunit of APC (Wang and Moerner, 2015), or the PCB bound at  $\beta$ -84 in PC (Squires and Moerner, 2017), revealing how far we are from fully understanding how the phycobilisome works.

### *iii. Chromatic acclimation*

Some cyanobacteria are able to dynamically modify the composition of their PBS depending on incoming light colour. Because this process is a short-term physiological response to an environmental cue and not a long-term evolutionary response, the historical designation “chromatic adaptation” has been progressively replaced with “chromatic acclimation” (CA; Kehoe and Gutu, 2006). However, the ability to perform CA can be considered as an adaptation to a changing light environment, although only a handful of studies focused on the adaptive value of CA (i.e. the gain of fitness of organisms able to perform CA compared to organisms that are not). Two studies based on the competition between CA-able strains and strains with a fixed pigmentation provided interesting insights into the dynamics and timescale of light condition changes over which CA might be advantageous, yet their results should be taken with caution as the strains able to perform chromatic acclimation and those with “fixed” pigmentation belonged to different species and displayed different ecologies, e.g. different salinity requirement and morphologies (single-celled vs filamentous; Stomp *et al.*, 2004, 2008). Thus, the observed fitness differences might possibly have arisen from ecological differences other than just photophysiology. Another study showed that chromatic acclimation allows keeping optimal photosynthesis quantum yield and thus could improve light use efficiency under different light colours, likely affecting the organism fitness (Campbell, 1996).

Three<sup>11</sup> types of CA have been evidenced to date, with different sensors, regulatory networks and outcomes. However, all result in the maximization of the overlap between the cells light absorption spectrum and the incident light spectrum, i.e. in the increase of the efficient cross-section of the light-harvesting antenna (see Montgomery, 2017; Kehoe and Gutu, 2006; Gutu and Kehoe, 2012 for complete reviews on CA). During Type II CA (CA2), PE synthesis is

---

<sup>11</sup> Four with the “Type I” group originally described by Nicole Tandeau de Marsac, which gathers strains that are *not* able to perform chromatic acclimation (Tandeau de Marsac and Cohen-bazire, 1977).

induced under green light (GL), but the cellular levels of PC remain constant whatever the light colour. During Type III CA (CA3), PE and PC genes are respectively up- and down-regulated under GL, and conversely under red light (RL; Tandeau de Marsac and Cohen-bazire, 1977; Gutu and Kehoe, 2012). Finally, the recently discovered Type IV CA (CA4) consists in changes in the chromophores attached to PE, more PUB being attached under blue light (BL) and more PEB under GL, the PE and PC levels remaining unchanged (Palenik, 2001; Everroad *et al.*, 2006).

CA2 is regulated through the CcaS/CcaR system (Hirose *et al.*, 2010). CcaS is a cyanobacteriochrome photoreceptor (a family of homologs of plant phytochromes that are found only in cyanobacteria) that binds PCB. Upon GL illumination, the PCB reversibly isomerizes from the GL-absorbing C15-Z state to the RL-absorbing C15-E state leading to the protonation of the PCB, which changes its spectral properties (Hirose *et al.*, 2013). This activates the kinase activity of CcaS, which both autophosphorylates and phosphorylates CcaR (Hirose *et al.*, 2008, 2010). Phosphorylated CcaR then binds to the promoter region of the *cpeC-cpcG2-cpeR1* operon, resulting in its upregulation. The *cpeBA* operon is in turn induced, likely by CpeR1. Conversely, CcaS has a phosphatase activity under RL, resulting in no phosphorylation of CcaR and no expression of the *cpeC-cpcG2-cpeR1* operon (Hirose *et al.*, 2010).

The regulation of CA3 has been extensively studied in *Fremyella diplosiphon*. It is more complex than for CA2 and involves at least three photoreceptors (RcaE, DpxA, and one unknown; Montgomery, 2017). RcaE (Regulator for CA) was the first discovered cyanobacteriochrome (Kehoe and Grossman, 1996), and has a kinase/phosphatase activity that is RL/GL regulated by the same photoconversion/protonation mechanism as CcaS (Hirose *et al.*, 2013). On the contrary of CcaS, RcaE has the kinase activity under RL, resulting in autophosphorylation and phosphorylation of the response regulator RcaF, which in turn increases the phosphorylation level of RcaC (Gutu and Kehoe, 2012). Highly phosphorylated RcaC binds more efficiently to the promoter region of genes and operons necessary for PC (*cpcB2-A2-H2-I2-D2*) and PCB (*pcyA*) synthesis, resulting in the up-regulation of these genes. Conversely, the phosphatase activity of RcaE is induced under GL, resulting in lowered phosphorylation of RcaC through RcaF, leading to low binding affinity to the promoter and low transcription levels of these genes and operons. In addition to the Rca system, the Cgi (control of green light induction) pathway controls the induction of the *cpeC-D-E-S-T-R* operon, which in turn induces the *cpeB-A* and *pebA-B* operons through CpeR (Gutu and Kehoe, 2012; Kehoe and Gutu, 2006). If the sensor for Cgi has not yet been identified, the translation initiation factor 3 (IF3 $\alpha$ ) has been found to be part of the regulatory pathway (Gutu *et al.*, 2013). Recently, the teal/yellow cyanobacteriochrome DpxA (decreased phycoerythrin expression) has been shown to have a yellow-activated autokinase activity, which leads to the repression of the accumulation of PE though unknown regulators (Wiltbank and Kehoe, 2016).

CA4, which is found only in marine *Synechococcus*, is in contrast much less understood. In particular, no homolog of known photoreceptors has been identified so far in these organisms. This phenotype has been linked with the occurrence in the genome of CA4-capable strains of a small genomic island, which exists under two different configurations, thus distinguishing CA4-A and CA4-B (Humily *et al.*, 2013). Both versions contain the two genes *fciA*, *fciB* (type four chromatic acclimation island). The CA4-A genomic island also contains *mpeZ*, which has been shown to code for a phycobilin lyase-isomerase (see I.3.d), and two other uncharacterized genes (*unk10* and *fciC*; Shukla *et al.*, 2012; Humily *et al.*, 2013). The CA4-B version only contains *fciA*, *fciB*, *unk10* and a gene encoding the putative phycobilin lyase *mpeW*. While *mpeZ* is more expressed under BL than under GL, it is the opposite for *mpeW* (Humily *et al.*, 2013). FciA and FciB, which have similarity to transcription factors of the AraC family, have been recently demonstrated to be master regulators of this process, and their activity likely regulated post-transcriptionally (Sanfilippo *et al.*, 2016). FciA appears to be an activator of the BL response, whereas FciB is probably a repressor under GL.

Initially defined as a reversible modification of the phycobilisome rod composition by different light colours, chromatic acclimation has recently been shown to extend to other cellular processes. As an example, Hirose and co-workers recently observed that some strains (such as *Synechocystis* PCC6803) do not have PE in their PBS, yet regulate the formation of PSI-associated atypical phycobilisomes through regulation of the linker CpcG2 (sometimes referred to as CpcL) by the CcaS/CcaR system (the regulatory system described for CA2; Hirose *et al.*, 2008, 2017). They proposed to call this process CA0, and suggested to redefine CA2 from a phenotype-based to a genotype-based definition similar to what has been done for CA4 (Humily *et al.*, 2013), with the regulation of PE levels only as observed in *Calothrix* PCC 7103 being designated as genuine CA2, and the regulation of both PE levels and CpcL levels as observed in *Calothrix* PCC 6303 being designated as CA2/0 hybrid (Hirose *et al.*, 2017). Similarly, the CA3 regulation is not limited to the PE/PC levels in the PBS, but extends to other key cellular processes such as cell morphology, filament morphogenesis, inorganic carbon uptake, response to oxidative stress and growth at low cell densities (Montgomery, 2017; Bussell and Kehoe, 2013). Finally, other light-dependent regulatory pathways have been recently uncovered such as the far-red light photoacclimation (FaRLiP), which allows the utilization of light wavelengths between 700 and 750 nm for photosynthesis by inducing the synthesis of atypical chlorophylls (Chl *d*, Chl *f*) and allophycocyanin<sup>12</sup> (ApcE2; see Ho *et al.*, 2017; Montgomery, 2017; Hernández-Prieto *et al.*, 2017), highlighting the diversity of light-colour regulated processes in cyanobacteria.

---

<sup>12</sup> On the contrary of other known phycobiliproteins, ApcE2 binds PCB non-covalently, resulting in a red-shifted absorption spectra (Miao *et al.*, 2016).

***f. Biotechnological applications***

Phycobiliproteins has been recognized early for their unique spectral and biochemical properties (Glazer 1994). Phycobiliproteins have been used as fluorescent labels for flow cytometry, cell imaging, immunoassays and study of biochemical processes such as protein folding among others, and artificial derivatives with new valuable spectral and/or biochemical properties are under active research (Glazer, 1994; Alvey *et al.*, 2011; Ge *et al.*, 2017; Zhao *et al.*, 2012; Miao *et al.*, 2016). More recently, the expanding fields of optogenetics<sup>13</sup> and synthetic biology used the regulatory machinery of Type II chromatic acclimation to precisely and dynamically control gene expression (Olson *et al.*, 2014; Venayak *et al.*, 2015; Miliadis-Argeitis *et al.*, 2016).

---

<sup>13</sup> Designated Method of the Year 2010 by Nature Methods (2011).

## II. The picocyanobacteria, key members of the phytoplankton

### 1. Phytoplankton

#### *a. Taxonomic, morphological and ecological diversity*

About 70% of our planet is covered by the oceans. These vast areas and volumes (on the order of  $10^9$  km<sup>3</sup>) probably constitute the biggest ecosystem on earth<sup>14</sup>, and are inhabited by a wealth of highly diverse organisms, from archaea to blue whales, with representatives of every life domain (Bacteria, Archaea, Eukarya). The diversity of marine life forms was recognized early. The unicellular eukaryotes, commonly known as “protists”, a term coined by Ernst Haeckel, comprise widely differing morphologies (**figure 17**; Haeckel, 1866). Since the advent of molecular phylogeny, this group has been recognized as paraphyletic and highly genetically diversified, with representatives from every supergroups within the Eukarya domain (Worden *et al.*, 2015).

Most of the marine organisms simply drift along with the currents and constitute the plankton, in opposition to the free-swimming nekton. Planktonic organisms cover at least 6 orders of magnitude in size, and have been classified from pico- to mega-plankton (**figure 18**). In addition to the taxonomic and size diversities, planktonic organisms present a wide range of ecology and trophic modes, from free-living to symbiotic and parasitic and from oxygenic photosynthesis performing organisms (*phytoplankton*), to heterotrophic or mixotrophic (i.e. phototrophic or heterotrophic depending on the environmental conditions) organisms (Yelton *et al.*, 2016). The phytoplankton is concentrated in the upper lit layer of the oceans, which is called the (eu)photic zone and can extend down to 250 m below sea surface in areas with the clearest waters such as the southern Pacific Ocean.

---

<sup>14</sup> If we exclude the Earth crust itself!



**Figure 17: Diverse marine planktonic organisms drawn by Ernst Haeckel in *Kunstformen der Natur*, 1904.** Top left, Copepod (Metazoa); top right, “protist” Phaeodaria (Rhizaria); bottom left, “protist” Acantharea (Rhizaria); bottom right, Leptomedusae (Metazoa). Images in the public domain, retrieved from Wikimedia Commons.

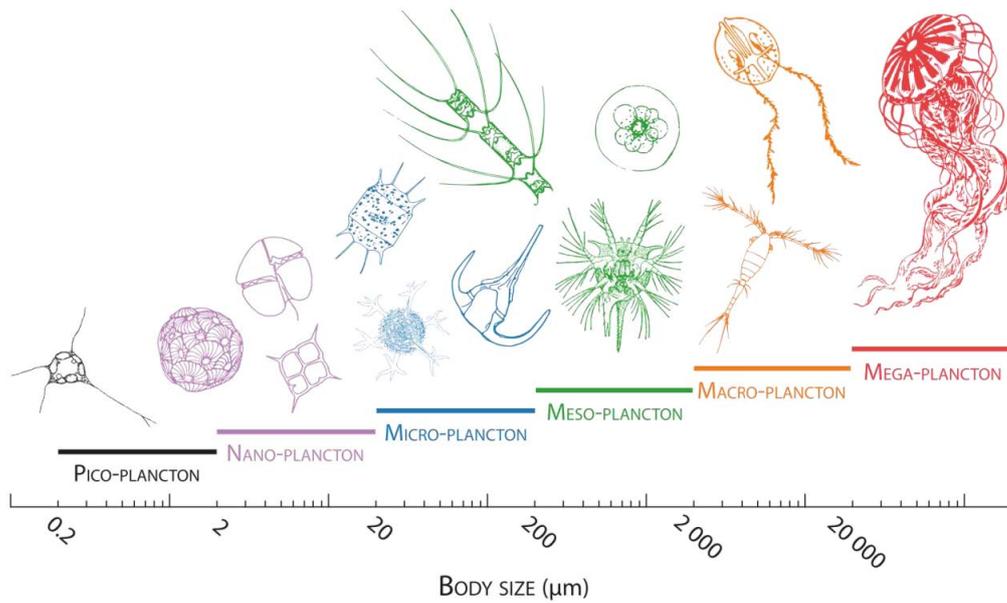


Figure 18: the different size classes of plankton. Figure reproduced from (Biard, 2015).

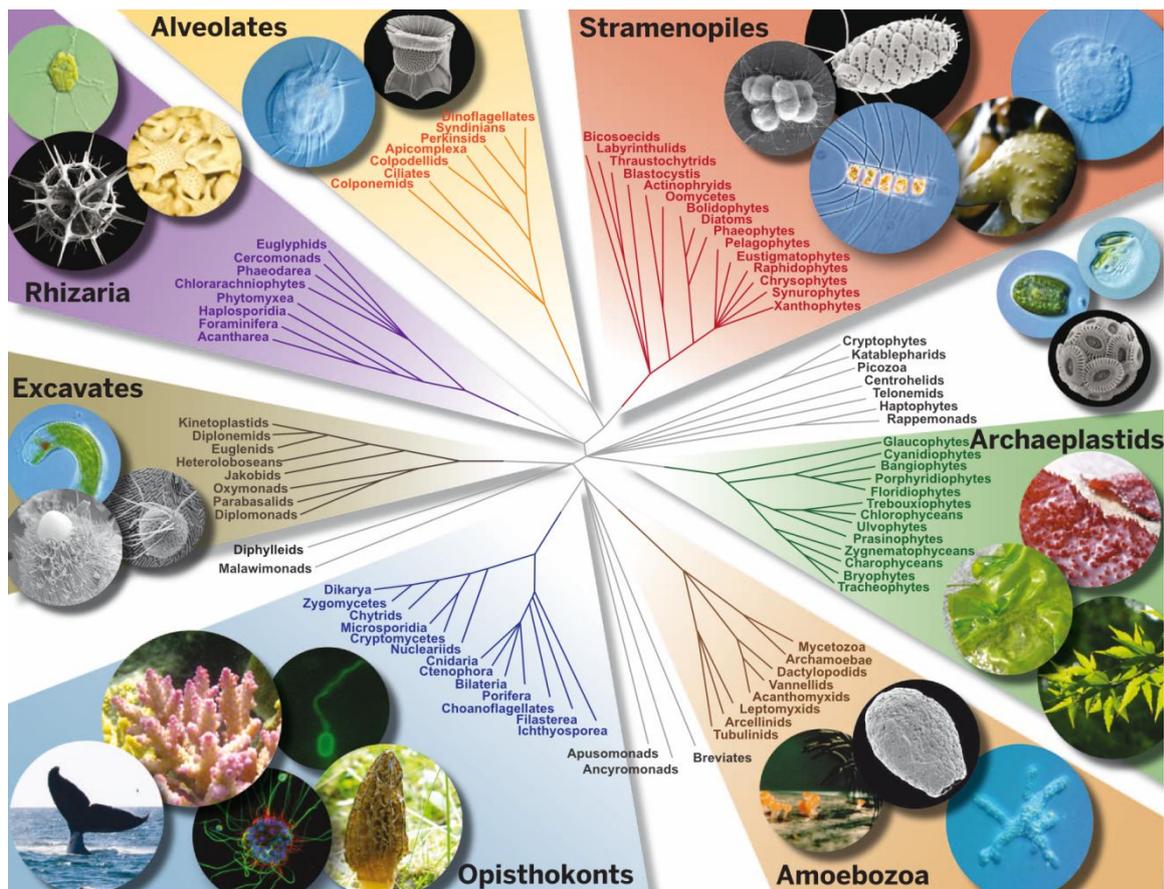
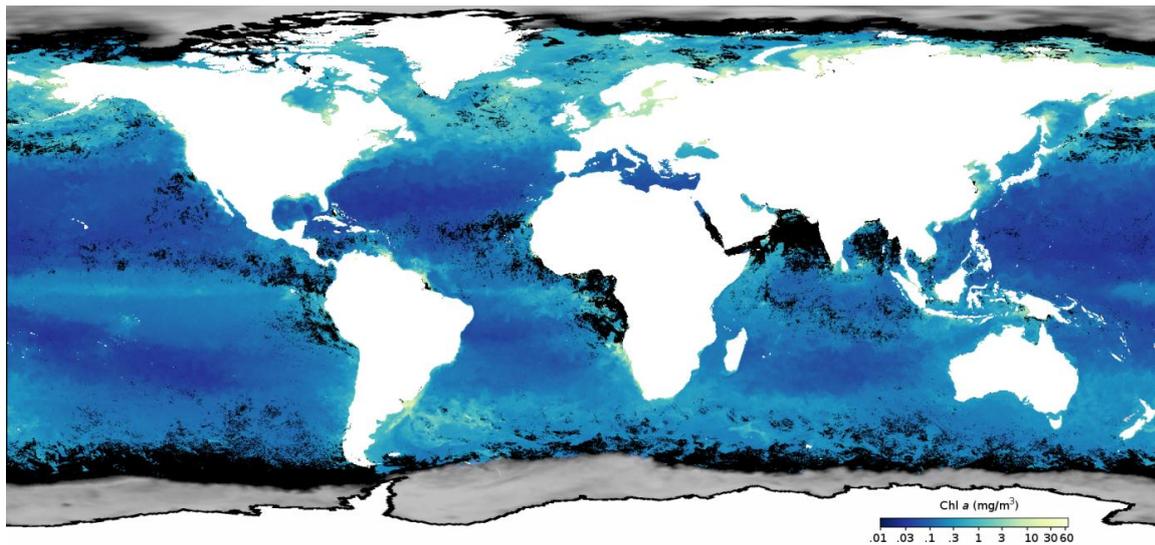


Figure 19: taxonomic and morphological diversity of single-celled eukaryotic marine plankton. Phylogenetic tree of the Eukarya domain of life, with seven supergroups highlighted. Pictures from representatives of each supergroup underline the morphological diversity of microbial and multicellular eukaryotes. Figure reproduced from (Worden *et al.*, 2015).

Phytoplanktonic organisms are only found in the smallest size fraction of the plankton (pico-, nano- and microplankton, **figure 18**). The microphytoplankton (20-200  $\mu\text{m}$ ) mostly comprises diatoms (Stramenopiles or heterokontophyta) and dinoflagellates (Alveolates, **figure 19**; Worden *et al.*, 2015; Caron *et al.*, 2012). Dinoflagellates have been shown to be phototrophs, mixotrophs or heterotrophs depending on the species. Dinoflagellates and diatoms dominate the nutrient-rich coastal areas. Diatoms in particular account for about 20% of global primary production and are key primary producers at high latitudes (Field *et al.*, 1998; Vargas *et al.*, 2015; Malviya *et al.*, 2016). Nanophytoplankton (2-20  $\mu\text{m}$ ) is primarily composed of cryptophytes and haptophytes, the phylogenetic relationship of which with other Eukarya is unclear. Recent studies suggest that they are two related groups from a monophyletic clade (Hacrobia) sister to stramenopiles, alveolata and rhizaria (**figure 19**; Worden *et al.*, 2015; Caron *et al.*, 2012). Most haptophytes are coccolithophores, which have a key role in global carbon biogeochemical cycle due to their calcified outer shell made of coccoliths, and could also have an important role in climate regulation through the emission of volatile sulfur compounds (Paasche, 2001). Finally, the picophytoplankton (0.8-2  $\mu\text{m}$ ) gathers representatives of both Eukarya and Bacteria life domains. The prokaryotic part is dominated by the two related genera *Synechococcus* and *Prochlorococcus* (see II.2), the picoeukaryotes are much more diverse and mostly belong to Chlorophyta, Haptophyta and Heterokontophyta (Vaulot *et al.*, 2008). Chlorophyta are early-diverging members of the green lineage, leading to the suggestion that the common ancestor of green algae and land plants was an ancestral green flagellate closely related to Chlorophyta (Tragin *et al.*, 2016). Chlorophyta are mostly found in coastal waters, and include the smallest known photosynthetic eukaryotes such as *Micromonas*, *Bathycoccus* and *Ostreococcus* (Vaulot *et al.*, 2008; Tragin *et al.*, 2016). Finally, anoxygenic phototrophs are also very diverse and widespread in marine plankton, including members from both Archeae and Bacteria (Béjà *et al.*, 2000; Lami *et al.*, 2007).

### ***b. Role in the biological pump***

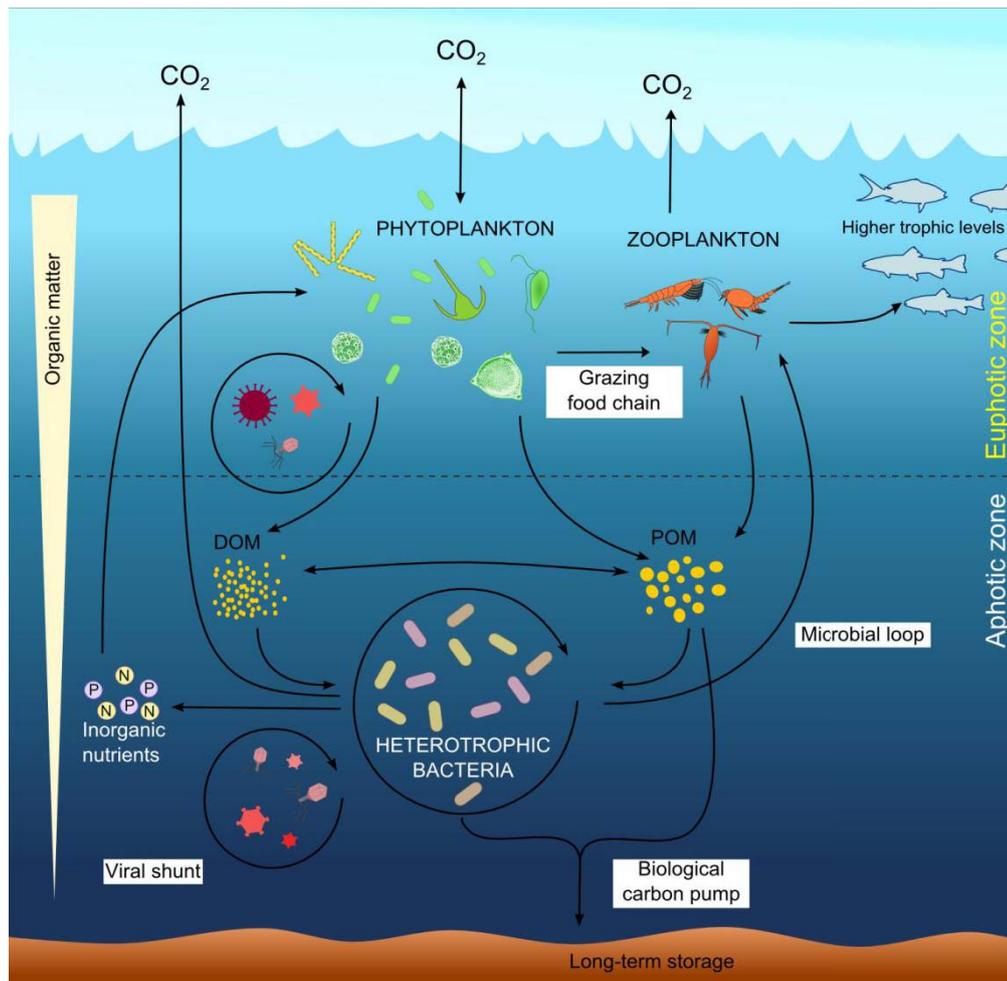
By using light energy to fix atmospheric  $\text{CO}_2$ , the phytoplankton has a key role in marine ecosystems and global biogeochemical cycles. It accounts for about half of our planet primary production, yet only represent 0.2% of the primary producer biomass (Field *et al.*, 1998). This comes from the very short turnover of carbon in the ocean, on the order of one week (Falkowski *et al.*, 1998).



**Figure 20: Marine primary production in August 2017.** Chlorophyll *a* concentration is a proxy for primary production. Black areas are missing data (sea ice or clouds). Grey areas correspond to sea ice extent on the 15<sup>th</sup> of August. Composite picture based on images made by Jesse Allen and Reto Stockli, NASA Earth Observatory Group, using data provided by the MODIS Land Science Team.

The phytoplankton organic matter can have different fates. The classical linear grazing food chain, with the phytoplankton being grazed upon by zooplankton, which in turn is consumed by fishes, only accounts for half of the phytoplankton organic matter losses (**figure 21**). The other half corresponds to excreted or dissolved organic matter (DOM), e.g. excreted proteins (Azam, 1998; Christie-Oleza *et al.*, 2015a; Becker *et al.*, 2014). About 10% of DOM aggregates into particles which, together with fecal pellets from zooplankton and other cellular debris, constitute the “marine snow”, which slowly sinks to the oceans floor (Guidi *et al.*, 2016; Agusti *et al.*, 2015). During the downward export, the marine snow can be consumed by microbial heterotrophs (bacteria and archaea), which re-mineralize it and release inorganic nutrients that can be reused by phytoplankton cells. These microbial heterotrophs are consumed by nanoplanktonic flagellates, which are in turn consumed by larger zooplankton, returning some of the organic carbon diverted from the main food chain to upper trophic levels in what is known as the “microbial loop” (**figure 21**; Azam *et al.*, 1983). Eventually, some of the marine snow that is not consumed reaches the oceans floor. This tiny fraction, which represents about 2% of the carbon primarily fixed by phytoplankton, can then be buried over geological timescales in oceanic sediments, effectively sequestering carbon and thus decreasing atmospheric CO<sub>2</sub> concentration. This sequestering is called the biological carbon pump (Guidi *et al.*, 2016). Finally, viruses can infect and lyse phytoplankton and microbial heterotrophs, thus releasing DOM. This process, contributing to the diversion of organic carbon from upper trophic levels, has been called the “viral shunt” (**figure 21**; Wilhelm and Suttle, 1999; Suttle, 2007). It could

represent up to 25% of the photosynthetically fixed organic carbon, and has been correlated with carbon export (Wilhelm and Suttle, 1999; Suttle, 2007; Guidi *et al.*, 2016). Membrane vesicles containing nucleic acids and proteins have been recently observed in *Prochlorococcus* cultures and *in situ*, and could also play a role in transferring organic matter from primary producers to heterotrophic bacteria and/or in the long-term export in sediments (Biller *et al.*, 2014b, 2017).

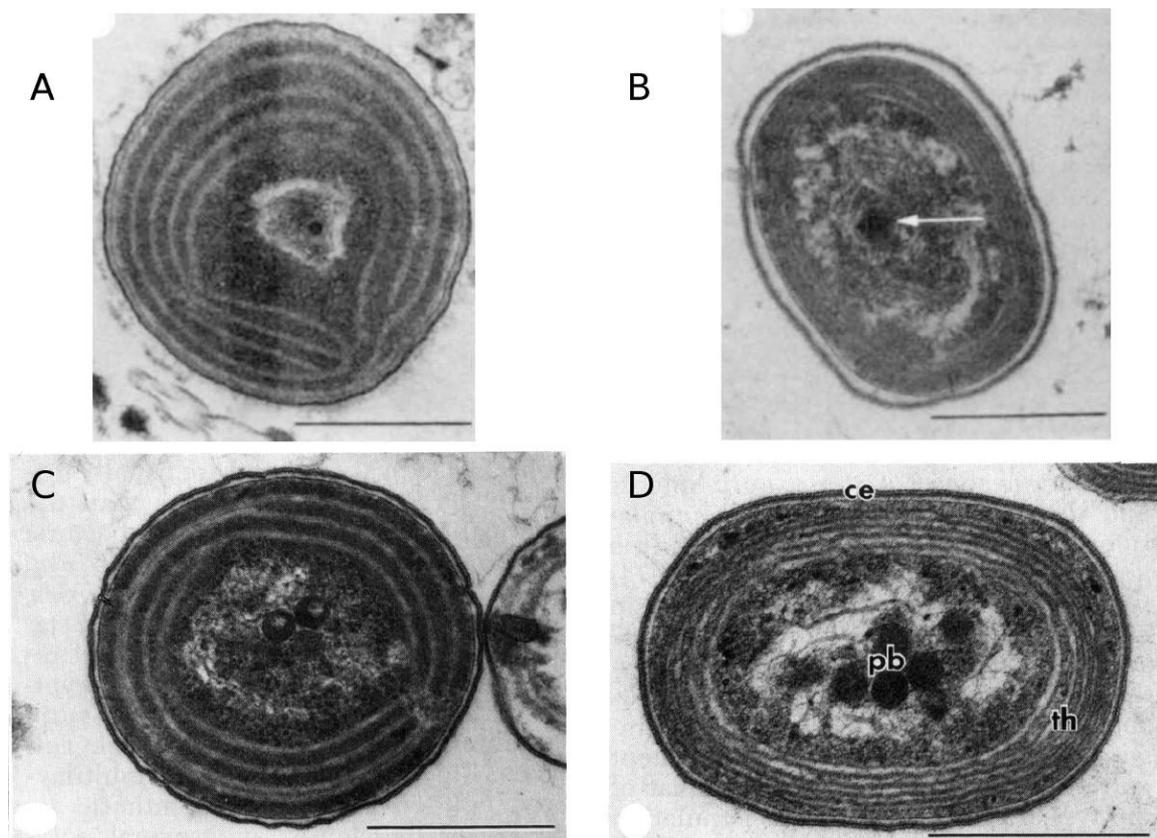


**Figure 21: The oceanic biological pump.** Phytoplankton has a central role as the primary producer of organic matter. Phytoplankton releases dissolved organic matter (DOM) and particulate organic matter (POM), and is consumed by zooplankton. DOM and POM are consumed by heterotrophic bacteria such as flavobacteria and roseobacters, which release nutrients that can be reincorporated into organic matter by phytoplankton. Zooplankton also consumes heterotrophic bacteria in what constitutes the “microbial loop”, returning organic matter to higher levels of the food chain. The “viral shunt” is the production of DOM and POM from phytoplankton and heterotrophic bacteria by viral-mediated cell lysis. After (Buchan *et al.*, 2014).

## 2. Marine picocyanobacteria

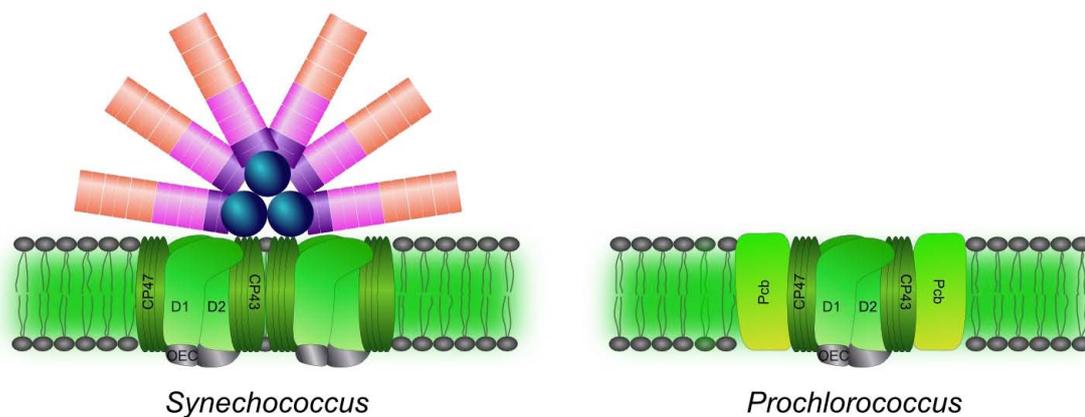
### *a. Discovery and physiology of marine picocyanobacteria*

Picophytoplankton and picocyanobacteria went long undetected because of their small size. In the 1970s, advances in sample collection and processing led to the recognition of the importance of this compartment (see Sieburth, 1978 for a short review). The advent of epifluorescence microscopy led to the discovery of small (0.9-2.2  $\mu\text{m}$ ), chroococcoid cyanobacteria containing phycoerythrin and thus fluorescing orange (in contrast to other phytoplankters fluorescing red), found from the surface down to 200-400 m depth in samples from various marine areas, and reaching concentrations of up to  $10^5$  cells/mL (**figure 22**; Waterbury *et al.*, 1979). The morphology (peripheral thylakoids) and physiology (division by binary fission in one plane) of this cyanobacteria led Waterbury and co-workers to assign them to a genus previously described in freshwater environments: *Synechococcus* (Waterbury *et al.*, 1979, 1986).



**Figure 22: Electron micrographs of *Synechococcus* (A, C) and *Prochlorococcus* (B, D).** Scale bar is 0.5  $\mu\text{m}$  for all four images. Note the larger spacing of thylakoids of *Synechococcus* compared to *Prochlorococcus*. (A, B) reproduced from (Chisholm *et al.*, 1988b); (B, D) reproduced from (Johnson and Sieburth, 1979).

A second major step in the knowledge of marine phytoplankton permitted by the application of flow-cytometry to oceanography at the beginning of the 1980s (Olson *et al.*, 1983; Chisholm *et al.*, 1988a). The “widespread distribution of very small cells (less than 0.8  $\mu\text{m}$ ) that fluoresce red when excited with blue or green light, indicating the presence of chlorophyll” was first reported in 1987 (Chisholm *et al.*, 1987, 1988a), and related to the observation 10 years before of small chroococcoid cyanobacteria with atypical thylakoids (Johnson and Sieburth, 1979). A first description of their abundance, distribution and pigment content was made in 1988 (Chisholm *et al.*, 1988b), but their successful isolation and formal description under the name *Prochlorococcus marinus* was only reported in 1992 (Chisholm *et al.*, 1992). Although initially thought based on their atypical pigment complement to be more closely related to the chlorophyll *b*-containing oxyphototrophs *Prochlorothrix* and *Prochloron* than to marine *Synechococcus* (Chisholm *et al.*, 1988b), the close affiliation of *Prochlorococcus* with the latter was then demonstrated by phylogenetic analyses (Chisholm *et al.*, 1992). With a cell size of 0.5-1.6  $\mu\text{m}$ , *Prochlorococcus* constitutes the tiniest oxyphotosynthetic organism known to date. Like *Synechococcus*, they have peripheral thylakoids, but they are much more appressed than in the latter (**figure 22**; Chisholm *et al.*, 1992).



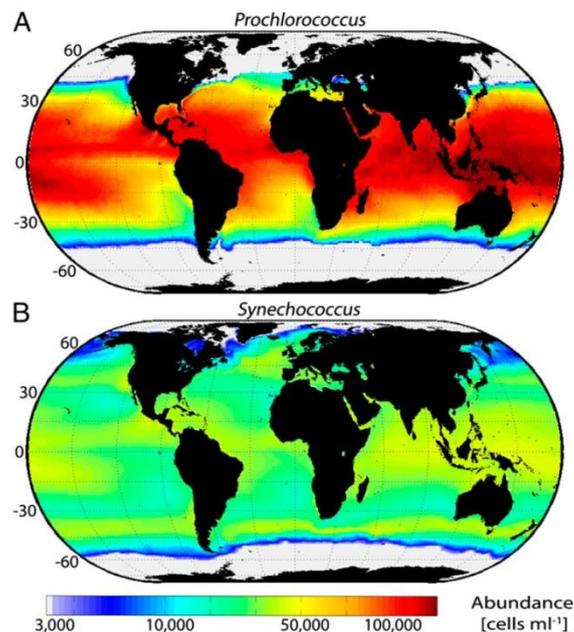
**Figure 23: Light-harvesting complexes associated to PSII in marine *Synechococcus* and *Prochlorococcus*.** *Synechococcus* main light-harvesting complex is a phycobilisome as typically found in cyanobacteria, whereas *Prochlorococcus* possesses an intrinsic light-harvesting antenna made of rings of prochlorophyte chlorophyll-binding proteins (Pcb), which mainly bind divinyl chlorophyll *a* and *b*.

Contrary to *Synechococcus* and other typical cyanobacteria, *Prochlorococcus* does not have phycobilisomes, but an original light-harvesting antenna intrinsic to the thylakoid membrane and made of “prochlorophyte chlorophyll binding” proteins (Pcb; **figure 23**). These antennas are made of different proteins binding divinyl Chl *a* and *b*, which are exclusively found

in *Prochlorococcus* (Chisholm *et al.*, 1988b, 1992). These chlorophylls present red-shifted absorption peaks compared to their monovinyl counterparts (i.e. typical Chl *a* and *b*), allowing them to absorb blue-light more efficiently. *Prochlorothrix* and *Prochloron* also possess Pcb-type antennas (binding monovinyl Chl *a* and *b*), and these three taxa are sometimes referred to as “Prochlorophyta”. However, phylogenetic analyses have demonstrated that this group was clearly polyphyletic and that PSII-associated Pcb, which are derived from PSI-associated iron-stress induced IsiA-like proteins, appeared multiple times during the evolution of cyanobacteria (Urbach *et al.*, 1992; Turner *et al.*, 1999; Partensky and Garczarek, 2003).

### **b. Distribution and ecological role**

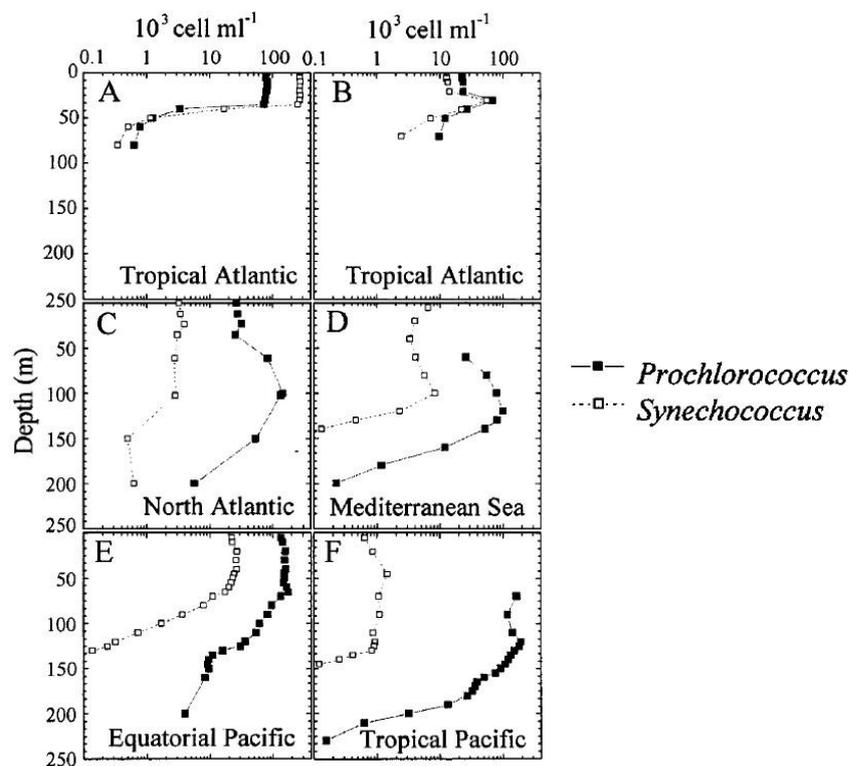
In spite of their small cell size, marine picocyanobacteria have a key role in marine biogeochemical cycles due to their ubiquity and high abundance in the field (Li, 1994; Flombaum *et al.*, 2013; Zwirgmaier *et al.*, 2008). With an estimated mean annual abundance of respectively  $2.9 \times 10^{27}$  and  $7.0 \times 10^{26}$  cells, *Prochlorococcus* and *Synechococcus* contribute respectively to 8.5% and 16.7% of the ocean net primary production (Flombaum *et al.*, 2013). Recently, Guidi and co-workers suggested a central role of *Synechococcus* and their phages in carbon export in subtropical oligotrophic areas (Guidi *et al.*, 2016).



**Figure 24: Global latitudinal distribution of *Prochlorococcus* and *Synechococcus*.** Figure reproduced from (Flombaum *et al.*, 2013).

*Prochlorococcus* exhibits a distribution restricted to a latitudinal band extending up to 50°N and 40°S, with anecdotal reports up to 60°N (figure 24; Partensky *et al.*, 1999b). This genus is particularly well adapted to warm, oligotrophic and stratified waters (Partensky *et al.*,

1999b; Flombaum *et al.*, 2013; Bouman *et al.*, 2011), where it is typically present at concentrations of the order of  $10^4$ - $10^5$  cells/mL and locally up to  $7 \times 10^5$  cells/mL (Campbell *et al.*, 1998; Olson *et al.*, 1990; Partensky *et al.*, 1999b; Zwiirgmaier *et al.*, 2008). It is also present in more nutrient-rich coastal waters, but is likely biologically excluded from such environments (outcompeted by other organisms and/or grazed by zooplankton; Biller *et al.*, 2014a; Partensky *et al.*, 1999b). Temperature is an important factor driving its distribution. *Prochlorococcus* is virtually absent below  $10^\circ\text{C}$ , and exhibits temperature optima around  $24$ - $29^\circ\text{C}$  depending on the phylogenetic clade (Partensky *et al.*, 1999b; Flombaum *et al.*, 2013; Biller *et al.*, 2014a; Johnson *et al.*, 2006). *Prochlorococcus* is detected throughout the whole euphotic zone, with a concentration increasing with depth until a maximal abundance between 50 and 100 m, and decreases afterwards with significant presence up to a depth of 200-250 m (**figure 25**; Partensky *et al.*, 1999b; Biller *et al.*, 2014a).



**Figure 25: Vertical distribution of *Prochlorococcus* and *Synechococcus* in different oceans and seas.** In mesotrophic waters, maximal abundances are often restricted to a relatively shallow mixed surface layer, with either predominance of *Synechococcus* (**A**) or similar abundances of *Prochlorococcus* and *Synechococcus* (**B**). In more oligotrophic areas, *Prochlorococcus* can be 10 (**C-E**) to 100 (**F**) times more abundant than *Synechococcus*, with a deep concentration maximum (**C, D**) or a relatively constant distribution throughout the water column (**E, F**). Figure reproduced from (Partensky *et al.*, 1999b).

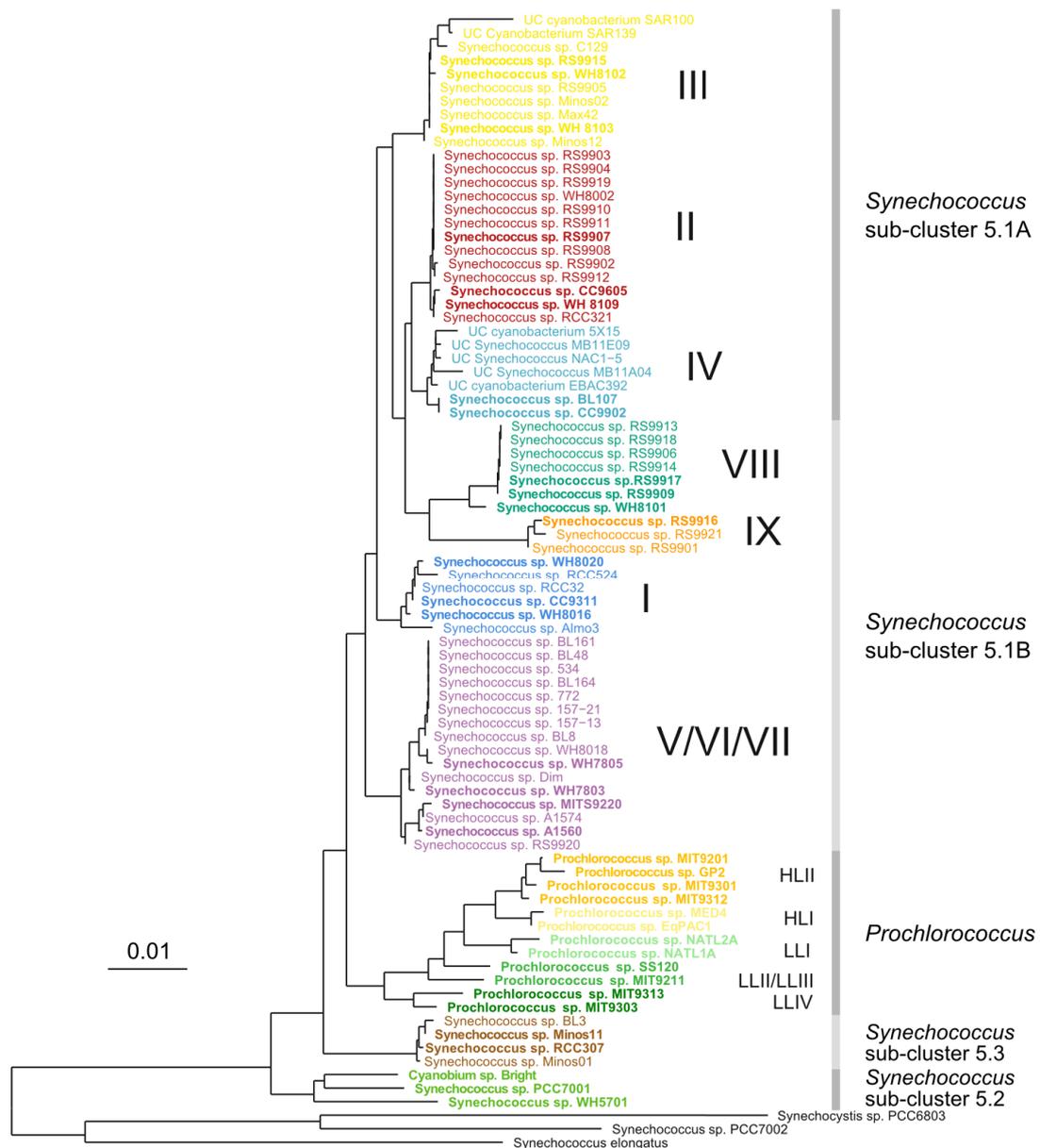
*Synechococcus* is globally more ubiquitous and widespread than *Prochlorococcus*. Indeed, it is found from nutrient-rich coastal areas to oligotrophic open-ocean waters and its latitudinal distribution extends from 50°S to 60°N, with some reports of very low abundance yet actively growing populations up to 80°N (**figure 24**; Partensky *et al.*, 1999b; Paulsen *et al.*, 2016; Not *et al.*, 2005; Huang *et al.*, 2012). *Synechococcus* is found in environments ranging from ~0°C to 30°C, but its abundance and growth rate are significantly decreased below 5-10°C (Li, 1998; Waterbury *et al.*, 1986; Zwirgmaier *et al.*, 2008; Fuller *et al.*, 2006; Hunter-Cevera *et al.*, 2016a). In addition to a wider temperature range, *Synechococcus* might also exhibit a wider growth irradiance range than *Prochlorococcus* (Six *et al.*, 2004, 2007a). *Synechococcus* abundance is on average one order of magnitude below *Prochlorococcus* abundance, but can reach higher maximal abundance in some specific environments such as the Arabian Sea or upwelling zones such as west Africa or in the Costa Rica Dome, where record concentrations of 1.5-3.7x10<sup>6</sup> cells/mL have been reported (Partensky *et al.*, 1999b; Saito *et al.*, 2005; Zwirgmaier *et al.*, 2008; Fuller *et al.*, 2006). Its vertical distribution appears to be more restricted than for *Prochlorococcus*, *Synechococcus* being rarely detected below 150 m (**figure 25**; Partensky *et al.*, 1999b).

### **c. Evolution and diversity of marine picocyanobacteria**

Phylogenetic analyses using a variety of markers concordantly demonstrated that *Prochlorococcus* and marine *Synechococcus* are two closely genetically related genera forming a monophyletic cluster (usually called “marine picocyanobacteria”) in the cyanobacteria radiation (**figure 26**; Urbach *et al.*, 1992; Fuller *et al.*, 2003; Dufresne *et al.*, 2008; Scanlan *et al.*, 2009; Scanlan, 2012; Shih *et al.*, 2013). However, even if closely related, both genera are highly diverse genetically.

Marine picocyanobacteria have been initially classified into different three marine clusters (MC-A to MC-C) based on taxonomic criteria (swimming ability, molecular G+C content, salt requirement for growth, light-harvesting pigments; Waterbury and Rippka, 1989). MC-A and MC-B were later grouped into *Synechococcus* cluster 5, with MC-B roughly corresponding to the new sub-cluster 5.2 (SC 5.2) and MC-A to the new sub-cluster 5.1 (SC 5.1; Herdman *et al.*, 2001). More recently, phylogenetic analyses combined with isolation of new strains demonstrated that four different monophyletic groups can be discriminated within the marine picocyanobacteria radiation (**figure 26**). SC 5.1 corresponds to obligate marine strains with PE as the main light-harvesting pigment, at the exception of the clade VIII which groups freshwater or euryhaline strains with only PC (Fuller *et al.*, 2003; Dufresne *et al.*, 2008; Hunter-Cevera *et al.*, 2016b). About 16-18 clades have been defined within SC 5.1 (Farrant *et al.*, 2016); see also II.3.b), and twice as many subclades (Mazard *et al.*, 2012). *Prochlorococcus* forms a group sister to SC 5.1, and between 8 and 16 different *Prochlorococcus* clades have been

recognized using different markers (**figure 26**; see also II.3.b; Urbach *et al.*, 1992; Scanlan, 2012; Shih *et al.*, 2013; Biller *et al.*, 2014a; Huang *et al.*, 2012). SC 5.2 mostly gathers freshwater and euryhaline strains with PC as the main light-harvesting pigment, but a few strains also contain PE (Chen *et al.*, 2004). Finally, SC 5.3 (formerly SC 5.1 clade X) is similar to SC 5.1 and groups marine strains with PE as the main light-harvesting (Dufresne *et al.*, 2008).

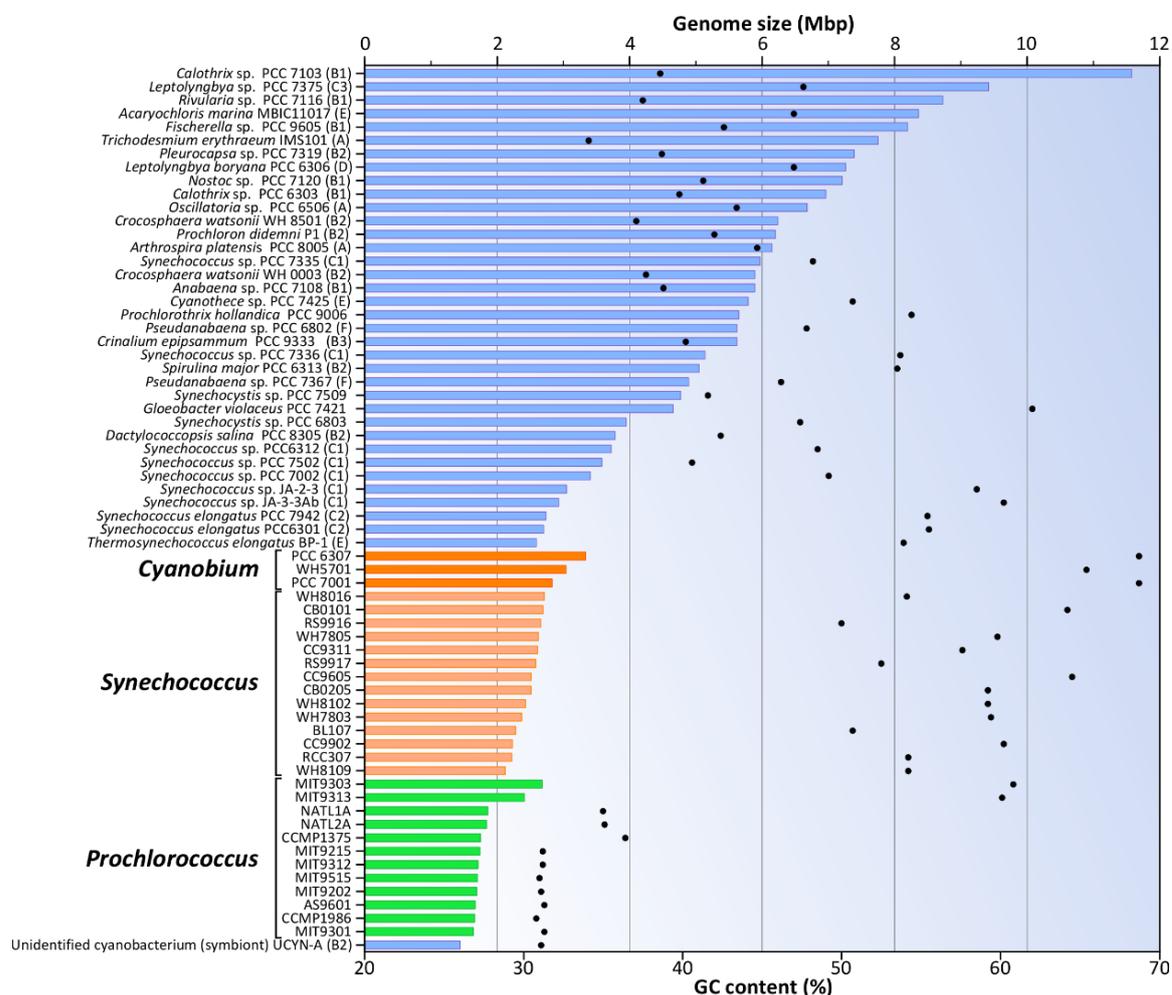


**Figure 26: Phylogenetic relationships amongst marine picocyanobacteria.** Strains with a sequenced genome are in bold. Tree made using the 16S rRNA gene and a neighbour-joining algorithm. Modified from (Scanlan, 2012).

However, metagenomes from freshwater reservoirs and the Baltic Sea suggest that some SC 5.3 strains might be euryhaline or have low salt requirements (Cabello-Yeves *et al.*, 2017; Celepli *et al.*, 2017). At least 6 clades have been recognized in this subcluster based on environmental 16S-23S rRNA sequences (Huang *et al.*, 2012). SC 5.2 and 5.3 are much less known than the *Prochlorococcus* and SC 5.1 lineages, and their relative positions within the picocyanobacteria radiation is unclear, some markers suggesting that SC 5.2 is more basal than *Synechococcus* SC 5.3 and 5.1 and *Prochlorococcus* (**figure 26**; Scanlan, 2012), while others suggest that SC 5.3 is more basal. Based on a simple molecular clock model, Dufresne and co-workers suggested that the separation of SC 5.1 *Synechococcus* and *Prochlorococcus* occurred about 150 Million years ago (Ma; Dufresne *et al.*, 2005). Using a more complex model and a Bayesian framework, Sanchez-Baracaldo and co-workers more recently suggested the most recent common ancestor (MRCA) of *Prochlorococcus* appeared between 543 and 684 Ma, the MRCA of SC 5.1 *Synechococcus* between 421 and 550 Ma, and the MRCA of marine picocyanobacteria about 1 Ga (Sánchez-Baracaldo *et al.*, 2014). A comparison of different methods, datasets and molecular clock calibration points concordantly suggested the MRCA of *Prochlorococcus* appeared 408 to 846 Ma, and the MRCA of SC 5.1 between 270 and 614 Ma (Sánchez-Baracaldo, 2015). Finally, analysis based on 192 orthologous genes suggested a divergence time of about 1.1 Ga for SC 5.1, 1 Ga for *Prochlorococcus* and 1.5 Ga for marine picocyanobacteria (Dvořák *et al.*, 2014). All these analyses suggest a very ancient origin for marine picocyanobacteria. Even if the precise dating remains unknown, it is probably safe to assume that SC 5.1 diverged at least 300 Ma and *Prochlorococcus* 400 Ma.

The evolution of marine picocyanobacteria was accompanied by an important reduction in the average genome size of its members compared to other cyanobacteria (**figure 27**; Dufresne *et al.*, 2005, 2008; Scanlan *et al.*, 2009). In a set of 97 genomes, of which 54 corresponded to marine *Synechococcus* and 43 to *Prochlorococcus*, the genome size of marine *Synechococcus* ranges from 2.1 million base pair (Mbp) in M16.1 or WH8109 to 3.3 Mbp in BIOS-E4-1, while *Prochlorococcus* have even smaller genomes (1.6-2.7 Mbp; Scanlan *et al.*, 2009). Accordingly, the number of genes encoded in these genomes ranges from 1,797 to 3,047 in *Prochlorococcus* sp. HLNC2 (clade HLIII) and MIT0701 (LLIV) respectively, and from 2,386 to 4,439 in *Synechococcus* sp. CC9902 (clade IV) and BIOS-E4-1 (CRD1) respectively (Cyanorak v.2; Farrant *et al.*, 2016; Doré *et al.*, *in prep*). The core genome (defined as the set of orthologous genes present in all genomes and supposed to correspond to the minimal set strictly necessary for maintaining cell function and integrity) of *Synechococcus* includes 1,218 genes, and 1,015 genes in *Prochlorococcus* (Farrant *et al.*, 2016; Doré *et al.*, *in prep*). This highly reduced set indicates an additional genome streamlining during *Prochlorococcus* evolution, which probably corresponds to an adaptation to its oligotrophic lifestyle through the minimization of the

resources necessary to sustain its growth (Partensky and Garczarek, 2010; Dufresne *et al.*, 2005; Scanlan *et al.*, 2009).



**Figure 27: Genome size (bar) and G+C content (plot) of marine picocyanobacteria compared to other cyanobacteria.** Figure from (Humily, 2013).

Although all *Prochlorococcus* and a significant proportion of *Synechococcus* isolates are considered as belonging to the same species using the threshold of less than 3% divergence in the 16S rRNA sequence commonly used in bacterial taxonomy, marine picocyanobacteria have a tremendous genetic diversity (Dufresne *et al.*, 2008; Biller *et al.*, 2014a; Chisholm, 2017). In fact, most sequenced isolates of marine picocyanobacteria would correspond to different species using the 94% threshold for genome average nucleotide identity commonly used to delineate species based on their genome (Konstantinidis and Tiedje, 2005; Biller *et al.*, 2014a; Dufresne *et al.*, 2008; Farrant *et al.*, 2016; Doré *et al.*, *in prep*). This diversity is reflected in the variable part of the genome. Genes are designated as accessory or unique if they are found in a restricted set of genomes or in only one genome, and the pan-genome corresponds to the global genetic repertoire

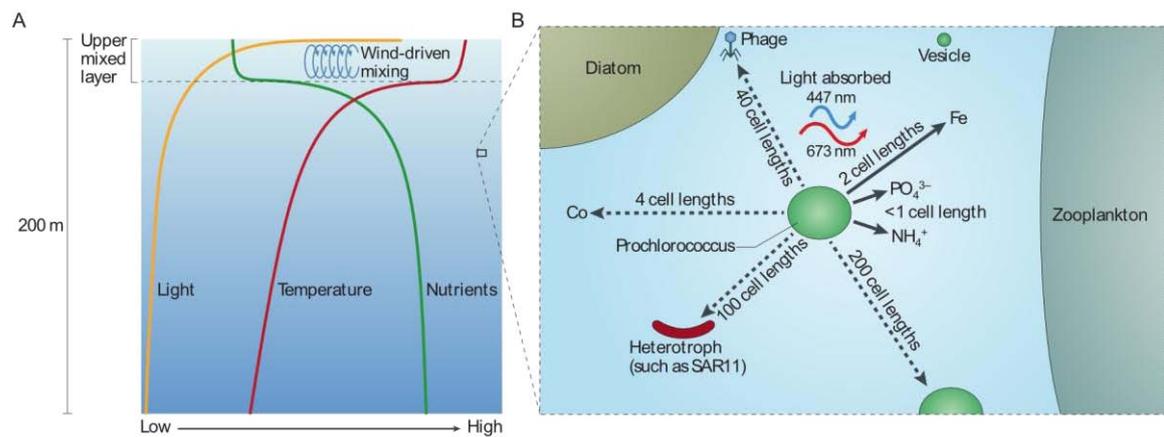
of a taxon and is defined as the set of all core, accessory and unique genes found in the genomes of all isolates. The pan-genome for the 54 *Synechococcus* complete genomes is of ~21,500 genes, and of ~7,500 genes for the 43 *Prochlorococcus* complete genomes (Farrant *et al.*, 2016; Doré *et al.*, *in prep*). However, these numbers do not reflect the genetic diversity found in these two genera, and each newly sequenced genome adds hundreds of new genes to this total (Biller *et al.*, 2014a). As an example, an analysis of complete genomes from cultured isolates, incomplete genomes generated from single cells isolated from the environment and consensus metagenomic assemblies showed a pan-genome size of >13,000 genes for 140 *Prochlorococcus* genomes, and projections suggest a pan-genome of ~85,000 genes for the whole *Prochlorococcus* genus (Biller *et al.*, 2014a).

Accessory genes are frequently found in hypervariable regions of the genomes, which present distinctive genomic characteristics (GC%, di-, tri- and tetranucleotide frequencies), and are thought to be easily laterally transferred in a population (Six *et al.*, 2007c; Dufresne *et al.*, 2008; Palenik *et al.*, 2006; Coleman *et al.*, 2006; Rocap *et al.*, 2003; Kashtan *et al.*, 2014, 2017; Biller *et al.*, 2014a; Doré *et al.*, *in prep*), although this has never been formally demonstrated. These genomic islands represent an efficient way of generating diversity in otherwise identical “genomic backbones” (Kashtan *et al.*, 2014, 2017). Genes encoded in these genomic islands have a suggested or demonstrated adaptive value for the colonization of new ecological niches (Six *et al.*, 2007c; Dufresne *et al.*, 2008; Martiny *et al.*, 2006, 2009a; Biller *et al.*, 2014a; Kashtan *et al.*, 2014, 2017; Berube *et al.*, 2015; Stuart *et al.*, 2009; Shukla *et al.*, 2012; Humily *et al.*, 2013; Paz-Yepes *et al.*, 2013; Palenik *et al.*, 2006).

### **3. Adaptive microdiversity in picocyanobacteria**

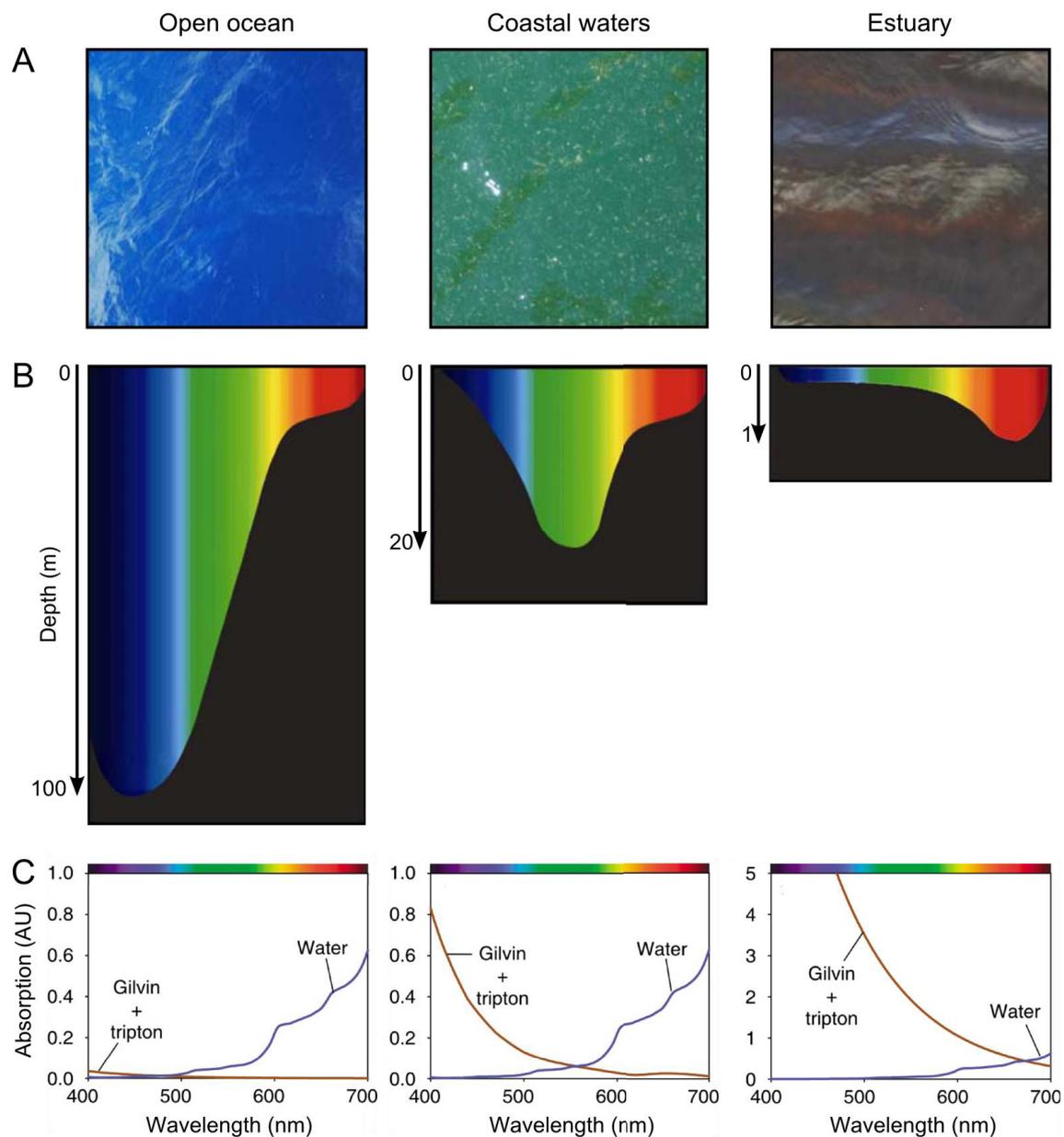
#### ***a. Environmental variability in the ocean***

Despite its apparent stability, the marine environment and in particular the euphotic zone is characterized by pronounced physical (light, temperature) and chemical (nutrients and other organic molecules) gradients. Additionally, it is inhabited by numerous organisms that interact either directly with each other in beneficial (as has been observed between heterotrophs and both *Synechococcus* and *Prochlorococcus*; Morris *et al.*, 2008; Christie-Oleza *et al.*, 2015b, 2017; Biller *et al.*, 2016), or antagonist relationships (e.g. grazing, viral infections; Biller *et al.*, 2014b; Guidi *et al.*, 2016; Fridman *et al.*, 2017) or allelopathic interactions, (Li *et al.*, 2010; Paz-Yepes *et al.*, 2013), or indirectly by competing for the same resources (Hutchins *et al.*, 1999; Johnson *et al.*, 2006). The tremendous genetic diversity found in marine picocyanobacteria is largely influenced by these factors, and thus can only be understood in an ecological context.



**Figure 28: Overview of the marine environment experienced by marine picocyanobacteria. (A)** Temperature, light and nutrient gradients in the water column. Light intensity decreases exponentially with depth. The surface part of the water column is consistently mixed by wind, waves and heat-driven turbulence, homogenizing temperature and nutrients. This mixed layer has a variable thickness, and can be only a few meters thick or extend to the whole water column. Below this mixed layer, temperature decreases with depth, whereas nutrients concentration increases due to the progressive remineralization of sinking particles by heterotrophs (see **figure 21**). **(B)** Typical distances between a *Prochlorococcus* cell and nutrients or other organisms (other *Prochlorococcus* cell, heterotrophic bacteria and phage). Figure reproduced from (Biller *et al.*, 2014a).

The euphotic zone is characterized by opposite gradients of temperature, light and nutrients (**figure 28**). Light intensity decreases exponentially throughout the water column. Temperature and nutrients concentration are relatively stable in the upper mixed layer, but respectively decreases and increases with depth. Marine picocyanobacteria growth is typically limited by lack of nutrients, typically nitrogen (Partensky and Garczarek, 2010; Partensky *et al.*, 1999b; Mary *et al.*, 2008), phosphorus (Coleman and Chisholm, 2007; Tsiola *et al.*, 2016) and iron (Ahlgren and Rocap, 2006; Sohm *et al.*, 2016; Farrant *et al.*, 2016), which are limiting or co-limiting in different areas of the world ocean. Thus, these organisms are faced with a trade-off between light requirements for photosynthesis (surface) and nutrients required for anabolism (depth). Marine picocyanobacteria living in the oligotrophic ocean experience a highly dilute environment, in which macro- and micronutrients are on average a few cells away (**figure 28**; Biller *et al.*, 2014a). In this context, the cell and genome size reduction observed in both *Synechococcus* and *Prochlorococcus* compared to other cyanobacteria is thought to be an adaptation to this dilute environment (Blank and Sánchez-Baracaldo, 2010; Biller *et al.*, 2014a; Dufresne *et al.*, 2005). Similarly, in *Prochlorococcus*, the replacement of the phycobilisome by an intrinsic Pcb-based light-harvesting antenna that requires less nitrogen and iron, and the replacement of sulfolipids by phospholipids are likely adaptations to its oligotrophic lifestyle (Ting *et al.*, 2002; Partensky *et al.*, 1999a; Biller *et al.*, 2014a).



**Figure 29: Gradients of light colour in different aquatic environments.** Left column, ALOHA Station in the Pacific ocean; middle column, coastal waters in the Baltic Sea; right column, turbid waters in lake Groote Moost. **(A)** Visual aspect of the aquatic environment; **(B)**, distribution of light wavelengths throughout the water column for the different environments; **(C)** model explaining the differential absorption of light colours with depth for various composition of the water. Pure water absorbs more longer than shorter light wavelengths, whereas gilvin (coloured dissolved organic matter) and tripton (coloured particulate organic matter) preferentially absorb short wavelengths. The light colour profile of oligotrophic waters corresponds mostly to pure water, whereas coastal waters are greener due to the increased gilvin and tripton concentration. Turbid lake or estuarine waters have a lot of dissolved and particulate organic matters, which give them their characteristic brown colour. Modified from (Stomp *et al.*, 2004, 2007; Stomp, 2008).

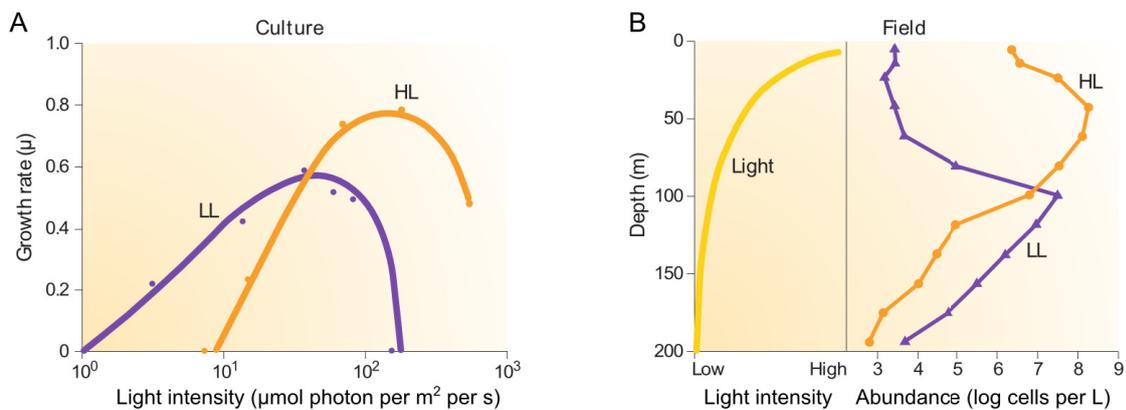
The stratification patterns vary both geographically and seasonally. As an example, in temperate environments, the increased temperatures of surface waters during summer induce a stratification of the water column, and phytoplanktonic growth becomes limited as the nutrients are absorbed during the season. During winter, the conjugation of cooling of surface water and recurrent storms homogenize the water column, destroying the stratification and enriching the surface waters by mixing with nutrient-rich deep waters. This nutrient enrichment induces favourable conditions and sustains phytoplanktonic growth during the next spring/summer, when the temperatures increase sufficiently to sustain positive net growth (Lindell and Post, 1995; Hunter-Cevera *et al.*, 2016a). On the contrary, the water column stratification is only occasionally disrupted in intertropical areas, in which *Prochlorococcus* is mostly found.

In addition to light intensity gradients, the different colours of light are also differentially filtered by the water column (**figure 29**; Stomp *et al.*, 2004, 2007). Red light is preferentially absorbed and scattered by water compared to blue light. On the contrary, dissolved and particulate organic matter preferentially absorbs blue light. Phytoplankton abundance also impacts the light colour, by absorbing blue and red light with chlorophylls. Thus, blue wavelengths (450-500 nm) penetrate the deepest in oligotrophic environments (typically down to 150-250 m), in which nutrients concentration are very low, and thus characterized by low phytoplankton abundance and biomass, and low concentration of organic matter. In contrast, green wavelengths penetrate the deepest (typically down to 20 m) in coastal waters, where the concentration of chlorophyll and organic matter is more important and filters out blue light and red light. Finally, in highly turbid lakes or estuaries, the load in organic matter filters is so high that light is very rapidly attenuated, and only the longest wavelengths penetrate to at most 1 m. It is also worth noting that the atmosphere also differentially filters light colours, as can be seen during sunsets and sunrises. Thus, light quality also varies with seasons and latitudes, both influencing the thickness of air layer crossed by light before reaching the ocean.

### ***b. Prochlorococcus genetic diversity is shaped by light and temperature***

The observation that two *Prochlorococcus* isolates had widely differing photophysiology, with different pigment concentrations and different light intensity optima, one growing optimally at light intensities at which the other was photoinhibited, led to hypothesize the existence of genetically distinct populations adapted to different light intensity niches (Campbell and Vault, 1993; Moore *et al.*, 1995). Two populations with high and low fluorescence were later observed *in situ*, presenting different distribution with depth (**figure 30**). Isolates from both populations corresponded to the two photophysiology previously described. Genetic analysis demonstrated that these two populations, although closely related, were

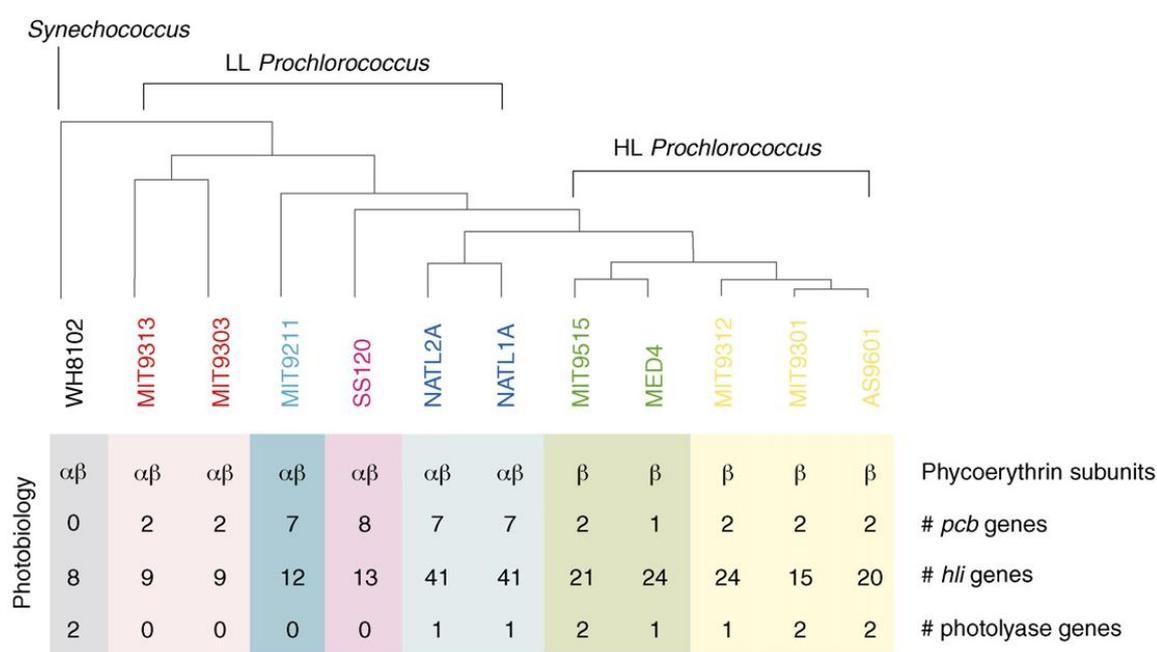
genetically distinct (**figure 26**). The concept of ecotype was created to designate these genetically and physiologically differentiated lineages from the same species, and these two populations were referred to as high-light (HL) and low-light (LL) ecotypes (Moore *et al.*, 1998); for a complete review of *Prochlorococcus* genetic diversity and ecotypes, see (Biller *et al.*, 2014a). However, the ecotype term implies the monophyly of the group considered. If this is true for the HL ecotype, the discovery of additional basal clades in the *Prochlorococcus* radiation made the “LL” ecotype paraphyletic (**figure 26**). Thus, it might be more appropriate to designate these different clades as different LL ecotypes. Accordingly, the different LL subclades have slightly differing photophysiology. In particular, the LLI ecotype, which is the nearest relative of HL *Prochlorococcus*, tolerate light shocks and can be found throughout the water column (Johnson *et al.*, 2006; Biller *et al.*, 2014a).



**Figure 30: Light ecotypes in *Prochlorococcus*.** (A) Light physiology of the HL and LL ecotypes. The LL ecotype is photoinhibited (no net growth) at the intensity at which the HL ecotype achieves maximal growth, and conversely that the LL ecotype grows well at light intensities below the HL ecotype compensation point. (B) Typical distribution of both light ecotypes in the environment. The HL ecotype dominates in the upper part of the water column, whereas the LL ecotype dominates below 100 m. Figure reproduced from (Biller *et al.*, 2014a).

These distinct photophysiology correspond to differences in the light-harvesting antenna of the different ecotypes. HL and LL strains present different pigment ratios and size of their Pcb antennas, allowing the latter to optimally absorb the blue photons present at depth (Partensky *et al.*, 1997; Bibby *et al.*, 2001; Moore *et al.*, 1995, 1998; Coleman and Chisholm, 2007; Garczarek *et al.*, 2007). From a genetic standpoint, all LL strains but LLIV present an extended Pcb gene repertoire, with up to eight Pcb copies (of two are associated to PSI, and the remaining six to PSII) compared to one or two (for each PS) in most HL strains (**figure 31**; Garczarek *et al.*, 2000; Coleman and Chisholm, 2007). Similarly, more high-light inducible proteins and photolyases, both of which are involved in photoprotection mechanisms, are found in HL strains (**figure 31**; Coleman and Chisholm, 2007). Interestingly, even though

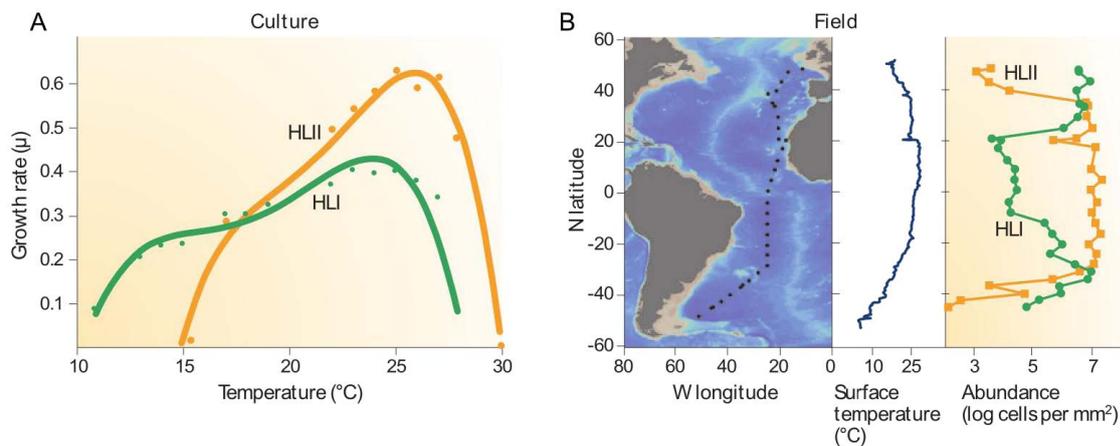
*Prochlorococcus* do not have phycobilisomes and lost the genes necessary for the synthesis of APC, PC and their associated linker proteins, all strains still have at least one some PE remnants. Indeed, while HL strains possess only a highly degenerated PE subunit ( $\beta$ -PEIII) and a few other PE biosynthesis genes (Hess *et al.*, 1999; Wiethaus *et al.*, 2010a, 2010b), LL strains have a comparatively extended gene set, encoding  $\alpha$ - and  $\beta$ -PEIII subunits, several putative phycobilin lyases (*cpeZ*, *cpeY*, *cpeF*, *cpeT*, *cpeS*), a putative linker protein (*ppeC*; absent from LLIV) and uncharacterized genes with a likely phycobilisome-related function (*unk3*, *unk6*, *unk13*; **figure 31**; Hess *et al.*, 1999). However, the function of these phycobiliproteins in LL *Prochlorococcus* is not well understood, although it has been suggested that they could serve directly in light harvesting or as a blue/green photosensor (Lokstein *et al.*, 1999; Steglich *et al.*, 2005).



**Figure 31: Copy number (#) and phyletic pattern of genes involved in photosynthesis in various *Prochlorococcus* isolates.** Pcb, prochlorophyte chlorophyll binding protein; hli, high-light inducible protein. Figure reproduced from (Coleman and Chisholm, 2007).

More recently, additional nested ecotypes differing in their temperature range and optima have been recognized within HL *Prochlorococcus* (**figure 32**; Johnson *et al.*, 2006; Biller *et al.*, 2014a). As for HL/LL ecotypes, the culture-based characterization of different isolates revealed shifted responses to temperature between different HL clades, with HLI having lower temperature preferenda than HLII strains. Accordingly, the different genotypes exhibit different distributions, with the high temperature HLII prevailing in intertropical zones (between 30°N and 30°S), and HLI above or below 30°N and 30°S (**figure 32**, **figure 34**; Johnson *et al.*, 2006; Biller *et al.*, 2014a). It is interesting to note that the HLI and -II are not entirely excluded from each

other thermal niche, but rather that the abundance ratio of these two ecotypes follows a log-linear relationship with temperature (Chandler *et al.*, 2016).



**Figure 32: Temperature ecotypes within the *Prochlorococcus* HL ecotype. (A)** Thermophysiology of HLI and HLII ecotypes. Note the shifted temperature optima between the two ecotypes. **(B)** Typical distribution of both temperature ecotypes in the environment. HLI dominates in colder environments, and HLII in warmer intertropical environments. Figure reproduced from (Biller *et al.*, 2014a).

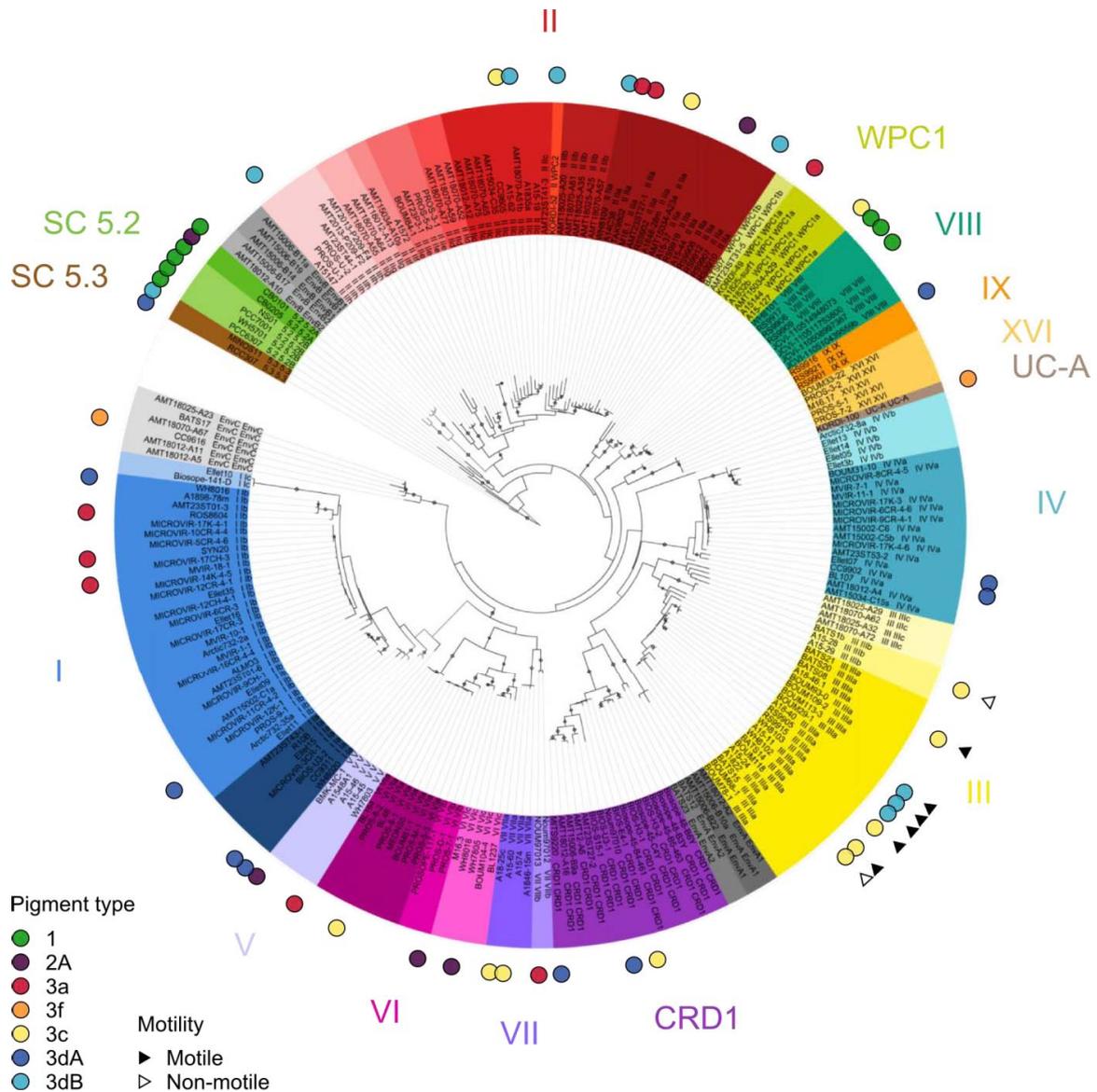
Some *Prochlorococcus* clades have been defined based on environmental sequences and lack cultured representatives (e.g. HLIII to –VI), and different studies suggested that these clades might correspond to nutrient (nitrogen, phosphate or/and iron) ecotypes (Biller *et al.*, 2014a). Recently, Kashtan and co-workers observed populations with a common “genomic backbone” and different sets of flexible genes, and suggested that adaptation at finer phylogenetic levels was mainly due to genomic islands (Kashtan *et al.*, 2014, 2017). Accordingly, the phyletic pattern (presence/absence of genes in a phylogenetic pattern, see **figure 31**), of genes involved in phosphate acquisition is incongruent with the core genome phylogeny of *Prochlorococcus*, suggesting that these genes are laterally transferred (Berube *et al.*, 2015).

### ***c. Synechococcus diversity***

#### ***i. Genetic (micro)diversity***

No ecotype as clear-cut as in *Prochlorococcus* have been defined in *Synechococcus*, due to i) the increased genetic diversity of *Synechococcus* compared to *Prochlorococcus*, ii) the different markers used in environmental studies have different resolutions, and some markers can fail at differentiating some clades from others, thus obfuscating potential differences in their

ecology, and iii) because the ecological preferences of the different *Synechococcus* clades seem less clear-cut and obvious than for *Prochlorococcus*.



**Figure 33: Major clades of the marine *Synechococcus* radiation.** Phylogenetic tree based on 230 *petB* sequences and inferred using ML. Circles represent bootstrap support >70%. Tree courtesy of Laurence Garczarek.

Still, a general picture has been emerging in the recent years about the ecology of the major clades of *Synechococcus*. The deep relationships between clades are not well resolved, and vary depending on the marker used, like the 16S rRNA (Fuller *et al.*, 2003), 16S-23S rRNA ITS (Ahlgren and Rocap, 2006), *rpoC1* (encoding the RNA polymerase gamma subunit; Palenik, 1994; Xia *et al.*, 2017c), *narB* (nitrate reductase; Paerl *et al.*, 2011) or more recently *petB* (cytochrome  $b_6$ ), which provides a very good phylogenetic resolution (Mazard *et al.*, 2012;

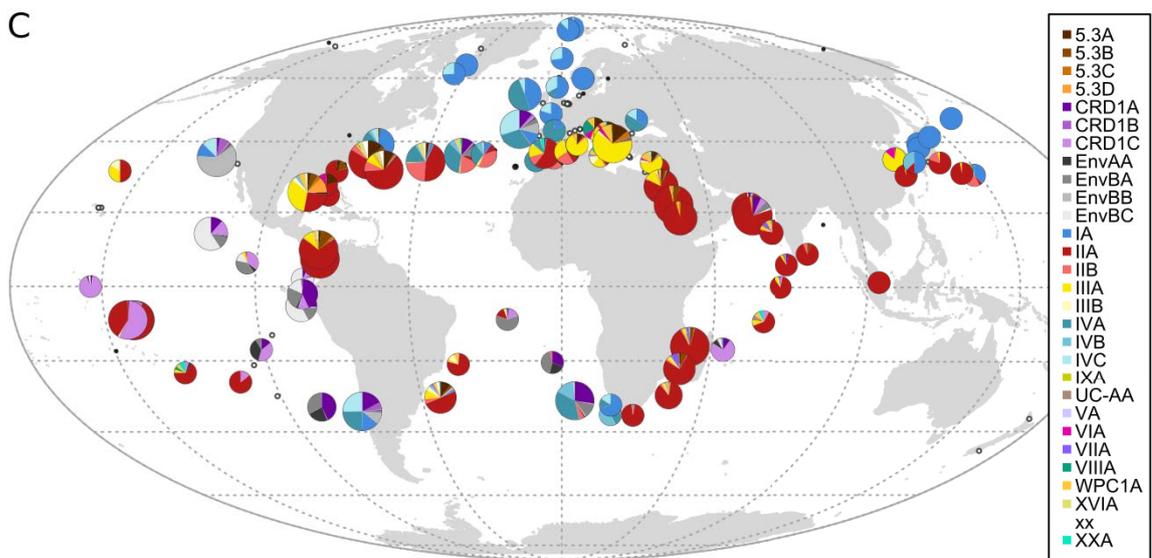
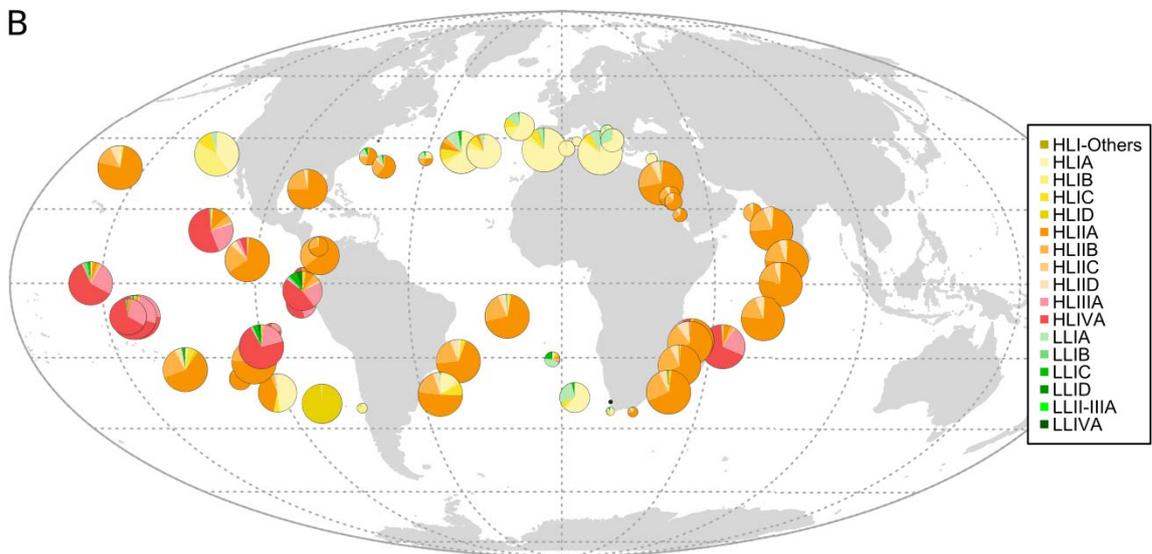
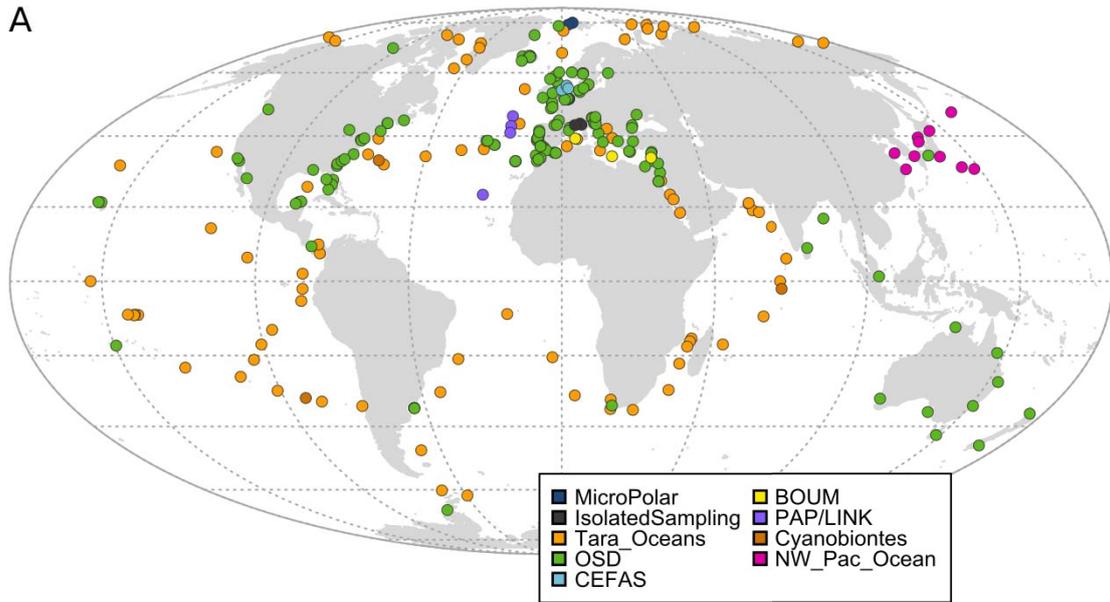
Farrant *et al.*, 2016). However, the clades II, III and IV are generally considered as belonging to the same deep-branching subcluster 5.1A, whereas clades I, CRD1, VI, VII, VIII and IX form the subcluster 5.1B (Dufresne *et al.*, 2008). It has been suggested that these two subclusters correspond to coastal (5.1B) or open-ocean (5.1A) “ecogroups” (Dufresne *et al.*, 2008), but new findings tend to infirm this hypothesis (Sohm *et al.*, 2016; Farrant *et al.*, 2016).

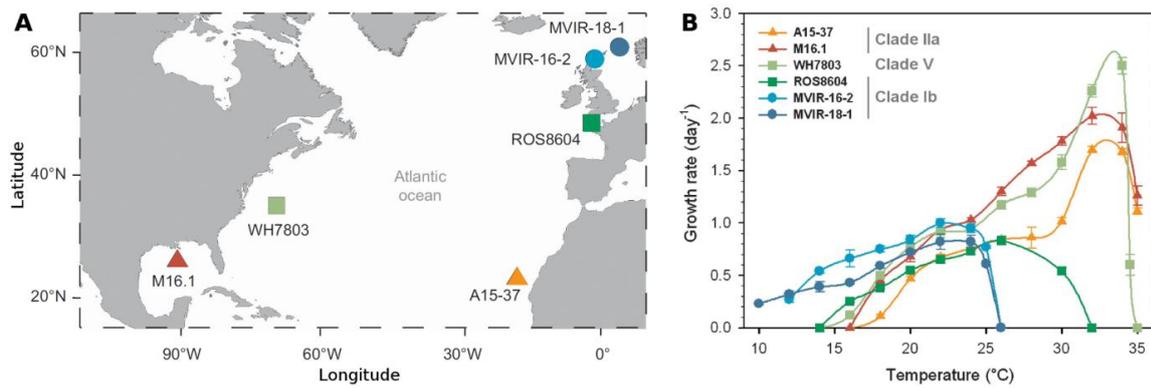
Similarly, *Synechococcus* from clade VIII, which have been isolated from coastal areas, have a large salinity range and are the only euryhaline strains within SC 5.1 (Dufresne *et al.*, 2008; Fuller *et al.*, 2003; Celepli *et al.*, 2017). Strikingly, isolates from clade VIII share a high number of genes with strains from the euryhaline SC 5.2. Additionally, genes specific of both groups are highly similar, and are found in specific genomic islands, suggesting lateral gene transfer between these distant lineages (Dufresne *et al.*, 2008). Another phenotypic trait shared between clade VIII and SC 5.2 strains is their pigmentation, as most isolates reported to date solely contain phycocyanin, or occasionally both phycocyanin and phycoerythrin-I (CB0250 in SC5.2 and HK01 in clade VIII; Xia *et al.*, 2017b), but never contain phycoerythrin-II (**figure 33**; see also II.3.c.ii).

If no specific trait has been found in other marine *Synechococcus* lineages, environmental studies suggested differences in the ecological niches of the major clades and sub-clades. Clades I and IV co-dominate in cold, mesotrophic waters above 30°N, with clade I particularly prevailing in the coldest arctic waters (**figure 34**; Sohm *et al.*, 2016; Farrant *et al.*, 2016; Paulsen *et al.*, 2016; Zwirgmaier *et al.*, 2008, 2007; Choi *et al.*, 2016). Clade II is most abundant and dominates in warm tropical and intertropical waters, corresponding to mesotrophic or iron-replete oligotrophic environment with a temperature above 20°C (**figure 34**; Sohm *et al.*, 2016; Farrant *et al.*, 2016; Mazard *et al.*, 2012; Zwirgmaier *et al.*, 2007, 2008). The distribution of clade III is less correlated with temperature, as it is found in environments with a wide range of temperatures (10-28°C; Zwirgmaier *et al.*, 2008), but is strongly associated with low phosphate concentrations such as in the Gulf of Mexico or the Mediterranean Sea (**figure 34**; Farrant *et al.*, 2016; Zwirgmaier *et al.*, 2007; Mazard *et al.*, 2012). Finally, clade CRD1, which was first described in the Costa Rica upwelling (Costa Rica Dome; Saito *et al.*, 2005), dominates in iron-depleted areas of the Pacific (Farrant *et al.*, 2016; Sohm *et al.*, 2016).

---

**Figure 34 (next page): Distribution of the different *Prochlorococcus* and *Synechococcus* clades in oceanic surface waters, assessed with the high-resolution phylogenetic marker *petB*. (A) Overview of the sampling cruises and sites for which *petB* data are available through metabarcoding or in silico read recruitment from metagenomes. (B, C) Distribution of *Prochlorococcus* (B) and *Synechococcus* (C) clades in the global ocean. Only Tara Oceans data have been used for *Prochlorococcus* as the primer set used in metabarcoding studies does not efficiently amplify *Prochlorococcus*. Unpublished figure courtesy of Hugo Doré.**





**Figure 35: thermal range of different *Synechococcus* strains isolated along a latitudinal gradient. (A)** Location of isolation sites of the strains used. **(B)** Thermal range and optima in which the different strains achieve growth. Figure modified from (Pittera *et al.*, 2014).

More recently, different environmental or culture-based studies revealed an important yet ecologically relevant microdiversity within each clades. By comparing the thermal acclimation of different isolates, Pittera and co-workers demonstrated that different isolates from different but also from the same clade could exhibit markedly different thermal range and optima (**figure 35**), and later linked these differences with adaptive differences in the thermostability of their PBS (Pittera *et al.*, 2014, 2016). Farrant and co-workers defined Ecologically Significant Taxonomic Units (ESTUs) as genetically related subgroups within clades that co-occur in the field (Farrant *et al.*, 2016). This definition is conceptually quite similar to *Prochlorococcus* ecotypes, with some important differences: i) ESTUs are solely based on observed *in situ* abundance patterns and ii) ESTUs, although grouping subclades based on their observed environmental niche, do not need be monophyletic groups, even though they are related since belonging to the same clade. Using this approach, they showed for instance that within clade II, the ESTU IIB was found in significantly colder waters (14.1-17.5°C) than ESTU IIA and co-occurred with the cold-adapted clade IV. Similarly, the different ESTUs within CRD1 exhibited different thermal ranges, with CRD1B being found in colder areas than CRD1C, and CRD1A tolerating a wide temperature range encompassing both CRD1B and -C thermal niches (Farrant *et al.*, 2016).

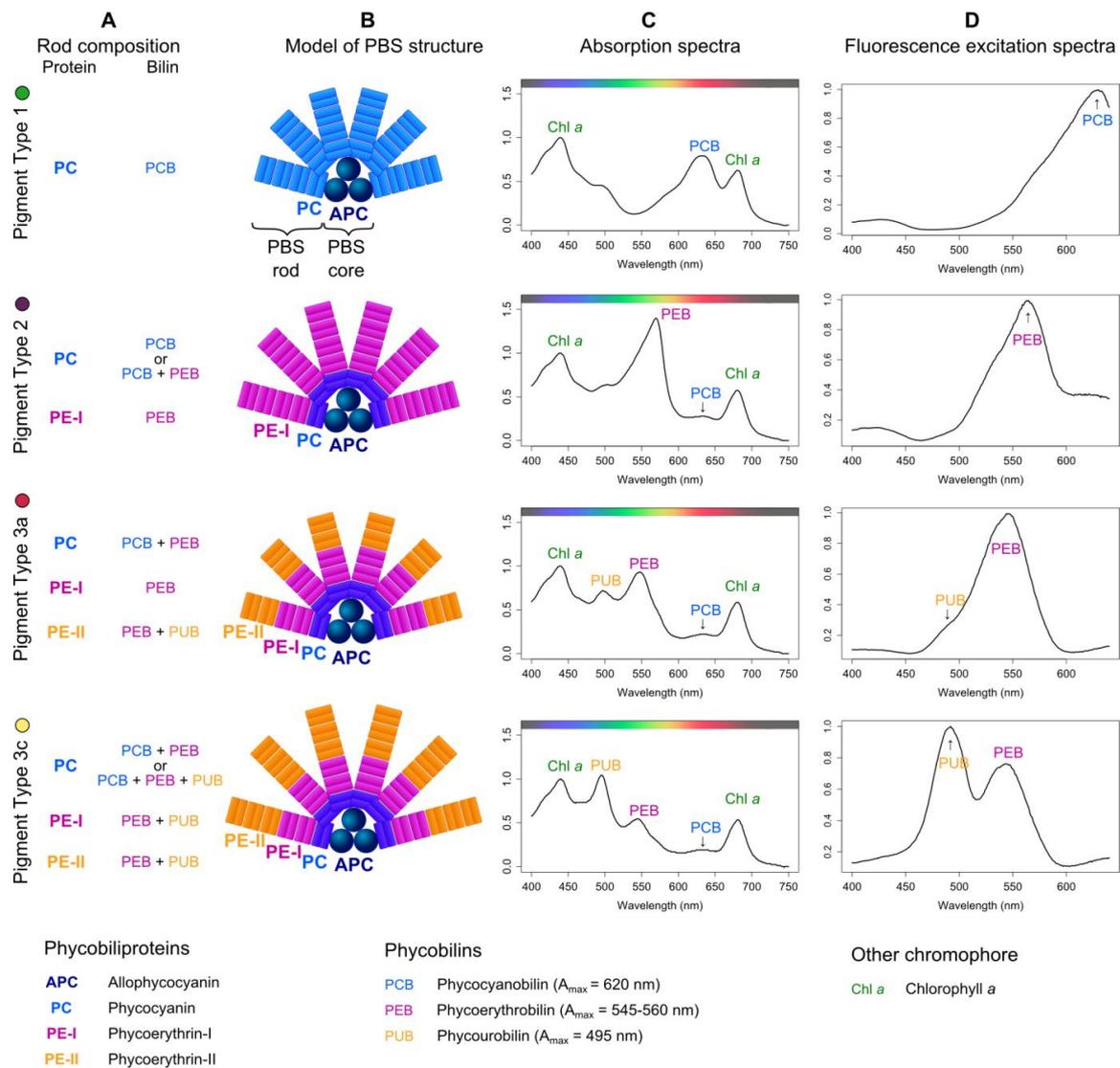
### ii. *Synechococcus* pigment types

The wide range of pigmentation displayed by *Synechococcus* strains was recognized early (**figure 36**; Waterbury *et al.*, 1986). Comparative physiology and biochemistry linked this diversity to differences in the phycobilisome composition, revealing that some strains possess rods made only of PC, both PC and PE-I, or both PC, PE-I and PE-II, and that the pigments attached onto the phycobiliprotein backbone can also vary between strains (Ong and Glazer, 1991; Sidler, 1994; Six *et al.*, 2007c; Everroad *et al.*, 2006).



**Figure 36: Picture of different cultures from strains with different pigment types.** Image courtesy of L. Garczarek.

In 2000, Palenik discovered that some strains were able to dynamically and reversibly change the ratio of PUB and PEB in their phycobilisome in response to blue light, a process that was later called Type IV Chromatic Acclimation (CA4; see I. 3.e.iii; Palenik, 2001; Everroad *et al.*, 2006). Six and co-workers defined three main pigment types (PT) based on the phycobiliproteins present in the phycobilisome rods: strains of PT 1 have rods solely made of PC, PT 2 have rods made of both PC and PE-I, and PT 3 corresponds to rods made of PC, PE-I and of the marine *Synechococcus* specific PE-II (**figure 37**; Six *et al.*, 2007c). Additionally, the chromophores attached on rod phycobiliproteins differ between PT: PT 1 only has PCB, PT 2 has both PCB and PEB, and PT 3 has PCB, PEB and PUB (**figure 37A**). Six and co-workers defined several subtypes within PT 3 based on the ratio of the two main chromophores PUB and PEB, as estimated from the ratio of fluorescence excitation at 495 nm and 545 nm with emission at 580 nm ( $Ex_{495:545}$ ). This ratio is low ( $Ex_{495:545} < 0.6$ ) in subtype 3a, which absorbs maximally green light, intermediate in subtype 3b ( $0.6 \leq Ex_{495:545} < 1.6$ ) and high ( $1.6 \leq Ex_{495:545}$ ) in subtype 3c, which maximally absorbs blue light (**figure 37C,D**; Six *et al.*, 2007c). Strains able to undergo CA4 are designated as 3d, and have a phenotype equivalent to PT 3a under GL, and equivalent to PT 3c under BL.

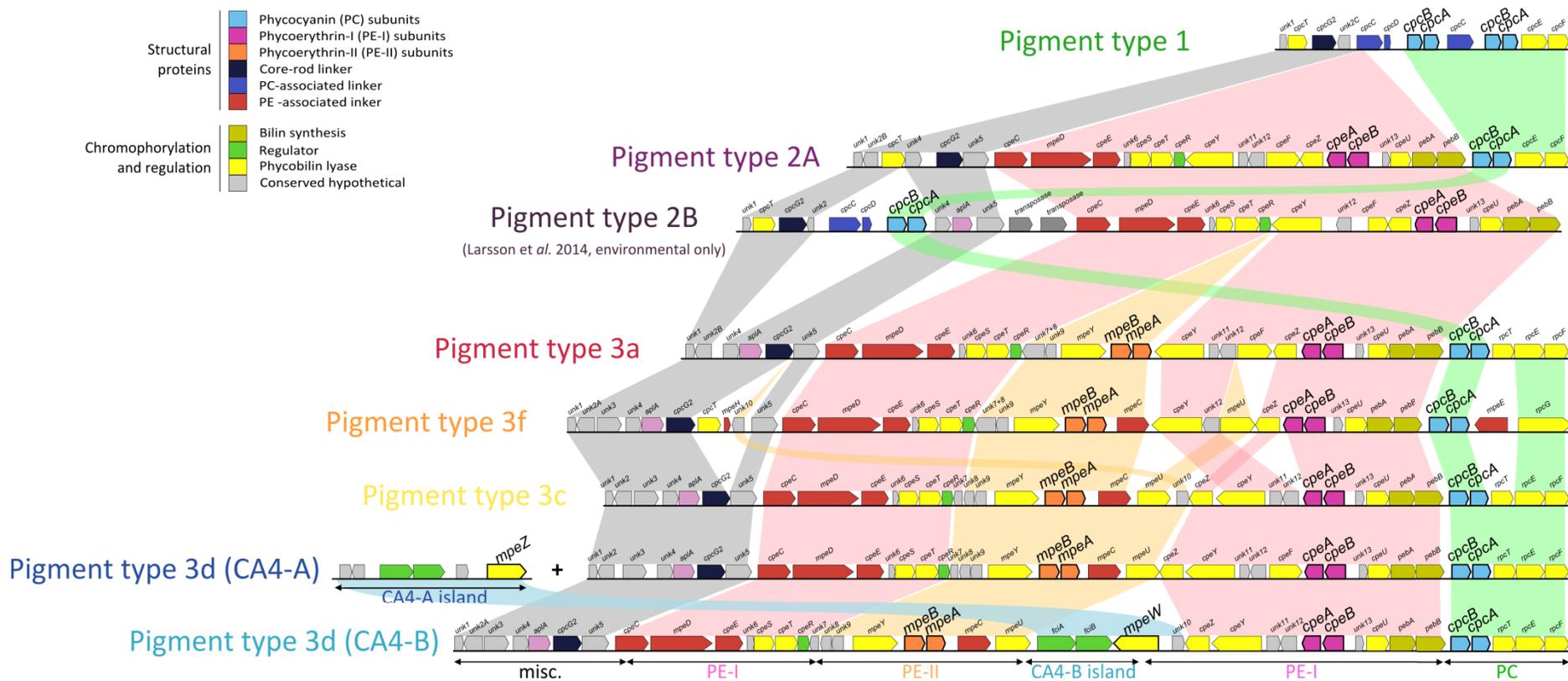


**Figure 37: Phycobilisome composition and spectral properties of different pigment types.** (A) Phycobiliproteins and phycobilins in the rods. (B) Model of PBS structure, as proposed in (Six *et al.*, 2007c). Note the progressive increase in rod length from PT 1 to PT 3c. (C) Whole-cell absorption spectra of strains from the corresponding pigment types (redrawn from Six *et al.*, 2007c). (D) Whole-cell fluorescence excitation with emission set at 680 nm.

By comparing the genomes of strains representative of the different pigment types, Six and co-workers found that all genes involved in the synthesis and regulation of phycobilisome rods were gathered into a dedicated genomic region, the gene content and organization of which corresponding to the different described pigment types, and which presents the genomic hallmarks of a genomic island (Six *et al.*, 2007c; Dufresne *et al.*, 2008). More recently, comparative genomic analysis showed that chromatic acclimation was correlated with the occurrence of a small specific genomic island (CA4 genomic island), that exists in two configurations (CA4-A and CA4-B), thus defining the two genotypes 3dA and 3dB (Humily *et al.*, 2013). Both CA4 genomic islands encode a tandem of regulators (*fciA* and *fciB*) as well as a

phycobilin-lyase (*mpeZ* in CA4-A and *mpeW* in CA4-B). Some strains possessing a CA4 genomic island are not able to perform CA4, and display a fixed Ex<sub>495:545</sub> similar to 3a, 3b or 3c strains. Interestingly, all 3b strains sequenced so far are in fact such “CA4-deficient” strains (Humily *et al.*, 2013; see also I. 3.e.iii).

Recently, Larsson and collaborators found a new PBS rod genomic region with an original gene order and complement in assembled metagenomes from the Baltic Sea. Based on its content, they predicted that it corresponds to strains having rods made of PC and PE-I, and referred to it as 2B (Larsson *et al.*, 2014). Similarly, Xia and collaborators described a new PBS rod genomic region from two newly sequences strains (KORDI-100 and CC9616), and designated it PT 3f even though these strains are seemingly indistinguishable from PT 3c strains (Xia *et al.*, 2017b).



**Figure 38: Organization of PBS rod genomic region for different pigment types.** Pigment type 2B has been only described in metagenomes from the Baltic Sea (Larsson *et al.*, 2014). Pigment type 3f has been recently described by Xia and co-workers, and is seemingly indistinguishable phenotypically from PT 3c (Xia *et al.*, 2017b).

A striking feature of *Synechococcus* pigment types is that there is no correspondence between pigmentation and vertical phylogeny, and closely related strains can have different PT, whereas distantly related strains can have the same PT, which led some authors to suggest that the PBS genomic region or part thereof could be easily transferred between strains (**figure 33**; Six *et al.*, 2007c; Dufresne *et al.*, 2008; Fuller *et al.*, 2003; Ahlgren and Rocap, 2006; Everroad and Wood, 2012). Moreover, the comparison of phylogenies from phycobiliproteins and other genes involved in the synthesis of PBS rods with the core genome phylogeny revealed that *Synechococcus* pigment types have a different evolutionary history than the rest of the genome, and suggested that pigmentation evolved toward the absorption of blue wavelength (Six *et al.*, 2007c; Everroad and Wood, 2012). Similarly, the comparison of freshwater cyanobacteria with phenotypes comparable to PT 1 and PT 2 suggest that the different PT correspond to adaptations to different light niches (Stomp *et al.*, 2008, 2004).

Studies of *Synechococcus* pigment type distribution in the environment used different complementary approaches, including flow cytometry (Olson *et al.*, 1990; Sherry and Wood, 2001; Jiang *et al.*, 2016), fluorescence excitation spectra (Lantoine and Neveux, 1997; Wood *et al.*, 1998; Campbell *et al.*, 1998; Wood *et al.*, 1999; Neveux *et al.*, 1999; Yona *et al.*, 2014), epifluorescence microscopy (Campbell and Iturriaga, 1988), strain isolation (Choi and Noh, 2009; Fuller *et al.*, 2003; Chen *et al.*, 2004), clone libraries of *cpcBA* or *cpeBA* (Haverkamp *et al.*, 2008b; Chung *et al.*, 2015; Liu *et al.*, 2014; Xia *et al.*, 2017c), metagenomics (Larsson *et al.*, 2014; Díez *et al.*, 2016) or a combination of these (Haverkamp *et al.*, 2008a; Hunter-Cevera *et al.*, 2016b). However, studies based on optical properties cannot differentiate in the field chromatic acclimators from fixed pigment types and rather report “High-PUB” (3c and BL-acclimated CA4) or “Low-PUB” (3a and GL-acclimated CA4) phenotypes, and more generally potentially fail at distinguishing *Synechococcus* from other PEB or PUB-rich organisms such as *Trichodesmium*. Similarly, genetic analysis solely based *cpcBA* or *cpeBA* could not differentiate all pigment types, and in particular could not account for 3dB (Humily *et al.*, 2014; Larsson *et al.*, 2014; Xia *et al.*, 2017c; Liu *et al.*, 2014; Chung *et al.*, 2015). These studies showed that PT1 is restricted to and dominates in estuarine or surface of brackish/coastal waters, environments that are characterized by a low salinity and/or high turbidity (Haverkamp *et al.*, 2008a; Hunter-Cevera *et al.*, 2016b; Fuller *et al.*, 2003; Neveux *et al.*, 1999; Chung *et al.*, 2015; Liu *et al.*, 2014; Larsson *et al.*, 2014; Jiang *et al.*, 2016), PT2 is found in coastal shelf waters or in the transition zones between brackish and marine environments (Haverkamp *et al.*, 2008a; Hunter-Cevera *et al.*, 2016b; Fuller *et al.*, 2003; Larsson *et al.*, 2014; Jiang *et al.*, 2016; Wood *et al.*, 1998; Chen *et al.*, 2004), and the PT3 “Low-medium PUB” and “High PUB” are found along the gradient from onshore mesotrophic waters characterized by green light dominance to offshore oligotrophic waters where blue light penetrates deepest (Lantoine and Neveux, 1997; Olson *et al.*, 1990;

Sherry and Wood, 2001; Wood *et al.*, 1999; Hunter-Cevera *et al.*, 2016b; Choi and Noh, 2009; Campbell and Iturriaga, 1988; Neveux *et al.*, 1999; Campbell *et al.*, 1998; Larsson *et al.*, 2014). Some authors also reported an increase in the PUB/PEB ratio with depth (Lantoiné and Neveux, 1997; Olson *et al.*, 1990; Wood *et al.*, 1999), but other reported a constant ratio throughout the water column, suggesting that this depends on the location, water column features and environmental parameters (Neveux *et al.*, 1999; Campbell and Iturriaga, 1988; Yona *et al.*, 2014).

Finally, some studies suggested that PT 3a strains were able to achieve higher growth rates than their 3c counterparts, and possessed distinct light irradiance ranges, with 3a and 3c putatively corresponding to “high-light” and “low-light” phenotypes respectively (Six *et al.*, 2004; Mackey *et al.*, 2017).

## **Objectives of this work**

Marine *Synechococcus* spp. present a wide range of pigmentation. This diversity comes from variations in the phycobiliprotein and chromophore composition of their light-harvesting antennae, called phycobilisomes. While the genomic bases of pigment diversity are relatively well understood, notably the occurrence in all pigment types of a large genomic island involved in phycobilisome rod synthesis and regulation, and a second small island specific of chromatic acclimators and involved in CA4, several aspects of pigment diversity remain unknown. Specifically, the work presented in this manuscript aimed at answering the following questions:

- How did the different pigment types evolve? What is the cause for the discrepancy observed between vertical phylogeny and pigmentation? Can we explain why CA4 exists in two different versions?
- What is the extent of genetic and genomic diversity within each pigment type? Can this diversity be related with adaptations to environmental factors?
- What is the global distribution of the different pigment types in the environment? What can we learn on their ecology, and how this related to *Synechococcus* ecology, from this distribution?
- Phycobilin lyases have a key role in determining the inherent optical properties of the phycobilisome, yet they only a few have been characterized in marine *Synechococcus*. By genetically and biochemically characterizing different members of this enzyme family, can we relate their biochemical properties with the diversity of genes and alleles encoding them? What are the molecular bases of CA4-B? Can we predict the function and specificity of all phycobilin lyases present in the genome of the different pigment types?



# CHAPTER I

---

## Genomics and evolution of *Synechococcus* pigment types

« Premising, then, that very much remains to be learned,  
I will now proceed to give an outline of some of the principal facts  
which I have so far been able to observe in a more or less satisfactory manner. »

H.C. Sorby, 1875

## I. Context of the work

It has long been observed that, in marine *Synechococcus*, there is no apparent link between pigmentation and vertical phylogeny, a given clade containing isolates with different pigment types and, conversely, a given pigment type being present in different and sometimes distant clades (Toledo *et al.*, 1999; Rocop *et al.*, 2002; Six *et al.*, 2007c; Dufresne *et al.*, 2008). The extensive analysis of the genomic diversity associated with the different pigment types further showed that most genes involved in the synthesis and regulation of PBS rods are grouped into a genomic region, the gene content and synteny of which matched the corresponding pigment type, independently of the strain clade (Six *et al.*, 2007c). This observation led Six and co-workers to hypothesize that large chunks of the PBS genomic region could have been subject to lateral transfers between members of the different clades, even after the differentiation of these lineages, but this hypothesis has never been tested. Moreover, their comparative study was based on a restricted set of 11 *Synechococcus* genomes, with in several cases only one representative of the pigment type described, and the occurrence of a small genomic island specifically associated with type IV chromatic acclimation (CA4) was not known at that time and even less that this island existed in two variants (Humily *et al.*, 2013). Similarly, a new pigment type (PT 3f) has been recently described based on the organization of the PBS genomic region and strain phenotype (Xia *et al.*, 2017b).

This chapter aims at exploring different aspects of the evolution of *Synechococcus* pigment types using the wealth of genomic and metagenomic data recently available. In particular, I tried to address the following questions:

- i. What is the extent of genetic and genomic diversity within each pigment type, and how does this diversity relate with environmental parameters other than light?
- ii. How did the different pigment types evolve? How and why did two different versions of CA4 evolve?
- iii. How can we reconcile pigment type and core genome phylogenies, i.e. how can we explain the evolutionary history of pigment types in the light of marine *Synechococcus* evolution? Can we observe traces of horizontal transfer of parts or of the whole PBS genomic region between different strains? Alternatively, can other mechanisms explain the observed distribution of pigment types among isolates and clades?

Florian Humily, a former PhD student in the MaPP team, prepared and screened the fosmid libraries in collaboration with Morgane Ratin and Dominique Marie (two Engineers in our group), notably for flow-cytometry sorting. Gregory Farrant, another former PhD student,

assembled the genomes of newly sequenced strains as part of his PhD. I personally did all bioinformatic analyses presented in this chapter, including the assembly and manual curation of fosmid regions, the comparative analysis of assembled fosmid regions and of the phycobilisome genomic region of newly sequenced strains under the supervision of Drs. Laurence Garczarek and Frédéric Partensky. I proposed the model for the evolution of the different pigment types and the idea that population-scale mechanisms could be important for explaining the discrepancy between pigment type and core genome phylogenies, which are described in (II. Genomics and evolution of *Synechococcus* pigment types). I personally did all bioinformatic analyses presented in (III. Metagenomes suggest adaptation of phycobiliproteins to temperature).

## **II. Genomics and evolution of *Synechococcus* pigment types**

### **1. Diversity and evolution of the phycobilisome gene region in relation to light niches in marine *Synechococcus* cyanobacteria**

#### **Diversity and evolution of the phycobilisome gene region in relation to light niches in marine *Synechococcus* cyanobacteria**

Théophile Grébert, Laurence Garczarek, Florian Humily, Dominique Marie, Morgane Ratin, Alban Devailly, Gregory K. Farrant, Isabelle Mary, Daniella Mella-Flores, Gwen Tanguy, Karine Labadie, Patrick Wincker, David M. Kehoe and Frédéric Partensky.

IN PREPARATION

## Introduction

*Synechococcus* picocyanobacteria constitute an ecologically important component of marine phytoplankton, accounting for about 15% of the global net primary production (Flombaum *et al.*, 2013). They are widespread in the marine environment from the equator to the polar circles and from estuaries to clear open ocean waters and therefore thrive over a wide range of nutrient concentrations, temperature and light regimes (Olson *et al.*, 1990; Wood *et al.*, 1998; Partensky *et al.*, 1999; Zwirgmaier *et al.*, 2008; Ahlgren and Rocap, 2012; Farrant *et al.*, 2016). Members of this group, together with the *Prochlorococcus* and *Cyanobium* genera, form the cyanobacterial cluster 5 (Herdman *et al.*, 2001), which has been divided into three groups, called subclusters (SC) 5.1, 5.2 and 5.3 (Dufresne *et al.*, 2008). While SC 5.2 contains both euryhaline and freshwater isolates, SC 5.1 contain mostly marine isolates, except SC5.1 clade VIII that groups euryhaline strains. Although SC 5.3 was till recently thought to contain only marine representatives, a recent report analyzing metagenome assembled genomes (MAG) from two freshwater reservoirs suggested that this group also occurs in freshwater environments (Cabello-Yeves *et al.*, 2017). Of all three SC, SC 5.1 is by far the predominant one at the global scale (Farrant *et al.*, 2016) and is very diverse genetically, with about 20 clades described so far (Ahlgren and Rocap, 2012; Huang *et al.*, 2012; Mazard *et al.*, 2012). The five most abundant clades (I-IV and CRD1) predominate in distinct niches of the oceans characterized by different temperatures and/or limiting nutrients (Zwirgmaier *et al.*, 2008; Sohm *et al.*, 2016; Farrant *et al.*, 2016). *Synechococcus* are also very diverse in terms of pigment content and a number of studies have suggested that predominant *Synechococcus* populations in a given niche possess phycobilisomes (PBS) with light harvesting properties that match the light color of this niche (Six *et al.*, 2007; Everroad and Wood, 2012; Xia *et al.*, 2017b, 2017a; Grébert *et al.*, in revision).

*Synechococcus* PBS are made of a conserved central core made of allophycocyanin (APC) from which radiate from 6 to 8 rods (Sidler, 1994; Six *et al.*, 2007; Arteni *et al.*, 2009; Liu *et al.*, 2013). These rods consist of hexameric donut-shaped structures made of phycobiliproteins. Phycobilin lyases covalently attach chromophores on phycobiliproteins, which then assemble into hexamers that are stacked into rods with the help of linker proteins (Yu and Glazer, 1982; Adir *et al.*, 2006; Schluchter *et al.*, 2010; Bretaudeau *et al.*, 2013). The phycobiliprotein and phycobilin composition of these rods is highly variable in marine *Synechococcus* (Ong and Glazer, 1991; Six *et al.*, 2007). The simplest ones contain only phycocyanin (PC) and bind the red-light absorbing phycocyanobilin (PCB,  $A_{\max}=620-650$  nm) as the sole chromophore. Cyanobacteria that possess such PBSs are designated as pigment type 1 (PT 1; Six *et al.*, 2007). Rods that contain PC and phycoerythrin-I (PE-I) and bind PCB and phycoerythrobilin (PEB) a green-light (GL) absorbing chromophore ( $A_{\max}=545-560$  nm) are found in pigment type 2 (PT 2).

Finally, in pigment type 3 (PT 3), rods contain three different phycobiliproteins (PC, PE-I and phycoerythrin-II; PE-II) and bind the three different pigments: PCB, PEB and the blue-light (BL) absorbing phycourobilin (PUB,  $A_{\max}=495$  nm, Ong *et al.*, 1984; Six *et al.*, 2007). Several subtypes are defined within PT 3 depending on the ratio of PUB and PEB, a practical proxy of which being the ratio of fluorescence excitation peaks at 495 and 545 nm ( $Exc_{495:545}$ ) with an emission set at 585 nm. Isolates of subtype 3a have a low ( $Exc_{495:545} < 0.6$ ) PUB:PEB ratio and are GL specialists, 3b have an intermediate ( $0.6 \leq Exc_{495:545} < 1.6$ ) ratio and 3c a high ( $Exc_{495:545} \geq 1.6$ ) ratio and are BL specialists (Six *et al.*, 2007). Additionally, some strains that correspond to subtype 3d are able to dynamically modulate their PUB:PEB ratio, a process known as type IV chromatic acclimation (CA4) that allows them to maximally absorb GL or BL (Palenik, 2001; Everroad *et al.*, 2006; Humily *et al.*, 2013; Shukla *et al.*, 2012).

A comparative genomics analysis based on the first 11 sequenced *Synechococcus* strains showed that most of the genes involved in the synthesis and regulation of PBS rods are grouped into a dedicated genomic region, hereafter called the PBS rod region, the gene content and organization of which correspond to the different pigment types (Six *et al.*, 2007). An interesting feature of genes contained in this region is that their evolutionary history seemingly differs from that of the core genome but fits well with strain PT (Six *et al.*, 2007), and a phylogenetic analysis indeed suggested a progressive adaptation of PBS to absorbing shorter wavelengths (Everroad and Wood, 2012). Recently, an original PBS rod region was assembled from metagenomes of the Baltic Sea (Larsson *et al.*, 2014), and analysis of its gene content suggested that it belonged to a new PT 2 phenotype, referred to as PT 2B, the previous isolate-based PT 2 being consequently renamed PT 2A. Moreover, the recently sequenced high-PUB strains KORDI-100 and CC9616 were found to exhibit a PBS rod region differing from PT 3c in terms of gene complement, order and alleles and have been classified as PT 3f (Mahmoud *et al.*, 2017; Xia *et al.*, 2017a; Grébert *et al.*, in revision). Besides the main PBS rod region, a small genomic island has been shown to be necessary for CA4 and to occur in two different configurations, CA4-A and CA4-B, defining the two distinct pigment genotypes 3dA and 3dB (Humily *et al.*, 2013; Shukla *et al.*, 2012; Sanfilippo *et al.*, 2016).

Here, we study the extent of genetic diversity of PBS genomic regions from newly sequenced marine *Synechococcus* isolates as well as from natural populations, as obtained using a targeted metagenomics technique (Humily *et al.*, 2014). We propose a scenario for the evolution of the different pigment types, and highlight the importance of population-scale mechanisms in shaping the distribution of pigment types among *Synechococcus* lineages.

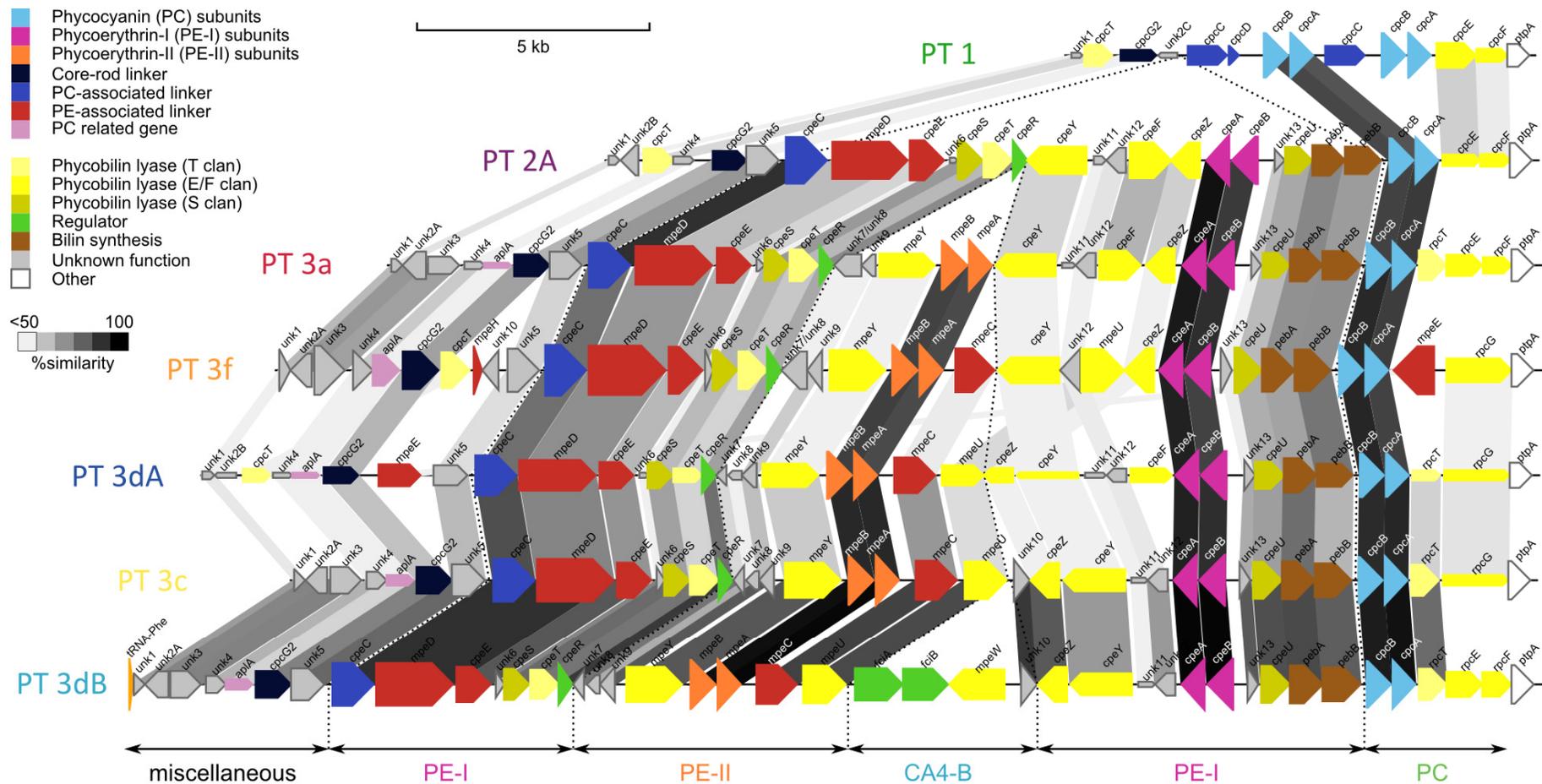
## Results

### *Comparative genomics and evolution of the phycobilisome rod region*

We analyzed the PBS rod region of 57 *Synechococcus* strains, of which 31 are newly sequenced strains. This set of strains covers not only a very wide genetic diversity, with representatives of all three subclusters (SC 5.1 to 5.3) and within SC 5.1 of all clades known so far except the uncultured EnvA and EnvB (Mazard *et al.*, 2012; Farrant *et al.*, 2016), but also a vast pigment diversity with representatives from all PTs except 2B, for which only metagenomes are available so far (Larsson *et al.*, 2014). Compared to the previous study on the comparative genomics of 11 *Synechococcus* marine by (Six *et al.*, 2007), our dataset includes two more pigment types (3dB and 3f), and many more representatives for each PT, allowing us to significantly refine previous analysis.

After careful gene annotation, we compared the gene complement and organization of the different PBS rod regions as well as the gene similarity among and between the different pigment types. The size of this region ranges from about 8 kb (PCC6307, SC 5.2/PT 1) to a little bit more than 30 kb (PROS-U-1, clade II/PT 3dB), and is always comprised between a tRNA for phenylalanine at the 5'-end and a tyrosine phosphatase at the 3'-end in tandem with a core bifunctional pantoate ligase/cytidylate kinase (Fig. 1). Globally, PBS rod genes content and synteny was remarkably conserved among representatives of a given PT. Pigment type 1 strains have the simplest rods and the shortest PBS rod region, comprising 2 to 4 copies of the *cpcBA* operon coding for  $\alpha$  and  $\beta$ -PC subunits (this operon is absent from the genome of strain CB0101, likely due to assembly errors), one rod-core linker gene (*cpcG2*), 3 or 4 rod linker genes (*cpcC* and *cpcD*), three phycobilin lyase genes (*cpcT* located near the 5'-end of the region and the *cpcEF* operon near the 3'-end) and two uncharacterized genes (*unk1* and *unk2*).

All but one copy of the *cpcBA* operons and associated PC rod linker genes (*cpcC/D*) are absent from all PT 2 strains, in which they are replaced by a suite of 16 to 18 genes, comprising genes necessary for the synthesis and regulation of PE hexamers, as was previously described for strain WH7805 (Six *et al.*, 2007). This PE-specific region include the *cpeBA* operon encoding the PE-I  $\alpha$  and  $\beta$ -subunits, three PE-associated linkers (*cpeC*, *mpeD* and *cpeE*), six putative lyases (*cpeF*, *cpeS*, *cpeT*, *cpeU*, *cpeY*, *cpeZ*), two enzymes involved in PEB synthesis (*pebA* and *pebB*; Frankenberg *et al.*, 2001), one putative regulator (*cpeR*; Cobley *et al.*, 2002) and a number of small, uncharacterized genes (*unk4-6* and *unk11-13*, except in A15-44, the only PT 2 strain to also possess *unk3*). Surprisingly, strain CB0205 and A15-44 are the only PT2 strain possessing the allophycocyanin-like gene *aplA* (Montgomery *et al.*, 2004) inserted between *unk4* and *cpcG2*, like in all PT 3 strains.



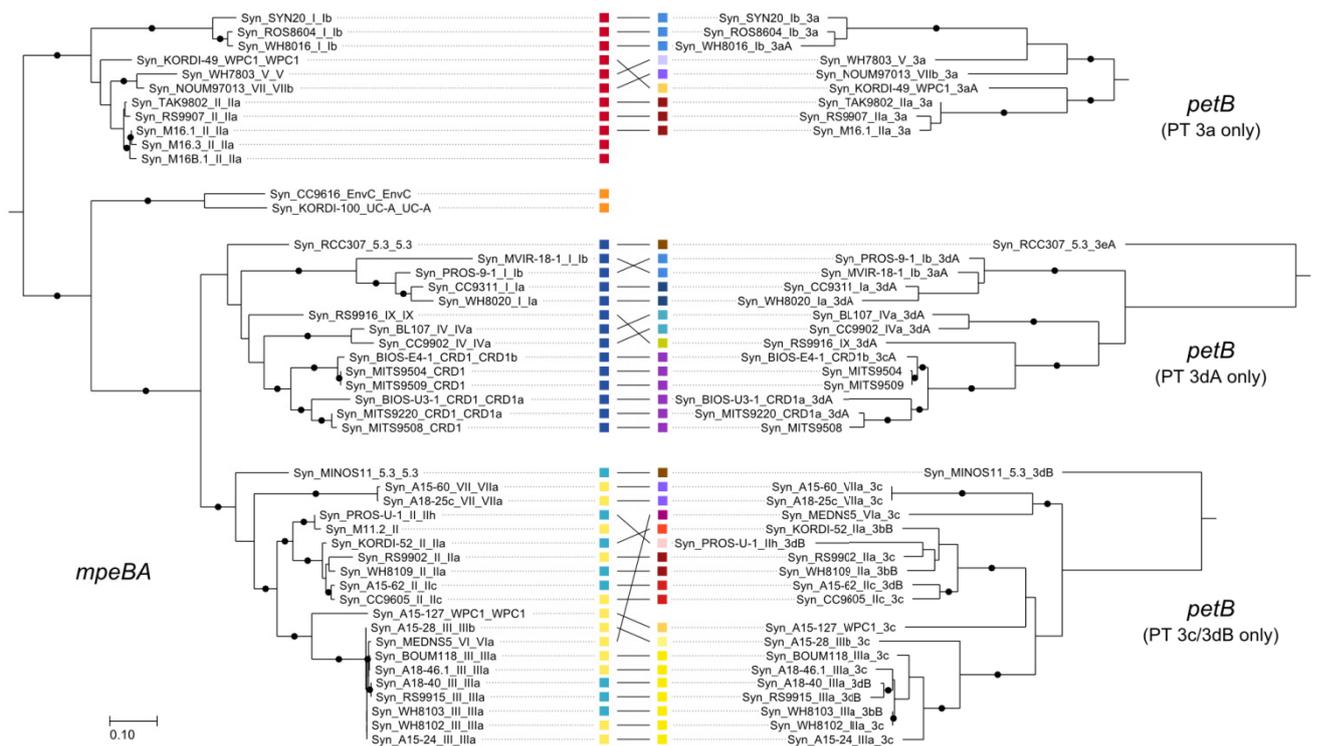
**Fig. 1: PBS rod genomic region for strains of different pigment types.** Regions are oriented from the phenylalanine tRNA (shown in orange only in the PT 3dB region) to the conserved tyrosine phosphatase *ptpA*. Genes are coloured according to their inferred function, their length being proportional to the gene size and their thickness proportional to the protein identity between strains of the same pigment type. Grey cross-links between regions are shaded according to the average protein identity between strains of the two pigment types. The strains represented here are WH5701 (PT 1), WH7805 (PT 2A), RS9907 (PT 3a), KORDI-100 (PT 3f), RS9916 (PT 3dA), WH8102 (PT 3c) and A15-62 (PT 3dB).

The main genomic differences between the PBS regions of PT 2 and 3a are the insertion between *cpeR* and *cpeY* of a small cluster containing only 5 genes: *unk7/8*, a fusion of the uncharacterized *unk7* and *unk8* genes found as separate genes in all other PT 3 subtypes, *unk9* another gene of unknown function, the phycobilin lyase gene *mpeY* (Grébert *et al.*, in prep.; see Chapter III) and the *mpeBA* operon encoding the  $\alpha$  and  $\beta$ -PE-II subunits. All PT 3 strains also possess an additional putative PE-II linker gene *mpeE* compared to PT2, but the location of this gene is highly variable (Table S1, Fig. S3-S7), as it can be found inserted in the PBS rod region or in a location seemingly unrelated to PBS synthesis, and it is even absent from the genome of the PT 3a strain Syn20.

Genomic differences between PT 3a and other PT 3 subtypes (i.e. 3c, 3dA, 3dB and the newly described 3f, Xia *et al.*, 2017a) are mainly located in the subregion between the PE-II (*mpeBA*) and PE-I (*cpeBA*) operons (Fig. 1). All PT 3 subtypes other than 3a have the distal PE-II associated linker *mpeC* (Six *et al.*, 2005) inserted downstream of *mpeBA*, as well as an extra gene (*mpeU*) that has recently been suggested to be a PEB lyase-isomerase (Mahmoud *et al.*, 2017). However, the gene order and complement between *mpeBA* and *cpeBA* is not conserved between subtypes, some lacking *cpeF* (PT 3f, 3c, 3dB; note that this gene was previously called *mpeV*; Wilbanks and Glazer, 1993; Six *et al.*, 2007) and/or *unk11* (3f, Fig. 1), some having the conserved hypothetical gene *unk10* (PT 3c and 3dB), which in 3dA strains is found in the CA4-A island (Humily *et al.*, 2013). The genomic difference between PT 3c and 3dB is strikingly small and lies in the insertion between *mpeU* and *unk10* in the latter genomes of three genes involved in chromatic acclimation and altogether forming the so-called CA4-B island (Humily *et al.*, 2013), which encompasses the putative regulators *fciA* and *fciB* (Sanfilippo *et al.*, 2016) and the PEB lyase gene *mpeW* (see Chapter III). Another notable difference between the PT 3 subtypes is the presence of the putative linker gene *mpeF* in all 3dA strains but RS9916 and RCC307, in which it is replaced by a distantly related linker gene, called *mpeG* (Table S1).

Other differences are more difficult to link with a specific pigment type. For example, the region located immediately upstream *ptpA* has two possible configurations: it can be either *rpcG*, a fusion gene that encodes a C84  $\alpha$ -PC PEB lyase-isomerase or the *rpcEF* operon thought to encode the two subunits of a C84  $\alpha$ -PC PEB lyase (Blot *et al.*, 2009). While all PT 3a (i.e. low PUB) strains possess a *rpcEF* operon, other PT 3 subtypes can have either *rpcEF* or *rpcG*, but this does not seem to match either a PT or a clade, with MINOS11, a SC 5.3/PT 3dB strain, even possessing both. This result is somewhat surprising because it was previously thought that the replacement of *rpcEF* by *rpcG* could confer *Synechococcus* cells a better adaptability to blue light environments (Blot *et al.*, 2009). Similarly, the distribution of the putative linker genes *mpeG* and *mpeH* (a truncated version of the former) cannot be linked with either a PT or a clade. Some additional variations with regard to the “typical” PBS rod region per PT are also worth

noting (Fig. S1-S7). This includes the PT 2A strain CB0205 that exhibits an insertion of 9 supplementary genes in the middle of the PBS region, of which one is similar to the *unk3* gene found in the 5' region of most PT 2 and PT 3 strains (Fig. S2). Another striking example is the position of the aforementioned *mpeE* linker gene that can be found at the 5'-end of the PBS rod region (Fig. 1, S5, S6, Table S1), at its 3'-end (Fig. S3, S4, Table S1), or elsewhere in the genome (Table S1). Finally, genes unrelated to PBS synthesis or regulation can sometimes be found inserted between the tRNA and the *unk1* gene at the 5'-end of the PBS rod region, questioning the exact 5'-boundary of this genomic region and the involvement of *unk1-4* genes into PBS synthesis or regulation. Interestingly, these insertions correspond to genes related to DNA metabolism or mobile genetic elements (restriction enzymes, resolvases, prophage proteins, recombinases) in strains WH8101 (PT 1/clade VIII), M16.1 and TAK9802 (both 3a/IIa), A18-46.1 (3c/IIIa) and RS9902 (3c/IIa), suggesting privileged insertion of foreign DNA at this position.



**Fig. 2: Correspondence between phylogenies for the *mpeBA* operon and for the marker gene *petB*, which reproduces the core genome phylogeny.** The pigment type for each strain is indicated by a coloured square in the *mpeBA* phylogeny, and its clade similarly indicated in the *petB* phylogeny.

Although the genomic organization of the PBS rod region is globally conserved, despite the few variations described above (Fig. S1-S7), the allelic diversity of genes gathered in this region is much less characterized (Fig. 1 and S8). Linker proteins for example show large variations between strains, some isolates presenting linkers with less than 70% of protein identity over their whole length with orthologs present in other strains. This result is somewhat surprising given their determinant role in PBS structure and the much higher conservation of  $\alpha$ - and  $\beta$ -subunits of phycobiliproteins (about 90% identity). The variability is even greater for phycobilin lyases and uncharacterized conserved proteins (“Unk”), the genes of some isolates presenting less than 60% protein identity to their orthologs in other strains.

Altogether, comparative genomic analyses show that gene order and organisation of the PBS region is globally specific of each PT. Similarly, phylogenetic analyses using phycobiliprotein subunits have been shown to group together strains according to their PT but independently from their clade phylogeny, the PC-encoding *cpcBA* operon allowing to discriminate PT 1, 2A, 2B and 3, while PE-I-encoding *cpeBA* operon separates strains of pigment type 2A, 3a, 3dA, 3f and the 3c+3dB group and PEII-encoding *mpeBA* operon provides an even better resolution for the PT 3 subtypes (Humily *et al.*, 2014; Xia *et al.*, 2017b; Grébert *et al.*, in revision; Everroad and Wood, 2012). Here, the addition of newly sequenced strains revealed that within each PT cluster, the topology actually resembles the topology based on core genes. Indeed, the direct comparison of the topology of trees based on *mpeBA* and on the core *petB* gene for each PT cluster shows that the *mpeBA* sequences from strains belonging to the same clade are more similar to one another than are for strains from different clades (Fig 2).

#### *Targeted metagenomics allows retrieving novel genomic organizations and distant alleles*

In order to investigate the diversity and genetic variability of PBS rod regions of natural *Synechococcus* populations, we used the targeted metagenomic approach developed by (Humily *et al.*, 2014). This method based on flow cytometry cell-sorting, WGA and fosmid library screening allowed us to analyze the PBS rod region from populations of the Mediterranean Sea, North Sea, English Channel and north-eastern Atlantic Ocean (Fig. 3A, Table S2). *Synechococcus* populations at these stations were quite diverse as assessed by the analysis of the community composition performed by PCR amplification, sequencing and taxonomic assignment of the *petB* gene (Fig. 3B). Stations from the North Sea (CEFAS cruise, fosmid library H) to the English Channel (Astan, fosmid library A) were exclusively composed of the cold-adapted clade I (mainly sub-clade Ib), the north-eastern Atlantic (AMT CTD5 used for library E and CTD8 and CTD10 used for library G) was co-dominated by clades I, CRD1 and EnvA/EnvB. Populations of





Although most of these PBS rod regions were syntenic with regions from characterized strains (see above; Fig. 1 and S2-S8), one of these regions exhibited a new gene organization related to PT 3dA (Fig. 4A). Indeed, this region included a CA4-A genomic island inserted between *unk3* and *unk4* at the 5'-end of the PBS rod region, while all previously described CA4-A genomic islands could be located anywhere in the genome but always outside the PBS rod region (Humily *et al.*, 2013). Genes in these unusual PBS rod regions had a tblastx best-hit to the 3dA/clade IV strains CC9902 or BL107, the PBS rod region and CA4-A genomic island of which is given as a reference in Fig. 4A. Interestingly, this new organization was found in contigs assembled from different sequencing libraries, originating from samples geographically as diverse as the Atlantic Ocean (library E, AMT station 5; library G, AMT stations 8 and 10), the Mediterranean Sea (library F, point B Villefranche-sur-mer), and the North Sea (library H, CEFAS stations 3, 7 and 13), suggesting that this new organization is biologically relevant and not a chimera produced during sample processing (WGA, sequencing, assembly).

Apart from this atypical example, all others assembled contigs presented a gene organisation similar to at least one of the sequenced strains (Fig. 1; S1-S7). Contigs corresponding to the canonical 3dA PBS rod region were found in assemblies from all but two libraries (C and D, from the Mediterranean samples BOUM station C and B), most of them with best-hits to clade I and IV strains (Fig. 4B). Some contigs from libraries A (Roscoff, English Channel, contigs A1 and A2) and H (CEFAS, North Sea, contigs H30, H31, H100, H101, H102) presented the same synteny as the altered PBS rod region of strain MVIR-18-1 (clade I) which lacks *mpeU* (Fig. 4B and Fig. S5; Humily *et al.*, 2013), a gene that has recently been shown to be necessary for reaching a high PUB/PEB ratio in CA4-A strain RS9916 (Mahmoud *et al.*, 2017). Accordingly, strain MVIR-18-1 displays a constant low PUB:PEB ratio like green light specialists (Humily *et al.*, 2013). It is noteworthy that the isolation site of this strain, i.e. the North Sea, is consistent with the geographic origin of these contigs (North Sea and English Channel), suggesting that MVIR-18-1-like populations are present in the environment at medium to high latitude.

Interestingly, libraries E (AMT Station 5) and G (AMT stations 8 and 10), in which *petB* pyrosequencing indicated the presence of *Synechococcus* from clade CRD1, also produced contigs matching the particular 3dA PBS rod region of some CRD1 strains such as BIOS-E4-1, MITS9504, MITS9508 and MITS9509 (Fig. 4B and Fig. S5). These strains have a partial (MITS9508) or highly degenerated (BIOS-E4-1, MITS9504 and MITS9509) *mpeY* sequence, miss the CA4 regulators *fciA* and *fciB* in their CA4-A genomic island and, at least in the case of BIOS-E4-1, exhibit a “fixed” BL-specialist phenotype whatever the growth light color (Humily *et al.*, 2013). This particular genomic organization has recently been found to be abundant and even dominate in CRD1-rich warm High Nutrient Low Chlorophyll (HNLC) areas, in particular

of the Pacific Ocean (Grébert *et al.*, in revision). Thus, finding such PBS rod region provide additional and more direct evidence that these “natural mutants” might be found in geographically very distant CRD1 populations.

PT 3a contigs were assembled solely from library H (North Sea, CEFAS cruise), and genes from these contigs were most similar to PT 3a/clade I genes (Fig. 4B). PT 3f contigs were retrieved from libraries of the north-eastern Atlantic (library G) as well as from the Mediterranean Sea (libraries C, I and D), even representing the totality of assembled contigs from the library D. This pigment type and genomic region organization have been previously described from strains KORDI-100 (Xia *et al.*, 2017a) and CC9616 (see above), which belong to the rare clades UC-A and XX, respectively. So, given that these clades were not found at the sampling station used to build these libraries (C, D, G and I), it is possible that other clades present at these stations exhibit this unusual pigment type.

Finally, a number of contigs corresponding to the phylogenetically indistinguishable PT 3c and 3dB group were found in libraries C, I, E, F and G, retrieved from Mediterranean Sea and north-eastern Atlantic Ocean (Fig. 4B). Among these contigs, five (C100, C40, G8D, I7, I21) could be unambiguously assigned to PT 3dB, due to the occurrence of a CA4-B genomic island between *mpeU* and *unk10*, and were found only in fosmid libraries C, G and I. Within PT 3dB, it must be noted that contig C100 resembles the PBS genomic region of strain MINOS11 (3dB/SC 5.3), with multiple genes with best-hit to this strain, and displays a *rpcG* gene copy between *unk10* and *cpeZ*, a specificity so far only observed for MINOS11. However, the CA4-B genomic island of this contig differs from that of MINOS 11 by the lack of *mpeW* (Fig. 4B). In contrast to contigs assigned to PT 3dA, those assigned to PT 3c or 3dB had best-hits to representatives of very diverse clades, including clades II, III, WPC1 and 5.3, a genetic diversity quite consistent with that observed with *petB* for these stations. More surprisingly, the genes constituting some contigs from fosmid library E (e.g., contig E102) were highly divergent from all strains sequenced so far and might thus belong to one of the clade for which no sequenced representative is available yet, such as EnvA and B, that together constitute more than 30 % of the *Synechococcus* community at this station (Fig. 3B).

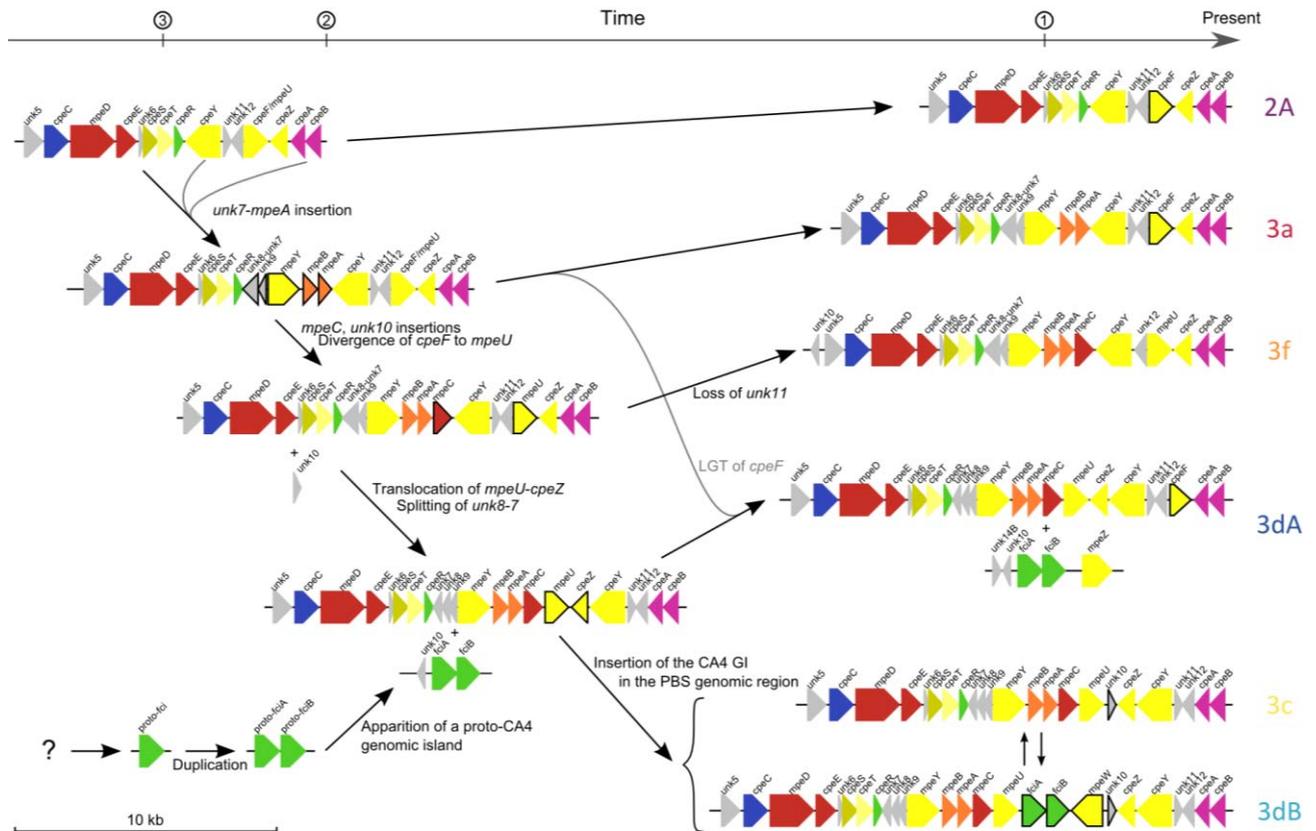
## Discussion

Most of the current knowledge on the genomic organization and genetic diversity of the PBS rod region relies on the comparison of the first 11 sequenced *Synechococcus* genomes (Six *et al.*, 2007) and from metagenomic assemblages from specific environments including i) the

Baltic sea, where essentially PT 2B/SC 5.2 was recovered (Larsson *et al.*, 2014), ii) freshwater reservoirs, dominated by PT 2A /SC 5.3 populations (Cabello-Yeves *et al.*, 2017) as well as iii) one station of the English channel from which PBS rod regions corresponding to 3dA/SC 5.1-clades I and IV were assembled (Humily *et al.*, 2014). Thus, the extent of the variability existing within each PT, both in terms of gene content and organisation as well as the level of allelic diversity of individual gene constituting the PBS rod region, remained largely unexplored, in particular in strictly marine environments. Here, by analysing a much larger set of *Synechococcus* genomes (57 genomes), covering the whole genetic and PT diversity known so far in this genus, as well as many partial environmental PBS rod regions (229 contigs) retrieved from a large variety of trophic regimes and light niches, we considerably increased the representativeness of PBS rod regions available.

These extensive analyses confirmed the remarkable conservation of the PBS rod region and that their gene content and organisation is highly specific of the different PT and subtypes. The largest differences occur between PT 1 and PT 2A, the acquisition of the capacity to synthesize PE-I requiring 16-18 additional genes (Fig. 1). The change from PT 2A to 3a was comparatively small, with only 5 supplementary genes, among which *mpeBA*, the operon encoding PE-II subunits, suggesting that many genes required to synthesize PE-II are common to both PE types. Yet, it is important to note that genes that appears to be orthologous between the different PBS rod regions might actually be distinct alleles with slightly different functions. For instance, the linker protein MpeD found in both PT 2 and PT 3 has an insertion of 17 amino acids near its N terminus only in all PT 3 strains, an insertion that has been suggested to be involved in PUB fixation in this PT (Six *et al.*, 2005, 05). A second example concerns *mpeY* that is present in all PT3 strains but was recently shown to have two distinct alleles, the first one found in PT 3a and PT 3dA, exhibiting a PEB lyase function, while the second allele found in PT 3c and PT 3dB acts as a lyase isomerase, binding a PUB at the same position of  $\alpha$  PE-II (Grébert *et al.*, in prep. ; see chapter III). In this context, the availability of many additional PBS rod regions, both from cultured isolates and from the environment, allowed us to have access to the allelic variability for each gene of this region, potentially indicative of functional variability (Fig. 1 and S8). While the diversity of phycobiliprotein subunit alleles shows that they are strongly conserved, with not less than 90% of identity (Fig S8), linker proteins and phycobilin lyases are much more variable, with percent identity to their best-hit among other members of the same group of orthologs as low as 80 and 70 %, respectively. Lyases of the E/F clan and uncharacterized conserved genes (“unk”) were the most variable, with sequences having less than 60% identity at the protein level, and thus constitute privileged targets to be biochemically characterized to fully understand the processes involved in the synthesis and regulation of PBS rods. This diversity might also indicate that the phycobilisome is adapted to environmental factors other than light (Pittera *et al.*, 2016).

The characterisation of PBS rod regions retrieved from the environment allowed us to discover a new organization relative to the canonical PT 3dA in which the CA4-A genomic island, usually located elsewhere in the genome, is inserted in the PBS rod region near the 5'-end (Humily *et al.*, 2013; Fig4A). Since all other genes of these atypical contigs exhibit best match and strong percent identity to the corresponding genes of PT 3dA strains, this new organization likely does not correspond to a new PT, but rather to a variant of PT 3dA that had not been observed before. This tends to confirm the different evolutionary histories between the two types of CA4 islands, the CA4-A being much more variable, both in terms of gene content (out of the 13 strains having a 3dA genotype of the PBS rod region, six present an incomplete CA4-A island; Fig. S5) and position in the genome, while the CA4-B island is always found at the same position in the genome and is complete in all 3dB strains sequenced so far (Humily *et al.*, 2013; Fig. S7). In this context, it is interesting to note that the contig C100, which apparently misses *mpeW* (Fig. 4), constitutes the first documented example of an incomplete CA4-B region that we predict should result in a blue-light specialist phenotype (see chapter III). As for 3dA, although we did not have access to the CA4-A island located in most cases outside the PBS rod region, we found two other examples where lyase genes were missing in the PBS rod region of natural *Synechococcus* populations, i.e. natural mutant genotypes that were previously observed for characterized strains known to be deficient in their ability to perform CA4. The first case corresponds to a series of contigs (E16A, G19A, G9B), all assigned to the CRD1 clade and lacking *mpeY*, a configuration also observed in the CRD1 strains BIOS-E4-1, MITS9504 and MITS9509, which also lack the CA4 regulators *fciA/fciB* in their CA4-A island. Interestingly, BIOS-E4-1 has been shown to exhibit a blue-light-specialist phenotype (Humily *et al.*, 2013; Grébert *et al.*, in revision). The second case are contigs lacking *mpeU*, a configuration also observed in the clade I strain MVIR-18-1, phenotypically characterized as being a green light specialist (Humily *et al.*, 2013; Mahmoud *et al.*, 2017; Grébert *et al.*, in revision). Thus, natural populations possessing these two types of PBS rod regions likely exhibited a blue and green light specialist phenotype, respectively. While the *mpeU*-lacking genotype seems to be infrequent *in situ* (Grébert *et al.*, in revision), the latter authors provided clear though indirect evidence for the widespread distribution of BIOS-E4-1-like deficient populations in the iron poor areas of the southern Pacific Ocean by recruiting several markers (*mpeA*, *mpeU* and *mpeZ*) from metagenomes. Thus, the present study provides a more direct evidence for the occurrence of BIOS-E4-1 like genotype in the field.



**Fig. 5: Putative evolutionary scenario for the apparition of the different *Synechococcus* pigment types.** This scenario is congruent with all individual phylogenies of genes in the phycobilisome rod genomic region. Note that the 5'- and 3'-end of the PBS rod region are cropped for better visualisation of the PE-I/PE-II sub-regions.

The mismatch between pigmentation and vertical phylogeny raises the question of how the different PTs have evolved and have been maintained independently from clade diversification. (Apt *et al.*, 1995) proposed a scenario for the evolution of phycobiliproteins by gene duplication, explaining the apparition of PE-II from PE-I in marine *Synechococcus*, and (Everroad and Wood, 2012) complemented it by proposing the occurrence of a directional evolution of *Synechococcus* phycobiliproteins towards higher PUB content. However, both scenarios only account for the evolution of phycobiliprotein  $\alpha$  and  $\beta$  subunits. Yet, these proteins are part of a complex supramolecular assembly and interact with numerous additional proteins, including linkers, phycobilin synthesis genes and phycobilin lyases, which all need to co-evolve. Thus, a more holistic approach is necessary to better understand the evolution of *Synechococcus* PTs. Here, we propose a scenario that is consistent with the evolutionary histories of the different genes of the PBS rod region, accounting for both variations in gene content and allelic diversity (Fig. 5). A first step from an ancestor with a PBS region similar to that PT 2A towards higher PUB content would involve a gene duplication event leading to PE-II from PE-I. This event was

accompanied by a duplication of the *cpeY* ancestor that eventually led to the formation of the *mpeY* ancestor (see Fig. 1 in Grébert *et al.*, in prep; chapter III). The origin of *unk7/8* and *unk9* is more difficult to assess. PT 3a would directly derive from this major duplication event. Phylogenetic analysis of *mpeBA* placed the recently described PT 3f in a cluster between those formed by 3a on the one hand and by 3dA and 3c/3dB on the other hand (Fig. 2). Consistently, the organization of the PT 3f PBS rod region is intermediate between 3a and the more complex 3c/3dB region, the differences between 3a and 3f being the addition of the linker gene *mpeC*, the replacement of *cpeF* by *mpeU*, as well as the loss of *unk11*, without any gene order rearrangement. This fits well with the “toward higher PUB” scenario proposed by (Everroad and Wood, 2012) since i) the supplementary linker MpeC potentially allows binding of an additional PUB-rich PE-II hexamer per PBS rod (Six *et al.*, 2005) and ii) the putative PEB lyase CpeF is replaced by MpeU, recently suggested to be a lyase-isomerase, adding a PUB on PE-II (Mahmoud *et al.*, 2017). We thus propose that the LCA of PT 3f, 3dA and 3c/3dB evolved from the PT 3 LCA by the acquisition of a linker protein gene (*mpeC*), and that a subsequent gene translocation event resulted in the apparition of the PT 3dA and 3c/3dB LCA (Fig. 5). As *cpeF* and *mpeU* are paralogs, we suggest that *mpeU* results from the divergence from a *cpeF*-like ancestor and that the occurrence of *cpeF* in PT 3dA would be due to a more recent lateral gene transfer from a 3a to the 3dA ancestor (Fig. 5). The evolution from the PT 3f, 3dA, 3c/3dB last common ancestor (LCA) to the 3dA and 3c/3dB LCA also involved a tandem translocation of *mpeU* and *cpeZ* between *mpeC* and *cpeY* (Fig. 5). The evolution from the 3dA and 3c/3dB LCA to 3dA involved the lateral transfer of *cpeF* from 3a and the acquisition of a CA4 genomic island, whereas the apparition of PT 3c/3dB required the insertion of *unk10* and the CA4 island into the PBS rod region. The separation of PT 3dA and 3c/3dB promoted the differentiation of the CA4 island into two versions. As pointed out previously, PT 3c and 3dB share the same allele for all PBS genes, and the only difference between them is the insertion of the CA4-B island in the PBS rod region. Thus, the conversion between these two pigment types is seemingly straightforward and possibly frequent. The apparition of the two types of CA4 islands is more difficult to trace, but phylogenetic analyses based on *mpeBA* (Fig. 2) or *mpeW/Y/Z* (Humily *et al.*, 2013; Grébert *et al.*, in prep.) provide evidence that PT 3dA and 3c/3dB share a common ancestor and that the two versions of the CA4 island, which both contains *fciA* and *fciB*, also derive from a common ancestor (Humily *et al.*, 2013). Hence, the proto-CA4 island probably occurred in the LCA of 3dA and 3c/3dB. The position of *mpeZ* in the *mpeW/Y/Z* phylogenetic tree suggests on the contrary that the apparition of the CA4-A island occurred after the divergence between PT 3dA and 3c/3dB, but this hypothesis does not match the observation that all sequenced 3dA strains so far possess at least remnants of a CA4-A island (see Fig. 1 in Grébert *et al.*, in prep; see chapter III).

The inconsistency between phylogenies obtained from PBS and core genes led different authors to suggest that frequent lateral gene transfers (LGT) of part or the whole PBS regions occurred during the evolution of *Synechococcus* (Everroad and Wood, 2012; Six *et al.*, 2007; Dufresne *et al.*, 2008; Haverkamp *et al.*, 2008). However, by adding several representatives of each PT/clade combination, we showed here that within each PT, phylogenies of PBS genes (e.g. *mpeBA*, Fig. 2) are finely structured and globally consistent with the core phylogeny. Thus, this suggests that transfers between distantly related lineages are actually quite rare, the only unambiguous such event being detected for MEDNS5 (PT 3c, clade VIa), for which *mpeBA* (as well as *cpeBA*, *mpeW/Y/Z* and *mpeU*; Grébert *et al.*, in revision; Humily *et al.*, 2014; Mahmoud *et al.*, 2017) sequences clustered with PT 3c/clade III *mpeBA* sequences (Fig. 2). This observation can be explained by population genetics, which states that incongruences between gene trees and species trees can arise when an ancestral polymorphism in a population (i.e., occurrence of several alleles for a same gene) is not fully sorted (i.e., resolved into monophyletic lineages) after a speciation event, because of the stochastic way in which lineages inherit alleles during speciation (Tajima, 1983; Galtier and Daubin, 2008). These incongruences, called incomplete lineage sorting (ILS), are predicted to occur for at least some genes in a genome (Degnan and Rosenberg, 2009) and are expected to be particularly important in prokaryotes since the coalescent time (i.e. time to the LCA) is predicted to be proportional to the effective population size (Abby and Daubin, 2007; Batut *et al.*, 2014). *Synechococcus* being the second most abundant photosynthetic organism in the oceans (Flombaum *et al.*, 2013), ILS is definitely expected to happen in this lineage. Assuming *Synechococcus* effective population size to be on the same order as that estimated for *Prochlorococcus* ( $10^{11}$  cells, Baumdicker *et al.*, 2012; Batut *et al.*, 2014; Kashtan *et al.*, 2014), and given a generation time on the order of days, the allelic fixation time would be of about 280 million years (My). This rough estimate is on the same order of timescale as the divergence between SC 5.3 and SC 5.2/ SC5.1 (700-1000 My ago) or between SC 5.2 and SC 5.1 (200-600 My ago, Sánchez-Baracaldo *et al.*, 2014; Sánchez-Baracaldo, 2015), further emphasizing the possibility of ILS being the major source of the apparent incongruence in PT distribution between clades. This new evolutionary scenario suggests that the different pigment types appeared before the diversification of SC 5.1 clades, and probably before the divergence of SC 5.1 and 5.3. In this regard, the basal position of the two 5.3 isolates in their respective PT clusters in the different phylogenies examined (Fig. 2; see also Dufresne *et al.*, 2008) reinforce this hypothesis.

In conclusion, the analysis the PBS rod region of newly sequenced *Synechococcus* isolates and retrieved from wild populations allowed us to refine previous findings about the relationships between gene content and organization of this region, allelic variability and *Synechococcus* PTs. We proposed a scenario for the evolution of the different PTs and present a

new hypothesis based on population genetics to explain the observed discrepancies between PT and core phylogenies. In this regard, analysing *Synechococcus* evolution from the perspective of its demographic history constitutes an interesting avenue for future studies.

## Materials and methods

### *Genome information*

Genomic regions used in this study were obtained from public complete or draft genomes as well as closed genomes of 31 newly marine *Synechococcus* genomes sequenced by the Genoscope and assembled as described by (Farrant *et al.*, 2015).

### *Fosmid libraries*

Samples for construction of the fosmid library were collected during oceanographic cruises CEFAS (North Sea), BOUM (Mediterranean Sea) and AMT (North-eastern Atlantic Ocean) as well as from three coastal stations belonging to the French long-term observatory of coastal environment (SOMLIT): ‘ASTAN’ located at 2.8 miles off Roscoff, ‘Point B’ at the entrance of the Villefranche-sur-mer Bay and ‘Boussole’ in the ligurian current at 32 miles off Nice. Details on the sampling conditions, dates and locations are provided in Table S1. 454 *petB* pyrosequencing, cell sorting, DNA extraction, whole genome amplification, fosmid library construction, screening and sequencing were performed as previously described (Humily *et al.*, 2014).

Sequencing reads were quality trimmed and filtered using custom scripts to remove bases with a quality score below 20, and reads shorter than 240 nt or with a mean quality score below 27 were discarded. Reads corresponding to the fosmid vector, the *E. coli* host or contaminants were removed using a custom adaptation of NCBI VecScreen (<https://www.ncbi.nlm.nih.gov/tools/vecscreen/about/>). Paired-reads were merged using FLASH v1.2.11 (Magoč and Salzberg, 2011), and merged and non-merged remaining reads were assembled using the CLC AssemblyCell software (CLCBio, Prismet, Denmark). Resulting contigs were scaffolded using SSPACE v3.0 (Boetzer *et al.*, 2011), and scaffolds shorter than 500 bp or with a sequencing coverage below 100x were removed. To reduce the number of contigs while preserving the genetic diversity, a second round of scaffolding was done using Geneious v6.1.8 (Biomatters, Auckland, New Zealand). Assembly statistics are reported in Table S2. Assembled scaffolds were manually examined to control for obvious WGA-induced as well as assembly chimeras. Annotation of PBS genes was performed manually using Geneious and the

Cyanorak v2.0 information system (<http://application.sb-roscoff.fr/cyanorak/>). Plotting of regions was done using BioPython (Cock *et al.*, 2009).

#### *Phylogenetic analyses*

Sequences were aligned using MAFFT v7.299b with the G-INS-i algorithm (default parameters; Katoh and Standley, 2013). ML phylogenies were reconstructed using PhyML v20120412 using both SPR and NNI moves (Guindon *et al.*, 2010). Phylogenetic trees were plotted using Python and the ETE Toolkit (Huerta-Cepas *et al.*, 2016).

#### **Acknowledgements**

This work was supported by the collaborative program METASYN with the Genoscope, the French “Agence Nationale de la Recherche” Programs SAMOSA (ANR-13-ADAP-0010), and the European Union’s Seventh Framework Programs, MicroB3 and MaCuMBA (grant agreements 287589 and 311975, respectively). Théophile Grébert was supported by the French Ministry of Higher Education and Research. We thank Fabienne Rigaud-Jalabert for sample collection at the Astan station and Thierry Cariou (‘Service d’Observation en Milieu LITtoral (SOMLIT)’, INSU-CNRS, Station Biologique de Roscoff) for providing physico-chemical parameters at the sampling site. We are most grateful to the Biogenouest Genomics core facilities for their technical support, notably Alexandra Dheilily, Delphine Naquin, and Oscar Lima for pyrosequencing in Rennes. Hélène Bergès, Arnaud Bellec, and Joelle Fourment from the CNRGV Platform are acknowledged for fosmid library screening. We are grateful to the ABIMS Platform (Station Biologique de Roscoff) for their help in setting up the genome database used in this study and for providing storage and calculation facilities for bioinformatics analyses.

**References**

- Abby S, Daubin V. (2007). Comparative genomics and the evolution of prokaryotes. *Trends Microbiol* **15**: 135–141.
- Adir N, Dines M, Klartag M, McGregor A, Melamed-Frank M. (2006). Assembly and disassembly of phycobilisomes. In: Microbiology Monographs. *Complex Intracellular Structures in Prokaryotes*. Springer, Berlin, Heidelberg, pp 47–77.
- Ahlgren NA, Rocap G. (2012). Diversity and distribution of marine *Synechococcus*: multiple gene phylogenies for consensus classification and development of qPCR assays for sensitive measurement of clades in the ocean. *Front Microbiol* **3**. e-pub ahead of print, doi: 10.3389/fmicb.2012.00213.
- Apt KE, Collier JL, Grossman AR. (1995). Evolution of the phycobiliproteins. *J Mol Biol* **248**: 79–96.
- Arteni AA, Ajlani G, Boekema EJ. (2009). Structural organisation of phycobilisomes from *Synechocystis* sp. strain PCC6803 and their interaction with the membrane. *Biochim Biophys Acta* **1787**: 272–279.
- Batut B, Knibbe C, Marais G, Daubin V. (2014). Reductive genome evolution at both ends of the bacterial population size spectrum. *Nat Rev Microbiol* **12**: nrmicro3331.
- Baumdicker F, Hess WR, Pfaffelhuber P. (2012). The infinitely many genes model for the distributed genome of bacteria. *Genome Biol Evol* **4**: 443–456.
- Blot N, Wu X-J, Thomas J-C, Zhang J, Garczarek L, Böhm S, *et al.* (2009). Phycourobilin in trichromatic phycocyanin from oceanic cyanobacteria is formed post-translationally by a phycoerythrobilin lyase-isomerase. *J Biol Chem* **284**: 9290–9298.
- Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**: 578–579.
- Bretaudeau A, Coste F, Humily F, Garczarek L, Le Corguillé G, Six C, *et al.* (2013). CyanoLyase: a database of phycobilin lyase sequences, motifs and functions. *Nucleic Acids Res* **41**: D396-401.
- Cabello-Yeves PJ, Haro-Moreno JM, Martin-Cuadrado A-B, Ghai R, Picazo A, Camacho A, *et al.* (2017). Novel *Synechococcus* genomes reconstructed from freshwater reservoirs. *Front Microbiol* **8**. e-pub ahead of print, doi: 10.3389/fmicb.2017.01151.

Cobley JG, Clark AC, Weerasurya S, Queseda FA, Xiao JY, Bandrapali N, *et al.* (2002). CpeR is an activator required for expression of the phycoerythrin operon (*cpeBA*) in the cyanobacterium *Fremyella diplosiphon* and is encoded in the phycoerythrin linker-polypeptide operon (*cpeCDESTR*). *Mol Microbiol* **44**: 1517–1531.

Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, *et al.* (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**: 1422–1423.

Degnan JH, Rosenberg NA. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol Evol* **24**: 332–340.

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

Everroad C, Six C, Partensky F, Thomas J-C, Holtzendorff J, Wood AM. (2006). Biochemical cases of type IV chromatic adaptation in marine *Synechococcus* spp. *J Bacteriol* **188**: 3345–3356.

Everroad RC, Wood AM. (2012). Phycoerythrin evolution and diversification of spectral phenotype in marine *Synechococcus* and related picocyanobacteria. *Mol Phylogenet Evol* **64**: 381–392.

Farrant GK, Doré H, Cornejo-Castillo FM, Partensky F, Ratin M, Ostrowski M, *et al.* (2016). Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proc Natl Acad Sci* **113**: E3365–E3374.

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

Flombaum P, Gallegos JL, Gordillo RA, Rincón J, Zabala LL, Jiao N, *et al.* (2013). Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc Natl Acad Sci* **110**: 9824–9829.

Frankenberg N, Mukougawa K, Kohchi T, Lagarias JC. (2001). Functional genomic analysis of the HY2 family of ferredoxin-dependent bilin reductases from oxygenic photosynthetic organisms. *Plant Cell* **13**: 965–978.

Galtier N, Daubin V. (2008). Dealing with incongruence in phylogenomic analyses. *Philos Trans R Soc Lond B Biol Sci* **363**: 4023–4029.

Grébert T, Doré H, Partensky F, Farrant GK, Emmanuel B, Picheral M, *et al.* (in revision). Light color acclimation: a key process in the global ocean distribution of *Synechococcus* cyanobacteria. *Proc Natl Acad Sci*.

Grébert T, Nguyen AA, Pokhrel S, Chen B, Ratin M, Karty JA, *et al.* (in prep.). Functional characterization of the MpeWYZ phycobilin lyase family provides key insights into color niche acclimation and adaptation in marine *Synechococcus*.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**: 307–321.

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

Herdman R, Castenholz RW, Itean I, Waterbury JB, Rippka R. (2001). Subsection I (Formerly Chroococcales Wettstein 1924, emend. Rippka, Deruelles, Waterbury, Herdman and Stanier 1979). In: Vol. 1 The archaea and the deeply branching and phototrophic bacteria. *Bergey's Manual of Systematics of Archaea and Bacteria*. D.R. Boone, R.W. Catsenholtz and G.M. Garrity, pp 493–514.

Huang S, Wilhelm SW, Harvey HR, Taylor K, Jiao N, Chen F. (2012). Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *ISME J* **6**: 285–297.

Huerta-Cepas J, Serra F, Bork P. (2016). ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol* **33**: 1635–1638.

Humily F, Farrant GK, Marie D, Partensky F, Mazard S, Perennou M, *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.

Humily F, Partensky F, Six C, Farrant GK, Ratin M, Marie D, *et al.* (2013). A gene island with two possible configurations is involved in chromatic acclimation in marine *Synechococcus*. *PLoS ONE* **8**: e84459.

Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A, *et al.* (2014). Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* **344**: 416–420.

Katoh K, Standley DM. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* **30**: 772–780.

Larsson J, Celepli N, Ininbergs K, Dupont CL, Yooseph S, Bergman B, *et al.* (2014). Picocyanobacteria containing a novel pigment gene cluster dominate the brackish water Baltic Sea. *ISME J* **8**: 1892–1903.

Liu H, Zhang H, Niedzwiedzki DM, Prado M, He G, Gross ML, *et al.* (2013). Phycobilisomes supply excitations to both photosystems in a megacomplex in Cyanobacteria. *Science* **342**: 1104–1107.

Magoč T, Salzberg SL. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinforma Oxf Engl* **27**: 2957–2963.

Mahmoud RM, Sanfilippo JE, Nguyen AA, Strnat JA, Partensky F, Garczarek L, *et al.* (2017). Adaptation to blue light in marine *Synechococcus* requires MpeU, an enzyme with similarity to phycoerythrobilin lyase isomerases. *Front Microbiol* **8**. e-pub ahead of print, doi: 10.3389/fmicb.2017.00243.

Mazard S, Ostrowski M, Partensky F, Scanlan DJ. (2012). Multi-locus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*. *Environ Microbiol* **14**: 372–386.

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**. e-pub ahead of print, doi: 10.3389/fmicb.2012.00285.

Montgomery BL, Casey ES, Grossman AR, Kehoe DM. (2004). Apla, a member of a new class of phycobiliproteins lacking a traditional role in photosynthetic light harvesting. *J Bacteriol* **186**: 7420–7428.

Olson RJ, Chisholm SW, Zettler ER, Armbrust EV. (1990). Pigments, size, and distributions of *Synechococcus* in the North Atlantic and Pacific Oceans. *Limnol Oceanogr* **35**: 45–58.

Ong LJ, Glazer AN. (1991). Phycoerythrins of marine unicellular cyanobacteria. I. Bilin types and locations and energy transfer pathways in *Synechococcus* spp. phycoerythrins. *J Biol Chem* **266**: 9515–9527.

Ong LJ, Glazer AN, Waterbury JB. (1984). An unusual phycoerythrin from a marine cyanobacterium. *Science* **224**: 80–83.

Palenik B. (2001). Chromatic adaptation in marine *Synechococcus* strains. *Appl Environ Microbiol* **67**: 991–994.

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

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

Pittera J, Partensky F, Six C. (2016). Adaptive thermostability of light-harvesting complexes in marine picocyanobacteria. *ISME J*. e-pub ahead of print, doi: 10.1038/ismej.2016.102.

Sánchez-Baracaldo P. (2015). Origin of marine planktonic cyanobacteria. *Sci Rep* **5**. e-pub ahead of print, doi: 10.1038/srep17418.

Sánchez-Baracaldo P, Ridgwell A, Raven JA. (2014). A neoproterozoic transition in the marine nitrogen cycle. *Curr Biol* **24**: 652–657.

Sanfilippo JE, Nguyen AA, Karty JA, Shukla A, Schluchter WM, Garczarek L, *et al.* (2016). Self-regulating genomic island encoding tandem regulators confers chromatic acclimation to marine *Synechococcus*. *Proc Natl Acad Sci* **113**: 6077–6082.

Schluchter WM, Shen G, Alvey RM, Biswas A, Saunée NA, Williams SR, *et al.* (2010). Phycobiliprotein biosynthesis in cyanobacteria: structure and function of enzymes involved in post-translational modification. *Adv Exp Med Biol* **675**: 211–228.

Shukla A, Biswas A, Blot N, Partensky F, Karty JA, Hammad LA, *et al.* (2012). Phycoerythrin-specific bilin lyase-isomerase controls blue-green chromatic acclimation in marine *Synechococcus*. *Proc Natl Acad Sci* **109**: 20136–20141.

Sidler WA. (1994). Phycobilisome and phycobiliprotein structures. In: *Advances in Photosynthesis. The Molecular Biology of Cyanobacteria*. Springer, Dordrecht, pp 139–216.

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.

Six C, Thomas J-C, Thion L, Lemoine Y, Zal F, Partensky F. (2005). Two novel phycoerythrin-associated linker proteins in the marine cyanobacterium *Synechococcus* sp. strain WH8102. *J Bacteriol* **187**: 1685–1694.

Sohm JA, Ahlgren NA, Thomson ZJ, Williams C, Moffett JW, Saito MA, *et al.* (2016). Co-occurring *Synechococcus* ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. *ISME J* **10**: 333–345.

Tajima F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**: 437–460.

Wilbanks SM, Glazer AN. (1993). Rod structure of a phycoerythrin II-containing phycobilisome. I. Organization and sequence of the gene cluster encoding the major phycobiliprotein rod components in the genome of marine *Synechococcus* sp. WH8020. *J Biol Chem* **268**: 1226–1235.

Wood AM, Phinney DA, Yentsch CS. (1998). Water column transparency and the distribution of spectrally distinct forms of phycoerythrin-containing organisms. *Mar Ecol Prog Ser* **162**: 25–31.

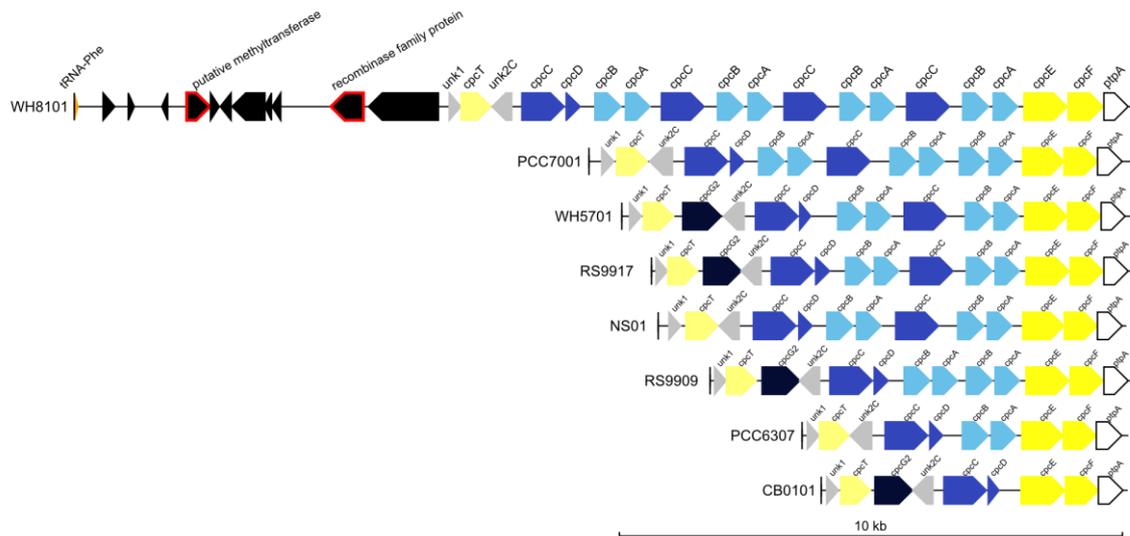
Xia X, Liu H, Choi D, Noh JH. (2017a). Variation of *Synechococcus* pigment genetic diversity along two turbidity gradients in the China Seas. *Microb Ecol* 1–12.

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

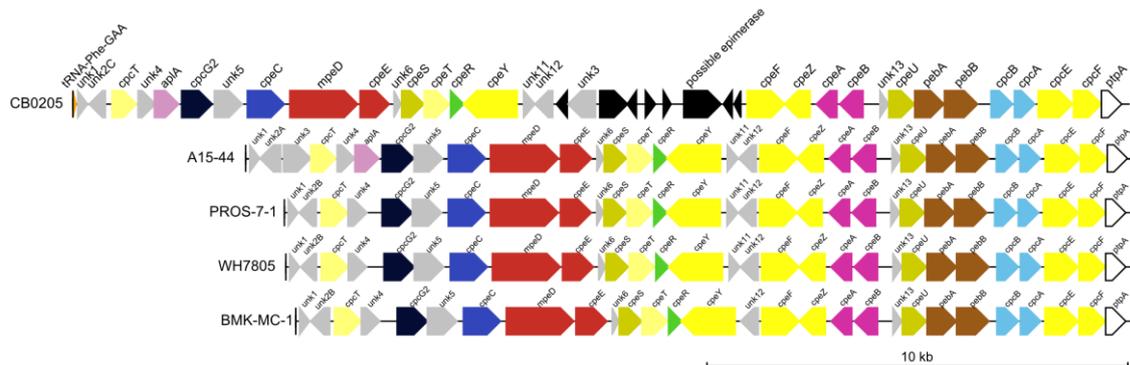
Yu MH, Glazer AN. (1982). Cyanobacterial phycobilisomes. Role of the linker polypeptides in the assembly of phycocyanin. *J Biol Chem* **257**: 3429–3433.

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

Supplementary figures



**Fig. S1: PBS rod genomic region for all strains of pigment type 1.** Regions are oriented from the phenylalanine tRNA (shown in orange only in the first genomic region) to the conserved tyrosine phosphatase *ptpA*. Genes are coloured according to their inferred function (see insert in Fig. 1) and their length is proportional to the gene size.



**Fig. S2: Same as Fig. S1 but for pigment type 2A.**

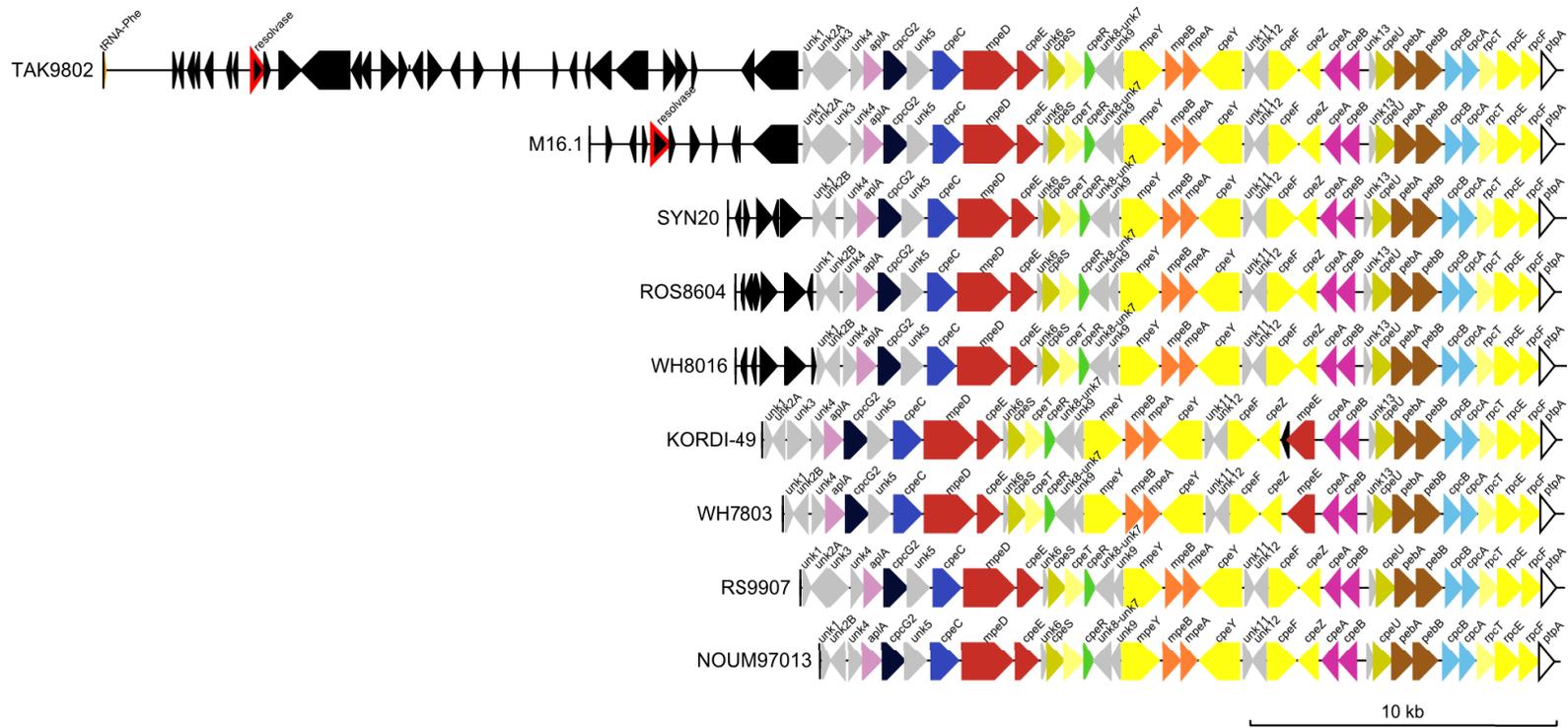


Fig. S3: Same as Fig. S1 but for pigment type 3a.

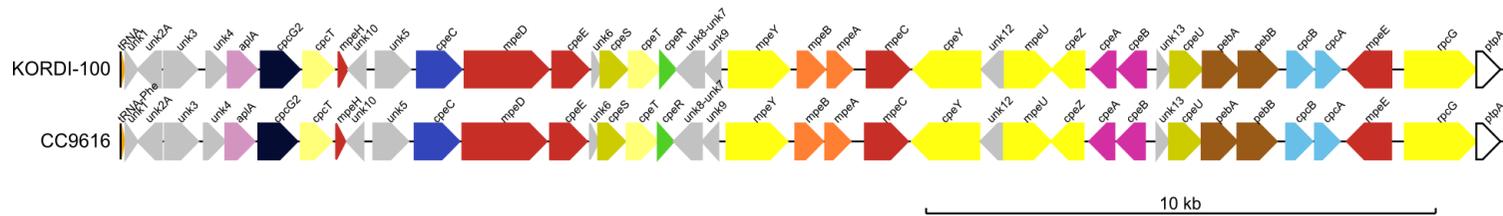


Fig. S4: Same as Fig. S1 but for pigment type 3f.

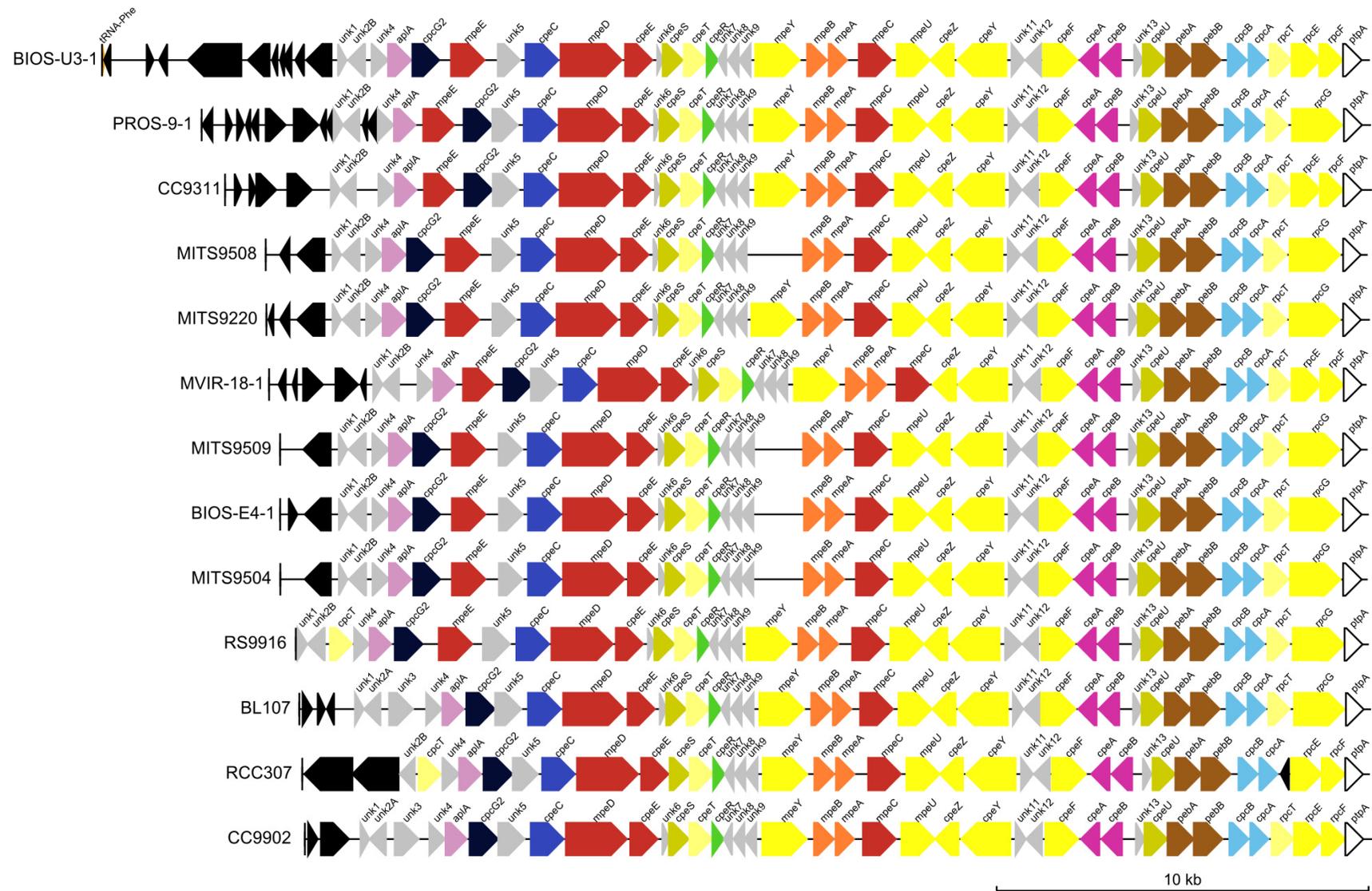


Fig. S5: Same as Fig. S1 but for pigment type 3dA.

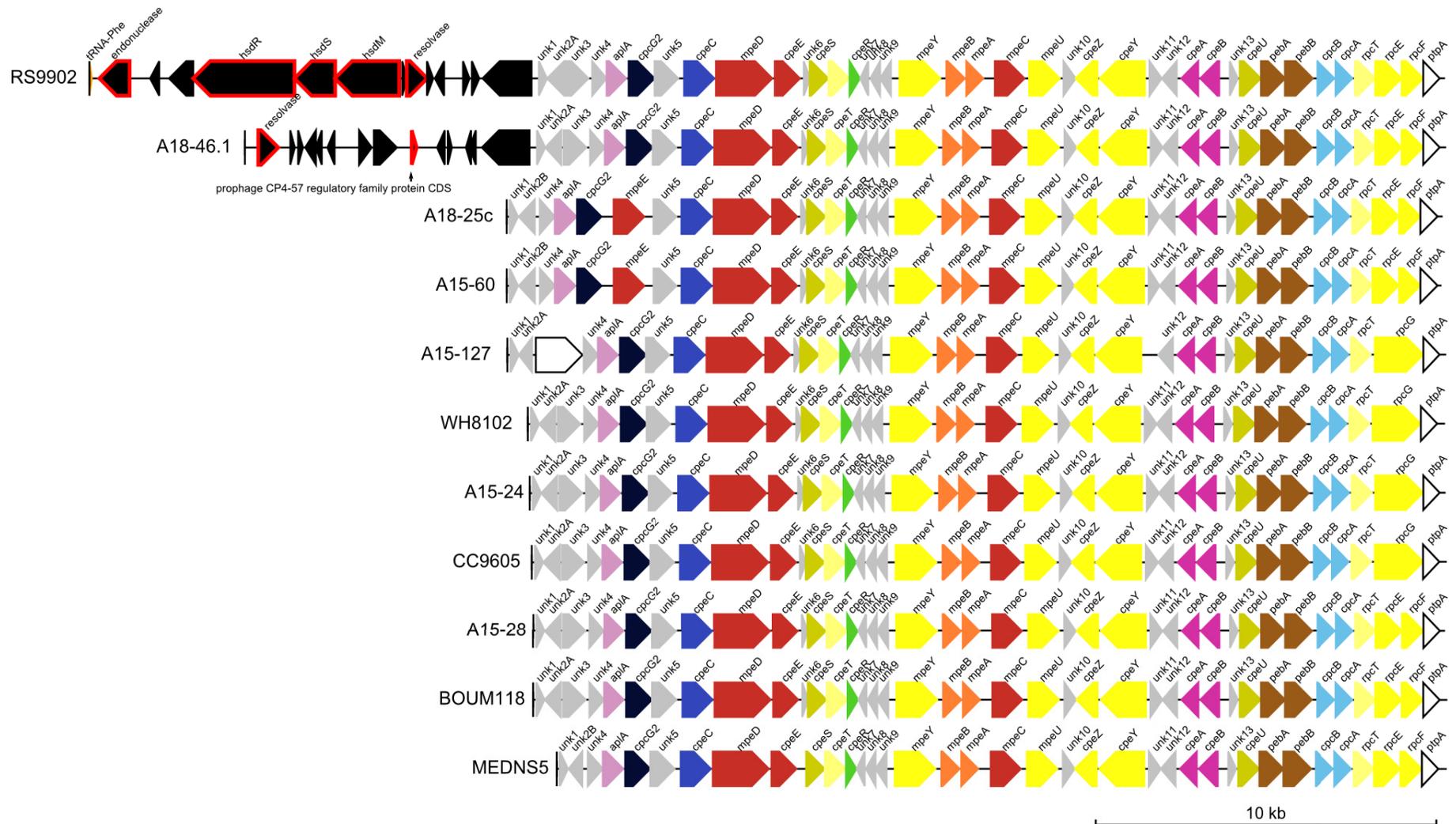


Fig. S6: Same as Fig. S1 but for pigment type 3c.

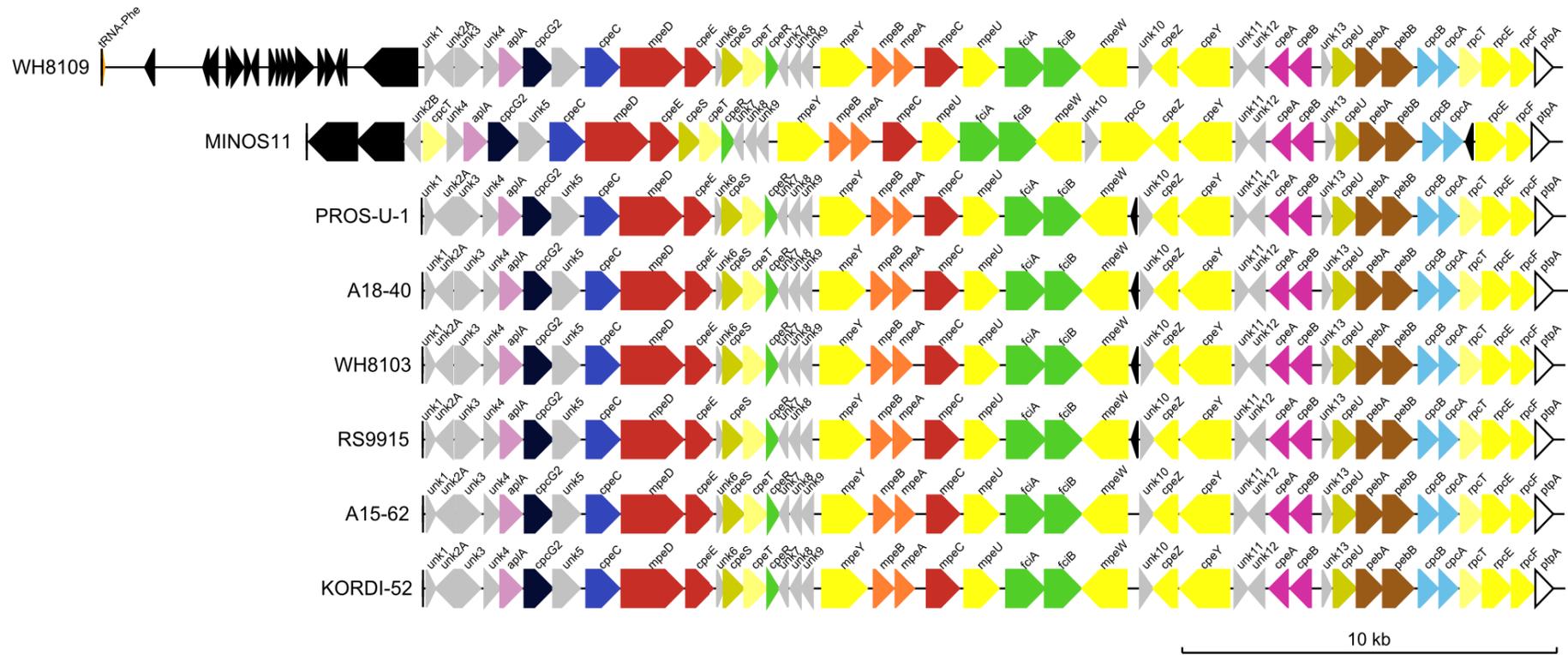
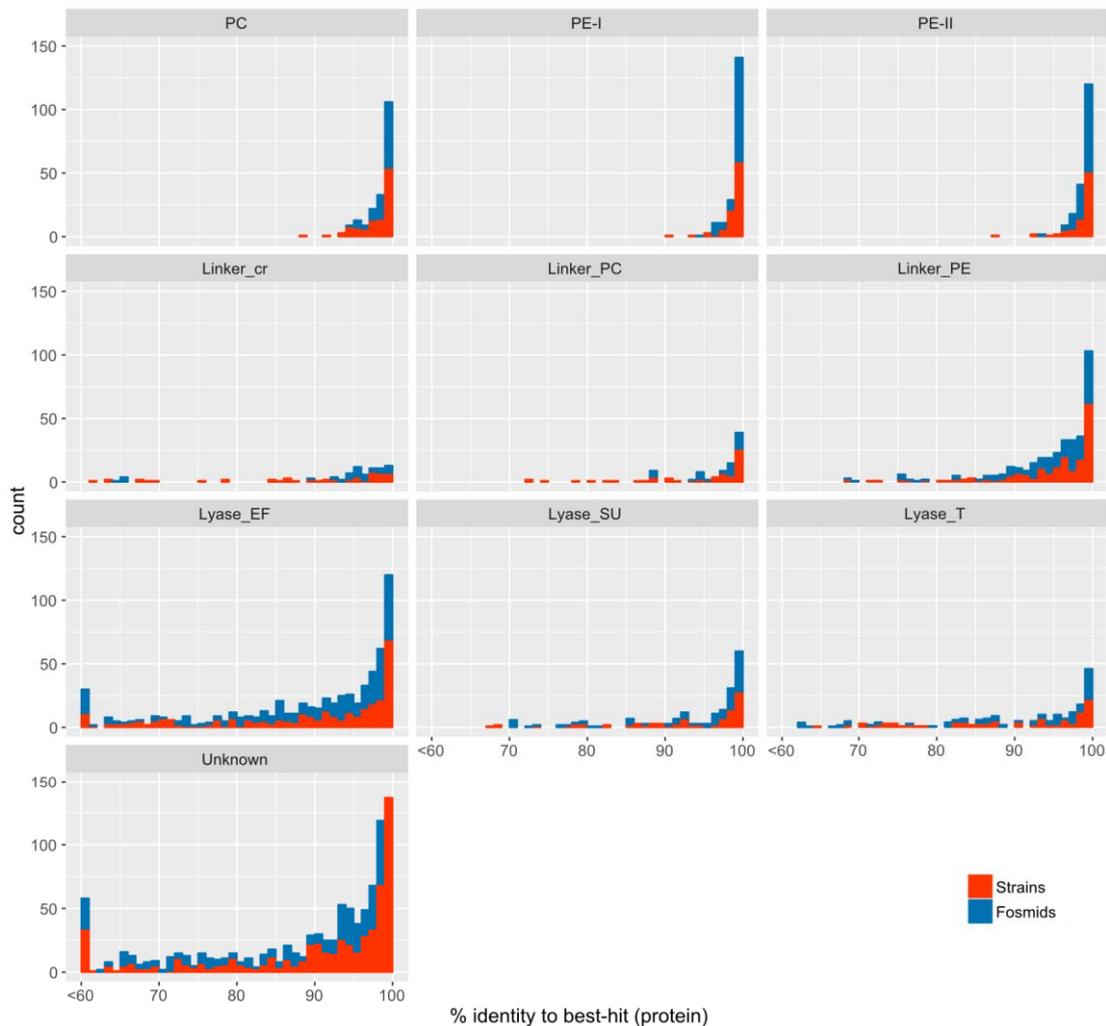


Fig. S7: Same as Fig. S1 but for pigment type 3dB.

## Supplementary figures



**Fig. S8: Genetic variability of the different proteins encoded in the PBS rod genomic region, grouped by functional categories.** For each group of orthologous proteins, identity is based on the blastp best-hit other than self-hit for strains, and blastp best-hit to all strains for fosmids, and results for each group of orthologs have then be gathered by functional categories as follows: PC, phycocyanin (C/RpcA/B); PE, -I, phycoerythrin-I (CpeA/B); PE-II, phycoerythrin-II (MpeA/B); Linker\_cr, linker rod-core (CpcG); Linker\_PC, PC-associated linker (CpcC, CpcD, CpeC); Linker\_PE, PE-associated linker (CpeE, MpeC, MpeD, MpeE, MpeH); lyase\_EF, lyase of the E/F structural clan (CpeF, CpeY, CpeZ, MpeU, MpeY, MpeW, MpeZ, RpcE, RpcF, RpcG, CpcE, CpcF); lyase\_SU, lyase of the S/U structural clan (CpeS, CpeU); lyase\_T, lyase of the T structural clan (CpcT, CpeT, RpcT); unknown, uncharacterized conserved hypothetical proteins (Unk1 through 13, Unk2A, Unk2B, Unk2C, Unk7/8, Unk14B).

## Supplementary tables

**Table S1: Presence or absence of genes encoding linker polypeptides in the different marine *Synechococcus* genomes. Syn, *Synechococcus*; Cya., *Cyanobium*; SC., Subcluster.**

| Genus | SC  | Clade | Strain    | PT  | APC                 |                                   | PC          |             |             | PEI         |             | PEII        |             |               |             |             |
|-------|-----|-------|-----------|-----|---------------------|-----------------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|---------------|-------------|-------------|
|       |     |       |           |     | <i>apcC</i><br>(Lc) | <i>apcE</i><br>(L <sub>CM</sub> ) | <i>cpcC</i> | <i>cpcD</i> | <i>cpcG</i> | <i>cpeC</i> | <i>cpeE</i> | <i>mpeD</i> | <i>mpeC</i> | <i>mpeE</i>   | <i>mpeF</i> | <i>mpeG</i> |
| Syn.  | 5.2 | 5.2   | WH5701    | 1   | CC                  | CC                                | 3           | RC          | 2           | -           | -           | -           | -           | -             | -           | -           |
| Syn.  | 5.2 | 5.2   | CB0101    | 1   | CC                  | CC                                | 2           | RC          | 2           | -           | -           | -           | -           | -             | -           | -           |
| Cya.  | 5.2 | 5.2   | NS01      | 1   | CC                  | CC                                | 3           | RC          | 2           | -           | -           | -           | -           | -             | -           | -           |
| Cya.  | 5.2 | 5.2   | PCC6307   | 1   | CC                  | CC                                | 2           | RC          | 2           | -           | -           | -           | -           | -             | -           | -           |
| Cya.  | 5.2 | 5.2   | PCC7001   | 1   | CC                  | CC                                | 3           | RC          | 2           | -           | -           | -           | -           | -             | -           | -           |
| Syn.  | 5.1 | VIII  | RS9909    | 1   | CC                  | CC                                | 2           | RC          | 2           | -           | -           | -           | -           | -             | -           | -           |
| Syn.  | 5.1 | VIII  | RS9917    | 1   | CC                  | CC                                | 2           | RC          | 2           | -           | -           | -           | -           | -             | -           | -           |
| Syn.  | 5.1 | VIII  | WH8101    | 1   | CC                  | CC                                | 2           | RC          | 2           | -           | -           | -           | -           | -             | -           | -           |
| Syn.  | 5.2 | 5.2   | CB0205    | 2   | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | -           | -             | -           | -           |
| Syn.  | 5.1 | II    | A15-44    | 2   | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | -           | -             | -           | -           |
| Syn.  | 5.1 | V     | BMK-MC-1  | 2   | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | -           | -             | -           | -           |
| Syn.  | 5.1 | VI    | PROS-7-1  | 2   | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | -           | -             | -           | -           |
| Syn.  | 5.1 | VI    | WH7805    | 2   | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | -           | -             | -           | -           |
| Syn.  | 5.1 | I     | ROS8604   | 3a  | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | -           | RC            | -           | -           |
| Syn.  | 5.1 | I     | SYN20     | 3a  | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | -           | -             | -           | -           |
| Syn.  | 5.1 | II    | M16.1     | 3a  | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | -           | RC            | -           | 1           |
| Syn.  | 5.1 | II    | RS9907    | 3a  | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | -           | NC            | -           | 1           |
| Syn.  | 5.1 | II    | TAK9802   | 3a  | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | -           | NC            | -           | 1           |
| Syn.  | 5.1 | V     | WH7803    | 3a  | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | -           | RC            | -           | -           |
| Syn.  | 5.1 | VII   | NOUM97013 | 3a  | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | -           | NC            | -           | -           |
| Syn.  | 5.1 | I     | WH8016    | 3aA | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | -           | NC            | -           | -           |
| Syn.  | 5.1 | WPC1  | KORDI-49  | 3aA | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | -           | RC            | -           | NC          |
| Syn.  | 5.1 | II    | CC9605    | 3c  | CC                  | CC                                | -           | -           | 3           | RC          | RC          | RC          | RC          | NC            | -           | 1           |
| Syn.  | 5.1 | II    | RS9902    | 3c  | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | RC          | NC            | -           | 1           |
| Syn.  | 5.1 | III   | A15-24    | 3c  | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | RC          | RC*           | -           | 1           |
| Syn.  | 5.1 | III   | A15-28    | 3c  | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | RC          | RC*           | -           | 1           |
| Syn.  | 5.1 | III   | A18-46.1  | 3c  | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | RC          | RC*           | -           | 1           |
| Syn.  | 5.1 | III   | BOUM118   | 3c  | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | RC          | RC*           | -           | 2           |
| Syn.  | 5.1 | III   | WH8102    | 3c  | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | RC          | RC*           | -           | 1           |
| Syn.  | 5.1 | VI    | MEDNS5    | 3c  | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | RC          | -             | -           | -           |
| Syn.  | 5.1 | VII   | A15-60    | 3c  | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | RC          | RC***<br>/ NC | -           | -           |
| Syn.  | 5.1 | VII   | A18-25c   | 3c  | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | RC          | RC***<br>/ NC | -           | -           |
| Syn.  | 5.1 | WPC1  | A15-127   | 3c  | CC                  | CC                                | -           | -           | 3           | RC          | RC          | RC          | RC          | RC*           | -           | NC          |
| Syn.  | 5.1 | 5.3   | RCC307    | 3dA | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | RC          | RC*           | -           | NC          |
| Syn.  | 5.1 | CRD1  | BIOS-U3-1 | 3dA | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | RC          | RC**          | NC          | -           |
| Syn.  | 5.1 | CRD1  | MIT59220  | 3dA | CC                  | CC                                | -           | -           | 3           | RC          | RC          | RC          | RC          | RC***         | NC          | -           |

| Genus | SC  | Clade | Strain    | PT  | <i>apcC</i><br>(Lc) | <i>apcE</i><br>(L <sub>CM</sub> ) | <i>cpcC</i> | <i>cpcD</i> | <i>cpcG</i> | <i>cpeC</i> | <i>cpeE</i> | <i>mpeD</i> | <i>mpeC</i> | <i>mpeE</i> | <i>mpeF</i> | <i>mpeG</i> | <i>mpeH</i> |
|-------|-----|-------|-----------|-----|---------------------|-----------------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Syn.  | 5.1 | I     | CC9311    | 3dA | CC                  | CC                                | -           | -           | 3           | RC          | RC          | RC          | RC          | RC          | NC          | -           | -           |
| Syn.  | 5.1 | I     | PROS-9-1  | 3dA | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | RC          | RC**        | NC          | -           | -           |
| Syn.  | 5.1 | I     | WH8020    | 3dA | CC                  | CC                                | -           | -           | 3           | RC          | RC          | RC          | RC          | RC**        | NC          | -           | -           |
| Syn.  | 5.1 | IV    | BL107     | 3dA | CC                  | CC                                | -           | -           | 3           | RC          | RC          | RC          | RC          | RC*         | NC          | -           | -           |
| Syn.  | 5.1 | IV    | CC9902    | 3dA | CC                  | CC                                | -           | -           | 3           | RC          | RC          | RC          | RC          | RC*         | NC          | -           | -           |
| Syn.  | 5.1 | IX    | RS9916    | 3dA | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | RC          | RC          | -           | NC          | 1           |
| Syn.  | 5.1 | I     | MVIR-18-1 | 3aA | CC                  | CC                                | -           | -           | 3           | RC          | RC          | RC          | RC          | RC          | NC          | -           | -           |
| Syn.  | 5.1 | CRD1  | BIOS-E4-1 | 3cA | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | RC          | RC***       | NC          | -           | -           |
| Syn.  | 5.1 | CRD1  | MITS9504  | 3cA | CC                  | CC                                | -           | -           | 3           | RC          | RC          | RC          | RC          | RC***       | NC          | -           | -           |
| Syn.  | 5.1 | CRD1  | MITS9508  | 3cA | CC                  | CC                                | -           | -           | 3           | RC          | RC          | RC          | RC          | RC***       | NC          | -           | -           |
| Syn.  | 5.1 | CRD1  | MITS9509  | 3cA | CC                  | CC                                | -           | -           | 3           | RC          | RC          | RC          | RC          | RC***       | NC          | -           | -           |
| Syn.  | 5.1 | 5.3   | MINOS11   | 3dB | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | RC          | RC*         | -           | NC          | -           |
| Syn.  | 5.1 | II    | A15-62    | 3dB | CC                  | CC                                | -           | -           | 3           | RC          | RC          | RC          | RC          | NC          | -           | -           | 1           |
| Syn.  | 5.1 | II    | PROS-U-1  | 3dB | CC                  | CC                                | -           | -           | 3           | RC          | RC          | RC          | RC          | NC          | -           | NC          | 1           |
| Syn.  | 5.1 | III   | A18-40    | 3dB | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | RC          | RC*         | -           | -           | 1           |
| Syn.  | 5.1 | III   | RS9915    | 3dB | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | RC          | RC*         | -           | -           | 1           |
| Syn.  | 5.1 | II    | KORDI-52  | 3bB | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | RC          | NC          | -           | NC          | 1           |
| Syn.  | 5.1 | II    | WH8109    | 3bB | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | RC          | NC          | -           | NC          | 1           |
| Syn.  | 5.1 | III   | WH8103    | 3bB | CC                  | CC                                | -           | -           | 2           | RC          | RC          | RC          | RC          | RC*         | -           | -           | 1           |
| Syn.  | 5.1 | EnvC  | CC9616    | 3f  | CC                  | CC                                | -           | -           | 3           | RC          | RC          | RC          | RC          | RC****      | -           | NC          | 1           |
| Syn.  | 5.1 | UC-A  | KORDI-100 | 3f  | CC                  | CC                                | -           | -           | 3           | RC          | RC          | RC          | RC          | RC****      | -           | -           | 1           |

RC1 = a few genes upstream of RC region

RC2 = upstream of *cpcG2*

RC3 = downstream of *cpcG2*

RC4 = downstream of *cpcA*

**Table S2: Characteristics of the samples used to build fosmid libraries.** Temp., temperature; Sal., salinity; *Syn.*, *Synechococcus*; *Proc.*, *Prochlorococcus*; PEuk, Picoeukaryotes; Med., Mediterranean; Vfr, Villefranche-sur-mer.

| Cruise            | Sample             | Region          | Station Lat (Dec N) | Station Lon (Dec E) | Date      | Time (GMT) | Sample depth (m) | Temp. (°C) | Sal. (psu) | DCM (m) | <i>Syn.</i> cells mL <sup>-1</sup> | <i>Proc.</i> cells mL <sup>-1</sup> | PEuk cells mL <sup>-1</sup> |
|-------------------|--------------------|-----------------|---------------------|---------------------|-----------|------------|------------------|------------|------------|---------|------------------------------------|-------------------------------------|-----------------------------|
| SOMLIT            | Astan              | English Channel | 48.7778             | -3.9375             | 29-May-12 | 14:40      | 1                | 12.55      | 35.242     | -       | 8.41E+02                           | -                                   | 3.44E+03                    |
| CEFAS (811)       | Station 3          | North Sea       | 54.2053             | -0.0133             | 08-May-11 | 17:08      | 6                | 10.76      | 30.82      | -       | 3.30E+05                           | -                                   | 3.29E+05                    |
| CEFAS (811)       | Station 7          | North Sea       | 55.6795             | 2.2694              | 10-May-11 | 5:40       | 6                | 11.10      | 31.31      | -       | 1.32E+06                           | -                                   | 1.49E+06                    |
| CEFAS (811)       | Station 13         | North Sea       | 54.7717             | 2.9985              | 11-May-11 | 5:40       | 7                | 11.28      | 30.98      | -       | 3.96E+05                           | -                                   | 5.92E+05                    |
| BOUM              | Station A          | Med. Sea        | 39.2543             | 5.1592              | 10-Jul-08 | 14:40      | 90               | 15.145     | 37.617     | 88      | 1.76E+03                           | 4.91E+04                            | 2.38E+03                    |
| BOUM              | Station B          | Med. Sea        | 34.1252             | 18.4555             | 04-Jul-08 | 09:00      | 12,5             | 24.948     | 38.436     | 141     | 6.89E+03                           | -                                   | 5.53E+02                    |
| BOUM              | Station C          | Med. Sea        | 33.7125             | 32.6527             | 29-Jun-08 | 12:45      | 75               | 17.914     | 39.41      | 108     | 4.48E+03                           | -                                   | 5.26E+02                    |
| BOUM              | Station C          | Med. Sea        | 33.7125             | 32.6527             | 28-Jun-08 | 12:45      | 100              | 17.614     | 39.405     | 108     | 1.32E+04                           | 3.35E+04                            | 3.65E+02                    |
| -                 | Vfr (point B)      | Med. Sea        | 43.6833             | 7.3167              | 02-Jun-09 | 09:00      | 40               | 15.68      | 37.85      | -       | 3.38E+04                           | 4.33E+03                            | 1.32E+03                    |
| -                 | Vfr (point B)      | Med. Sea        | 43.6833             | 7.3167              | 02-Jun-09 | 09:00      | 30               | 17.326     | 3775       | -       | 3.32E+04                           | -                                   | 1.61E+03                    |
| -                 | Nice (Boussole)    | Med. Sea        | 43.3667             | 7.9000              | 16-Jun-09 | nd         | 45               | nd         | nd         | -       | 1.80E+04                           | -                                   | 3.95E+03                    |
| -                 | Marseille (Frioul) | Med. Sea        | 43.2417             | 5.2917              | 04-Jun-09 | 07:00      | 40               | 15.17      | 37.851     | -       | 3.49E+04                           | 1.87E+03                            | 2.09E+03                    |
| -                 | Marseille (Frioul) | Med. Sea        | 43.2417             | 5.2917              | 04-Jun-09 | 07:00      | 10               | 19.71      | 37.78      | -       | 1.89E+04                           | -                                   | 3.37E+03                    |
| Atlantic transect | Station 5          | Atlantic Ocean  | 45.9897             | -20.0014            | 21-Jul-11 | 18:30      | Surf             | 16.64      | 35.79      | 43      | 5.35E+04                           | 1.51E+04                            | 1.95E+04                    |
| Atlantic transect | Station 8          | Atlantic Ocean  | 43.0030             | -19.9911            | 22-Jul-11 | 22:00      | 25               | 16.9       | 35.89      | 35      | nd                                 | nd                                  | nd                          |
| Atlantic transect | Station 10         | Atlantic Ocean  | 41.0009             | -20.0086            | 24-Jul-11 | 16:30      | 86               | 14.37      | 35.92      | 55      | nd                                 | nd                                  | nd                          |
| Atlantic transect | Station 28         | Atlantic Ocean  | 23.3423             | -27.1358            | 31-Jul-11 | 12:40      | Surf             | 24.13      | 37         | 113     | nd                                 | nd                                  | nd                          |

**Table S3: Comparative statistics for the assembly and scaffolding of PBS genomic rod regions performed from 8 fosmid libraries**

| Sampling station                     | ASTAN            | BOUM - C   |         | BOUM - B      | Atlantic CTD5 | Villefranche |     | Atlantic CTD8 | Atlantic CTD10 | CEFAS 3    | CEFAS 7 | CEFAS 13 | BOUM - A   | Boussole |
|--------------------------------------|------------------|------------|---------|---------------|---------------|--------------|-----|---------------|----------------|------------|---------|----------|------------|----------|
| Sampling depth                       | Surface          | 75m        | 100m    | 12.5m         | Surface       | 30m          | 40m | 25m           | 86m            | 6m         | 6m      | 7m       | 90m        | 45m      |
| Number of fosmids pooled             |                  | 25         |         | 25            | 25            | 25           |     | 25            |                | 19         |         |          | 20         |          |
| Fosmid library code                  | A                | C          |         | D             | E             | F            |     | G             |                | H          |         |          | I          |          |
| Number of raw reads                  | 2x<br>6930930    | 2x 2956459 |         | 2x<br>1304833 | 2x<br>1551515 | 2x 1379469   |     | 2x 1747044    |                | 2x 1564879 |         |          | 2x 2300094 |          |
| Number of reads after filtering      | 2x<br>3832836    | 2x 1730953 |         | 2x<br>798848  | 2x<br>879142  | 2x 878425    |     | 2x 1040306    |                | 2x 950149  |         |          | 2x 1139978 |          |
| After assembly and elongation        | n(contigs)       | 80         | 56      | 17            | 47            | 56           |     | 31            |                | 39         |         |          | 37         |          |
|                                      | Avg length       | 4851.9     | 4797.75 | 4655.9        | 6957.4        | 3393.1       |     | 9250.5        |                | 4103.4     |         |          | 7881.8     |          |
|                                      | Max length       | 36709      | 18596   | 11678         | 38615         | 11971        |     | 25008         |                | 18439      |         |          | 23359      |          |
|                                      | N50 <sup>a</sup> | 7958       | 7088    | 8667          | 12279         | 4702         |     | 15021         |                | 6325       |         |          | 13438      |          |
| After manual curation and annotation | n(contigs)       | 60         | 51      | 12            | 37            | 55           |     | 25            |                | 37         |         |          | 33         |          |
|                                      | Avg length       | 4853       | 3942    | 4171          | 5605          | 2923         |     | 7789          |                | 3557       |         |          | 4638       |          |
|                                      | Max length       | 34378      | 18594   | 11496         | 27051         | 9987         |     | 23575         |                | 18422      |         |          | 21952      |          |
|                                      | N50 <sup>a</sup> | 7480       | 6134    | 7837          | 8055          | 3935         |     | 11140         |                | 6144       |         |          | 7961       |          |
| After final scaffolding              | n(contigs)       | 43         | 29      | 6             | 29            | 45           |     | 22            |                | 27         |         |          | 28         |          |
|                                      | Avg length       | 6278       | 6339    | 7238          | 7054          | 3494         |     | 9008          |                | 4636       |         |          | 5342       |          |
|                                      | Max length       | 29599      | 20295   | 14507         | 27051         | 13626        |     | 24826         |                | 18419      |         |          | 21839      |          |
|                                      | N50 <sup>a</sup> | 8521       | 8718    | 11496         | 12856         | 4393         |     | 15202         |                | 7915       |         |          | 9005       |          |

<sup>a</sup> contig size over which the sum of contig sizes corresponds to 50 % of the assembly

## 2. Future work

The work presented in the previous pages brings several important new insights about the evolution of *Synechococcus* phycobilisomes, namely a novel evolutionary scenario and a lower bound for the date of the appearance of the different pigment types. It also suggests that population-scale mechanisms are necessary to explain the distribution and maintenance of pigment types in the different *Synechococcus* clades. However, these hypotheses still need careful testing before this draft article can be submitted for publication. In particular, I still have to statistically evaluate the different evolutionary hypotheses (Planet, 2006). This has been previously done by Everroad and Wood (2012) for the evolution of phycoerythrin-I and -II using the approximately unbiased test of Shimodaira (Shimodaira, 2002), a test that revealed positive directional selection towards higher PUB/PEB ratio. This analysis could be repeated with the extended dataset that we now have, to i) check Everroad and Wood model in the light of the newly described PT 3f and ii) test whether the nested clade delineation within the different phycobiliproteins is significant. To this end, I will sequence the *mpeBA* operon (encoding PE-II subunits) from a selection of strains in order to obtain sequences for yet unrepresented pigment type/clade combinations. An interesting example of this is *Synechococcus* sp. strain MV1715 (Hunter-Cevera *et al.*, 2016b), an PT 3a representative kindly provided to us by Dr. Kristen Hunter-Cevera and one of the rare subcluster 5.3 isolates available in culture.

Another important but still unaddressed question is the mechanism of transfer of the PBS genomic region. Viruses are recognized as important drivers of microbial evolution and as potential genetic transfer agents (Lindell *et al.*, 2004, 2005; Fridman *et al.*, 2017). Moreover, genes encoding photosynthetic genes have been widely observed in cyanophages, including genes involved in PBS synthesis such as the lyase genes *cpeS* or *cpeT* (Sullivan *et al.*, 2006; Ledermann *et al.*, 2016; Dammeyer *et al.*, 2008; Gasper *et al.*, 2017). Still, no complete or partial phycobilisome genomic region has been reported from viral metagenomes to date, but it is unclear whether it is because they are absent from the viral metagenome datasets currently available or whether they have not been looked for. I will screen through such datasets for potential PBS genomic regions, as it could potentially reveal an important factor of gene transfer.

### **III. Metagenomes suggest adaptation of phycobiliproteins to temperature**

#### **1. Background**

By carefully comparing the physiology of six different *Synechococcus* strains, Pittera and co-workers showed that the thermostability of their phycobilisomes (PBS) varied in relation with the average annual sea temperature at the isolation sites of these strains, both *in vivo* and *in vitro* (Pittera *et al.*, 2016). They related these different PBS stabilities with differences in the molecular flexibility of individual phycobiliproteins constituting the rods, and showed that phycocyanin (PC) was the most sensitive of all rod phycobiliprotein to temperature increases. Finally, they identified two amino acid substitutions that were predicted to be responsible for major differences in protein flexibility in a region ensuring contact between  $\alpha$ - and  $\beta$ -subunits. These substitutions are the replacement of an alanine (A43  $\alpha$ -PC) and an asparagine (N42  $\beta$ -PC) in warm-adapted strains with a glycine (G43  $\alpha$ -PC) and a serine (S42  $\beta$ -PC) in cold-adapted strains. This led them to hypothesize that these specific substitutions might play a key role in the thermostability of the PBS and thus potentially represent adaptations to differing thermal niches. Here, I wanted to test whether a similar pattern could be observed in PC subunits from natural *Synechococcus* populations from different environments. To this end, I applied the metagenomic read recruitment approach used in (Grébert *et al.*, in revision; see Chapter II) to recruit reads corresponding to *cpcBA*, the operon encoding PC  $\beta$ - and  $\alpha$ -subunits, in the extensive *Tara* Oceans metagenomic dataset, which altogether represents 137 samples ranging from  $-0.8^{\circ}\text{C}$  (sampling at the deep chlorophyll maximum (87 m depth) of TARA\_085, located in the Southern Ocean) to  $30.4^{\circ}\text{C}$  (surface (5 m depth) of TARA\_045, Indian Ocean). I then extracted from these reads the proportion of amino acids at each position along CpcA and CpcB.

#### **2. Materials and methods**

##### *Metagenomic samples*

This study focused on 137 metagenomic samples of the worldwide oceans, of which 109 collected during the 2.5-year *Tara* Oceans circumnavigation (2009-2012, Sunagawa *et al.*, 2015) and 28 during the *Tara* Polar Circle expedition (2013, circumnavigation around the Arctic). Sample processing and features are the same as described in (Farrant *et al.*, 2016; see also Chapter II). Temperature was either retrieved from (Farrant *et al.*, 2016) for *Tara* Oceans or extracted from PANGAEA for the yet-unpublished *Tara* Polar Circle part of the dataset (Tara Oceans Consortium and Tara Oceans Expedition, 2017).

*Databases: reference and outgroup sequences*

A reference database comprising the full-length *cpcBA* operon nucleotide sequences was generated based on culture isolates with characterized pigment type (see Dataset 1 of Grébert *et al.*, in revision). It comprised 83 *cpcBA* sequences, of which 64 unique. A reference alignment was generated with MAFFT L-INS-i v6.953b (Kato and Standley, 2013), and a reference tree with PhyML v. 20120412 (GTR+I+G, 10 random starting trees, best of SPR and NNI moves, 500 bootstraps; Guindon *et al.*, 2010).

A database of outgroups was also built, comprising paralogous phycobiliprotein sequences coming from marine *Synechococcus* or *Prochlorococcus* (*cpeBA*, *mpeBA*, *apcA*, *apcB*, *apcD*, *apcF*, *ppeBA*) as well as *cpcBA* sequences from other marine and freshwater organisms retrieved from public databases.

*Read assignment and estimation of PT abundance*

Reads were preselected using BLAST+ (Camacho *et al.*, 2009) with relaxed parameters (blastn, maximum E-value of 1e-5, minimum percent identity 60%, minimum 75% of read length aligned), using reference sequences as subjects. The selection was then refined by a second BLAST+ round against databases of outgroups: reads with a best-hit to outgroup sequences were excluded from downstream analysis. Selected reads were then aligned to the marker reference alignment with MAFFT v.7.299b (--addfragments --adjustdirectionaccurately) and assigned to a given pigment type base on *pplacer* placement in the reference *cpcBA* phylogenetic tree (Matsen *et al.*, 2010). Amino acids counts at each position along CpcB and CpcA were then extracted from reads covering entirely the corresponding nucleotide codon with custom scripts using *BioPython* v1.65 (Cock *et al.*, 2009). Only samples with more than 20 *cpcBA* reads assigned to PT 3 and entirely covering the considered codons were taken into account to have reliable proportion estimates.

### 3. Results and discussion

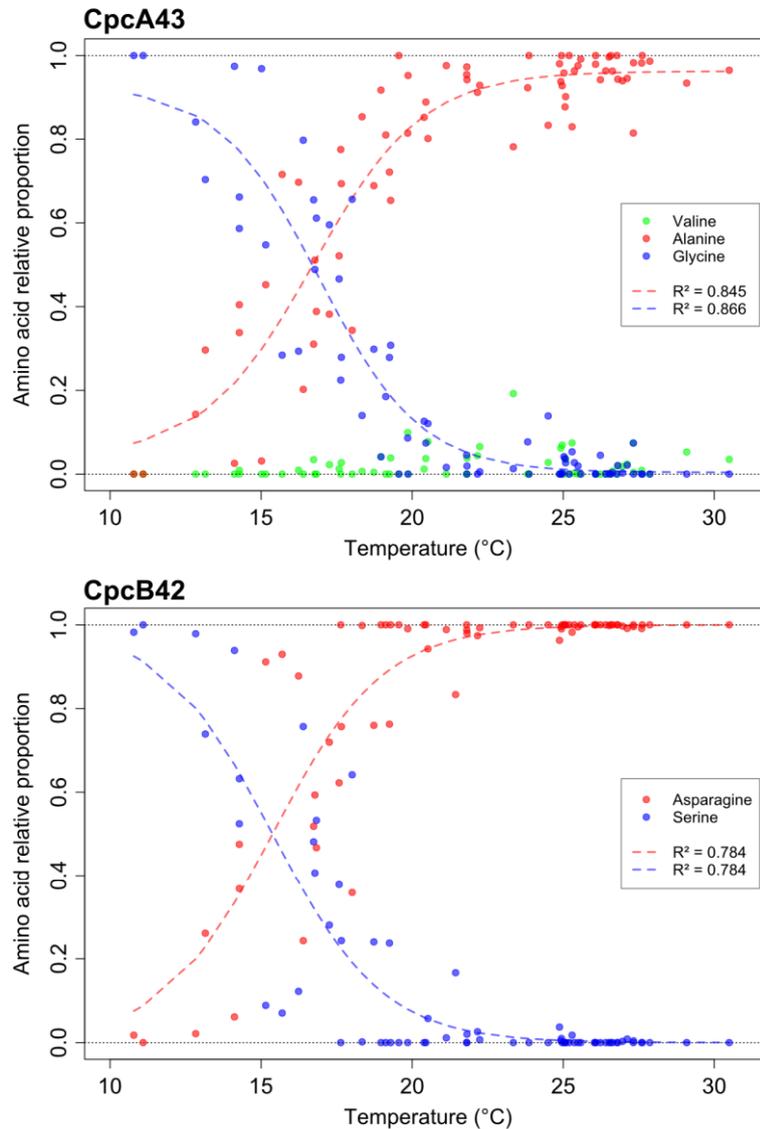
A total of 96 stations had more than 20 reads assigned to PT 3 and covering the 43<sup>rd</sup> codon of *cpcA*, and 95 stations for the 42<sup>nd</sup> codon of *cpcB*. The temperature at these stations ranged from 10.8°C (deep chlorophyll maximum of TARA\_155) to 30.5°C (surface of TARA\_45). Amino acids observed in natural populations at CpcA43 were alanine (A), glycine (G), and valine (V, **figure 39**). Although the latter substitution has not been described by Pittera and coworkers (Pittera *et al.*, 2016), we found two sequenced strains (CC9616 and KORDI-100, both PT 3f) that carry a valine at this position. In agreement with their analysis, the only two amino acids observed at CpcB42 in natural *Synechococcus* populations were serine (S) and asparagine (N; **figure 39**). The relative proportion of each amino acid at these positions in the

different metagenomic samples was found to vary with temperature (**figure 39**), with the substitutions A43G (CpcA) and N42S (CpcB) increasing in frequency with temperature. Interestingly, A was the only amino acid found at CpcA43 at eight stations, and conversely G was the only amino acid at two stations. Similarly, S was the only amino acid observed at CpcB42 at one station, and N at 32 stations (**figure 39**). The shift from one amino acid to the other seems to occur between 5-10°C and 25°C, with only CpcB42N and mostly CpcA43A above this temperature. CpcA43V seems much more limited in frequency and temperature range than CpcA43G and CpcA43A, in agreement with the limited distribution of PT 3f in the environment (see Chapter II). The observed shifts in amino acids are well described by logistic curves ( $R^2$  between 0.78 and 0.86). The midpoint for CpcA43 (16.8°C) is slightly higher than for CpcB42 (15.4°C), indicating that populations present in locations within this temperature range might possess the “hot” variant CpcB42N and the “cold” variant CpcA43G.

Other positions on CpcA and CpcB had substitutions with frequencies varying with temperature, and might represent adaptation to temperature. However, correlations were not as good as for CpcA43 and CpcB42, notably because of the low number of reads retrieved from cold stations (even though the *Tara* Polar Circle dataset was included), e.g. for substitutions at CpcB21 and CpcA49 that have a midpoint below 15°C. It is important to note that, in this analysis, reads have to entirely cover a specific codon, and not just map onto the genetic marker used. This greatly reduces the number of reads available for analysis, and in our case reaches the limits of the sequencing effort. Finding such positions on phycoerythrin-I or -II (PE-I/II) is even more challenging, as PE subunits present different alleles for the different pigment types (see part II of this chapter and Chapter II). Thus, reads have to be sorted according to the allele they correspond to, and substitutions can only be looked for within these different alleles, further dividing the number of reads available for analysis.

Overall, our results based on natural *Synechococcus* populations nicely complement Pittera and coworkers’ comparative analyses based on isolated strains (Pittera *et al.*, 2016) and suggest that different combinations of substitutions at CpcA43 and CpcB42 could provide progressive adaptation to different thermal niches. Finding other temperature-dependent substitutions in phycobiliproteins, although possibly more speculative than for CpcA43 and CpcB42, could guide future biochemical studies on phycobilisome stability. An interesting additional analysis could be to test for signal of adaptive selection at individual protein sites and in a phylogeny-aware maximum likelihood framework and implemented for example in the PAML/CodeML software, as proposed by (Yang and Nielsen, 2002). This has recently been used to relate functional differences in the temperature optima of the photosynthetic apparatus of different thermophilic *Synechococcus* to specific adaptive protein changes (Pedersen and Miller,

2017). This approach could complement the *in situ* variation observed here by providing additional arguments for the adaptive value of the different substitutions.



**Figure 39: Amino acids observed at CpcA43 (top) and CpcB42 (bottom) in phycocyanin alpha- and beta-subunit sequences retrieved from the Tara Oceans dataset.** Each point corresponds to the proportion of a given amino acid in a metagenomic sample. Sigmoidal fit and associated  $R^2$  are shown.

The intra-allelic level of variation such as demonstrated here is currently rarely taken into account in large-scale (meta)genomic studies, yet reveals an ecologically-relevant additional layer of complexity to the gene presence/absence observations (see Chapter III for an example of allelic variation affecting the light-harvesting phenotype and in turn the ecology of marine *Synechococcus*). In a recent preprint, Delmont and co-workers conducted a systematic study of the single-nucleotide variations and amino-acid substitutions observed in a globally abundant

population of *Pelagibacteraceae* (Delmont *et al.*, 2017). This new approach allows one to estimate the number of synonymous and non-synonymous nucleotide substitutions in a wild population, which informs about the selection processes and strength in an ecologically-relevant context. Their analysis revealed that about 20% of the amino-acid substitutions observed in core genes corresponded to the replacement of one amino acid by another with very different biochemical properties (e.g. replacement of a positively charged amino acid with a negatively charged one), which they interpreted as indicative of functional diversification (Delmont *et al.*, 2017). Based on the clustering of metagenomes according to the observed substitution profiles, they further suggest that temperature is the major driver of functional diversification of *Pelagibacteraceae* proteins. This approach, which is diametrically opposed to that used in our short analysis (analyzing amino acid variants for the whole proteome to find global patterns *vs.* testing a hypothesis about the distribution in the environment of a particular substitution of interest), is only possible for organisms that are naturally abundant and with a well characterized genetic diversity. Indeed, low abundance generally translates into a low number of reads in metagenomes and thus in a lack of statistical power, whereas low sampling of the genetic diversity of the organism of interest implies divergent or unrepresented alleles that could potentially impede downstream analyses. However, the current stringency of these two criteria will decrease with the continued increase in the number and depth of (meta)genomic data. In this context, the ubiquitous marine picocyanobacteria *Synechococcus* and *Prochlorococcus* constitute ideal candidates for systematic analyses of allelic variation in relation to their environment, including temperature, but also nutrients and light, and such analyses will certainly provide novel insights into the molecular mechanisms of niche adaptation in these ecologically relevant microorganisms.



# CHAPTER II

---

Distribution of  
*Synechococcus*  
pigment types

## I. Context of the work

A key feature of *Synechococcus* pigmentation is that the evolutionary history of this phenotypic trait is incongruent with that of *Synechococcus* as a species. If the previous chapter in part focused on understanding the possible mechanisms explaining this discrepancy, this also impacts our understanding of the ecology of *Synechococcus* and its pigment types. Indeed, a practical consequence is that one cannot infer the pigment content of a strain by knowing the clade it belongs to. Thus, the distribution in the environment of the various pigment types has long been studied based on spectral properties of bulk water (Lantoine and Neveux, 1997; Wood *et al.*, 1998; Hoge *et al.*, 1998; Campbell *et al.*, 1998; Taylor *et al.*, 2013; Yona *et al.*, 2014), or of picocyanobacterial cells (Olson *et al.*, 1990; Sherry and Wood, 2001; Campbell, 1996; Chung *et al.*, 2015). These are useful and cost-effective approaches, but they cannot differentiate all *Synechococcus* pigment types, as strains capable of chromatic acclimation are indistinguishable from 3c strains when grown under BL and from 3a strains when grown under GL. Moreover, bulk approaches also measure fluorescence from other PEB and PUB containing organisms such as *Trichodesmium* and *Crocospaera*. In the last decades, molecular-based approaches have been increasingly employed, using the genes encoding PC (*cpcBA*; Haverkamp *et al.*, 2008a; Larsson *et al.*, 2014; Liu *et al.*, 2014; Chung *et al.*, 2015; Díez *et al.*, 2016) or PE-I (*cpeBA*; Xia *et al.*, 2017b, 2017c, 2017a) subunits as markers. These provided new insights into the distribution of some but not all pigment types. In particular, these markers do not differentiate 3c from 3dB populations (see previous chapter). An original targeted metagenomic method has been developed and successfully used by Humily and co-workers, and consists in cloning whole PBS region from flow-cytometry sorted cells into fosmids (Humily *et al.*, 2014). This approach allows comparing results obtained from pyrosequencing of *cpcBA* and *cpeBA* with the gene order and complement of sequenced regions, and is of particular interest for discovering novel or divergent yet potentially complete PBS rod genomic regions (see previous chapter; Humily *et al.*, 2014). However, it is particularly time-intensive for both the production of the fosmid library and its screening, and is thus not suitable for large-scale studies. Additionally, it will probably be quickly outdated by single-cell approaches as recently used by Kashtan and co-workers for *Prochlorococcus* (Kashtan *et al.*, 2014, 2017).

This chapter aimed at providing the first analysis of all marine *Synechococcus* pigment types at a global scale, and to use the information contained in the distribution patterns to gain new insights into the ecology of *Synechococcus* pigment types. I personally designed and conducted all the bioinformatic analyses presented in this chapter, under the supervision of Drs. Laurence Garczarek and Frédéric Partensky.

## **II. Light color acclimation: a key process in the global ocean distribution of *Synechococcus* cyanobacteria**

An important step not presented in the following manuscript was to find suitable marker gene(s), i.e. one or several gene(s) that present a phylogeny consistent with pigment types and that would allow quantifying all pigment types in a metagenomic dataset. To this end, I designed and implemented an algorithm to find in a set of phylogenetic trees, those in which user-defined groups of sequences form monophyletic groups. This allowed me to search our in-house database Cyanorak v2 (<http://application.sb-roscoff.fr/cyanorak/>), which gathers information about 54 public and private genomes of marine *Synechococcus*. In particular, genes predicted in these 54 genomes are grouped into “clusters of orthologs” based on their similarity between strains, and a phylogenetic tree is automatically inferred for each cluster. By extensively searching this set of trees, I have been able to show that no single marker gene can predict all *Synechococcus* pigment types. Indeed, *cpcB* and *cpcA* are the only possible markers for delineating pigment types 1, 2a, 2b and 3; but cannot differentiate subtypes within pigment type 3. Moreover, no single marker can fully resolve all pigment type 3 subtypes (3a, 3c, 3f, 3dA and 3dB) since, in the best case, pigment type 3c and 3dB are mixed in a monophyletic group (see figure 1 of the paper presented in the following pages). Genes able to resolve 3a, 3f, 3dA and (3c + 3dB) as monophyletic groups are *mpeB*, *mpeA* (encoding PE-II subunits), *mpeD* (encoding a PE-II linker protein), and *mpeY*, *cpeY* and *cpeZ*, encoding (putative) phycobilin lyases. In this set of six genes, we chose to use the *mpeBA* operon, as i) simulation studies confirmed its high resolution power, ii) this marker is paralogous and relatively similar to *cpcBA*, and we supposed this would reduce potential biases between these two markers, and iii) some supplementary sequences had been published for this gene, thus enriching our reference database (Everroad and Wood 2012). Finally, this analysis showed that no gene present in both 3c and 3dB strains could clearly separate these pigment types into two monophyletic clusters, and we chose to use the 3dB-specific gene *mpeW*, which is present in the CA4-B genomic island, as it allowed to get PT 3c counts by subtraction.

This analysis demonstrated that the minimal set of genes necessary to predict all *Synechococcus* pigment type comprises the three different markers, *cpcBA*, *mpeBA* and *mpeW*. The use of these three genetic markers to recruit metagenomic reads from the extensive *Tara* Oceans dataset allowed us to quantify the abundance and distribution of the different pigment types in a global dataset, and further insights were obtained through the comparison of the abundance profiles of additional marker genes. The work presented in the following pages has been accepted for publication in the Proceedings of the National Academy of Science of the United States of America with minor changes, and is currently under revision.

# Light color acclimation: a key process in the global ocean distribution of *Synechococcus* cyanobacteria

Théophile Grébert<sup>a</sup>, Hugo Doré<sup>a</sup>, Frédéric Partensky<sup>a</sup>, Gregory K. Farrant<sup>a,1</sup>, Emmanuel Boss<sup>b</sup>, Marc Picheral<sup>c</sup>, Lionel Guidi<sup>c</sup>, Stéphane Pesant<sup>d,e</sup>, Patrick Wincker<sup>f</sup>, Silvia G. Acinas<sup>g</sup>, David M. Kehoe<sup>h</sup> and Laurence Garczarek<sup>a,2</sup>

<sup>a</sup>Sorbonne Universités-Université Paris 06 & Centre National de la Recherche Scientifique (CNRS), UMR 7144, Marine Phototrophic Prokaryotes Team, Station Biologique, CS 90074, 29688 Roscoff cedex, France; <sup>b</sup>Maine In-situ Sound and Color Lab, University of Maine, Orono, ME 04469; <sup>c</sup>Sorbonne Universités-Université Paris 06 & CNRS, UMR 7093, Observatoire océanologique, 06230 Villefranche-sur-mer, France; <sup>d</sup>PANGAEA, Data Publisher for Earth and Environmental Science, University of Bremen, Bremen, Germany; <sup>e</sup>MARUM, Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany; <sup>f</sup>Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), Institut de Génomique, Genoscope, 91057 Evry, France; <sup>g</sup>Department of Marine Biology and Oceanography, Institute of Marine Sciences (ICM), Consejo Superior de Investigaciones Científicas (CSIC), Barcelona ES-08003, Spain; <sup>h</sup>Department of Biology, Indiana University, Bloomington, IN 47405 <sup>1</sup>Present address: Food Safety, Environment, and Genetics, Matis Ltd., 113 Reykjavik, Iceland. <sup>2</sup>To whom correspondence should be addressed. Email: laurence.garczarek@sb-roscoff.fr.

Submitted to Proceedings of the National Academy of Sciences of the United States of America

**Marine *Synechococcus* cyanobacteria are major contributors to global oceanic primary production and exhibit a unique diversity of photosynthetic pigments, allowing them to exploit a wide range of light niches. However, the relationship between pigment content and niche partitioning has remained largely undetermined due to the lack of a single-genetic marker resolving all pigment types (PT). Here, we developed and employed a novel and robust method based on three distinct marker genes to estimate the relative abundance of all known *Synechococcus* PTs from metagenomes. Analysis of the Tara Oceans dataset allowed us, for the first time, to reveal the global distribution of *Synechococcus* PTs and to define their environmental niches. Green-light specialists (PT 3a) dominated in warm, green equatorial waters, whereas blue-light specialists (PT 3c) were particularly abundant in oligotrophic areas. Type IV chromatic acclimators (CA4-A/B), which are able to dynamically modify their light absorption properties to maximally absorb green or blue light, were unexpectedly the most abundant PT in our dataset and predominated at depth and high latitudes. We also identified local populations in which CA4 might be inactive due to the lack of specific CA4 genes, notably in warm high nutrient low chlorophyll areas. Major ecotypes within clades I-IV and CRD1 were preferentially associated with a particular PT, while others exhibited a wide range of PTs. Altogether, this study provides unprecedented insights into the ecology of *Synechococcus* PTs and highlights the complex interactions between vertical phylogeny, pigmentation and environmental parameters that shape *Synechococcus* community structure and evolution.**

marine cyanobacteria | metagenomics | light quality | phycobilisome  
| Tara Oceans

## Introduction

Marine *Synechococcus* is the second most abundant phytoplankton group in the world's oceans and constitutes a major contributor to global primary production and carbon cycling (1, 2). This genus displays a wide genetic diversity and several studies have shown that among the ~20 clades defined based on various genetic markers, five (clades I-IV and CRD1) predominate *in situ* and can be broadly associated with distinct sets of physicochemical parameters (3–5). In a recent study, we further defined Ecologically Significant Taxonomic Units (ESTUs), i.e. organisms belonging to the same clade and co-occurring in the field, and highlighted that the three main parameters affecting the *in situ* distribution of these ESTUs were temperature and availability of iron and phosphorus (6). Yet, marine *Synechococcus* also display a wide pigment diversity, suggesting that light could also influence their ecological distribution, both qualitatively and quantitatively (7, 8).

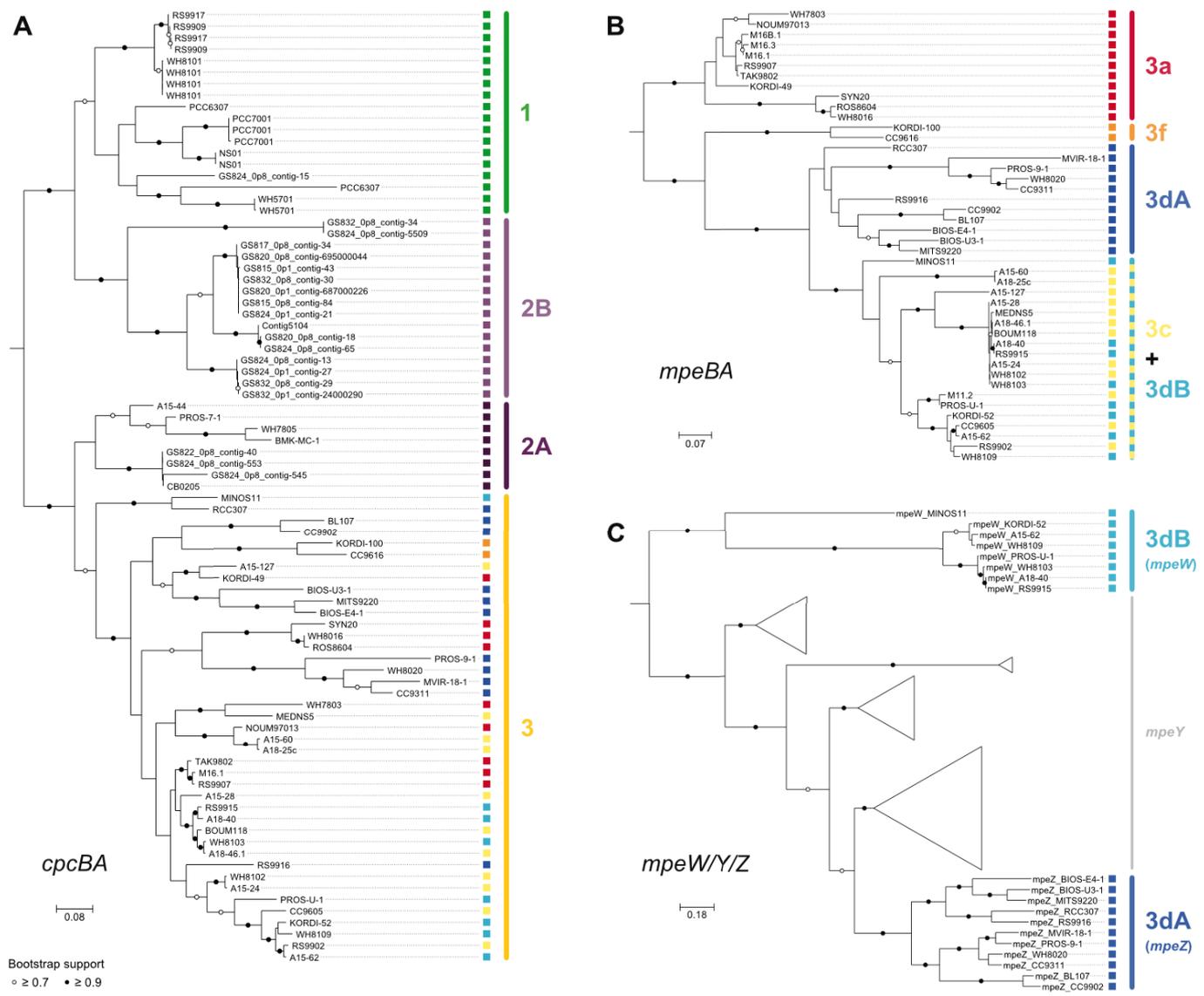
This diversity comes from differences in the composition of their main light-harvesting antennae, called phycobilisomes (PBS, 7–9). These water-soluble macromolecular complexes consist of a central core anchoring at least six radiating rods made of several distinct phycobiliproteins, i.e. proteins to which specific enzymes (phycobilin lyases) covalently attach chromophores called phycobilins (7, 10). Although the PBS core is conserved in all marine *Synechococcus*, rods have a very variable composition, and three main pigment types (PTs) are usually distinguished (Figs. S1, 7, 11). In PT 1, PBS rods are solely made of phyco-cyanin (PC, encoded by the *cpcBA* operon) and bear the red-light absorbing phycocyanobilin (PCB;  $A_{max} = 620$  nm) as the sole chromophore. In PT 2, rods are made of PC and phycoerythrin I (PE-I, encoded by *cpeBA*) and attach both PCB and the green-light (GL) absorbing phycoerythrobilin (PEB;  $A_{max} = 550$  nm). All other marine *Synechococcus* belong to PT 3 and have rods made of PC, PE-I and PE-II (encoded by *mpeBA*) that bind PCB, PEB and the blue-light (BL) absorbing phycourobilin (PUB;  $A_{max} = 495$  nm; Fig. S1). Several subtypes can be defined within PT3, based on the fluorescence excitation ratio at 495 nm and 545 nm (hereafter  $Ex_{495:545}$ , Fig. S1), a proxy for the PUB:PEB ratio. This ratio is low ( $Ex_{495:545} < 0.6$ ) in subtype 3a (GL specialists), intermediate in subtype 3b ( $0.6 \leq Ex_{495:545} < 1.6$ ) and high ( $Ex_{495:545}$

## Significance

Understanding the functional diversity of specific microbial groups on a global scale is critical yet poorly developed. By combining the considerable knowledge accumulated through recent years on the molecular bases of photosynthetic pigment diversity in marine *Synechococcus*, a major phytoplanktonic organism, with the wealth of metagenomic data provided by the Tara Oceans expedition, we have been able to reliably quantify all known pigment types along the transect and provide the first global distribution map. Unexpectedly, cells able to dynamically change their pigment content to match the ambient light color were ubiquitous and predominated in many environments. Altogether, our results enlighten the role of adaptation to light quality on niche partitioning in a key primary producer.

## Reserved for Publication Footnotes

137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204



**Fig. 1.** : Maximum likelihood phylogenetic trees of (A) *cpcBA* operon, (B) *mpeBA* operon and (C) the *mpeW/Y/Z* gene family. The *cpcBA* tree includes both strains with characterized pigment type (PT) and environmental sequences (prefixed with GS) assembled from metagenomes of the Baltic Sea (38). Circles at nodes indicate bootstrap support (black: > 90 %; white: > 70 %). Note that for PT 2B clade, only environmental sequences are available. The PT associated with each sequence is indicated as a colored square. The scale bar represents the number of substitutions per nucleotide position.

> 1.6) in subtype 3c (BL specialists; 7, 11). Additionally, some strains of subtype 3d are able to change their PUB:PEB ratio depending on ambient light color, a process called type IV chromatic acclimation (hereafter CA4), allowing them to maximally absorb BL or GL (11–14). Comparative genomic analyses further showed that genes involved in the synthesis and regulation of PBS rods are gathered into a dedicated genomic region, the content and organization of which correspond to the different PTs (7). Similarly, chromatic acclimation has been correlated with the presence of a small specific genomic island (CA4 genomic island) that exists in two distinct configurations (CA4-A and CA4-B, 11). Both contain two regulators (*fciA* and *fciB*) and a phycobilin lyase (*mpeZ* in CA4-A or *mpeW* in CA4-B), thus defining two distinct CA4 genotypes: 3dA and 3dB (11, 14, 15). Finally, some strains possess a complete or partial CA4 genomic island but are not able to perform CA4, displaying a fixed Ex<sub>495:545</sub> corresponding to 3a, 3b or 3c phenotypes (11).

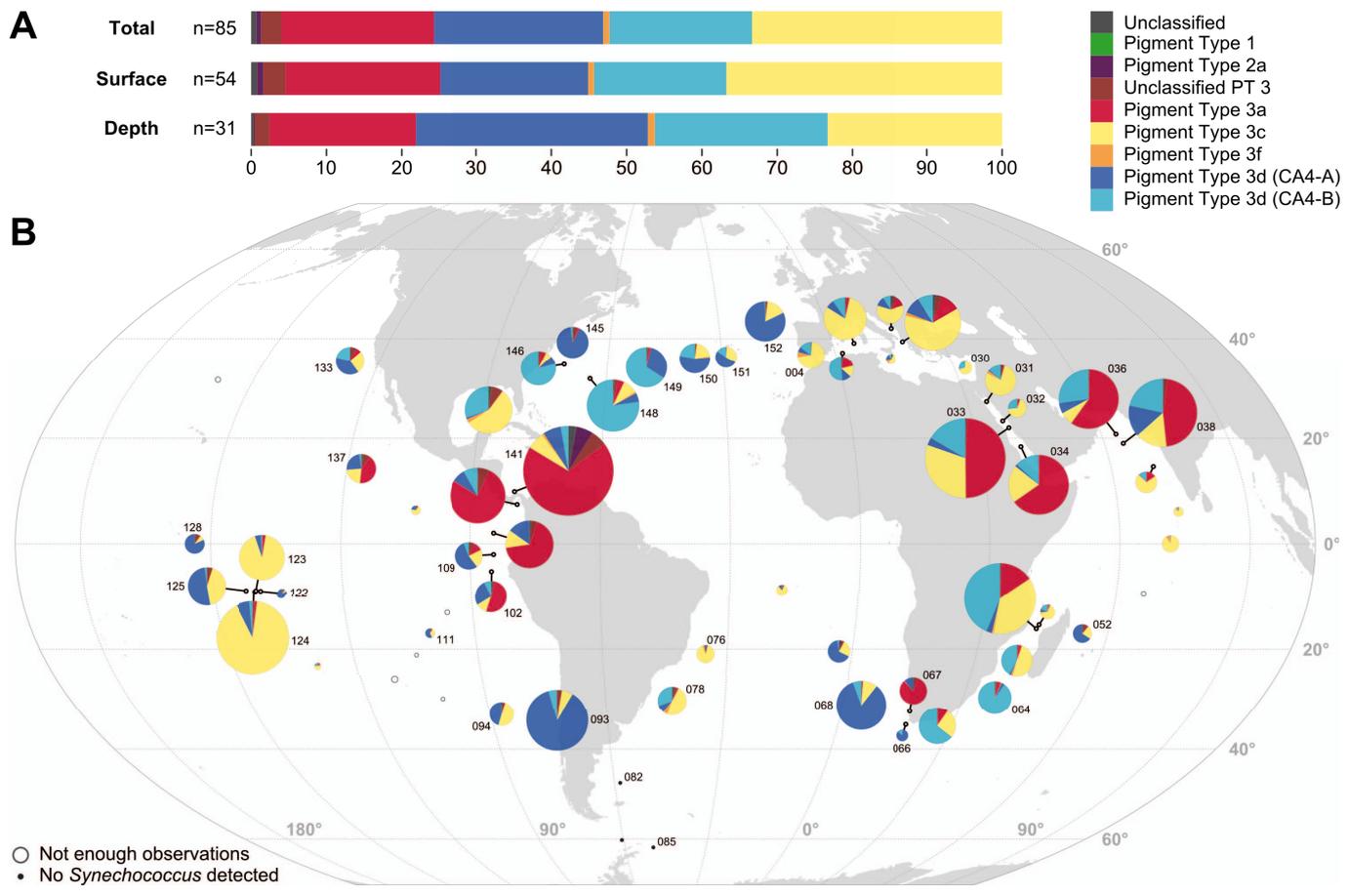
As there is no correspondence between pigmentation and core genome phylogeny (7, 16, 17), deciphering the relative abun-

dance and niche partitioning of *Synechococcus* PTs in the environment requires specific approaches. In the past 30 years, studies have been based either on i) proxies of the PUB:PEB ratio as assessed by flow cytometry (18–20), fluorescence excitation spectra (21–27), epifluorescence microscopy (28), or ii) phylogenetic analyses of *cpcBA* or *cpeBA* (17, 29–34). However, analyses based on optical properties could only describe the distribution of high- and low-PUB populations without being able to differentiate GL (3a) or BL (3c) specialists from CA4 cells (3d) acclimated to GL or BL, while genetic analysis solely based on *cpcBA* and/or *cpeBA* could not differentiate all PTs. For instance, only two studies have reported CA4 populations *in situ* either in the western English Channel (17) or in sub-polar waters of the western Pacific Ocean (29) but none of them were able to differentiate CA4-B from high PUB (i.e. 3c) populations. As a consequence, the global relative abundance of CA4 as well as the link between genetic and pigment diversity has remained largely unclear.

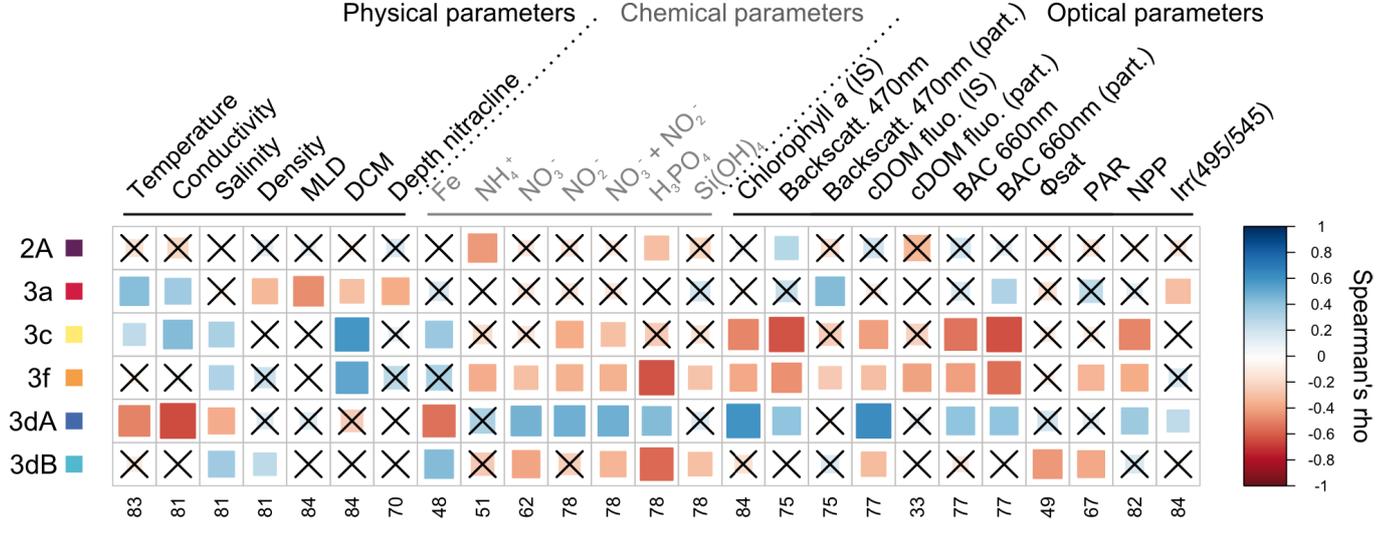
Here, we analyzed 109 metagenomic samples collected from all major oceanic basins during the 2.5-yr *Tara* Oceans (2009-

205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272

273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340



**Fig. 2.** Distribution of *Synechococcus* pigment types (PT). (A) Relative abundance of each PT in the whole dataset (Total), in surface and at the DCM (Deep Chlorophyll Maximum). (B) Map showing the global distribution of all *Synechococcus* PTs in surface waters along the Tara Oceans transect. Diameters of pies are proportional to the number of *cpcBA* reads normalized by the sequencing effort. Stations with less than 30 *cpcBA* or *mpeBA* reads are indicated by open circles and those with no *cpcBA* reads by black dots. Numbers next to pies correspond to Tara Oceans stations.



**Fig. 3.** Correlation analysis between *Synechococcus* pigment types (PT) and environmental parameters measured along the Tara Oceans transect for all sampled depths. The scale shows the degree of correlation (blue) or anti-correlation (red) between two variables. Non-significant correlations (adjusted *P* value > 0.05) are indicated by crosses. Number of observations for each environmental parameter is indicated at the bottom. Abbreviations: MLD, mixed layer depth; DCM, deep chlorophyll maximum; IS, *in situ*; Backscatt., backscattering; part., particulate; cDOM fluo, colored dissolved organic matter fluorescence; BAC, beam attenuation coefficient; Φ<sub>sat</sub>, satellite-based non-photochemical quenching (NPQ)-corrected quantum yield of fluorescence (proxy for iron limitation; 6); PAR, photosynthetically active radiation; NPP, net primary production; Irr<sub>495:545</sub>, ratio of downwelling irradiance at 495 nm and 545 nm.

2011) expedition (35) using an original bioinformatic pipeline combining a *mi*Tag approach (6, 36) to recruit single reads from multiple PBS gene markers, and placement of these reads in reference trees to assign them to a given PT. This pipeline has



545 global view of how a major photosynthetic organism adapts to  
546 natural light color gradients in the ocean.

## 547 Results

### 548 A novel, robust approach for estimating pigment types abundance from metagenomes

549 We developed a multi-marker approach combining phylogenetic information retrieved from three different genes or operons  
550 (*cpcBA*, *mpeBA* and *mpeW*, Fig. 1 and Datasets 1-2) to overcome  
551 the issue of fully resolving the whole range of PTs. While *cpcBA*  
552 discriminated PT 1, 2 and 3 (Fig. 1A), only the *mpeBA* operon,  
553 a PT3 specific marker, was able to distinguish the different PT 3  
554 subtypes (Fig. 1B), though as for *cpeBA* it could not differentiate  
555 PT 3dB (CA4-B) from BL specialists (i.e. PT 3c; 11, 29). The  
556 *mpeW* marker was thus selected to specifically target PT 3dB  
557 and, by subtraction, enumerate PT 3c (Fig. 1C). Using the *cpcBA*  
558 marker, members of PT 2 were split into two clear-cut clusters,  
559 2A and 2B (Fig. 1A), the latter corresponding to a purely environmental  
560 PT identified from assembled metagenomes of the Baltic  
561 Sea (38). Strains KORDI-100 and CC9616 also clustered apart  
562 from other strains in the *mpeBA* phylogeny, suggesting that they  
563 have a divergent evolutionary history from other PT 3 members  
564 (Fig. 1B). This is supported by the diverged gene content and  
565 order of their PBS rod genomic region and these strains were  
566 recently referred to as PT 3f, even though they have a similar  
567 phenotype as PT 3c (Ex<sub>495:545</sub> ratio  $\geq 1.6$ ; 30). To investigate the  
568 phylogenetic resolution of small fragments of these three markers,  
569 simulated reads (150 bp as compared to 164 bp in average  
570 for *Tara* Oceans cleaned/merged reads) were generated from all  
571 sequences in our reference databases and assigned to a PT using  
572 our bioinformatic pipeline. Inferred and known PTs were then  
573 compared. The percentage of simulated reads assigned to the  
574 correct PT was between 93.2% and 97.0% for all three markers,  
575 with less than 2.1-5.6% of reads that could not be classified and  
576 an error-rate below 2%, showing that all three markers display  
577 a sufficient resolution to reliably assign the different PTs (Fig. S2B,  
578 D and F).

579 To ensure that the different markers could be quantitatively  
580 compared in a real dataset, we examined the correlations between  
581 estimates of PT abundances using the different markers in the  
582 109 metagenomes analyzed in this study. Total *cpcBA* counts  
583 were highly correlated ( $R^2=0.994$ ,  $n=109$ , Fig. S3A) with total  
584 *Synechococcus* counts obtained with the *petB* gene, which was  
585 previously used to study the phylogeography of marine picocyanobacteria (6),  
586 and the correlation slope was not significantly different from 1 (slope: 1.040;  
587 Wilcoxon's paired difference test  $p$ -value=0.356). *cpcBA* is thus as good as *petB*  
588 at capturing the total population of *Synechococcus* reads. Moreover, counts  
589 of *cpcBA* reads assigned to PT 3 and total *mpeBA* counts (specific for PT 3)  
590 were also strongly correlated ( $R^2=0.996$ ,  $n=109$ , Fig. S3B),  
591 and not skewed from 1 (slope of 0.991, Wilcoxon's  $p$ -value=0.607),  
592 indicating that *mpeBA* and *cpcBA* counts can be directly compared.  
593 Although no redundant information for PT 3dB is available with the three  
594 selected markers, another marker targeting 3dB (*ficiAB*) was tested and produced  
595 results similar to *mpeW* (Fig. S3C). These results demonstrate that our  
596 multi-marker approach can be used to reliably and quantitatively infer  
597 the different *Synechococcus* PTs from short metagenomic reads, with PT 1,  
598 2A, 2B abundances being assessed by *cpcBA* normalized counts, PT 3a,  
599 3f and 3dA by *mpeBA* normalized counts, PT 3dB by *mpeW* normalized  
600 counts and PT 3c by the difference between *mpeBA* normalized counts for  
601 3c + 3dB and *mpeW* normalized counts. We thus used this approach on  
602 the *Tara* Oceans metagenomes, generated from 109 samples collected at  
603 65 stations located in the major oceanic basins (Fig. 2).

### 604 CA4 populations are widespread and predominate at depth and high latitudes

605 The latitudinal distribution of *Synechococcus* inferred from  
606 *cpcBA* counts is globally consistent with previous studies (2, 6,  
607 39), with *Synechococcus* being present in most oceanic waters,  
608 but quasi absent ( $< 20$  *cpcBA* counts) beyond 60° S (Southern  
609 Ocean stations TARA\_082 to TARA\_085; Fig. 2B). Overall, the  
610 number of *cpcBA* recruited reads per station was between 0 and  
611 8,151 ( $n=63$ , median: 449, mean: 924, sd: 1478) for surface and  
612 0 and 3,200 ( $n=46$ , median:170, mean: 446, sd: 664) for deep  
613 chlorophyll maximum (DCM) samples, respectively. Stations with  
614 less than 30 *cpcBA* reads were excluded from further analysis.

615 As expected from *Tara* cruise sampling being mostly performed  
616 in oceanic waters, PT 1 and 2, both known to be mostly abundant  
617 in coastal waters (Fig. 2A-B; 29, 38, 40, 41; see also refs in Dataset 3),  
618 were almost absent from this dataset (total of 15 and 513 *cpcBA* reads,  
619 respectively). While PT 2A was mostly found at the surface at one  
620 station off Panama (TARA\_141, 417 out of 6,637 reads at this station,  
621 Fig. 2B), PT 2B was virtually absent (total of 3 *cpcBA* reads) from  
622 our dataset and might thus be confined to the Baltic Sea (38). This  
623 low abundance of PT 1 and 2 precluded the correlation analysis  
624 between their distribution and physico-chemical parameters. PT 3 was  
625 by far the most abundant along the *Tara* Oceans transect, accounting  
626 for  $99.1 \pm 1.4\%$  (mean  $\pm$  sd) of *cpcBA* reads at stations with  $\geq 30$   
627 *cpcBA* read counts. Interestingly, several PT 3 subtypes often co-  
628 occurred at a given station.

629 PT 3a (green light specialists) totaled 20.3% of read counts, with  
630 similar abundance between surface (20.5%) and DCM (19.4%)  
631 samples, and was particularly abundant in intertropical oceanic  
632 borders and regional seas, including the Red Sea, the Arabian Sea  
633 and the Panama/Gulf of Mexico area (Fig. 2B). Correlation analyses  
634 show that this PT is consistently associated with high temperatures  
635 but also with greenish (as estimated from a low blue to green  
636 downwelling irradiance ratio: Irr<sub>495:545</sub>), particle-rich waters  
637 (high particle backscattering at 470 nm and beam attenuation  
638 coefficient at 660 nm; Fig. 3). Still, in contrast with previous  
639 studies that reported the distribution of low-PUB populations (18,  
640 21, 23–26), this PT does not seem to be restricted to coastal  
641 waters, explaining its absence of correlation with chlorophyll  
642 concentration and colored dissolved organic matter (cDOM).

643 Blue light specialists (PT 3c) appear to be globally widespread,  
644 with the exception of high latitude North Atlantic waters, and  
645 accounted for 33.4% of reads, with a higher relative abundance  
646 at the surface (36.8%) than at the DCM (23.3%, Fig. 2A). This  
647 PT is dominant in transparent, oligotrophic, iron-replete areas  
648 such as the Mediterranean Sea as well as South Atlantic and Indian  
649 Ocean gyres (Figs. 2B and 4C). In the South Pacific, PT 3c was  
650 also found to be predominant in the Marquesas Islands area  
651 (TARA\_123 and 124), where the coast proximity induced a local  
652 iron enrichment (6). Consistently, PT 3c was found to be  
653 positively associated with iron concentration, high temperature  
654 and DCM depth and anti-correlated with chlorophyll fluorescence,  
655 nitrogen concentrations, net primary production (NPP) as well  
656 as other related optical parameters, such as backscattering at  
657 470 nm and beam attenuation coefficient at 660 nm (Fig. 3).  
658 Despite its rarity, PT 3f seems to thrive in a similar environment,  
659 with the highest relative abundances in the Indian Ocean and  
660 Mediterranean Sea (Figs. 2B and 4C). Its occurrence in the latter  
661 area might explain its strong anti-correlation with phosphorus  
662 availability.

663 The most striking result of this study was the widespread occurrence  
664 of both CA4 types, 3dA and 3dB, which represented 22.6% and  
665 18.9% of reads respectively, and could locally account for up  
666 to 95% of the total *Synechococcus* population (Figs. 2, 4C and  
667 S4). In contrast to blue and green light specialists, both CA4  
668 types were proportionally less abundant at the surface (19.8%  
669 and 17.5%, for 3dA and 3dB, respectively) than at depth (30.9%  
670 and 38.0%, for 3dA and 3dB, respectively).

681 and 22.9%). Interestingly, PT 3dA and 3dB generally displayed  
682 complementary distributions along the Tara Oceans transect (Fig.  
683 2B). PT 3dA was predominant at high latitude in the northern  
684 hemisphere as well as in other vertically mixed waters such as  
685 in the Chilean upwelling (TARA\_093) or in the Agulhas current  
686 (TARA\_066 and 68; Fig 2B). Accordingly, PT 3dA distribution  
687 seems to be driven by low temperature, high nutrient and highly  
688 productive waters (high NPP, chlorophyll *a* and optical param-  
689 eters), a combination of physico-chemical parameters almost  
690 opposite to those observed for blue light specialists (PT 3c; Fig.  
691 3). In contrast, PT 3dB shares a number of characteristics with  
692 PT 3c, including the anti-correlation with nitrogen concentration  
693 and association with iron availability (as indicated by both a pos-  
694 itive correlation with [Fe] and negative correlation with the iron  
695 limitation proxy  $\Phi_{\text{sat}}$ ; Fig. 3), consistent with their widespread  
696 occurrence in iron replete oceanic areas. Also noteworthy, PT  
697 3dB was one of the sole PT (with 3f) to be associated with low  
698 photosynthetically available radiation (PAR).

#### 700 Niche partitioning of *Synechococcus* populations rely on a 701 subtle combination of ESTU and PT niches

702 We previously showed that temperature, iron and phospho-  
703 rus availability constituted major factors influencing the divers-  
704 ification and niche partitioning of *Synechococcus* ESTUs (i.e.,  
705 genetically related subgroups within clades that co-occur in the  
706 field; 6). Yet, these results cannot be extended to PTs since the  
707 pigment content does not follow the vertical phylogeny (7). In  
708 order to decipher the respective roles of genetic and pigment  
709 diversity in *Synechococcus* community structure, we examined the  
710 relationships between ESTUs and PTs *in situ* abundances through  
711 correlation and NMDS analyses (Fig. 4A-B) and compared their  
712 respective distributions (Figs. 4C and S4).

713 Interestingly, all PTs are either preferentially associated with  
714 or excluded from a subset of ESTUs. PT 2A is found at low  
715 abundance at a few stations along the Tara Oceans transect and,  
716 when present, it is seemingly associated with the rare ESTU  
717 5.3B (Fig. 4A), an unusual PT/genotype combination so far only  
718 observed in metagenomes from freshwater reservoirs (42). PT 3a  
719 is associated with ESTUs EnvBC (occurring in low iron areas)  
720 and IIA, the major ESTU in the global ocean (Fig. 4A), a result  
721 consistent with NMDS analysis, which shows that PT 3a is found  
722 in assemblages dominated by these two ESTUs (indicated by red  
723 and grey backgrounds in Fig. 4B), as well as with independent  
724 observations on cultured strains (Dataset 3). PT 3c is associated  
725 with ESTU IIIA (the dominant ESTU in P-depleted areas), as  
726 observed on many isolates (Dataset 3), and is also linked, like  
727 PT 3f, with ESTUs IIIB and WPC1A, both present at lower  
728 abundance than IIIA in P-poor waters (Fig. 4A). PT 3f is also  
729 associated with the newly described and low-abundance ESTU  
730 XXA (previously EnvC; Fig. S5; 4, 6). Both PT 3f and ESTU  
731 XXA were rare in our dataset but systematically co-occurred, in  
732 agreement with the fact that the only culture representative of the  
733 latter clade belongs to PT 3f (Dataset 3).

734 PT 3dA appears to be associated with all ESTUs from clades  
735 CRD1 (specific to iron-depleted areas) as well as with those  
736 representative of coastal and cold waters (IA, IVA, IVC), but  
737 is strikingly anti-correlated with most other major ESTUs (IIA,  
738 IIIA and -B, WPC1A and 5.3B; Fig. 4A). This pattern is opposite  
739 to PT 3dB that is preferentially found associated with ESTU  
740 IIA, IIIB and 5.3A, but not in CRD1A or -C (Fig. 4A). Thus,  
741 it seems that the two types of CA4 are found in distinct and  
742 complementary sets of ESTUs. Interestingly, our analysis might  
743 suggest the occurrence of additional PTs not isolated so far, since  
744 a number of reads (0.7% and 2.7% of *cpcBA* and *mpeBA* counts,  
745 respectively, Fig. 2A) could not be assigned to any known PTs.  
746 For instance, while most CRD1C seem preferentially associated  
747 with PT 3dA, a fraction of the population could only be assigned  
748 at the PT 3 level (Fig. 4A). Similarly, a number of reads could not

749 be assigned to any known PT in stations rich in ESTU 5.3A and  
750 XXA, although one cannot exclude that this observation might  
751 be due to a low number of representative strains, and thus PT  
752 reference sequences, for these ESTUs.

753 The preferred association of PTs with specific ESTUs is also  
754 well illustrated by some concomitant shifts of PTs and ESTU  
755 assemblages. For instance, in the wintertime North Atlantic  
756 Ocean, the shift from 3dB-dominated stations on the western side  
757 (TARA\_142 and TARA\_146-149) to 3dA-dominated stations near  
758 European coasts (TARA\_150 to 152) and North of Gulf stream  
759 (TARA\_145) is probably related to the shift in ESTU assemblages  
760 occurring along this transect, with ESTU IIA being gradually re-  
761 placed by ESTU IVA (Fig. 4C; see also 6). Similarly, the takeover  
762 of CRD1C by IIA in the Marquesas Island area (TARA\_123  
763 to 125), which is iron-enriched with regard to surrounding high  
764 nutrient low chlorophyll (HNLC) waters (TARA\_122 and 128),  
765 perfectly matched the corresponding replacement of PT 3dA by  
766 3c. However, in several other cases, PT shifts were not associated  
767 with a concomitant ESTU shift or vice versa. One of the most  
768 striking examples of these dissociations is the transect from the  
769 Mediterranean Sea to the Indian Ocean, where the entry in the  
770 northern Red Sea through the Suez Canal triggered a sharp shift  
771 from a IIIA to a IIA-dominated community (TARA\_030 and  
772 031) which was not accompanied by any obvious change in PT.  
773 Conversely, a sharp rise in the relative abundance of PT 3a was  
774 observed in the southern Red Sea/northeastern Indian Ocean  
775 (TARA\_033 to 038) without changes in the large dominance  
776 of ESTU IIA. Altogether, this strongly suggests that a subtle  
777 combination of ESTU and PTs respective niche occupancy is  
778 responsible for the observed niche partitioning of *Synechococcus*  
779 populations.

#### 780 Deficient chromatic acclimators are dominant in HNLC ar- 781 eas

782 Although our results clearly indicate that CA4 cells represent  
783 a large proportion of the *Synechococcus* community in a wide  
784 range of ecological niches, this must be somewhat tempered by  
785 the fact that, in culture, about 30% of the strains possessing  
786 a CA4-A or B genomic island are not able to chromatically  
787 acclimate (Dataset 3; 11). Some of these natural mutants have  
788 an incomplete CA4 genomic island (Fig. S6K). For example,  
789 strains WH8016 (ESTU IA) and KORDI-49 (WPC1A) both lack  
790 the CA4-A specific lyase-isomerase *MpeZ*, an enzyme shown to  
791 bind a PUB molecule on PE-II (14), and display a green light  
792 specialist phenotype (PT 3a;  $EX_{495:545} = 0.4$ ) whatever the ambient  
793 light color (11). However, since they possess a PT 3a *mpeBA*  
794 allele, reads from field WH8016- or KORDI-49-like cells are  
795 adequately counted as PT 3a (Fig. S6K). Another CA4-deficient  
796 strain, BIOS-E4-1 (ESTU CRD1C), possesses *mpeZ* and a 3dA  
797 *mpeBA* allele but lacks the CA4 regulators *FciA* and *FciB* as  
798 well as the putative lyase *MpeY* and exhibits a fixed blue light  
799 specialist phenotype (PT 3c;  $EX_{495:545} = 1.7$ ; Fig. S6K; 11, 15). Thus,  
800 reads from such natural *Synechococcus* CA4-incapable mutants  
801 in the field are counted as 3dA using the *mpeBA* marker. Lastly,  
802 the strain MVIR-18-1 possesses a complete CA4-A island and a  
803 3dA *mpeBA* allele but lacks *mpeU*, a gene necessary for blue light  
804 acclimation (Fig. S6K; 43). While MVIR-18-1 displays a fixed  
805 green light phenotype, reads from such *Synechococcus* are also  
806 erroneously counted as 3dA.

807 To assess the significance of these genotypes in the field,  
808 we compared the normalized read counts obtained for 3dA  
809 with *mpeBA*, *fciAB*, *mpeZ*, *mpeU* and *mpeY* (Fig. S6A-J). Over-  
810 all this analysis revealed a high consistency between these dif-  
811 ferent markers ( $0.860 < R^2 < 0.986$ ), indicating that most *mpeZ*-  
812 containing populations also contained 3dA alleles for *fciAB*,  
813 *mpeY*, *mpeU* and *mpeBA* and are therefore likely able to per-  
814 form CA4. However, a number of stations, all located in HNLC  
815 areas (TARA\_094, 111 and 122 to 128 in the Pacific Ocean and  
816

TARA\_052 located northwest of Madagascar, Fig. 2B), displayed more than 10-fold higher *mpeBA*, *mpeU* and *mpeZ* counts than *fciAB* and *mpeY* counts (Fig. S6A, B, E, F, H, I). This indicates that a large proportion or even the whole population (TARA\_122 and 124) of 3dA in these HNLC areas is probably lacking the *FciA/B* regulators and *MpeY* and thus, like strain BIOS-E4-1 (Fig. S6K), might be stuck in the blue light specialist phenotype (PT 3c; 11). Conversely, station TARA\_067 exhibited consistently more than twice the *fciAB* and *mpeZ* counts compared to *mpeBA*, *mpeY* or *mpeU* (Fig. S6B-E, G, H) and was a clear outlier when comparing pigment type and clade composition (Fig. S7). This suggests that the proportion of PT 3dA might have been underestimated at this station, as a significant proportion of this population probably corresponds to PT 3a genotypes that have acquired a CA4-A island by lateral gene transfer, as is seemingly the case for strains WH8016 and KORDI-49. Finally, no station exhibited markedly lower *mpeU* counts compared to all other genes, indicating that the strain MVIR-18-1 genotype is probably rare in the oceans.

It must be noted that two out of the six sequenced CA4-B strains (WH8103 and WH8109) also have a deficient CA4 phenotype and display a constant, intermediate  $Ex_{495:545}$  ratio (0.7 and 1, respectively), despite any obvious PBS- or CA4-related gene deletion (11). Accordingly, the plot of 3dB normalized read counts obtained with *mpeW* vs. *fciAB* shows no clear outlier (Fig. S3C).

## Discussion

Marine *Synechococcus* display a large pigment diversity, with different PTs preferentially harvesting distinct regions of the light spectrum. Previous studies based on optical properties or on a single genetic marker could not differentiate all PTs (17, 29–31), and thus neither assess their respective realized environmental niches (37) nor the role of light quality on global *Synechococcus* distribution. Here, we showed that a *mi*Tag approach combining three genetic markers can be used to reliably predict all major PTs. Applied to the extensive Tara Oceans dataset, this original approach, which avoids PCR amplification and cloning biases, allowed us to describe for the first time the distribution of the different *Synechococcus* PTs at the global scale and to refine our understanding of their ecology.

PT 3 was found to be largely dominant over PT 1 and 2 along the Tara Oceans transect, and biogeography and correlation analyses with environmental parameters provided several novel and important insights concerning niche partitioning of PT 3 subtypes. Green (PT 3a) and blue (PT 3c) light specialists were both shown to dominate in warm areas but display clearly distinct niches, with 3a dominating in *Synechococcus*-rich stations located on oceanic borders, while 3c predominated in purely oceanic areas where the global abundance of *Synechococcus* is low. These results are in agreement with the prevailing view of an increase in the PUB:PEB ratio from green onshore mesotrophic waters to blue offshore oligotrophic waters (18, 19, 21–26, 28, 29, 44, 45). Similarly, we showed that PT 3dB, which could not be distinguished from PT 3c in previous studies (17, 29–31), prevails in more coastal and/or mixed temperate waters than do 3c populations. The realized environmental niche of the second type of CA4 (PT 3dA) is the best defined of all PTs as it is clearly associated with nutrient-rich waters and with the coldest stations of our dataset, occurring at high latitude, at depth and/or in vertically mixed waters (e.g., TARA\_068, 093 and 133). This result is consistent with a recent study demonstrating the dominance of 3dA in sub-Arctic waters of the Northwest Pacific Ocean (29), suggesting that the prevalence of 3dA at high latitude can be generalized. Altogether, while little was previously known about the abundance and distribution of CA4 populations in the field, here we show that they are ubiquitous, dominate in a wide range

of niches, are present both in coastal and oceanic mixed waters, and overall are the most abundant *Synechococcus* PT.

The relationship between ESTUs and PTs shows that some ESTUs are preferentially associated with only one PT, while others present a much larger pigment diversity. ESTU IIA, the most abundant and ubiquitous ESTU in the field (5, 6) displays the widest PT diversity (Fig. 4B), a finding confirmed by clade II isolates spanning the largest diversity of pigment content, with representative strains of PT 2, 3a, 3c and 3dB within this clade (Dataset 3; see also 7, 11, 46–48). This suggests that this ESTU can colonize all light color niches, an ability which might be partially responsible for its global ecological success. Our current results do not support the previously observed correlation between clade III and PT 3a (29) since the two ESTUs defined within this clade (IIIA and B) were associated with PT 3c and/or 3f. This discrepancy could be due either to the different methods used in these studies or to the occurrence of genetically distinct clade III populations in coastal areas of the northwestern Pacific Ocean and along the Tara Oceans transect. However, the pigment phenotype of strains isolated to date is more consistent with our findings (Dataset 3; 16, 41).

In contrast to most other PTs, the association between PT 3dA and ESTUs was found to be nearly exclusive in the field, as ESTUs from clades I, IV, CRD1 and EnvA were not associated with any other PT, and reciprocally PT 3dA is only associated with these clades (Fig. 4A). An interesting exception to this general rule was observed in the Benguela upwelling (TARA\_067), where the dominant ESTU IA population possesses both a 3a *mpeBA* allele and *fciA/B* and *mpeZ* genes (Figs. S6K and S7), suggesting that cells, which were initially green light specialists (PT 3a), have inherited a complete CA4-A island through lateral gene transfer at this station. Interestingly, among the seven clade I strains sequenced to date, three possess a 3a *mpeBA* allele, among which WH8016 also has a CA4-A island but only partial (lacking *mpeZ*) and therefore not functional (11). It is thus difficult to conclude whether the lateral transfer of this island, likely a rare event since it was only observed in populations of the Benguela upwelling, has conferred these populations the ability to perform CA4.

Another striking result of this study was the unsuspected importance of populations that have likely lost the ability to chromatically acclimate, specifically in warm HNLC areas, which cover wide expanses of the South Pacific Ocean (49). Interestingly, populations living in these ultra-oligotrophic environments have a different genetic basis for their consistently elevated PUB phenotype than do typical blue light specialists (i.e., PT 3c), since they have lost the CA4 regulators *fciA/B* and accumulated mutations in *mpeY*, a yet uncharacterized member of the phycobilin lyase family, as observed in strain BIOS-E4-1 (Fig. S6K; 11). This finding, consistent with the previous observation that the south Pacific is dominated by “high PUB” *Synechococcus* (22), is further supported by the recent sequencing of three isolates from the Equatorial Pacific, strains MITS9504, MITS9509 (both CRD1C) and MITS9508 (CRD1A; 50), all of which contain, like BIOS-E4-1, a 3dA *mpeBA* allele, a CA4-A island lacking *fciA/B* and a partial (MITS9508) or highly degenerated (2 other MIT strains) *mpeY* gene sequence (Fig. S6K). Thus, these natural CA4-A mutants seem to have adapted to blue, ultra-oligotrophic waters by inactivating a likely energetically costly acclimation mechanism (positive selection), although we cannot exclude that it might be a consequence of the lower selection efficiency associated to the reduced effective population size of *Synechococcus* in such an extreme environment (genetic drift). If, as we hypothesize, all *Synechococcus* cells counted as 3dA at these stations are CA4-deficient, these natural mutants would represent about 15% of the total 3dA population. In contrast, CRD1-A populations of the eastern border of the Pacific Ocean (TARA\_102, 109–110, 137) are likely true CA4 populations as they possess all CA4 genes

(Fig. S6K). In conclusion, our study provided unprecedented insights into the distribution, ecology and adaptive value of all known *Synechococcus* PTs. Surprisingly, the sum of 3dA and 3dB constitutes about 40% of the total *Synechococcus* counts in the *Tara* Oceans dataset, making chromatic acclimators (PT 3d) the most globally abundant PT, even when taking into account potential CA4-deficient natural mutants. In addition, this PT made up 95% of the *Synechococcus* population at high latitudes and was present in every one of the five major ESTUs in the field (I, II, III, IV and CRD1). This suggests that chromatic acclimation likely confers a strong adaptive advantage compared to strains with a fixed pigmentation, particularly in vertically mixed environments and at depth at stations with a stratified water column. The previously unsuspected abundance of CA4 populations could partially explain previous contradicting observations that the PUB:PEB ratio either increases with depth (18, 21, 26) or remains constant throughout the water column (22, 27, 28, 51). The occurrence of natural CA4 mutants and evidence for lateral transfer of the CA4 genomic island suggest that not only temperature and nutrient availability but also light quality co-exert a selective pressure on *Synechococcus* evolution, and that changes in pigment diversity occur in response to changes in light niches by acquisition or loss of specific PBS synthesis and/or regulation genes, as previously observed for phosphorus and nitrogen transport genes in *Prochlorococcus* (52–54). Still, the complex interactions between PTs, vertical phylogeny and environmental parameters remain unclear and more work is needed to refine our understanding of the balance between the forces shaping community composition. Is it temperature and/or nutrient availability that primarily controls community clade composition, and after that the within-clade populations displaying the pigmentation that fits best the light condition are selected, or does light quality select for a given PT and then other environmental parameters for the most adapted clades? Part of the answer lies in the observation that in some environments and more specifically at the boundaries of *Synechococcus* environmental niche(s), where the harshest conditions are encountered, both pigment and clade diversities are drastically reduced. This could indicate either a genetic bottleneck and founder effect or that extreme selective pressure only allows the fittest organism to survive. However, this concomitant reduction of genetic and pigment diversity rather tends to support a co-selection by light quality and other environmental parameters. On the contrary, the diverse PTs occurring within some clades, as well as the co-occurrence of different PTs at most stations compared to more clear-cut clade shifts (e.g., in the Red Sea/Indian Ocean) might indicate that light quality is not the strongest selective force or that light changes are too transient to allow the dominance and fixation of a particular PT in a population. Future experimental work exploring the fitness effects of ESTU members and pigmentation under different controlled environmental conditions (including temperature, nutrients and light) might help to clarify their respective effects on the diversification of this ecological important photosynthetic organism.

## Materials and Methods

### Metagenomic samples

This study focused on 109 metagenomic samples corresponding to 65 stations from the worldwide oceans collected during the 2.5-year *Tara* Oceans circumnavigation (2009–2011). Sample processing and features are the same as described in (6). Sequencing depths ranged from  $16 \times 10^6$  to  $258 \times 10^6$  reads per sample after quality control and paired-reads merging, and corresponding fragments lengths averaged  $164 \pm 20$  bp (median: 168 bp).

### Databases: reference and outgroup sequences

A reference database comprising the full-length gene or operon nucleotide sequences was generated for each marker used in this study (*cpcBA*, *mpeBA* and *mpeW*) based on culture isolates with characterized pigment type (Dataset 1). These databases comprised 83 *cpcBA* sequences (64 unique), including 18 PT 1, 5 PT 2A, 19 PT 2B and 39 PT 3, 41 *mpeBA* sequences (all unique), including 11 PT 3a, 2 PT 3f, 11 PT 3dA and 17 PT 3dB and 5 unique

*mpeW* sequences. For each marker, a reference alignment was generated with MAFFT L-INS-i v6.953b (55), and a reference phylogenetic tree was inferred with PhyML v. 20120412 (GTR+I+G, 10 random starting trees, best of SPR and NNI moves, 500 bootstraps; 56) and drawn using the ETE Toolkit (57).

A database of outgroups was also built, comprising paralogous sequences coming from marine *Synechococcus* or *Prochlorococcus* as well as orthologous sequences from other marine and freshwater organisms retrieved from public databases. For *cpcBA* and *mpeBA*, the outgroup databases comprised *apcA*, *apcB*, *apcD*, *apcF* and *cpeBA* from marine *Synechococcus*, *ppeBA* from *Prochlorococcus*, *cpcBA* and *cpeBA* from other non-picrocyanobacterial organisms as well as either *mpeBA* or *cpcBA* from marine *Synechococcus*, respectively (Datasets 1–2). For *mpeW*, the outgroup database was made of paralogous genes (*mpeZ*, *mpeY* and *cpeY*) from marine *Synechococcus* or *Prochlorococcus*, as no ortholog could be identified in public databases. Similarly, for *mpeY* and *mpeZ*, the outgroup database comprised *cpeY*, *mpeW* as well as *mpeZ* and *mpeY*, respectively. The outgroup database for *mpeU* comprised *cpeF* paralogous sequences from marine *Synechococcus* and *Prochlorococcus*. No outgroup database was used for *fciaB*, as no paralog or even distantly related sequence were found either in marine *Synechococcus* and *Prochlorococcus* or in public databases.

### Read assignment and estimation of PT abundance

Reads were preselected using BLAST+ (58) with relaxed parameters (blastn, maximum E-value of  $1e-5$ , minimum percent identity 60%, minimum 75% of read length aligned), using reference sequences as subjects; the selection was then refined by a second BLAST+ round against databases of outgroups: reads with a best-hit to outgroup sequences were excluded from downstream analysis. Selected reads were then aligned to the marker reference alignment with MAFFT v.7.299b (--addfragments --adjustdirectionaccurately) and placed in the marker reference phylogenetic tree with *pplacer* (59). For each read, *pplacer* returns a list of possible positions (referred to as placements) at which it can be placed in the tree and their associated "likelihood weight ratio" (LWR, proxy for the probability of the placement; see *pplacer* publication and documentation for more details). Reads were then assigned to a pigment type using a custom classifier written in Python. Briefly, internal nodes of the reference tree were assigned a pigment type based on the pigmentation of descending nodes (PT of child reference sequences if the same for all of them, "unclassified" otherwise). For each read, placements were assigned to their nearest ascending or descending node based on their relative position on the edge, and the lowest common ancestor (LCA) of the set of nodes for which the cumulated LWR was greater than 0.95 (LCA of possible placements at 95% probability) was then computed. Finally, the read was assigned to the pigment type of this LCA. Different combinations of read assignment parameters (LCA at 90%, 95% or 100%; assignment of placements to the ascending, descending or nearest node) were also assessed, and resulted either in higher rates of unassigned reads or unacceptable error rates (Fig. S2).

Read counts were normalized by adjusted marker length: for each marker and each sequence file, counts were normalized by  $(L - \ell + 1)$ , with  $L$  the length of the marker gene (*cpcBA* mean length: 1053.7bp; *mpeBA* mean length: 1054.6bp; *mpeW* mean length: 1193.3bp) and  $\ell$  the mean length of reads in the sequence file. Finally, the abundance of PT1, 2A and 2B was defined as the normalized *cpcBA* read counts of these PT, the abundance of PT3a, 3f and 3dA as the normalized *mpeBA* read counts of these PT, 3dB as the normalized *mpeW* count and 3c as the difference between the normalized *mpeBA* (3c + 3dB) read count and the PT3dB count assessed with *mpeW*. The abundance of unclassified sequences was also taken into account.

Detailed *petB* counts for clade and ESTU abundances were obtained from (6).

### Read assignment simulations

For each marker, simulated reads were generated from one reference sequence at a time using a sliding window of 100, 125 or 150bp (*TARA* mean read length: 164.2bp; median 169bp) and steps of 5 bp. Simulated reads were then assigned to a pigment type with the aforementioned bioinformatic pipeline, using all reference sequences except the one used to simulate reads. Inferred pigment types of simulated fragments were then compared to known pigment types of originating reference sequence.

### Statistical analyses

All environmental parameters used for statistical analyses are the same as in (6), except the blue to green irradiance ratio that was modeled as described in the supplementary materials and methods. Hierarchical clustering and NMDS analyses of stations were performed using R (60) packages cluster v1.14.4 (61) and MASS v7.3–29 (62), respectively. PT contingency tables were filtered by considering only stations with more than 30 *cpcBA* reads and 30 *mpeBA* reads, and only PT appearing in at least 2 stations and with more than 150 reads in the whole dataset. Contingency tables were normalized using Hellinger transformation that gives lower rates to rare PT. The Bray–Curtis distance was then used for ordination (isoMDS function; maxit, 100; k, 2). Correlations were performed with R package Hmisc3.17-4 with Benjamini & Hochberg multiple comparison adjusted p-value (63).

### Acknowledgements

We warmly thank Dr. Annick Bricaud for fruitful discussions on biooptics, members of the ABiMS platform (Roscoff) for providing us an efficient storage and computing facility for our bioinformatics analyses as well as the NERC Biomolecular Analysis Facility (NBAF, Centre for Genomic Research, University of Liverpool) for sequencing some *Synechococcus* genomes used in this study. This work was supported by the French "Agence Nationale de la Recherche" Programs SAMOSA (ANR-13-ADAP-0010) and France Génomique (ANR-10-INBS-09), the French Government "Investissements d'Avenir" programs OCEANOMICS (ANR-11-BTBR-0008), the European Union's Seventh Framework Programs FP7 MicroB3 (grant agreement 287589) and MaCuMBA

(grant agreement 311975), UK Natural Environment Research Council Grant NE/I00985X/1 and the Spanish Ministry of Science and Innovation grant MicroOcean PANGENOMICS (GL2011-26848/BOS). We also thank the support and commitment of the Tara Oceans coordinators and consortium, Agnès b. and E. Bourgois, the Veolia Environment Foundation, Région Bretagne, Lorient Agglomération, World Courier, Illumina, the EDF Foundation, FRB, the Prince Albert II de Monaco Foundation, the Tara schooner and its captains and crew. Tara Oceans would not exist without continuous support from 23 institutes (<http://oceans.taraexpeditions.org>).

- Guidi L, et al. (2016) Plankton networks driving carbon export in the oligotrophic ocean. *Nature* 532(7600):465–470.
- Flombaum P, et al. (2013) Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc Natl Acad Sci USA* 110(24):9824–9829.
- Zwirgmaier K, et al. (2008) Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ Microbiol* 10(1):147–161.
- Mazard S, Ostrowski M, Partensky F, Scanlan DJ (2012) Multi-locus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*. *Environ Microbiol* 14(2):372–386.
- Sohm JA, et al. (2016) Co-occurring *Synechococcus* ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. *ISME J* 10(2):333–345.
- Farrant GK, et al. (2016) Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proc Natl Acad Sci USA* 113(24):E3365–E3374.
- Six C, et al. (2007) Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.: a comparative genomics study. *Genome Biol* 8(12):R259.
- Alberte RS, Wood AM, Kursar TA, Guillard RRL (1984) Novel Phycocyanins in Marine *Synechococcus* spp. *Plant Physiol* 75(3):732–739.
- Ong LJ, Glazer AN (1991) Phycocyanins of marine unicellular cyanobacteria. I. Bilin types and locations and energy transfer pathways in *Synechococcus* spp. phycocyanins. *J Biol Chem* 266(15):9515–9527.
- Sidler WA (1994) Phycobilisome and Phycobiliprotein Structures. *The Molecular Biology of Cyanobacteria*, Advances in Photosynthesis. (Springer, Dordrecht), pp 139–216.
- Humily F, et al. (2013) A gene island with two possible configurations is involved in chromatic acclimation in marine *Synechococcus*. *PLoS ONE* 8(12):e84459.
- Palenik B (2001) Chromatic adaptation in marine *Synechococcus* strains. *Appl Environ Microbiol* 67(2):991–994.
- Everroad C, et al. (2006) Biochemical bases of type IV chromatic adaptation in marine *Synechococcus* spp. *J Bacteriol* 188(9):3345–3356.
- Shukla A, et al. (2012) Phycocyanin-specific bilin lyase-isomerase controls blue-green chromatic acclimation in marine *Synechococcus*. *Proc Natl Acad Sci* 109(49):20136–20141.
- Sanfilippo JE, et al. (2016) Self-regulating genomic island encoding tandem regulators confers chromatic acclimation to marine *Synechococcus*. *Proc Natl Acad Sci* 113(21):6077–6082.
- Toledo G, Palenik B, Brahamsha B (1999) Swimming marine *Synechococcus* strains with widely different photosynthetic pigment ratios form a monophyletic group. *Appl Environ Microbiol* 65(12):5247–5251.
- Humily F, et al. (2014) Development of a targeted metagenomic approach to study a genomic region involved in light harvesting in marine *Synechococcus*. *FEMS Microbiol Ecol* 88(2):231–249.
- Olson RJ, Chisholm SW, Zettler ER, Armbrust EV (1990) Pigments, size, and distributions of *Synechococcus* in the North Atlantic and Pacific Oceans. *Limnol Oceanogr* 35(1):45–58.
- Sherry ND, Michelle Wood A (2001) Phycocyanin-containing picocyanobacteria in the Arabian Sea in February 1995: diel patterns, spatial variability, and growth rates. *Deep Sea Res Part II* 48(6–7):1263–1283.
- Jiang T, et al. (2016) Temporal and spatial variations of abundance of phycocyanin- and phycocyanin-rich *Synechococcus* in Pearl River Estuary and adjacent coastal area. *J Ocean Univ China* 15(5):897–904.
- Lantoine F, Neveux J (1997) Spatial and seasonal variations in abundance and spectral characteristics of phycocyanins in the tropical northeastern Atlantic Ocean. *Deep Sea Res Part I* 44(2):223–246.
- Neveux J, Lantoine F, Vaulot D, Marie D, Blanchot J (1999) Phycocyanins in the southern tropical and equatorial Pacific Ocean: Evidence for new cyanobacterial types. *J Geophys Res Oceans* 104(C2):3311–3321.
- Wood AM, Phinney DA, Yentsch CS (1998) Water column transparency and the distribution of spectrally distinct forms of phycocyanin-containing organisms. *Mar Ecol Prog Ser* 162:25–31.
- Campbell L, et al. (1998) Response of microbial community structure to environmental forcing in the Arabian Sea. *Deep Sea Res Part II Top Stud Oceanogr* 45(10):2301–2325.
- Hoge FE, Wright CW, Kana TM, Swift RN, Yungel JK (1998) Spatial variability of oceanic phycocyanin spectral types derived from airborne laser-induced fluorescence emissions. *Appl Opt* 37(21):4744–4749.
- Wood AM, Lipsen M, Coble P (1999) Fluorescence-based characterization of phycocyanin-containing cyanobacterial communities in the Arabian Sea during the Northeast and early Southwest Monsoon (1994–1995). *Deep Sea Res Part II Top Stud Oceanogr* 46(8):1769–1790.
- Yona D, Park MO, Oh SJ, Shin WC (2015) Distribution of *Synechococcus* and its phycocyanin pigment in relation to environmental factors in the East Sea, Korea. *Ocean Sci J* 49(4):367–382.
- Campbell L, Iturriaga R (1988) Identification of *Synechococcus* spp. in the Sargasso Sea by immunofluorescence and fluorescence excitation spectroscopy performed on individual cells. *Limnol Oceanogr* 33(5):1196–1201.
- Xia X, et al. (2017) Phylogeography and pigment type diversity of *Synechococcus* cyanobacteria in surface waters of the northwestern Pacific Ocean. *Environ Microbiol* 19(1):142–158.
- Xia X, Liu H, Choi D, Noh JH (2017) Variation of *Synechococcus* pigment genetic diversity along two turbidity gradients in the China Seas. *Microb Ecol*:1–12.
- Xia X, Guo W, Tan S, Liu H (2017) *Synechococcus* assemblages across the salinity gradient in a salt wedge estuary. *Front Microbiol* 8:1254.
- Liu H, Jing H, Wong THC, Chen B (2014) Co-occurrence of phycocyanin- and phycocyanin-rich *Synechococcus* in subtropical estuarine and coastal waters of Hong Kong. *Environ Microbiol Rep* 6(1):90–99.
- Chung C-C, Gong G-C, Huang C-Y, Lin J-Y, Lin Y-C (2015) Changes in the *Synechococcus* assemblage composition at the surface of the East China Sea due to flooding of the Changjiang river. *Microb Ecol* 70(3):677–688.
- Haverkamp T, et al. (2008) Diversity and phylogeny of Baltic Sea picocyanobacteria inferred from their ITS and phycobiliprotein operons. *Environ Microbiol* 10(1):174–188.
- Sunagawa S, et al. (2015) Structure and function of the global ocean microbiome. *Science* 348(6237):1261359.
- Logares R, et al. (2014) Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environ Microbiol* 16(9):2659–2671.
- Pearman PB, Guisan A, Broennimann O, Randin CF (2008) Niche dynamics in space and time. *Trends Ecol Evol* 23(3):149–158.
- Larsson J, et al. (2014) Picocyanobacteria containing a novel pigment gene cluster dominate the brackish water Baltic Sea. *ISME J* 8(9):1892–1903.
- Paulsen ML, et al. (2016) *Synechococcus* in the Atlantic Gateway to the Arctic Ocean. *Front Mar Sci* 3:191.
- Haverkamp THA, et al. (2008) Colorful microdiversity of *Synechococcus* strains (picocyanobacteria) isolated from the Baltic Sea. *ISME J* 3(4):397–408.
- Hunter-Cevera KR, Post AF, Peacock EE, Sosik HM (2015) Diversity of *Synechococcus* at the Martha's Vineyard Coastal Observatory: Insights from culture isolations, clone libraries, and flow cytometry. *Microb Ecol* 71(2):276–289.
- Cabello-Yeves PJ, et al. (2017) Novel *Synechococcus* genomes reconstructed from freshwater reservoirs. *Front Microbiol* 8:1151.
- Mahmoud RM, et al. (2017) Adaptation to blue light in marine *Synechococcus* requires MpeU, an enzyme with similarity to phycocyanobilin lyase isomerases. *Front Microbiol* 8:243.
- Choi DH, Noh JH (2009) Phylogenetic diversity of *Synechococcus* strains isolated from the East China Sea and the East Sea. *FEMS Microbiol Ecol* 69(3):439–448.
- Veldhuis MJW, Kraay GW (1993) Cell abundance and fluorescence of picoplankton in relation to growth irradiance and nitrogen availability in the Red Sea. *Neth. J. Sea Res* 31(2):135–145.
- Ahlgren NA, Rocap G (2006) Culture isolation and culture-independent clone libraries reveal new marine *Synechococcus* ecotypes with distinctive light and N physiologies. *Appl Environ Microbiol* 72(11):7193–7204.
- Bemal S, Anil AC (2016) Genetic and ecophysiological traits of *Synechococcus* strains isolated from coastal and open ocean waters of the Arabian Sea. *FEMS Microbiol Ecol* 92(11).
- Everroad RC, Wood AM (2012) Phycocyanin evolution and diversification of spectral phenotype in marine *Synechococcus* and related picocyanobacteria. *Mol Phylogenet Evol* 64(3):381–392.
- Morel A, et al. (2007) Optical properties of the "clearest" natural waters. *Limnol Oceanogr* 52(1):217–229.
- Cubillos-Ruiz A, Berta-Thompson JW, Becker JW, van der Donk WA, Chisholm SW (2017) Evolutionary radiation of lanthipeptides in marine cyanobacteria. *Proc Natl Acad Sci USA* 114(27):E5424–E5433.
- Katano T, Nakano S (2006) Growth rates of *Synechococcus* types with different phycocyanin composition estimated by dual-laser flow cytometry in relationship to the light environment in the Uwa Sea. *J Sea Res* 55(3):182–190.
- Martiny AC, Huang Y, Li W (2009) Occurrence of phosphate acquisition genes in *Prochlorococcus* cells from different ocean regions. *Environ Microbiol* 11(6):1340–1347.
- Martiny AC, Coleman ML, Chisholm SW (2006) Phosphate acquisition genes in *Prochlorococcus* ecotypes: evidence for genome-wide adaptation. *Proc Natl Acad Sci USA* 103(33):12552–12557.
- Martiny AC, Kathuria S, Berube PM (2009) Widespread metabolic potential for nitrite and nitrate assimilation among *Prochlorococcus* ecotypes. *Proc Natl Acad Sci USA* 106(26):10787–10792.
- Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* 30(4):772–780.
- Guindon S, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59(3):307–321.
- Huerta-Cepas J, Serra F, Bork P (2016) ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol* 33(6):1635–1638.
- Camacho C, et al. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.
- Matsen FA, Kodner RB, Armbrust EV (2010) pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinform*

1225  
1226  
1227  
1228  
1229  
1230  
1231  
1232  
1233  
1234  
1235  
1236  
1237  
1238  
1239  
1240  
1241  
1242  
1243  
1244  
1245  
1246  
1247  
1248  
1249  
1250  
1251  
1252  
1253  
1254  
1255  
1256  
1257  
1258  
1259  
1260  
1261  
1262  
1263  
1264  
1265  
1266  
1267  
1268  
1269  
1270  
1271  
1272  
1273  
1274  
1275  
1276  
1277  
1278  
1279  
1280  
1281  
1282  
1283  
1284  
1285  
1286  
1287  
1288  
1289  
1290  
1291  
1292

60. R Core Team (2014) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria) Available at: <http://www.R-project.org/>.

61. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K (2017) *cluster: Cluster Analysis Basics and Extensions*.

62. Venables WN, Ripley BD (2002) *Modern Applied Statistics with S* (Springer, New York).

Fourth Available at: <http://www.stats.ox.ac.uk/pub/MASS4>.

63. Harrell FE (2016) *Hmisc: Harrell Miscellaneous* Available at: <http://CRAN.R-project.org/package=Hmisc>.

1293  
1294  
1295  
1296  
1297  
1298  
1299  
1300  
1301  
1302  
1303  
1304  
1305  
1306  
1307  
1308  
1309  
1310  
1311  
1312  
1313  
1314  
1315  
1316  
1317  
1318  
1319  
1320  
1321  
1322  
1323  
1324  
1325  
1326  
1327  
1328  
1329  
1330  
1331  
1332  
1333  
1334  
1335  
1336  
1337  
1338  
1339  
1340  
1341  
1342  
1343  
1344  
1345  
1346  
1347  
1348  
1349  
1350  
1351  
1352  
1353  
1354  
1355  
1356  
1357  
1358  
1359  
1360

# Submission PDF

## Supplementary Materials and Methods

## Modeling of the blue to green irradiance ratio (Ir495/Ir545) at Tara Oceans stations.

We used the clear sky surface irradiance model of Frouin and McPherson in Fortran and translated to Matlab by Werdell (see Frouin et al., 1989 and Tanre et al., 1979 for the analytical formula used) using the date, latitude and longitude of each station, assuming sunny sky and at noon.

The spectral light distribution averaged over the mixed layer was computed from:

$$\langle Ir(\lambda) \rangle = \frac{\int_0^{MLD} E(\lambda, 0^-) e^{-k(\lambda, chl)z} dz}{MLD} = \frac{I(\lambda, 0^-)}{MLD k(\lambda, chl)} \{1 - e^{-k(chl)MLD}\}$$

where:

- *chl* denotes the average chlorophyll value in the mixed layer. [*chl*] was based on a fluorometer that was calibrated against HPLC data and corrected for non-photochemical quenching,
- MLD is the mixed layer depth that was computed based on a temperature threshold criterion
- $k(\lambda, chl)$  is the diffuse attenuation coefficient at wavelength  $\lambda$  (495 or 545 using a 10 nm bandwidth). This parameter was computed using Morel and Maritorena (2001)'s equation:

$$k(\lambda, chl) = k_w(\lambda) + \chi(\lambda)[chl]^{e(\lambda)}$$

$k_w$ ,  $\chi$  and  $e$  are provided in Table 2 of Morel and Maritorena (2001) and have the following values for the wavelengths of interest:

| Wavelength [nm] | $k_w(\lambda)$ [ $m^{-1}$ ] | $\chi(\lambda)$ | $e(\lambda)$ |
|-----------------|-----------------------------|-----------------|--------------|
| 495             | 0.01885                     | 0.06907         | 0.68947      |
| 545             | 0.05212                     | 0.04253         | 0.65591      |

If the sampling depth was below the MLD, the irradiance was computed as follows:

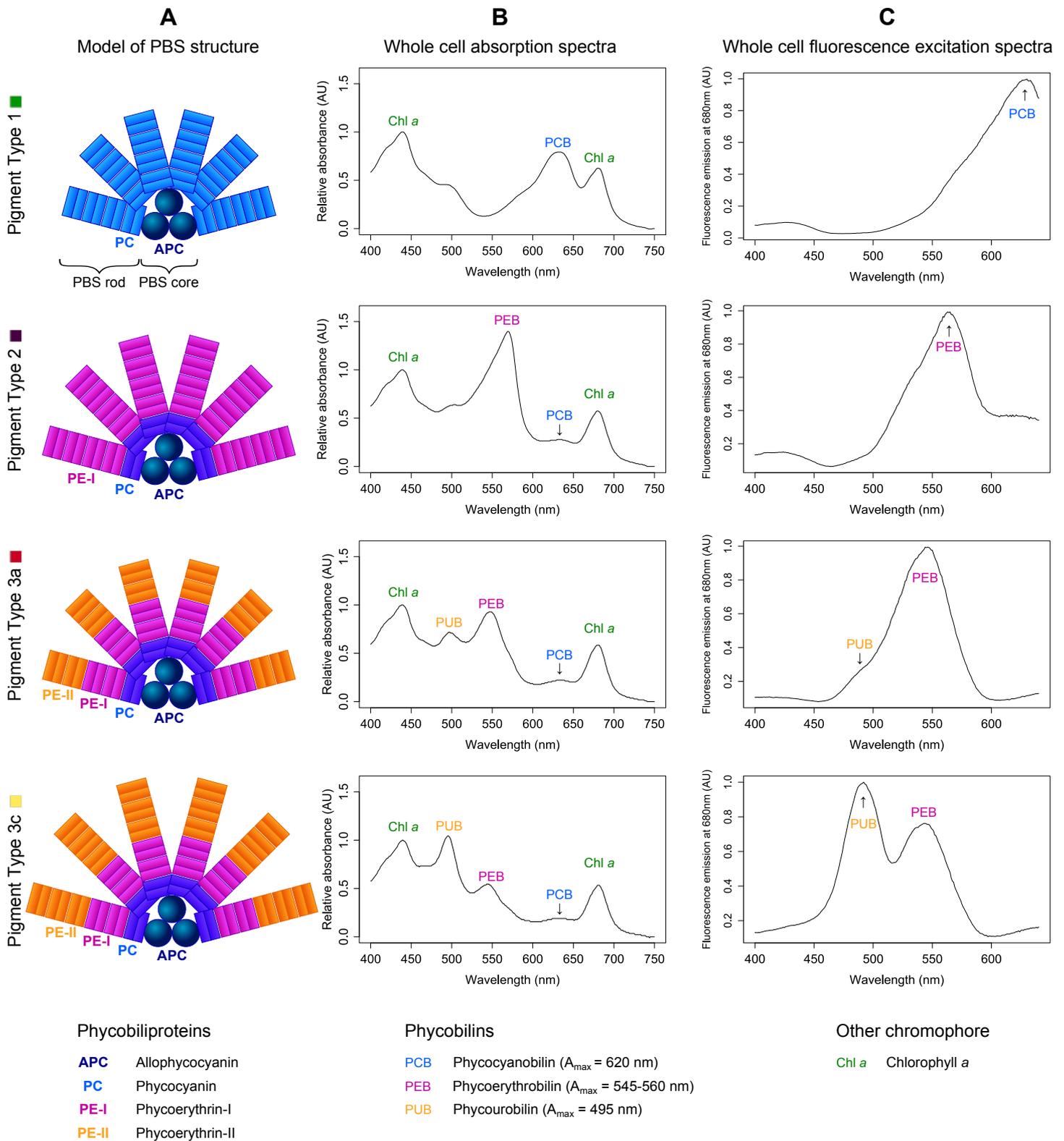
$$Ir(\lambda, sampling\ depth) = I(\lambda, 0^-) e^{-k(\lambda, chl) sampling\ depth}$$

The ratio was then computed as  $Ir(495)/Ir(545)$ .

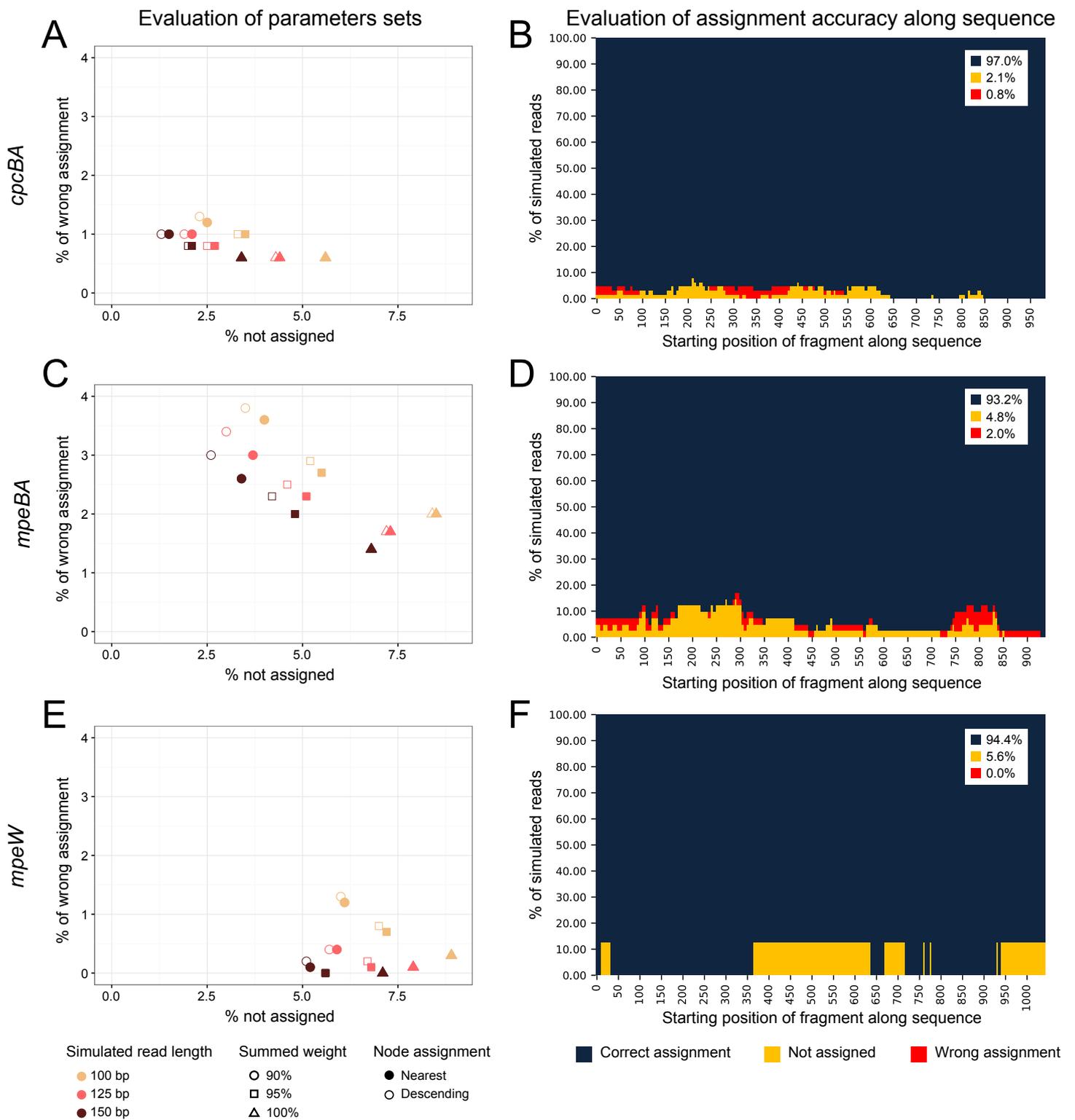
### References:

- Frouin, R., D. W. Ligner, and C. Gautier, 1989: A Simple analytical formula to compute clear sky total and photosynthetically available solar irradiance CC at the ocean surface. *J. Geophys. Res.*, 94, 9731-9742.
- Morel, A. and S. Maritorena, 2001: Bio-optical properties of oceanic waters: A reappraisal. *J. Geophys. Res.*, 106, 7163–7180.
- Tanre, D., M. Herman, P.-Y. Deschamps, and A. De Lefte, 1979: Atmospheric modeling for Space measurements of ground reflectances, including bi-directional properties. *Appl. Optics*, 18, 21,3587-21,3597.

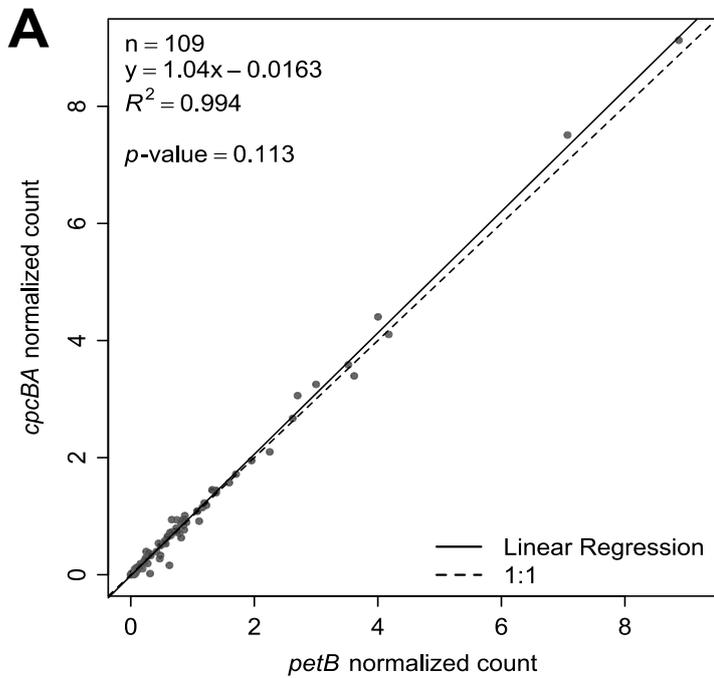
## Supplementary Figures



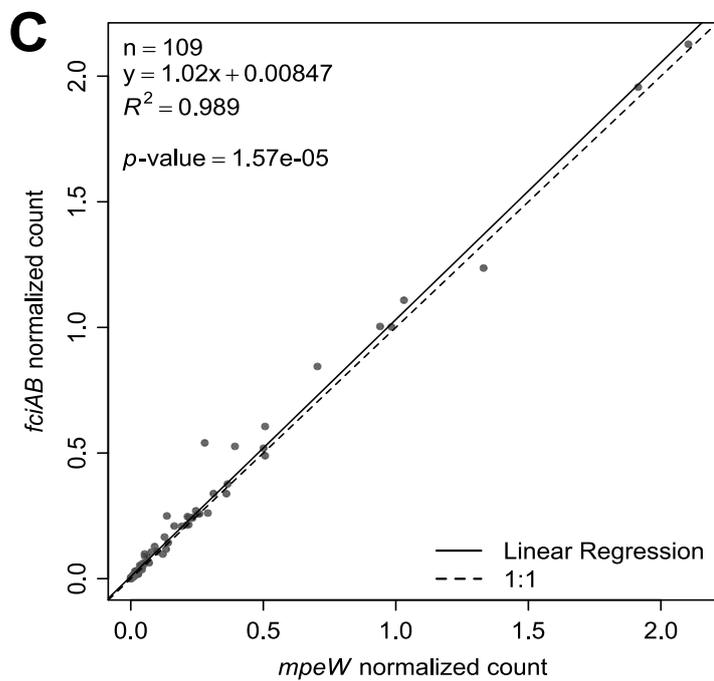
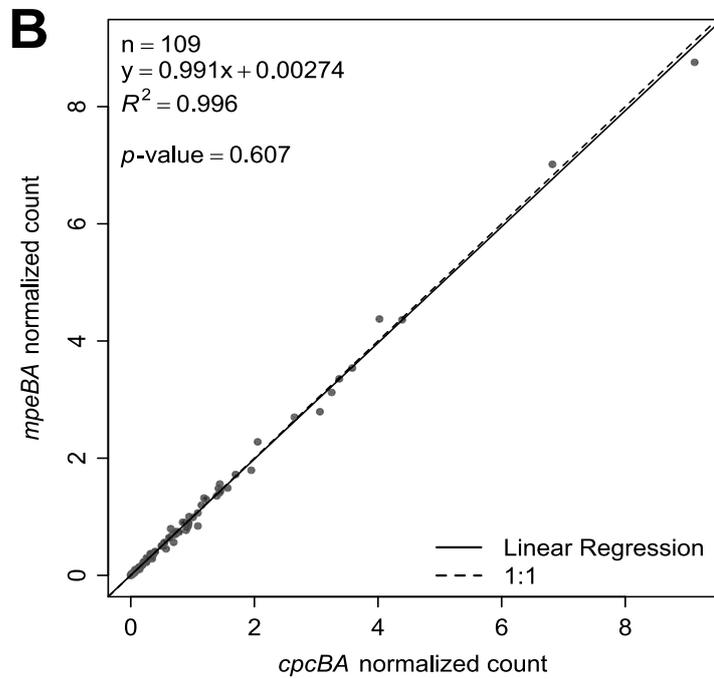
**Figure S1:** Biochemical composition and biooptical properties of phycobilisomes (PBS) of the main *Synechococcus* pigment types (PT). (A) Models of PBS structure, highlighting the conserved core and variable rods of increasing complexity from PT1 to PT3 (Redrawn after Six *et al.*, 2007). (B) Whole cell absorption spectra of the different PTs (Reproduced after Six *et al.*, 2007). Chromophores responsible of each absorption peaks are indicated. (C) Whole cell fluorescence excitation spectra with emission at 680 nm. Note that for chromatic acclimators (PT 3d), the PBS structure is similar to other PT 3 but that the excitation ratio at 495 nm and 545 nm ( $Ex_{495:545}$ ) varies from 0.6 in green light to 1.6 in blue light (not shown).

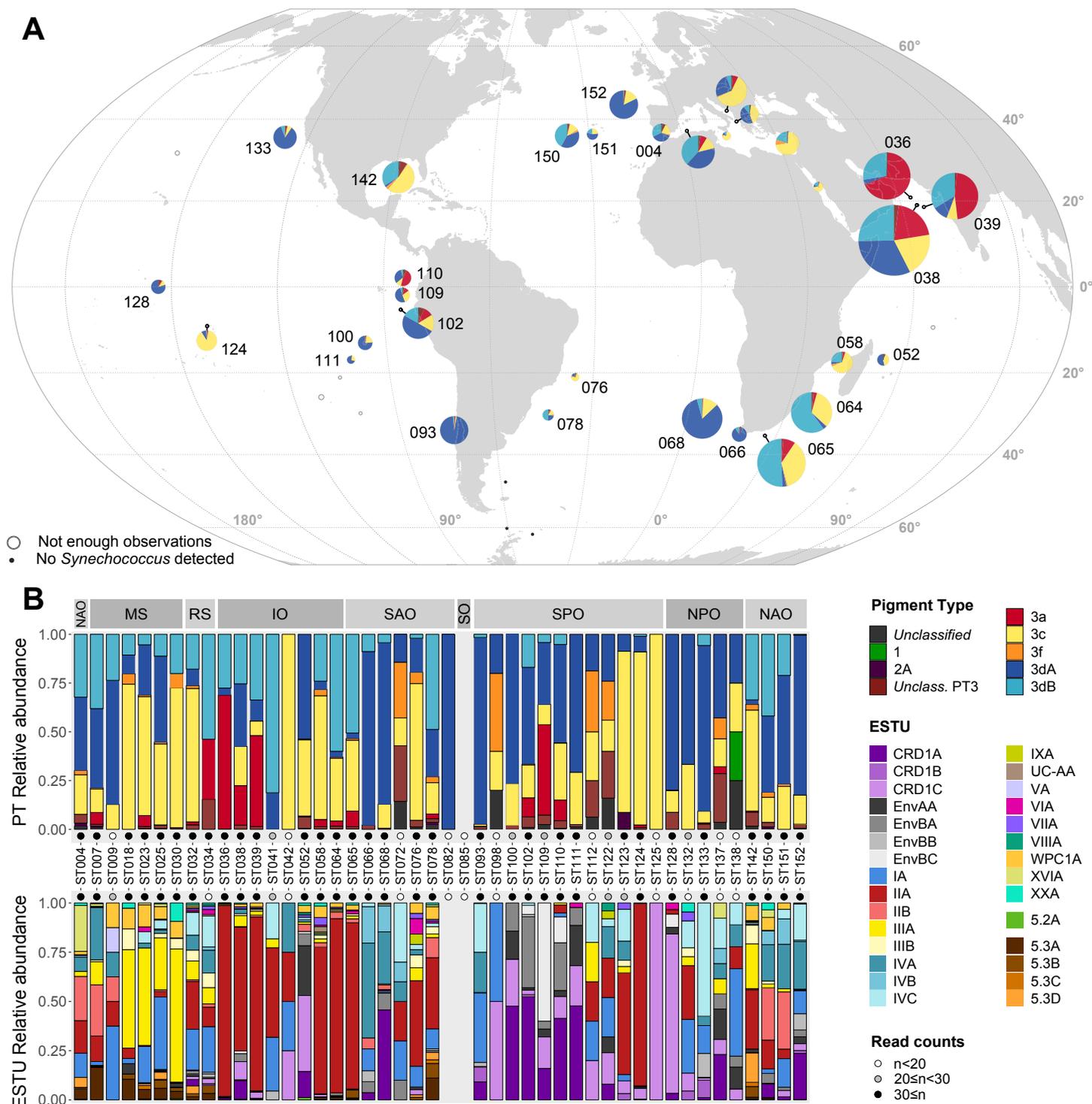


**Figure S2:** Evaluation of the assignment pipeline and the resolution power of the different markers used in this study. Simulated reads were generated from the reference dataset and assigned using a custom-designed pipeline (see materials and methods). (A, C, E) Evaluation of different sets of parameters tested for read assignment for the different markers: *cpcBA* (A), *mpeBA* (C) and *mpeW* (E). 100 (yellow), 125 (pink) and 150 bp (dark red) long reads were simulated. For each read, pplacer returns a list of possible positions in the tree, each associated with a likelihood weight. From these placements, we considered only those that reached a summed likelihood weight of either 90% (circle), 95% (square) or 100% (triangle). The assignment was then performed based on the phenotype of either the nearest node (solid symbol) in the tree or the descending (child) node (empty symbol). (B, D, F) Evaluation the resolution power along *cpcBA* (B), *mpeBA* (D) or *mpeW* (F) for 150 bp simulated reads assigned using the parameters selected for *Tara Oceans* metagenomic read assignment (i.e., nearest node assignment and summed weight of 95%). Note that *Tara Oceans* reads had a mean length of 164 bp.

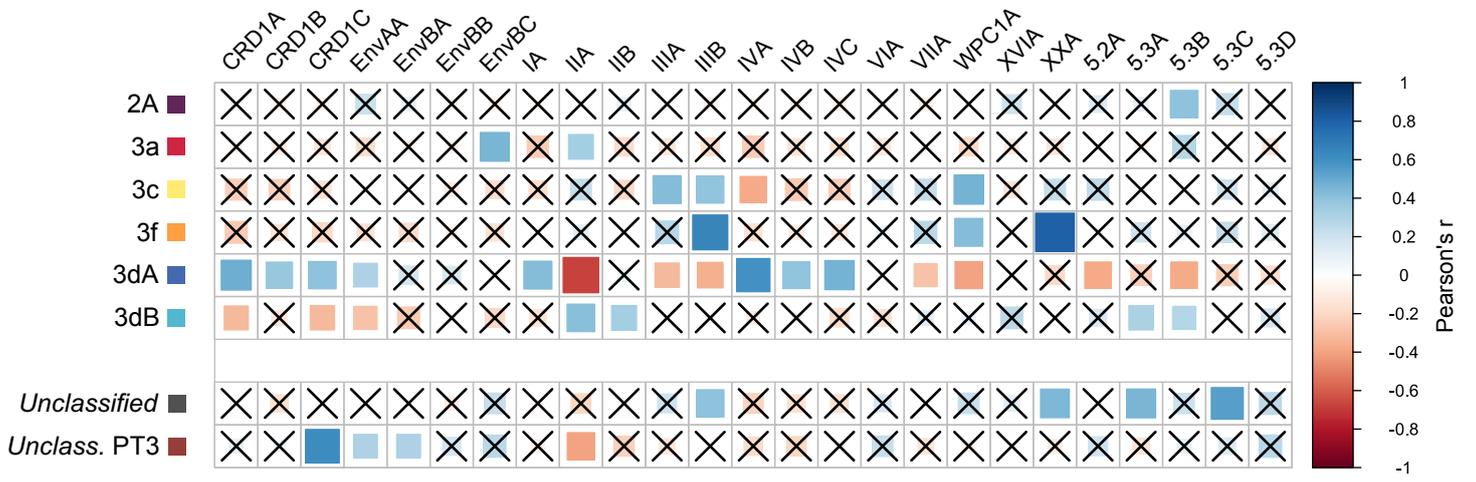


**Figure S3:** Correlations between the number of reads recruited using the main markers used in this study. (A) Correlation between *petB* (vertical phylogeny) and *cpcBA* counts used to discriminate pigment types (PT) 1, 2 and 3. (B) Correlation between PT 3 counts using *cpcBA* and total *mpeBA* counts. Note that *mpeBA* is a PT3 specific marker and is used to discriminate PTs 3a, 3dA, 3f and 3c +3dB. (C) Correlation between PT 3dB counts using *fciAB*, a PT 3dB- and 3dA-specific marker and total *mpeW* counts, a PT 3dB-specific marker.

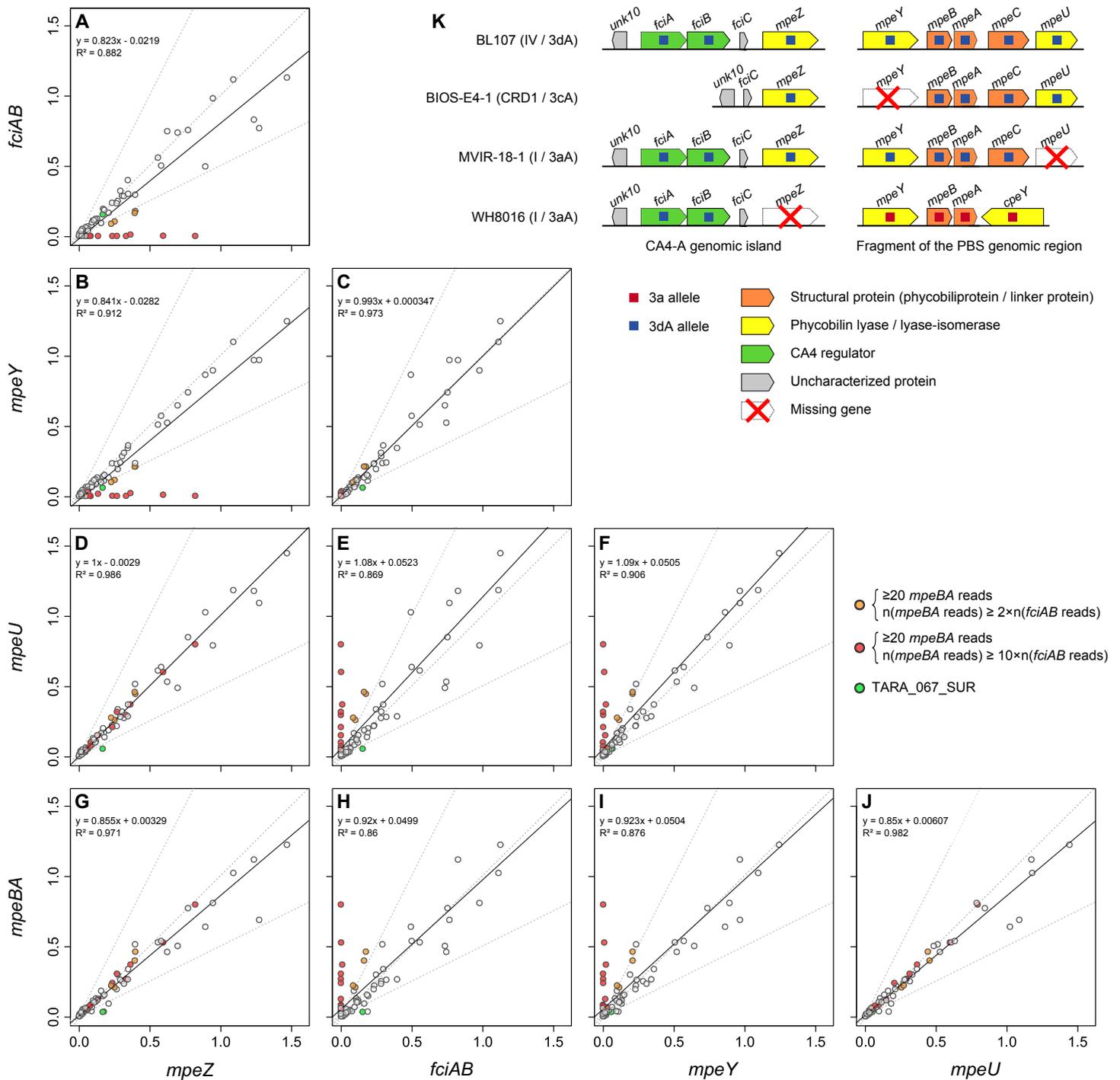




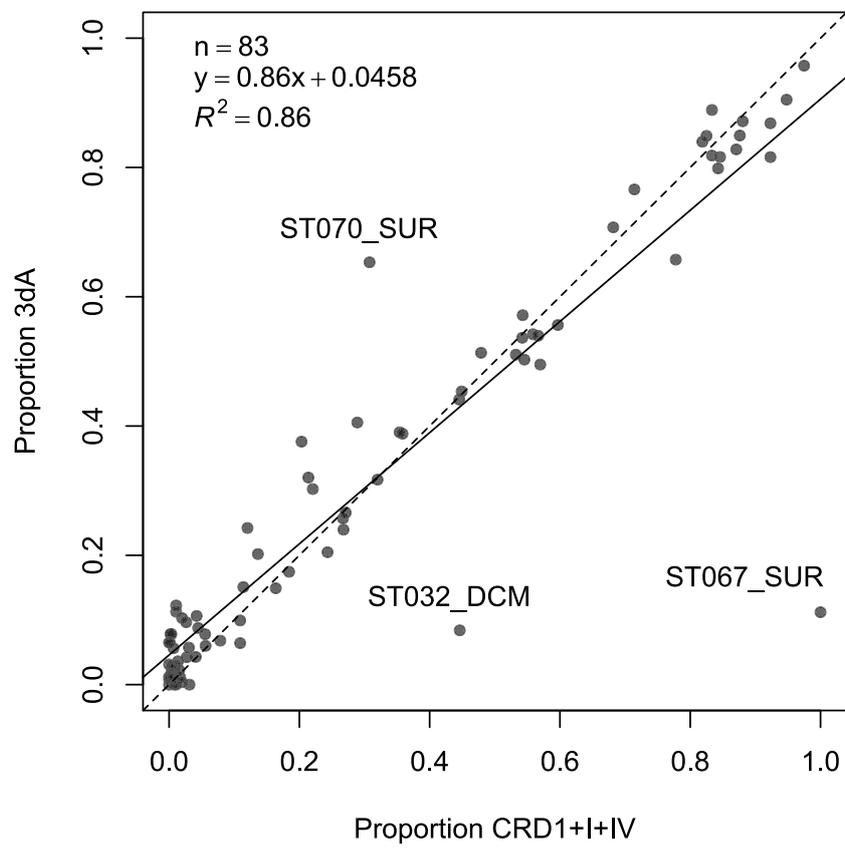
**Figure S4:** Distribution of *Synechococcus* pigment types (PTs) at depth (Deep Chlorophyll Maximum). (A) Map showing the global distribution of all *Synechococcus* PTs at depth along the *Tara* Oceans transect. Diameters of pies are proportional to the number of *cpcBA* reads normalized by the sequencing effort. Stations with less than 30 *cpcBA* or *mpeBA* reads are indicated by open circles and those with no *cpcBA* reads by black dots. Numbers next to pies correspond to *Tara* Oceans stations. (B) PTs and ESTU relative abundance at depth for sampling station along the *Tara* Oceans transect. Oceanic provinces are indicated in the top gray panels. NAO, North Atlantic Ocean; MS, Mediterranean Sea; RS, Red Sea; IO, Indian Ocean; SAO, South Atlantic Ocean; SO, Southern Ocean; SPO, South Pacific Ocean; NPO, North Pacific Ocean.



**Figure S5:** Same as Fig. 4A but for all ESTUs. Unclass., unclassified.



**Figure S6:** Focus on pigment type (PT) 3dA natural mutants, exhibiting an altered gene content with regard to typical PT 3dA. (A-J) Correlation between the number of reads assigned as PT 3dA using different markers (all present in single gene copy in typical 3dA). Each circle corresponds to a *Tara* Oceans station and depth. Orange circles: stations with at least 20 *mpeBA* reads assigned to PT 3dA and at least twice more 3dA counted with *mpeBA* than with *fciAB*, corresponding to the surface sample of stations TARA\_070, TARA\_110 and TARA\_137 and the DCM of stations TARA\_038, TARA\_058 and TARA\_110. Red circles: same but with more than 10-fold 3dA counted with *mpeBA* than *fciAB*, corresponding to the surface sample of stations TARA\_052, TARA\_094, TARA\_111 and TARA\_122 to TARA\_128, and DCM of stations TARA\_052, TARA\_100, TARA\_111 and TARA\_128. Green circle: surface of station TARA\_067. (K) CA4-A genomic island and fragment of the phycobilisome (PBS) genomic region for a typical, CA4-able 3dA strain (strain BL107), and 3 CA4-deficient strains, which are stuck either in blue light phenotype (similar to strain BIOS-E4-1), or green light phenotype (as strains MVIR-18-1 and WH8016). Note that KORDI-49 and WH8016 strains have identical PBS gene complement and genomic arrangement. The complete PBS genomic region of the BL107 strain can be found in Six *et al.*, 2007. Note that for readability, surface of station TARA\_093 has been omitted since it has the highest normalized counts (2.7-3.2) for all markers and exhibited a good agreement between markers (ratio close to 1:1).



**Figure S7:** Correlation between the proportion of clades I, IV and CRD1, as assessed with *petB*, and the proportion of pigment type 3dA, as assessed with *mpeBA*, at each station.



# CHAPTER III

---

## Genetic approaches to *Synechococcus* pigment types

« It may also come as a surprise to some phytoplanktologists and even some bacteriologists that the bacterioplankton are more than just little nondescript bags of enzymes. »

J.M. Sieburth, 1978

## I. Context of the work

The current version (v2) of the Cyanorak information system, a database and analysis platform dedicated to the comparative genomics of *Prochlorococcus* and *Synechococcus* (Dufresne *et al.*, 2008), gathers information from 97 complete genomes, of which 43 *Prochlorococcus*, 51 *Synechococcus* and 3 *Cyanobium* (Doré *et al.*, *in prep.*). Of the 27,432 predicted clusters of orthologs containing from 1 (unique) to 143 (multi-copy) sequences, 3,344 are annotated as “uncharacterized” or “unknown”, and 21,996 of the remaining 24,088 are annotated as “hypothetical” or “putative”, yet these annotations constitute a major progress compared to the one publicly available for these genomes. This exemplifies the current lack of knowledge about the exact function of the vast majority of genes in non-model organisms, which now constitutes a major bottleneck in comparative (meta)genomics. As an example, in their analysis of the ocean microbiome, Sunagawa *et al.* found that 40% of the core orthologous groups (genes found in all samples) were of unknown function, compared to only 9% in the human gut microbiome. The proportion was even greater in non-core orthologous groups (Sunagawa *et al.*, 2015). In marine *Synechococcus*, the proportion of genes of unknown function even reaches 95% for non-core genes (genes not found in every *Tara* Ocean metagenome; H. Doré, *personal communication*). Recent advances in bacterial comparative genomics foreshadow the advent of new powerful tools for statistically relating genetic variation (gene presence/absence or even allelic variation of core genes) with measureable phenotypes (qualitative traits such as antibiotic resistance and/or quantitative traits such as cell size). However, these necessitate hundreds or thousands of closely related genomes (e.g. epidemic outbreaks of pathogens such as *Staphylococcus aureus*; see Chen and Shapiro, 2015 for a concise review) and will not be amenable to such a highly diverse group as *Synechococcus* without extreme sequencing effort. Still, top-down approaches can provide important insights into the function of genes or gene variants. For *Synechococcus* and *Prochlorococcus*, such approaches (often used in combination) include comparative genomics and physiology (Dufresne *et al.*, 2008; Biller *et al.*, 2014a; Berube *et al.*, 2015), transcripts level analysis under different environmental conditions using microarrays (Tetu *et al.*, 2009; Tai *et al.*, 2009; Stuart *et al.*, 2009; Blot *et al.*, 2011; Gierga *et al.*, 2012; Dupont *et al.*, 2012; Tetu *et al.*, 2013), RT-qPCR (Mella-Flores *et al.*, 2012), and/or RNASeq (Sanfilippo *et al.*, 2016; Doron *et al.*, 2016; Thompson *et al.*, 2016), proteomics (Christie-Oleza *et al.*, 2017, 2015b, 2015a; Varkey *et al.*, 2016; Mackey *et al.*, 2017; Christie-Oleza *et al.*, 2017), as well as environmental genomics (Kashtan *et al.*, 2014). If these approaches are useful for gaining a general idea of a gene function(s) and very powerful by their high throughput, they do not allow one to fully

characterize the function of a gene. In particular, such approaches cannot describe an enzyme substrate(s) specificity and/or kinetics.

In the specific case of phycobilisome (PBS) biosynthesis, comparative genomics of strains representative of the different pigment types revealed the presence in all *Synechococcus* strains of one large genomic region that gathers most genes involved in the synthesis and regulation of PBS rods, the gene content and organization of this region being specific of the pigment type of the strain (Six *et al.*, 2007a; Humily *et al.*, 2013; see also Chapter I). Yet, many genes of the PBS region remain poorly characterized. In particular, while several putative phycobilin lyases or lyase-isomerases have been identified in this region, the specificity of several of them remains unclear because of the relatively low similarity of the different families of phycobilin lyases and the lack of knowledge on functionally important protein domains or motifs (Blot *et al.*, 2009; Schluchter *et al.*, 2010; Bretaudeau *et al.*, 2013). Moreover, most phycobilin lyases/lyase-isomerases from marine *Synechococcus* belong to the E/F structural clan (see Introduction), for which no 3D structure has been published to date, explaining why protein structure modelling for members of this clan using e.g. Phyre2 ([www.sbg.bio.ic.ac.uk/~phyre2/](http://www.sbg.bio.ic.ac.uk/~phyre2/)) gives highly variable and thus unreliable results.

Using heterologous expression in *E. coli*, Blot and co-workers showed that RpcG is a phycoerythrobilin (PEB) lyase-isomerase, i.e. an enzyme both attaching PEB and isomerizing it into PUB, specific of cysteine-83 (C83) of the  $\alpha$ -subunit of phycocyanin (PC; Blot *et al.*, 2009). Using the same approach as well as inactivation mutants in the CA4-A capable *Synechococcus* sp. RS9916, Shukla and co-workers demonstrated that MpeZ is a PEB lyase-isomerase, attaching PEB at cysteine-83 (C83) of the  $\alpha$ -subunit of phycoerythrin II ( $\alpha$ -PEII) and isomerizing it into PUB (Shukla *et al.*, 2012). Mahmoud and collaborators characterization of *mpeU* inactivation mutants revealed that this gene is necessary for optimally absorbing BL, suggesting that it probably encodes a phycobilin lyase-isomerase (Mahmoud *et al.*, 2017). Finally, in a yet unpublished study, Sanfilippo and co-workers characterized MpeY in RS9916, revealing that it is likely a phycobilin-lyase acting on the same site as MpeZ ( $\alpha$ -PEII C83), and that the chromophore ultimately attached at C83 can be predicted from the transcript ratio of *mpeZ* and *mpeY* (Sanfilippo *et al.*, *in prep.*). This led them to propose a model in which the two enzymes compete for the same site. Several other lyases from other cyanobacteria, including *Prochlorococcus*, freshwater cyanobacteria or even cyanophages have been characterized and have homologs in marine *Synechococcus* that are generally considered to have the same specificity as their characterized counterparts (Schluchter *et al.*, 2010; Bretaudeau *et al.*, 2013).

Altogether, among the up to 12-13 phycobilin lyases encoded in *Synechococcus* genomes, less than half have been genetically and/or biochemically characterized, and none of

these is involved in CA4-B. The objectives of this chapter were thus to i) functionally characterize the putative phycobilin-lyases *mpeW* and *mpeY* in a CA4-B strain and unveil the role of these enzymes in the CA4-B process (Humily *et al.*, 2013), a work presented hereafter as a manuscript (Grébert *et al.*, *in prep.*) that should be submitted to the *Journal of Biological Chemistry* as soon as the few missing data are available (see II.2 of this chapter); and ii) to experimentally test hypotheses about the evolution and lateral transfer of CA4 between different *Synechococcus* backgrounds, a work still in progress that is presented as a suite of results (see III of this chapter). It will be completed in the months to come and will should be valorized as a collaborative paper led by David M. Kehoe's team (Indiana University, Bloomington, IN, USA).

The work presented in this chapter has been made possible thanks to an international collaboration with Prof. David M. Kehoe (Kehoe lab, Indiana University, Bloomington, IN, USA), in whose lab I spent 5 months (Nov. 2016-March 2017) supported by a Fulbright grant, Prof. Wendy M. Schluchter Lab at the University of New Orleans (New Orleans, LA, USA) and the Marine Photosynthetic Prokaryote (MaPP) team at the Station Biologique de Roscoff (UMR7144 CNRS/UPMC, Station Biologique, Roscoff, France) in which I did most of my thesis. Prof. David M. Kehoe is a specialist of chromatic acclimation and light-related processes in cyanobacteria and an expert in genetics of cyanobacteria. During my stay in his lab, I personally conducted under his supervision and with expert advices from his post-doctoral fellow Dr. Bo Chen the gene inactivation described in the first part of this chapter as well as the artificial lateral transfer of a CA4-B island into PT 3c strains (hereafter dubbed the “FrankenSyn” experiment, after the *Frankenstein* novel from Mary Shelley) described in the second part of this chapter. Prof. Wendy M. Schluchter is a specialist of the biochemistry of phycobilin-lyases and phycobiliprotein synthesis in freshwater (*F. diplosiphon*, *Synechococcus* sp.) and marine (*Synechococcus* sp.) cyanobacteria. Suman Pokhrel and Adam A. Nguyen, two PhD students from her lab, conducted all heterologous expression experiments, as well as all purification of phycobiliproteins by HPLC, spectral characterization and digestion of phycobiliprotein subunits evoked in the first part of this chapter. The collaboration also involved the mass spectrometry platform of Indiana University Bloomington led by Dr. Jonathan A. Karty, who developed a specific LC/UV-Vis/MS-MS method for the analysis of the chromophorylation of phycobiliproteins and analyzed the phycobiliprotein subunits separated by the Schluchter Lab. In Roscoff, I personally performed the extraction of phycobilisomes from wild-type and mutants *Synechococcus* and I performed comparative genomic analyses. Additionally, in collaboration with Morgane Ratin (Research Engineer in the MaPP team) I performed the mutant complementation experiments, all of which are described in the first part of this chapter and were made under the supervision of Drs. Laurence Garczarek and Frédéric Partensky.

## **II. Characterization of two enzymes involved in phycoerythrin-II chromophorylation**

### **1. Functional characterization of the MpeWYZ phycobilin lyase family provides key insights into color niche acclimation and adaptation in marine *Synechococcus***

#### **Functional characterization of the MpeWYZ phycobilin lyase family provides key insights into color niche acclimation and adaptation in marine *Synechococcus***

Théophile Grébert<sup>1</sup>, Adam A. Nguyen<sup>3,4</sup>, Suman Pokhrel<sup>3,4</sup>, Bo Chen<sup>2</sup>, Morgane Ratin<sup>1</sup>, Jonathan A. Karty<sup>5</sup>, Laurence Garczarek<sup>1</sup>, Wendy M. Schluchter<sup>3,4</sup>, David M. Kehoe<sup>2</sup>, Frédéric Partensky<sup>1</sup>

<sup>1</sup>Sorbonne Universités, Université Pierre et Marie Curie University Paris 06, CNRS, UMR 7144, Station Biologique, Plankton Group, 29688 Roscoff, France;

<sup>2</sup>Department of Biology, Indiana University, Bloomington, Indiana, 47405 U.S.A.;

<sup>3</sup>Department of Biological Sciences, University of New Orleans, New Orleans, LA 70148;

<sup>4</sup>Department of Chemistry, University of New Orleans, New Orleans, LA 70148;

<sup>5</sup>METACyt Biochemical Analysis Center, Department of Chemistry Indiana University, Bloomington, Indiana 47405 U.S.A.

IN PREPARATION



**Abstract**

Marine *Synechococcus* play a central role in oceanic ecosystems and primary production. Their ecological success owe in part to their ability to colonize and exploit a large array of light spectral niches, thanks to a wide range of photosynthetic pigmentations. Within the globally dominant pigment type 3 (PT 3), some subtypes specialized in harvesting green or blue light, whereas other developed the ability to dynamically modify their light absorption spectrum to optimally collect either colour. This process called Type IV Chromatic Acclimation (CA4) has been linked to the occurrence of a small genomic island existing in two configurations (CA4-A and -B). As the enzymes covalently attaching chromophores on phycobiliproteins, phycobilin lyases are the fundamental determinants of pigmentation. In particular, the MpeWYZ family appears pivotal as MpeY is the only lyase found in all PT3 and its paralogs MpeW and MpeZ are specifically found in the CA4 genomic island. Here we characterized two members of this family and demonstrated their critical role in CA4-B. While MpeW attaches the green-light absorbing phycoerythrobilin to cysteine-83 of the  $\alpha$ -subunit of phycoerythrin II, MpeY additionally isomerizes it into the blue-light absorbing phycourobilin. Together with the previous characterization of MpeZ and MpeY in a CA4-A strain, our results uncover the functional diversity of the MpeWYZ phycobilin lyase family, providing important insights into the molecular mechanisms controlling both adaptation and acclimation to light colour in marine *Synechococcus*, a key ecological trait of this ubiquitous phytoplankton organism.

## Introduction

Marine *Synechococcus* are the second most abundant phototrophs in the oceans, significantly contributing to oceanic primary production and carbon cycling (Flombaum *et al.*, 2013; Guidi *et al.*, 2016). These picocyanobacteria are found from the equator up to 80°N, from the coast to the open ocean and over a wide part of the lit layer (Partensky *et al.*, 1999; Zwirgmaier *et al.*, 2008; Paulsen *et al.*, 2016). They exhibit a broad range of photosynthetic pigments, allowing them to optimally exploit this large variety of light environments (Six *et al.*, 2007b). This diversity comes from differences in the composition of their light-harvesting antennae called phycobilisomes (PBS). These complexes are made of chromophorylated phycobiliproteins that form a central core and six or eight radiating rods. The latter consist of  $\alpha/\beta$  heterodimers of phycobiliproteins that assemble into  $(\alpha\beta)_6$  hexamers after chromophorylation by phycobilin lyases. Hexamers are then stacked into rods thanks to linker proteins. In pigment type (PT) 1, rods are only made of phycocyanin (PC) and only bear the red-light absorbing phycocyanobilin (PCB,  $A_{\max}=650$  nm), whereas in PT 2 they are made of both PC and phycoerythrin I (PE-I), bearing both PCB the green-light (GL) absorbing phycoerythrobilin (PEB,  $A_{\max}=550$  nm; Six *et al.*, 2007b; Ong and Glazer, 1991). However, most oceanic *Synechococcus* correspond to PT 3 (Grébert *et al.*, in revision; see Chapter II), in which rods are made of PC, PE-I and phycoerythrin-II (PE-II) and bear PCB, PEB and the blue-light (BL) absorbing phycourobilin (PUB,  $A_{\max}=495$  nm; Ong *et al.*, 1984). PEB and PUB are quantitatively the most abundant chromophores in PT 3, and different subtypes have been defined based on the relative ratio of fluorescence excitation at 495 nm and 545 nm with emission at 580 nm ( $\text{Exc}_{495:545}$ ), a proxy for their PUB:PEB ratio. This ratio can be low ( $\text{Exc}_{495:545}<0.6$ ), intermediate ( $0.6\leq\text{Exc}_{495:545}<1.6$ ) or high ( $1.6\leq\text{Exc}_{495:545}$ ), corresponding respectively to PT 3a, 3b and 3c (Six *et al.*, 2007b). Genes coding for PBS rod components are gathered into a specific genomic region, hereafter the PBS region, whose gene content and organization has been shown to be specific of these PTs (Six *et al.*, 2007b). Finally, some strains, corresponding to PT 3d, have a  $\text{Exc}_{495:545}$  that is variable with ambient light color, a process called type IV acclimation (CA4) that allows them to optimally absorb GL or BL depending on which light color predominates (Palenik, 2001; Everroad *et al.*, 2006). Specifically, three out of eleven chromophore-binding cysteine residues on PE-II and PE-I bear PEB in GL and PUB in BL in the CA4-A strain *Synechococcus* sp. RS9916 (Shukla *et al.*, 2012). CA4 is widespread among *Synechococcus* populations, and has been linked with the occurrence of a small specific genomic island (GI) existing in two distinct configurations (Humily *et al.*, 2013; Grébert *et al.*, in revision). Each version contains the two regulators *fcIA/B* and one member of the phycobilin lyase family, *mpeZ* or *mpeW*, defining the two distinct genotypes CA4-A and CA4-B, respectively (Humily *et al.*, 2013).

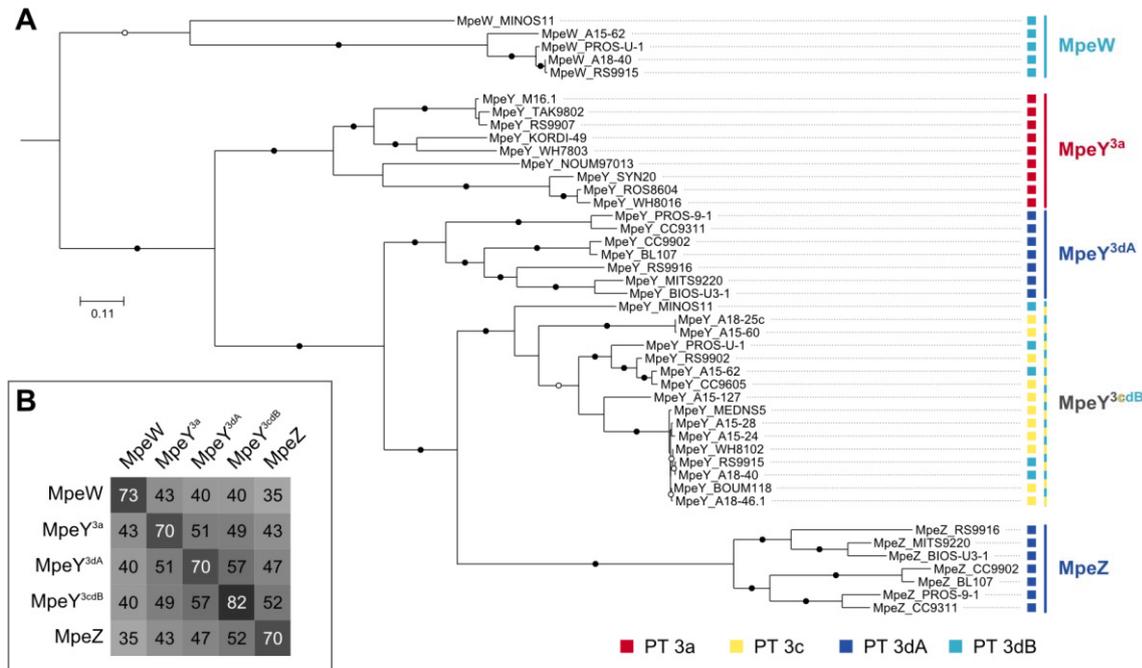
Phycobilin lyases are enzymes responsible for the covalent attachment of chromophores at specific cysteine residues on phycobiliprotein subunits and are thus key players in *Synechococcus* pigmentation. Indeed, chromophorylation is necessary for phycobiliprotein assembly, and a proper phycobilin arrangement is crucial for efficient energy transfer along the rods (Glazer, 1989). However, among the dozen (putative) phycobilin lyases present in marine *Synechococcus* genomes, only a handful have been biochemically characterized (Six *et al.*, 2007b). RpcG from *Synechococcus* sp. WH8102 (PT 3c) was the first one, and was shown to concomitantly attach PEB to cysteine 84 (C84) of the  $\alpha$ -subunit of PC ( $\alpha$ PC) and to isomerize it into PUB (Blot *et al.*, 2009). Such dual-function enzymes are called phycobilin lyase-isomerases. Hereafter, we specifically refer to lyase-isomerases for enzymes attaching PEB and isomerizing it into PUB and lyase for enzymes attaching PEB, both types belonging to the phycobilin lyase family. More recently, two enzymes implied in CA4-A have been characterized in the 3dA strain RS9916. MpeZ is a phycobilin lyase-isomerase adding PUB at C83 of  $\alpha$ PE-II (Shukla *et al.*, 2012), whereas MpeY was recently shown to be a phycobilin lyase acting on the same site (Sanfilippo *et al.*, in prep). Whereas *mpeZ* is more expressed under BL through the action of *fciA/B* (Humily *et al.*, 2013; Shukla *et al.*, 2012; Sanfilippo *et al.*, 2016), *mpeY* appears to be constitutively expressed whatever the light color, the cellular ratio of the two enzymes ultimately controlling the chromophore bound at  $\alpha$ PE-II C83 (Sanfilippo *et al.*, in prep). On the contrary, the function of enzymes involved in CA4-B was so far much less understood.

Here we demonstrate that in 3dB strains, *mpeW* encodes a phycobilin lyase acting on C83 of  $\alpha$ PE-II and is coupled with an allele of *mpeY* coding for a lyase-isomerase specific of the same cysteine position, revealing the functional heterogeneity of the different *mpeY* alleles. Altogether, our results highlight the importance of the MpeWYZ family in the adaptation to different light color niches.

## Results

Based on its conserved position in the PBS region, the gene exhibiting typical HEAT lyase repeat domains and located upstream of the operon coding for PEII subunits has been called *mpeY* in all *Synechococcus* PT 3 (Six *et al.*, 2007b). A phylogenetic analysis of the different members of the MpeWYZ family revealed that this gene is in fact polyphyletic, with 3 clear-cut clusters, each corresponding to a different PT (Fig. 1A; Grébert *et al.*, in revision): PT 3a (hereafter allele *mpeY*<sup>3a</sup>), PT 3dA (*mpeY*<sup>3dA</sup>) and the undifferentiated PT 3c and 3dB, which possess the same allele (*mpeY*<sup>3cdB</sup>). MpeW, MpeZ and the different MpeY alleles have on average 35-57% amino acid identity over their whole length, compared to an average of 70-82% identity within each allele (MpeY) or gene (MpeW and MpeZ; Fig. 1B). Since the biochemical

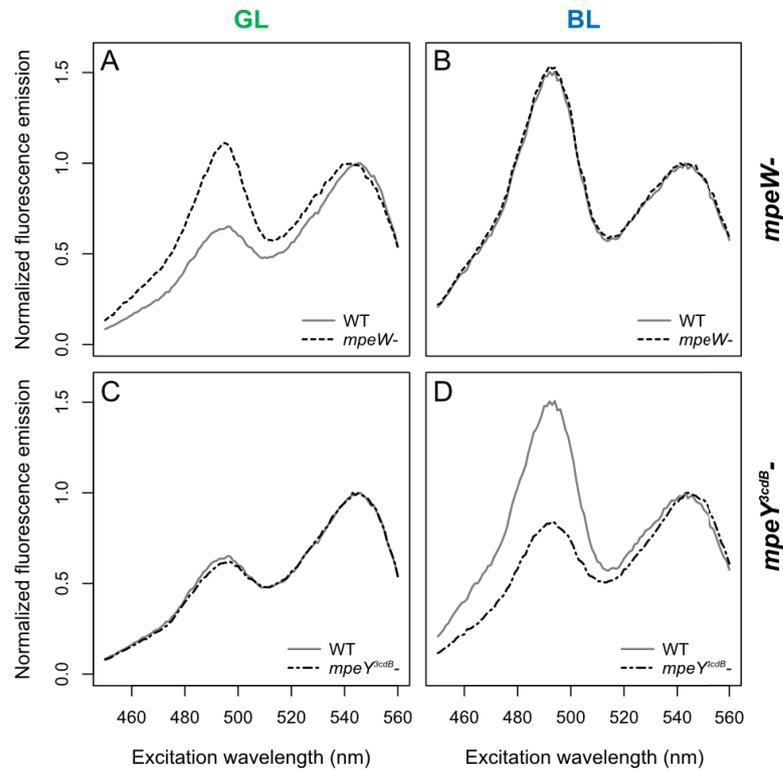
characterization of MpeZ and MpeY<sup>3dA</sup> showed that they are respectively a lyase-isomerase and a lyase both acting on C83 of  $\alpha$ -PEII couple (Shukla *et al.*, 2012), we hypothesized that in CA4-B MpeW is a C83  $\alpha$ -PEII PEB lyase and MpeY<sup>3cdB</sup> a C83  $\alpha$ -PEII PEB lyase-isomerase, i.e. that the different MpeY alleles have distinct lyase or lyase-isomerase functions.



**Fig. 1: Diversity of the MpeWYZ phycobilin lyase family.** (A) Maximum likelihood phylogeny (protein sequences, LG+I+G+F model). CpeY sequences were used as outgroup to root the tree. Nodes with bootstrap support >70% and >90% are indicated by empty and filled black circles, respectively. (B) Mean pairwise percentage of identity between the protein sequences of the different genes and alleles, calculated based on full-length alignments.

We created *mpeW* and *mpeY<sup>3cdB</sup>* insertion mutants in the CA4-B strain *Synechococcus* sp. A15-62 (hereafter A15-62). Comparison of fluorescence excitation spectra revealed that for GL-grown cultures,  $\text{Exc}_{495:545}$  was increased in the *mpeW*- mutant compared to the wild-type (WT), whereas no difference could be seen under BL (Fig. 2A,B). Conversely, no difference could be seen between  $\text{Exc}_{495:545}$  of *mpeY<sup>3cdB</sup>*- and WT under GL, while the  $\text{Exc}_{495:545}$  ratio was decreased in BL-grown *mpeY<sup>3cdB</sup>*- compared to WT (Fig. 2C,D). It is worth noting that both mutants still exhibited some levels of chromatic acclimation, with an  $\text{Exc}_{495:545}$  varying from 1.53 in BL to 1.11 in GL for *mpeW*- and from 0.62 in GL to 0.84 in BL for *mpeY<sup>3cdB</sup>*- (compared to 0.65 and 1.51 for the WT in GL and BL respectively). This represents ~40% and ~26% of the variation observed in the WT for *mpeW*- and *mpeY<sup>3cdB</sup>*- respectively (Fig. 2). Based on these results, we hypothesized that *mpeY<sup>3cdB</sup>* is required for PUB attachment, while *mpeW* would be

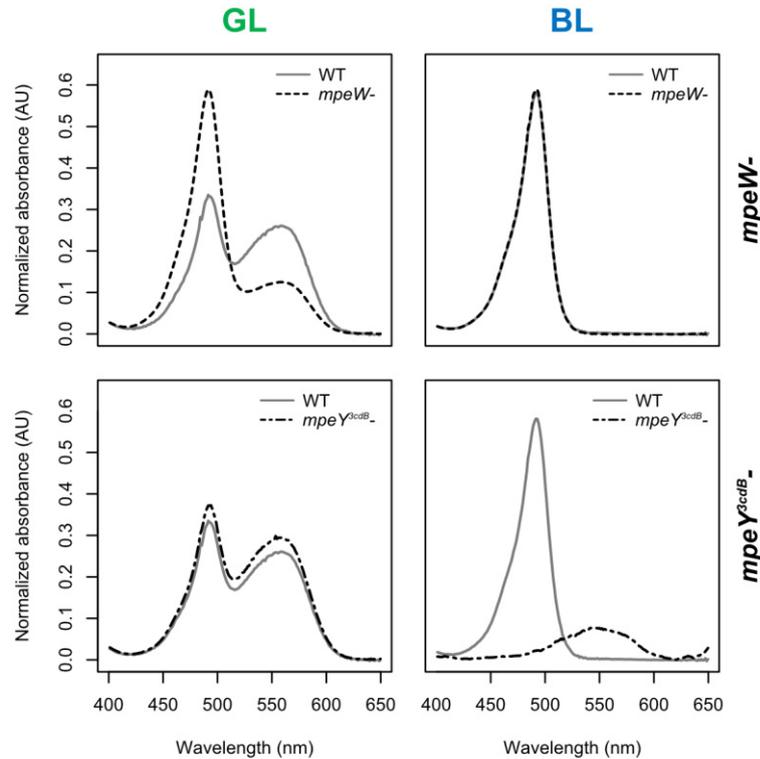
required for PEB attachment. To identify the biochemical changes responsible for the light-absorption differences, PBS were purified from WT and mutants and the PE-I and PE-II  $\alpha$  and  $\beta$  subunits separated.



**Fig. 2: Whole-cell fluorescence excitation spectra (emission set at 580 nm) for *Synechococcus* sp. A15-62 wild-type (WT), *mpeW* inactivation mutant (*mpeW*<sup>-</sup>) and *mpeY*<sup>3cdB</sup> inactivation mutant (*mpeY*<sup>3cdB</sup><sup>-</sup>).** (A, B) WT and *mpeW*<sup>-</sup>; (C, D) WT and *mpeY*<sup>3cdB</sup><sup>-</sup>. Cultures were grown under green light (A, C) or blue light (B, D). Fluorescence emission was normalized at 545 nm.

The absorption spectra for individual PE-I and PE-II subunits from A15-62 WT and *mpeW*<sup>-</sup> and *mpeY*<sup>3cdB</sup><sup>-</sup> mutants only differed for the  $\alpha$ -PEII subunit (MpeA). MpeA from the *mpeW*<sup>-</sup> mutant had a lowered PEB peak (550 nm), whereas MpeA from the *mpeY*<sup>3cdB</sup><sup>-</sup> mutant had a lowered PUB peak (495 nm) compared to the WT (Fig. 3). The absorption spectra of other phycobiliprotein subunits were similar (Fig. S1). This demonstrates that both MpeW and MpeY<sup>3cdB</sup> are involved in  $\alpha$ -PEII (MpeA) chromophorylation, with MpeW being necessary for PEB attachment and MpeY<sup>3cdB</sup> for PUB attachment. The absorption spectra for individual PE-I and PE-II subunits of A15-62 WT perfectly matched the one for RS9916 WT subunits in both BL and GL. Furthermore, PE subunits isolated from the *mpeW*<sup>-</sup> mutant presented the same absorption spectra as PE subunits from the RS9916 *mpeY*<sup>3dA</sup><sup>-</sup> mutant, whereas absorption spectra from subunits from the *mpeY*<sup>3cdB</sup><sup>-</sup> mutant matched absorption spectra for RS9916 *mpeZ* mutant.

This suggests that the PE chromophorylation is identical in CA4-A and CA4-B, and that MpeY<sup>3cdB</sup> and MpeW act on the same site as MpeZ and MpeY3dA, i.e.  $\alpha$ -PEII C83 (Shukla *et al.*, 2012; Sanfilippo *et al.*, in prep). Altogether, this suggest that MpeW is necessary for PEB attachment on  $\alpha$ -PEII C83, and conversely MpeY<sup>3cdB</sup> is required for PEB attachment and isomerization into PUB at the same site.



**Fig. 3: Absorbance spectra of HPLC-purified  $\alpha$ -PE-II subunits (MpeA) for *Synechococcus* sp. A15-62 wild-type (WT), *mpeW* inactivation mutant (*mpeW*-) and *mpeY<sup>3cdB</sup>* inactivation mutant (*mpeY<sup>3cdB</sup>*-) grown under GL or BL. See Fig. S1 for absorbance spectra of other phycoerythrin subunits.**

The activity of A15-62 MpeW and MpeY<sup>3cdB</sup> was then tested by heterologous expression in *Escherichia coli* along with genes necessary for PEB biosynthesis (*pebS* and *hol*, Shukla *et al.*, 2012) and six-histidine-tagged (HT) potential phycobiliprotein substrates: MpeA (PEII  $\alpha$ -subunit), MpeB (PEII  $\beta$ -subunit) and CpeA (PEI  $\alpha$ -subunit) cloned from the same *Synechococcus* strain. This showed that MpeW is a PEB lyase, whereas MpeY<sup>3cdB</sup> is a PEB lyase-isomerase, both specific of MpeA (S. Pokhrel, A. Nguyen and W. M. Schluchter, unpublished). Heterologous expression with site-directed mutants of MpeA from *Synechococcus* sp. RS9916 further revealed that both MpeW and MpeY<sup>3cdB</sup> are specific for C83  $\alpha$ -PEII (S. Pokhrel, A. Nguyen and W. M. Schluchter, unpublished).

## Discussion

Marine *Synechococcus* PT 3 are the only cyanobacteria to possess PE-II, a property that confers them a remarkable diversity of pigment content. The ecological success of this genus is closely tied with this innovation, as evidenced by the predominance and ubiquity of PT 3 in the world ocean, an environment dominated by either blue or green light (Grébert *et al.*, in revision). Acquisition of PE-II likely occurred by duplication and divergence of the corresponding PE-I subunits, but also implied the concomitant acquisition of enzymes necessary for proper PE-II chromophorylation (Six *et al.*, 2007b; Everroad and Wood, 2012; Grébert *et al.*, in prep, see Chapter D). In this context, *mpeY* appears to be pivotal as the only phycobilin lyase specific to all PE-II-containing *Synechococcus* (Six *et al.*, 2007b; Grébert *et al.*, in prep). In addition, the two *mpeY* homologs, *mpeW* and *mpeZ*, have been linked with CA4, a process allowing cells to acclimate to green and blue light that is widespread among *Synechococcus* populations, further highlighting the key importance of this protein family for *Synechococcus* pigmentation and ecology (Humily *et al.*, 2013; Grébert *et al.*, in revision).

Here, we show that MpeW, specific of CA4-B, is a phycobilin lyase attaching PEB at C83  $\alpha$ PE-II, and that MpeY<sup>3cdB</sup>, found in both CA4-B and 3c, is a phycobilin lyase-isomerase that binds PUB at the same position. We further demonstrated that both enzymes are necessary for CA4-B. Our results imply that all members of the MpeWYZ enzyme family are specifically acting on C83  $\alpha$ PE-II (Shukla *et al.*, 2012; Sanfilippo *et al.*, in prep). Surprisingly, the MpeY sub-family appeared to be functionally heterogeneous: even though all MpeY have the same site specificity, MpeY<sup>3a</sup> and MpeY<sup>3dA</sup> are phycobilin lyases, whereas MpeY<sup>3cdB</sup> are phycobilin lyase-isomerases. Recently, Larkin and Martiny described two microbial diversification processes: the “Renaissance model”, corresponding to the acquisition of a new trait axis (e.g., ability to grow using a new carbon source for a heterotroph) and the “Maestro model”, corresponding to a shift of the growth optima along a pre-existing trait (e.g., the optimal irradiance growth of a photosynthetic organism; Larkin and Martiny, 2017). The functional diversification observed here fits very well with the latter model, with different alleles of MpeY exhibiting a relatively low level of sequence divergence, resulting in large phenotypic changes (*i.e.*, best absorption of GL or BL), which in turn strongly affect the ecology of this organism (Grébert *et al.*, in revision). It was previously shown that phycobilin lyases can be split into three large clans (*i.e.*, E/F, T and S/U clans; Schlachter *et al.*, 2010; Bretaudeau *et al.*, 2013). All members of the T and S/U clans have the same specificity for a given conserved chromophore-binding cysteine residue ( $\beta$ -155 and  $\beta$ -82, respectively), and the different members of each clan target different phycobiliproteins (e.g., *cpcT* and *cpeT* respectively target C82  $\beta$ -PC and C82  $\beta$ -PEI; Gasper *et al.*, 2017; Zhao *et al.*, 2007). Similarly, RpcG, CpeY and MpeWYZ are members of the E/F clan all acting on  $\alpha$ -82

of respectively PC, PE-I and PE-II (Blot *et al.*, 2009; Bretaudeau *et al.*, 2013; Schluchter *et al.*, 2010). Here, we show that difference in activity (lyase vs lyase-isomerase) may arise at lower evolutionary distance between close homologs. This suggests a nested adaptation of phycobilin lyases, with first adaptation to a site (clan or superfamily), then to a phycobiliprotein (gene family), and eventually a fine-tuning of activity as either lyase or isomerase (alleles).

Another major outcome of this study is to highlight the functional and evolutionary differences between CA4-A and CA4-B processes. It has recently been shown that in CA4-A, the lyase MpeY<sup>3dA</sup> and the lyase-isomerase MpeZ compete for the chromophorylation of  $\alpha$ PE-II C83 (Sanfilippo *et al.*, in prep). While *mpeY*<sup>3dA</sup> is constitutively expressed whatever the light color, *mpeZ* synthesis is induced upon BL illumination, the transcript ratio of the two enzymes ultimately controlling chromophorylation at  $\alpha$ PE-II C83 (Humily *et al.* 2013; Sanfilippo *et al.*, in prep). Our results suggest that CA4-B is phenotypically but also mechanistically similar to CA4-A, with the lyase MpeW and the lyase-isomerase MpeY<sup>3cdB</sup> competing for the chromophorylation of  $\alpha$ PE-II C83, with the difference of *mpeW* being induced upon GL illumination (Humily *et al.*, 2013). This provides a possible evolutionary scenario for the occurrence of two CA4 types in the *Synechococcus* radiation. We hypothesize that the 3dA “basal state” corresponds to a GL specialist phenotype and that CA4-A confers cells the ability to change their spectral properties upon BL illumination by inducing the expression of the phycobilin lyase-isomerase MpeZ. Conversely, the 3dB “basal state” would correspond to a BL specialist phenotype, CA4-B providing cells the ability to acclimate to GL by inducing the synthesis of the phycobilin lyase MpeW (Humily *et al.*, 2013; Sanfilippo *et al.*, 2016; Sanfilippo *et al.*, in prep). Thus, in both cases, CA4 appears to be a “plug-and-play” mechanism that comes on top of the existing chromophorylation. This view is further supported by the gathering of genes necessary for CA4 (regulators and phycobilin lyases) in a small dedicated genomic island (Sanfilippo *et al.*, 2016; Humily *et al.*, 2013; Shukla *et al.*, 2012), and the absence of any PBS gene content or allelic difference between PTs 3c and 3dB, (Fig. 1; Grébert *et al.*, in prep; Grébert *et al.*, submitted, see Chapters I and II), suggesting that CA4-B might be more easily transferred through horizontal gene transfer than CA4-A. It is also interesting to note that, as previously observed for *mpeZ* and *mpeY*<sup>3dA</sup> mutants in the CA4-A strain RS9916 (Shukla *et al.*, 2012; Sanfilippo *et al.*, in prep), both *mpeW* and *mpeY*<sup>3cdB</sup> CA4-B mutants still presented some degrees of chromatic acclimation (Fig. 2). In both lyase-isomerase mutants (MpeZ in CA4-A and MpeY<sup>3cdB</sup> in CA4-B), *Exc*<sub>495:545</sub> is strongly decreased in BL (~0.8) compared to the WT (~1.5-1.6), but higher than in GL (~0.6 for both mutants and WT; Fig. 2C, D; Shukla *et al.*, 2012). Similarly, both lyase mutants (MpeY<sup>3dA</sup> in CA4-A and MpeW in CA4-B) had an increased *Exc*<sub>495:545</sub> in GL relative to the WT (~1.1 vs ~0.6), but still lower than under BL (1.5-1.6 for both; Fig.2A,B; Sanfilippo *et al.*, in prep). This suggests that other genes are involved in CA4. In the absence of any obvious other CA4-specific

phycobilin lyase, an interesting candidate could be the conserved hypothetical gene *unk10*, which is present in both versions of the CA4 genomic island and strongly overexpressed under BL in the CA4-A strain RS9916 (Humily *et al.*, 2013; Shukla *et al.*, 2012).

The discovery that all members of such a large phycobilin lyase gene family (MpeWZ and 3 MpeY alleles) act on the same site of the  $\alpha$ -PEII subunit, also has consequences on the search for potential candidates for the chromophorylation of other phycobiliprotein sites, for which no lyase have been characterized yet. Indeed, comparative genomics only revealed a limited set of candidates lyase genes that is seemingly insufficient to explain all possible combinations of sites ( $\alpha$ 75,  $\alpha$ 82,  $\alpha$ 140,  $\beta$ 50/61,  $\beta$ 82,  $\beta$ 159), phycobiliproteins (PC, PE-I, PE-II) and phycobilins (PCB, PEB, PUB) observed in marine *Synechococcus* (Six *et al.*, 2007b). Results from this study suggest that limited allelic variations could be sufficient to confer a novel function (isomerase) to a lyase, a finding that could explain how such a restricted set of candidate genes can generate all these combinations. An alternative hypothesis would be that some of the chromophores attachment could be autocatalytic or involving proteins with no homology to known lyases. Future characterization of lyase candidates or conserved hypothetical proteins encoded in the PBS genomic regions from different pigment types will help disentangling the complex mechanisms behind *Synechococcus* diverse pigmentations, a trait broadly impacting the ecology of this key phytoplankton.

## Materials and methods

### *Strains and growth conditions*

*Synechococcus* strain A15-62 was obtained from the RCC (RCC number 2374). It was originally isolated near Cape Verde (Atlantic Ocean) from 30m deep (Mazard *et al.*, 2012). Both wild-type and mutant strains were grown at 22°C in PCR-S11 (ref) in clear polystyrene CytoOne (StartLab, USA) flasks with 20-30  $\mu$ E white light. A15-62 mutants were maintained with 50  $\mu$ g/mL kanamycin. Strains were acclimated for at least two weeks under 20  $\mu$ E. m-2.s-1 blue light (BL) or green light (GL) provided by LED ramps (Luxeon Rebel LED LXML-PB01-0040 and LXML-PM01-0100 for BL and GL, respectively; Alpheus, France).

### *Plasmid construction*

Plasmids used are listed in Table S1 and primers in Table S2. pCH02, pTG\_A15-62\_mpeW and pTG\_A15-62\_mpeY were made by PCR amplification of a 820-885 bp internal fragment of RS9915 mpeW, A15-62 mpeW and A15-62 mpeY using primers pairs Syn\_RS9915\_mpeW\_143F\_EcoRI/962R\_EcoRI, Syn\_A15-62\_mpeW\_126F\_BamHI/999R\_SphI

and Syn\_A15-62\_mpeY\_123F\_EcoRI/983R\_SphI. PCR fragments were inserted into similarly cut pMUT100 that was either extracted from DH5 $\alpha$  *E. coli* (pCH02) or PCR-amplified using primer pairs pMUT100\_1291F\_SphI/5535R\_BamHI (pTG\_A15-62\_mpeW) or pMUT100\_1291F\_SphI/5535R\_EcoRI (pTG\_A15-62\_mpeY).

#### *mpeW and mpeY disruption*

Conjugation between *E. coli* MC1061 containing either pTG\_A15-62\_mpeW or pTG\_A15-62\_mpeY and *Synechococcus* A15-62 was performed as previously described (Shukla *et al.*, 2012; Brahamsha, 1996). Individual colonies were picked and tested for *mpeW* or *mpeY* disruption by PCR amplification.

#### *PBS isolation*

Phycobilisomes were purified as previously described (Six *et al.*, 2007a, 2005), starting from 10L of cultures. Briefly, cells were harvested by centrifugation, washed twice and resuspended in 0.65M phosphate buffer, and broken twice using a French press system. Membranes and hydrophobic pigments were removed by 1h incubation with 5% w/v Triton X-100 and centrifugation. Red aqueous layer was loaded onto 0.25-1.0M discontinuous sucrose gradient in phosphate buffer, and centrifuged overnight at 22,500 rpm in a SW28 rotor (Beckman Coulter) at 22°C. Colored bands were collected and frozen at -20°C until analysis. All steps were carried at room temperature. HPLC separation of phycobiliprotein was performed as described previously (Shukla *et al.*, 2012).

#### *Sequences used in this study*

Sequences used in this study (MpeW, Y, Z) can be found in Table S3.

**References**

- Blot N, Wu X-J, Thomas J-C, Zhang J, Garczarek L, Böhm S, *et al.* (2009). Phycourobilin in trichromatic phycocyanin from oceanic cyanobacteria is formed post-translationally by a phycoerythrobilin lyase-isomerase. *J Biol Chem* **284**: 9290–9298.
- Brahamsha B. (1996). A genetic manipulation system for oceanic cyanobacteria of the genus *Synechococcus*. *Appl Environ Microbiol* **62**: 1747–1751.
- Breitaudeau A, Coste F, Humily F, Garczarek L, Le Corguillé G, Six C, *et al.* (2013). CyanoLyase: a database of phycobilin lyase sequences, motifs and functions. *Nucleic Acids Res* **41**: D396-401.
- Everroad C, Six C, Partensky F, Thomas J-C, Holtzendorff J, Wood AM. (2006). Biochemical cases of type IV chromatic adaptation in marine *Synechococcus* spp. *J Bacteriol* **188**: 3345–3356.
- Everroad RC, Wood AM. (2012). Phycoerythrin evolution and diversification of spectral phenotype in marine *Synechococcus* and related picocyanobacteria. *Mol Phylogenet Evol* **64**: 381–392.
- Flombaum P, Gallegos JL, Gordillo RA, Rincón J, Zabala LL, Jiao N, *et al.* (2013). Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc Natl Acad Sci* **110**: 9824–9829.
- Gasper R, Schwach J, Hartmann J, Holtkamp A, Wiethaus J, Riedel N, *et al.* (2017). Auxiliary metabolic genes - distinct features of cyanophage-encoded T-type phycobiliprotein lyase  $\theta$ CpeT. *J Biol Chem* jbc.M116.769703.
- Glazer AN. (1989). Light guides. Directional energy transfer in a photosynthetic antenna. *J Biol Chem* **264**: 1–4.
- Grébert T, Doré H, Partensky F, Farrant GK, Emmanuel B, Picheral M, *et al.* (in revision). Light color acclimation: a key process in the global ocean distribution of *Synechococcus* cyanobacteria. *Proc Natl Acad Sci*.
- 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**: 465–470.
- Humily F, Partensky F, Six C, Farrant GK, Ratin M, Marie D, *et al.* (2013). A gene island with two possible configurations is involved in chromatic acclimation in marine *Synechococcus*. *PLoS ONE* **8**: e84459.

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

Mazard S, Ostrowski M, Partensky F, Scanlan DJ. (2012). Multi-locus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*. *Environ Microbiol* **14**: 372–386.

Ong LJ, Glazer AN. (1991). Phycoerythrins of marine unicellular cyanobacteria. I. Bilin types and locations and energy transfer pathways in *Synechococcus* spp. phycoerythrins. *J Biol Chem* **266**: 9515–9527.

Ong LJ, Glazer AN, Waterbury JB. (1984). An unusual phycoerythrin from a marine cyanobacterium. *Science* **224**: 80–83.

Palenik B. (2001). Chromatic adaptation in marine *Synechococcus* strains. *Appl Environ Microbiol* **67**: 991–994.

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

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

Sanfilippo JE, Nguyen AA, Karty JA, Shukla A, Schluchter WM, Garczarek L, *et al.* (2016). Self-regulating genomic island encoding tandem regulators confers chromatic acclimation to marine *Synechococcus*. *Proc Natl Acad Sci* **113**: 6077–6082.

Schluchter WM, Shen G, Alvey RM, Biswas A, Saunée NA, Williams SR, *et al.* (2010). Phycobiliprotein biosynthesis in cyanobacteria: structure and function of enzymes involved in post-translational modification. *Adv Exp Med Biol* **675**: 211–228.

Shukla A, Biswas A, Blot N, Partensky F, Karty JA, Hammad LA, *et al.* (2012). Phycoerythrin-specific bilin lyase-isomerase controls blue-green chromatic acclimation in marine *Synechococcus*. *Proc Natl Acad Sci* **109**: 20136–20141.

Six C, Joubin L, Partensky F, Holtendorff J, Garczarek L. (2007a). UV-induced phycobilisome dismantling in the marine picocyanobacterium *Synechococcus* sp. WH8102. *Photosynth Res* **92**: 75–86.

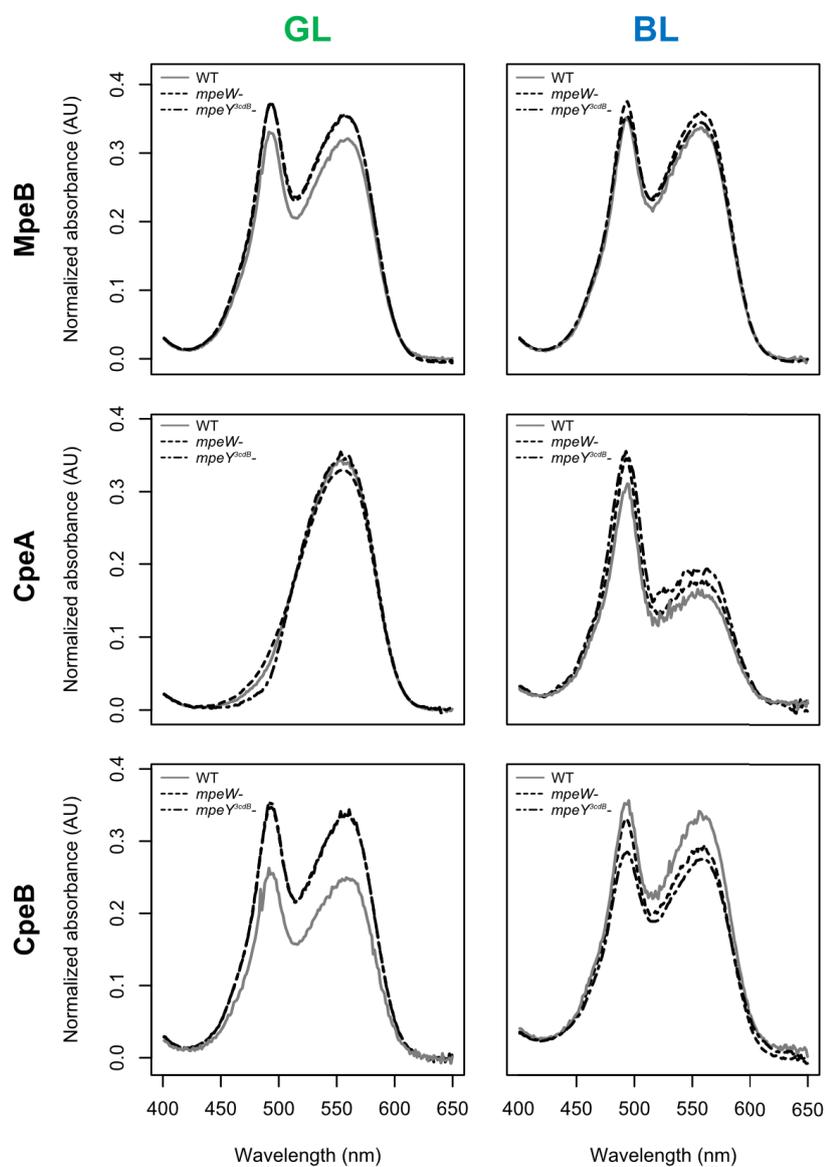
Six C, Thomas J-C, Garczarek L, Ostrowski M, Dufresne A, Blot N, *et al.* (2007b). Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.: a comparative genomics study. *Genome Biol* **8**: R259.

Six C, Thomas J-C, Thion L, Lemoine Y, Zal F, Partensky F. (2005). Two novel phycoerythrin-associated linker proteins in the marine cyanobacterium *Synechococcus* sp. strain WH8102. *J Bacteriol* **187**: 1685–1694.

Zhao K-H, Zhang J, Tu J-M, Böhm S, Plöschner M, Eichacker L, *et al.* (2007). Lyase activities of CpcS- and CpcT-like proteins from *Nostoc* PCC7120 and sequential reconstitution of binding sites of phycoerythrocyanin and phycocyanin beta-subunits. *J Biol Chem* **282**: 34093–34103.

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

## Supplementary figures



**Fig. S1: Absorbance spectra of HPLC-purified  $\beta$ -PE-II,  $\alpha$ -PE-I and  $\beta$ -PE-I subunits (MpeB, CpeA, CpeB respectively) for *Synechococcus* sp. A15-62 wild-type (WT), *mpeW* inactivation mutant (*mpeW*-) and *mpeY3<sup>cdB</sup>* inactivation mutant (*mpeY3<sup>cdB</sup>*-) grown under GL and BL.**

## Supplementary tables

Table S1: plasmids used in this study.

| Plasmid             | Description   | Source     |
|---------------------|---|------------|
| pBCLacZ             | pJS1 derivative with <i>lacZ</i> inserted between BamHI and EagI sites for easy screening | This study |
| pTG_A15-62_mpeW     | pMUT100 derivative, for <i>mpeW</i> disruption in <i>Synechococcus</i> strain A15-62      | This study |
| pTG_A15-62_mpeWcomp | pBCLacZ derivative for expression of A15-62 <i>mpeW</i> in <i>Synechococcus</i>           | This study |
| pTG_A15-62_mpeY     | pMUT100 derivative, for <i>mpeY</i> disruption in <i>Synechococcus</i> strain A15-62      | This study |
| pTG_A15-62_mpeYcomp | pBCLacZ derivative for expression of A15-62 <i>mpeY</i> in <i>Synechococcus</i>           | This study |

Table S2: primers used in this study

| Primer                     | Sequence (5' to 3')               |
|----------------------------|-----------------------------------|
| pMUT100_1291F_SphI         | TAAGCATGCCACCTCGCTAACGGATTCAC     |
| pMUT100_5535R_BamHI        | ATTGGATCCTTCTTAGACGTCAGGTGGCA     |
| pMUT100_5535R_EcoRI        | AATGAATTCTTCTTAGACGTCAGGTGGCA     |
| pMUT100_5368F_seq          | GCGACACGGAAATGTTGAAT              |
| pMUT100_1395R_seq          | GGACGCGATGGATATGTTCT              |
| Syn_A15-62_mpeW_126F_BamHI | ATAGGATCCACGGTTGGGAGCTTGTTCTTCA   |
| Syn_A15-62_mpeW_999R_SphI  | CACGCATGCTAAGCCGAGTTGAGACACCAAACA |
| Syn_A15-62_mpeY_123F_EcoRI | AATGAATTCTGGGGCAAGTAATAGCGAAG     |
| Syn_A15-62_mpeY_983R_SphI  | TAAGCATGCACGACGGCAGTCAGGTAA       |
| Syn_A15-62_unk10_38R_BamHI | ATAGGATCCAGAACCTGAAGAAAGCGTATG    |
| Syn_A15-62_fciB_892F_EagI  | AATCGGCCGCGCTATTACTGCATTCGGA      |
| Syn_A15-62_unk9_13R_ApaI   | AATATAGGGCCCGGGAGACTGGCATGATACAA  |
| Syn_A15-62_mpeB_56R        | ATAAAGGAGCCGCTGGAATC              |

**Table S3: Strains and sequence accession number used in this study.**

| Strain                             | Pigment type | genome       | PBS_region | CA4_island | <i>mpeY</i> <sup>3a</sup> | <i>mpeY</i> <sup>3dA</sup> | <i>mpeY</i> <sup>3cdB</sup> | <i>mpeY</i> <sup>3f</sup> | <i>mpeW</i> | <i>mpeZ</i> | <i>cpeY</i> |
|------------------------------------|--------------|--------------|------------|------------|---------------------------|----------------------------|-----------------------------|---------------------------|-------------|-------------|-------------|
| <i>Synechococcus</i> sp. ROS8604   | 3a           | -            | -          | -          | MG014591                  | -                          | -                           | -                         | -           | -           | MG014542    |
| <i>Synechococcus</i> sp. SYN20     | 3a           | -            | -          | -          | MG014588                  | -                          | -                           | -                         | -           | -           | MG014539    |
| <i>Synechococcus</i> sp. M16.1     | 3a           | -            | -          | -          | MG014596                  | -                          | -                           | -                         | -           | -           | MG014548    |
| <i>Synechococcus</i> sp. RS9907    | 3a           | -            | -          | -          | MG014589                  | -                          | -                           | -                         | -           | -           | MG014540    |
| <i>Synechococcus</i> sp. TAK9802   | 3a           | -            | -          | -          | MG014587                  | -                          | -                           | -                         | -           | -           | MG014538    |
| <i>Synechococcus</i> sp. WH7803    | 3a           | CT971583     | -          | -          | genome                    | -                          | -                           | -                         | -           | -           | genome      |
| <i>Synechococcus</i> sp. NOUM97013 | 3a           | -            | -          | -          | MG014593                  | -                          | -                           | -                         | -           | -           | MG014545    |
| <i>Synechococcus</i> sp. RS9902    | 3c           | -            | -          | -          | -                         | -                          | MG014590                    | -                         | -           | -           | MG014541    |
| <i>Synechococcus</i> sp. CC9605    | 3c           | CP000110     | -          | -          | -                         | -                          | genome                      | -                         | -           | -           | genome      |
| <i>Synechococcus</i> sp. A15-24    | 3c           | -            | -          | -          | -                         | -                          | MG014604                    | -                         | -           | -           | MG014558    |
| <i>Synechococcus</i> sp. A18-46.1  | 3c           | -            | -          | -          | -                         | -                          | MG014598                    | -                         | -           | -           | MG014551    |
| <i>Synechococcus</i> sp. BOUM118   | 3c           | -            | -          | -          | -                         | -                          | MG014597                    | -                         | -           | -           | MG014549    |
| <i>Synechococcus</i> sp. WH8102    | 3c           | BX548020     | -          | -          | -                         | -                          | genome                      | -                         | -           | -           | genome      |
| <i>Synechococcus</i> sp. A15-28    | 3c           | -            | -          | -          | -                         | -                          | MG014603                    | -                         | -           | -           | MG014557    |
| <i>Synechococcus</i> sp. MEDNS5    | 3c           | -            | -          | -          | -                         | -                          | MG014595                    | -                         | -           | -           | MG014547    |
| <i>Synechococcus</i> sp. A15-60    | 3c           | -            | -          | -          | -                         | -                          | MG014602                    | -                         | -           | -           | MG014555    |
| <i>Synechococcus</i> sp. A18-25c   | 3c           | -            | -          | -          | -                         | -                          | MG014600                    | -                         | -           | -           | MG014553    |
| <i>Synechococcus</i> sp. A15-127   | 3c           | -            | -          | -          | -                         | -                          | MG014601                    | -                         | -           | -           | MG014554    |
| <i>Synechococcus</i> sp. MINOS11   | 3dB          | -            | -          | KF177889   | -                         | -                          | KF177891                    | -                         | CA4_island  | -           | KF177890    |
| <i>Synechococcus</i> sp. A15-62    | 3dB          | -            | -          | KF177881   | -                         | -                          | KF177883                    | -                         | CA4_island  | -           | KF177882    |
| <i>Synechococcus</i> sp. PROS-U-1  | 3dB          | -            | -          | KF177895   | -                         | -                          | KF177897                    | -                         | CA4_island  | -           | KF177896    |
| <i>Synechococcus</i> sp. A18-40    | 3dB          | -            | -          | -          | -                         | -                          | MG014599                    | -                         | MG014586    | -           | MG014552    |
| <i>Synechococcus</i> sp. RS9915    | 3dB          | -            | -          | KF177898   | -                         | -                          | KF177900                    | -                         | CA4_island  | -           | KF177899    |
| <i>Synechococcus</i> sp. KORDI-52  | 3bB          | CP006271     | -          | -          | -                         | -                          | genome                      | -                         | genome      | -           | genome      |
| <i>Synechococcus</i> sp. WH8109    | 3bB          | ACNY00000000 | -          | -          | -                         | -                          | genome                      | -                         | genome      | -           | genome      |
| <i>Synechococcus</i> sp. WH8103    | 3bB          | LN847356     | -          | KF177901   | -                         | -                          | KF177904                    | -                         | CA4_island  | -           | genome      |
| <i>Synechococcus</i> sp. BIOS-U3-1 | 3dA          | -            | -          | KF177886   | -                         | KF177888                   | -                           | -                         | -           | CA4_island  | KF177887    |

| Strain                             | Pigment type | genome       | PBS_region | CA4_island | <i>mpeY</i> <sup>3a</sup> | <i>mpeY</i> <sup>3dA</sup> | <i>mpeY</i> <sup>3cdB</sup> | <i>mpeY</i> <sup>3f</sup> | <i>mpeW</i> | <i>mpeZ</i> | <i>cpeY</i> |
|------------------------------------|--------------|--------------|------------|------------|---------------------------|----------------------------|-----------------------------|---------------------------|-------------|-------------|-------------|
| <i>Synechococcus</i> sp. MITS9220  | 3dA          | -            | -          | -          | -                         | MG014594                   | -                           | -                         | -           | MG014605    | MG014546    |
| <i>Synechococcus</i> sp. CC9311    | 3dA          | CP000435     | -          | -          | -                         | genome                     | -                           | -                         | -           | genome      | genome      |
| <i>Synechococcus</i> sp. WH8020    | 3dA          | CP011941     | -          | KF177904   | -                         | KF177905                   | -                           | -                         | -           | CA4_island  | genome      |
| <i>Synechococcus</i> sp. PROS-9-1  | 3dA          | -            | -          | -          | -                         | MG014592                   | -                           | -                         | -           | MG014606    | MG014543    |
| <i>Synechococcus</i> sp. BL107     | 3dA          | AATZ00000000 | -          | -          | -                         | genome                     | -                           | -                         | -           | genome      | genome      |
| <i>Synechococcus</i> sp. CC9902    | 3dA          | CP000097     | -          | -          | -                         | genome                     | -                           | -                         | -           | genome      | genome      |
| <i>Synechococcus</i> sp. RS9916    | 3dA          | AAUA00000000 | -          | -          | -                         | genome                     | -                           | -                         | -           | genome      | genome      |
| <i>Synechococcus</i> sp. MVIR-18-1 | 3aA          | -            | KF177893   | KF177892   | -                         | PBS_region                 | -                           | -                         | -           | CA4_island  | PBS_region  |
| <i>Synechococcus</i> sp. WH8016    | 3aA          | AGIK00000000 | -          | -          | -                         | genome                     | -                           | -                         | -           | genome      | genome      |
| <i>Synechococcus</i> sp. KORDI-49  | 3aA          | CP006270     | -          | -          | -                         | genome                     | -                           | -                         | -           | genome      | genome      |
| <i>Synechococcus</i> sp. BIOS-E4-1 | 3cA          | -            | -          | KF177884   | -                         | missing                    | -                           | -                         | -           | CA4_island  | KF177885    |
| <i>Synechococcus</i> sp. RCC307    | 3eA          | CT978603     | -          | -          | -                         | genome                     | -                           | -                         | -           | -           | genome      |
| <i>Synechococcus</i> sp. CC9616    | 3f           | AZXL00000000 | -          | -          | -                         | -                          | -                           | genome                    | -           | -           | genome      |
| <i>Synechococcus</i> sp. KORDI-100 | 3f           | CP006269     | -          | -          | -                         | -                          | -                           | genome                    | -           | -           | genome      |

## 2. Future work

The study presented in the previous pages is in an advanced stage, but still needs some experimental work and data collection, including:

- complementing *Synechococcus* mutants generated in this study to confirm the phenotype and rule out potential polar effects of the inactivation;
- the extraction and molecular characterization of phycobilisomes and phycobiliprotein subunits from a PT 3a (low PUB:PEB) strain to confirm the attachment of PEB at  $\alpha$ -PEII C83 in this pigment type;
- the identification by LC/UV-Vis/MS-MS of the chromophores and their attachment site on phycobiliproteins from WT and mutant *Synechococcus* grown under GL and BL;
- the confirmation of the chromophores and their binding site in heterologous expression by LC/UV-Vis/MS-MS;
- testing for potential in vitro activity of MpeW and MpeY<sup>3cdB</sup> on other phycobiliprotein subunits than  $\alpha$ -PEII ( $\alpha$ -PEI and  $\beta$ -PEII).

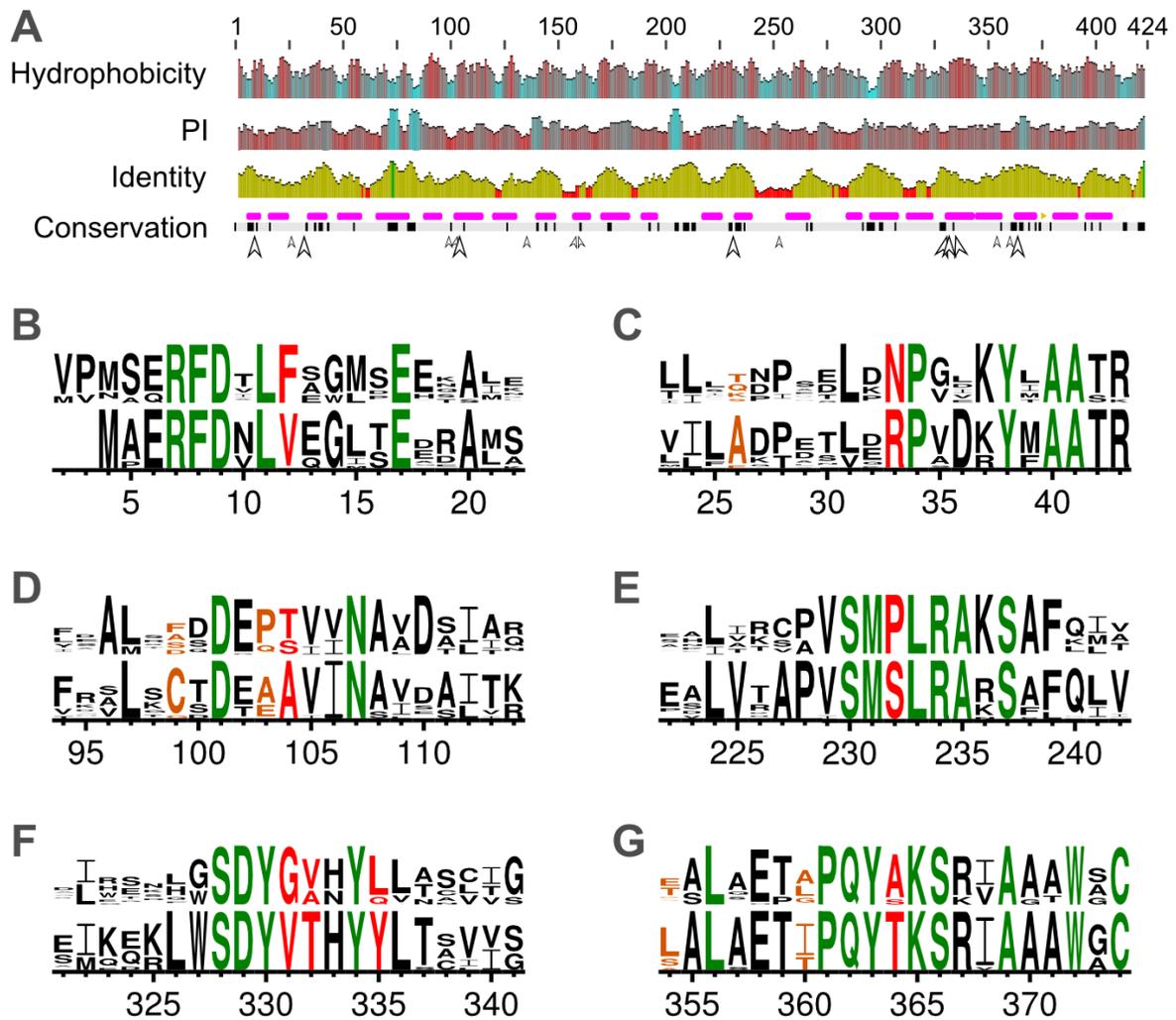
All these experiments are either ongoing or planned, and should be done by the end of 2017/early 2018. Analysis of newly collected data will be done diligently, and the manuscript submitted shortly after that.

## 3. Identification of motifs involved in the lyase-isomerase activity

To investigate the putative amino acids responsible for the difference of activity between lyases and lyase-isomerases of the MpeWYZ family, we aligned and compared all available sequences of this family (**figure 40, Table S1**). MpeY<sup>3a</sup> were included in these analyses, as protein similarity, phylogenetic position, and 3a whole-cell phenotype consistently suggest they are PEB-lyases. Phycobilin-lyases (MpeW, MpeY<sup>3dA</sup> and the putative lyase MpeY<sup>3a</sup>) and lyase-isomerases (MpeY<sup>3cdB</sup> and MpeZ) differed at 17 amino acid positions, among which 8 positions (indicated by big arrows on **figure 40A**), distributed within 6 distinct motifs (**figure 40B-G**), corresponded to residues most conserved in both functional classes: F12V, N33R, (S/T)104A, P232S, G331V, (A/V)332T, (Q/L)335Y and (A/S)364T. These mutations are found in fairly conserved regions, possibly indicating a functional/catalytic role (**figure 40A-G, table s2**). To further narrow down the list of residues potentially involved in the lyase and isomerase function, we looked at the equivalent positions in MpeY sequences of the newly described PT 3f, which exhibits a similar phenotype to PT 3c (i.e., high PUB:PEB ratio, Xia *et al.*, 2017b). Indeed, MpeY<sup>3f</sup> is likely to be a phycobilin lyase-isomerase attaching PUB at  $\alpha$ -PEII C83 since the presence of PUB at this particular residue appears to be necessary to reach such a high PUB:PEB

ratio (Shukla *et al.*, 2012; MpeY<sup>3cdB</sup> mutants). The comparison revealed that residues at positions 12, 33, 232 and 332 are identical or similar between lyases and MpeY<sup>3f</sup>, whereas the other four (104, 331, 335 and 364) were identical or similar between isomerases and MpeY<sup>3f</sup>, suggesting that the latter four residues might be involved in the isomerase function (**table s2**). Similarly, the MpeWYZ homolog CpeY has been demonstrated to be a phycobilin-lyase adding PEB at the equivalent position on PE-I in the freshwater cyanobacterium *Fremyella diplosiphon* PCC 7601 ( $\alpha$ -PEI C82; Biswas *et al.*, 2011). As this position bears PEB in 3a as well as in BL- and GL-grown 3d strains (Ong and Glazer 1991; Shukla *et al.*, 2012), we hypothesized that CpeY might also be a phycobilin-lyase in marine *Synechococcus*. Although CpeY sequences of marine *Synechococcus*, *Prochlorococcus* and *F. diplosiphon* are distantly related from marine *Synechococcus* MpeWYZ, residue at position 331 proved to be a highly conserved glycine in all lyases and putative lyases (331G) and to differ from the conserved valine found at this position in all isomerases and putative isomerases of this dataset (331V; **table s2**). Finally, the motif (N/D)HCQG(N/K) previously demonstrated to be necessary for isomerase activity in the distant homologs PecF, a phycocyanobilin isomerase found in the subaerophytic cyanobacterium *Fischerella* sp. PCC 7603 (Zhao *et al.*, 2005b), and RpcG, a PEB lyase-isomerase present in some marine *Synechococcus* (Blot *et al.*, 2009) does not seem to be conserved in lyase isomerases of the MpeWYZ family. However, residues homologs to (N/D)HCQG(N/K) are highly conserved in both MpeWYZ lyases and lyase isomerases and form the 362QY(A/S/T)KS(K/R) motif (**figure 40G**). Both PecF and RpcG act on the cysteine binding site equivalent to C83  $\alpha$ PE-II but on phycoerythrocyanin and phycocyanin, respectively. Thus, this suggests that this motif might also be important in the MpeWYZ enzymes function. Interestingly, the sole residue of this motif that differs between lyases (364A/S) and lyase-isomerases (364T) is found at the equivalent position to the cysteine residue of the (N/D)HCQG(N/K) motif, which is critical for the isomerase function (Zhao *et al.*, 2005b). Altogether, residues 104, 331-335 and 364 appear to be particularly important for the isomerase function (**figure 40B-G**). No 3D structure has been published to date for phycobilin lyases of the E/F clan. By replacing the motifs identified here in their protein context, such structure would help identifying residues key in both the lyase and isomerase activities of these enzymes.

These results will constitute the basis for a future study focusing on the identification of amino acids key to phycobilin lyase/lyase-isomerase specificity and activity, which will be led by the Schluchter Lab in collaboration with the Kehoe Lab and the MaPP team.



**Figure 40: Sequence comparison of lyases and lyase-isomerases of the MpeWYZ family. (A)** Sequence properties of the MpeWYZ family. Mean hydrophobicity, isoelectric point (PI), and identity at each position along the alignment were computed using Geneious 6.1.8. with a sliding window of 5 residues. Residues conserved in all sequences are represented in black. Helix and strands inferred with Jpred 4 are shown in pink. Arrowhead represent positions of residues that differ between lyases and lyase-isomerases, with big arrowheads representing residues that are the most conserved and small arrowheads, more variable residues. **(B-G)** Sequence logo for lyases (top) and lyase-isomerases (bottom). Residues conserved in all sequences are in green, while residues differing between lyases and lyase-isomerases are in red if conserved within each class and in orange if more variable.

**Table S1: Strains and sequence accession number used in this study.**

| Strain                             | Pigment type | genome       | PBS_region | CA4_island | <i>mpeY</i> <sup>3a</sup> | <i>mpeY</i> <sup>3dA</sup> | <i>mpeY</i> <sup>3cdB</sup> | <i>mpeY</i> <sup>3f</sup> | <i>mpeW</i> | <i>mpeZ</i> | <i>cpeY</i> |
|------------------------------------|--------------|--------------|------------|------------|---------------------------|----------------------------|-----------------------------|---------------------------|-------------|-------------|-------------|
| <i>Synechococcus</i> sp. ROS8604   | 3a           | -            | -          | -          | MG014591                  | -                          | -                           | -                         | -           | -           | MG014542    |
| <i>Synechococcus</i> sp. SYN20     | 3a           | -            | -          | -          | MG014588                  | -                          | -                           | -                         | -           | -           | MG014539    |
| <i>Synechococcus</i> sp. M16.1     | 3a           | -            | -          | -          | MG014596                  | -                          | -                           | -                         | -           | -           | MG014548    |
| <i>Synechococcus</i> sp. RS9907    | 3a           | -            | -          | -          | MG014589                  | -                          | -                           | -                         | -           | -           | MG014540    |
| <i>Synechococcus</i> sp. TAK9802   | 3a           | -            | -          | -          | MG014587                  | -                          | -                           | -                         | -           | -           | MG014538    |
| <i>Synechococcus</i> sp. WH7803    | 3a           | CT971583     | -          | -          | genome                    | -                          | -                           | -                         | -           | -           | genome      |
| <i>Synechococcus</i> sp. NOUM97013 | 3a           | -            | -          | -          | MG014593                  | -                          | -                           | -                         | -           | -           | MG014545    |
| <i>Synechococcus</i> sp. RS9902    | 3c           | -            | -          | -          | -                         | -                          | MG014590                    | -                         | -           | -           | MG014541    |
| <i>Synechococcus</i> sp. CC9605    | 3c           | CP000110     | -          | -          | -                         | -                          | genome                      | -                         | -           | -           | genome      |
| <i>Synechococcus</i> sp. A15-24    | 3c           | -            | -          | -          | -                         | -                          | MG014604                    | -                         | -           | -           | MG014558    |
| <i>Synechococcus</i> sp. A18-46.1  | 3c           | -            | -          | -          | -                         | -                          | MG014598                    | -                         | -           | -           | MG014551    |
| <i>Synechococcus</i> sp. BOUM118   | 3c           | -            | -          | -          | -                         | -                          | MG014597                    | -                         | -           | -           | MG014549    |
| <i>Synechococcus</i> sp. WH8102    | 3c           | BX548020     | -          | -          | -                         | -                          | genome                      | -                         | -           | -           | genome      |
| <i>Synechococcus</i> sp. A15-28    | 3c           | -            | -          | -          | -                         | -                          | MG014603                    | -                         | -           | -           | MG014557    |
| <i>Synechococcus</i> sp. MEDNS5    | 3c           | -            | -          | -          | -                         | -                          | MG014595                    | -                         | -           | -           | MG014547    |
| <i>Synechococcus</i> sp. A15-60    | 3c           | -            | -          | -          | -                         | -                          | MG014602                    | -                         | -           | -           | MG014555    |
| <i>Synechococcus</i> sp. A18-25c   | 3c           | -            | -          | -          | -                         | -                          | MG014600                    | -                         | -           | -           | MG014553    |
| <i>Synechococcus</i> sp. A15-127   | 3c           | -            | -          | -          | -                         | -                          | MG014601                    | -                         | -           | -           | MG014554    |
| <i>Synechococcus</i> sp. MINOS11   | 3dB          | -            | -          | KF177889   | -                         | -                          | KF177891                    | -                         | CA4_island  | -           | KF177890    |
| <i>Synechococcus</i> sp. A15-62    | 3dB          | -            | -          | KF177881   | -                         | -                          | KF177883                    | -                         | CA4_island  | -           | KF177882    |
| <i>Synechococcus</i> sp. PROS-U-1  | 3dB          | -            | -          | KF177895   | -                         | -                          | KF177897                    | -                         | CA4_island  | -           | KF177896    |
| <i>Synechococcus</i> sp. A18-40    | 3dB          | -            | -          | -          | -                         | -                          | MG014599                    | -                         | MG014586    | -           | MG014552    |
| <i>Synechococcus</i> sp. RS9915    | 3dB          | -            | -          | KF177898   | -                         | -                          | KF177900                    | -                         | CA4_island  | -           | KF177899    |
| <i>Synechococcus</i> sp. KORDI-52  | 3bB          | CP006271     | -          | -          | -                         | -                          | genome                      | -                         | genome      | -           | genome      |
| <i>Synechococcus</i> sp. WH8109    | 3bB          | ACNY00000000 | -          | -          | -                         | -                          | genome                      | -                         | genome      | -           | genome      |
| <i>Synechococcus</i> sp. WH8103    | 3bB          | LN847356     | -          | KF177901   | -                         | -                          | KF177904                    | -                         | CA4_island  | -           | genome      |
| <i>Synechococcus</i> sp. BIOS-U3-1 | 3dA          | -            | -          | KF177886   | -                         | KF177888                   | -                           | -                         | -           | CA4_island  | KF177887    |





**Table S2: Residues sets differing between lyases and isomerases from the MpeWYZ family and homolog residues from CpeY. *Syn.*, *Synechococcus*; *Proc.*, *Prochlorococcus*.**

| Position | MpeWY lyase         | MpeYZ isomerase        | MpeY <sup>31</sup> (isomerases) | <i>Syn.</i> CpeY (lyases) | <i>Proc.</i> CpeY (lyases) | <i>F. diplosiphon</i> CpeY (lyase) |
|----------|---------------------|------------------------|---------------------------------|---------------------------|----------------------------|------------------------------------|
| 12       | F                   | I, V                   | F                               | H                         | -                          | -                                  |
| 26       | Q, K, R, S, T       | A, E                   | S                               | X                         | D, N, Q, S                 | D                                  |
| 33       | N                   | R                      | N                               | A, L, S, T                | L                          | E                                  |
| 99       | A, S, D, F          | C, G                   | G                               | A, S, D, H                | N, S                       | D                                  |
| 103      | Q, P, T             | A, S, E                | I                               | Y                         | N, Y                       | Y                                  |
| 104      | S, T                | A                      | A                               | L, M, T                   | I, L, T                    | T                                  |
| 135      | P, S, T, N          | E                      | D                               | X                         | H, N, K, F                 | -                                  |
| 158      | -                   | I, Q, K, E, D          | I, V                            | X                         | X                          | L                                  |
| 159      | -                   | A, E, D, H, K, N, Q, S | D                               | X                         | X                          | E                                  |
| 232      | P                   | S                      | P                               | V, A, S, T                | Y, F, T, S                 | V                                  |
| 252      | A, K, R, D, G       | I, C, M, L, F          | -                               | -                         | Q, -                       | -                                  |
| 331      | G                   | V                      | V                               | G                         | G                          | G                                  |
| 332      | A, V                | T                      | I                               | A                         | A, S                       | A                                  |
| 335      | Q, L                | Y                      | Y                               | F, C                      | F                          | H                                  |
| 354      | E, D, G, K, Q, R, T | A, S, L                | S, T                            | E, D, S, Q, N             | F, S, Y, H                 | E                                  |
| 360      | A, E, L, G          | I, T                   | I                               | R                         | W                          | Q                                  |
| 364      | A, S                | T                      | T                               | X                         | M, I                       | Q                                  |



### **III. The “FrankenSynechococcus” experiment: how easy is it to gain CA4?**

#### **1. Introduction**

Since its discovery in 2001 (Palenik 2001), several studies have been focused on the characterization of CA4 (Everroad *et al.*, 2006; Shukla *et al.*, 2012; Humily *et al.*, 2013; Sanfilippo *et al.*, 2016). From a genomic standpoint, Six and co-workers (Six *et al.*, 2007c) observed a specific gene complement and organization of the PBS rod genomic region in the CA4-capable strains BL107, CC9902, CC9311, RS9916. These features were also observed in strain RCC307, which Six and coworkers thought to have completely lost its ability to chromatically acclimate, but that was more recently shown to exhibit a reduced CA4 capacity (Humily *et al.*, 2003). The latter authors also discovered that strains able to perform CA4 systematically possessed a small specific genomic island, which they found to exist in two different versions (CA4-A and CA4-B), both comprising a putative phycobilin-lyase (*mpeZ* in CA4-A and *mpeW* in CA4-B) and two putative regulators (*fciAB*). Using a combination of genetic and biochemical approaches, Shukla and co-workers demonstrated that MpeZ is a phycobilin lyase-isomerase and that this gene is necessary for the green-to-blue transition in CA4-A in the 3dA strain RS9916 (Shukla *et al.*, 2012). The characterization of *fciA* and *fciB* inactivation mutants in RS9916 showed that FciA and FciB have diametric effects on CA4, the first mutant displaying a constitutive green light phenotype and the second one a constitutive blue light phenotype (Sanfilippo *et al.*, 2016). This work also showed that CA4-A activates genes under blue light (BL) only, leading Sanfilippo and co-workers to suggest that the 3dA ancestor might have been a green light (GL) specialist (PT 3a; Sanfilippo *et al.*, 2016).

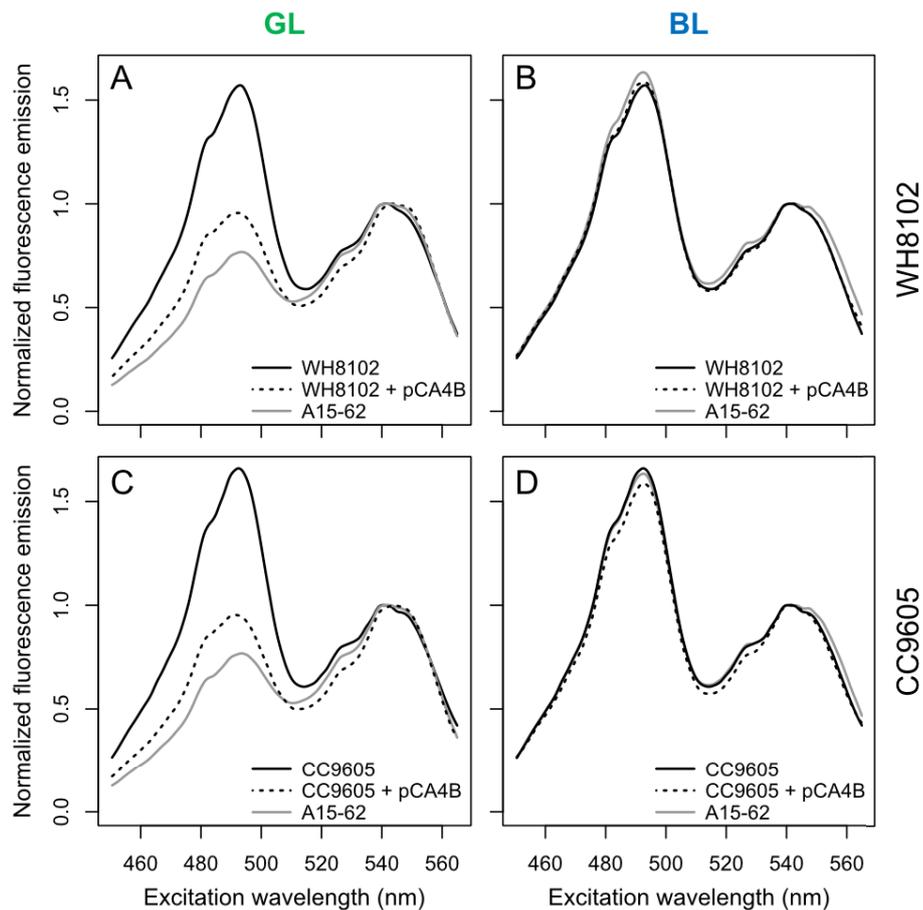
Comparative analyses with an extended set of genomes showed that while 3dA strains have an original PBS rod genomic region, with specific gene complement, order and alleles, 3dB strains are in all ways (gene complement, order and alleles) identical to 3c strains, except for the insertion of the CA4-B island in the middle of the PBS region near the 3'-end of the PEII genomic sub-region (see Chapter I). This, together with the observation that the CA4-B specific gene *mpeW* is activated under green light (Humily *et al.*, 2013), led us to hypothesize that the PT 3dB genotype might derive from a blue light specialist genotype (PT 3c).

Altogether, results published so far clearly demonstrated that the CA4 genomic island is *necessary* for CA4. The goal of the present study was to check whether it could be *sufficient* for CA4. Prof. David M. Kehoe designed a study to test this hypothesis, by cloning either version of the CA4 genomic island in a plasmid autonomously replicating in marine *Synechococcus* and transferring these constructs into strains displaying a fixed pigmentation. Here, I present some

preliminary results that I obtained during my 5-month stay in the Kehoe Lab, where I have tested the inclusion of a CA4-B genomic island into two PT 3c (i.e. high PUB) strains. The twin experiment, consisting in the inclusion of a CA4-A island into several PT 3a (i.e. low-PUB) strains has also been performed by Bo Chen, currently a postdoc in the Kehoe Lab, but his results are not shown here.

## 2. Results

Colonies of transformants were obtained for both *Synechococcus* PT 3c strains CC9605 and WH8102. After PCR verification of the presence of the plasmid (containing the CA4-B island from the 3dB strain A15-62), transformants were acclimated for at least 15 days in GL or BL before measuring their spectral properties.



**Figure 41: Whole cell fluorescence excitation spectra with emission at 585 nm for two different “FrankenSynechococcus” transformants acclimated under green (A, C) or blue light (B, D). (A, B) *Synechococcus* sp. WH8102 wild-type (WH8102) and with a plasmid containing *Synechococcus* sp. A15-62 CA4-B genomic island (WH8102 + pCA4B). Fluorescence of A15-62 is provided as a reference for CA4-B. (C, D) Same as (A, B) for strain *Synechococcus* sp. CC9605.**

The fluorescence excitation spectra (with emission at 585 nm) of these transformants of both strains showed no difference with the spectrum from the wild-type (WT) CC9605 and WH8102 strains when grown in BL, while transformants had a decreased Exc495:545 ratios (proxy for the PUB:PEB ratio) compared to the WT in GL (**figure 41**). Comparison of BL- and GL-acclimated transformants with WT cells of the 3dB strain A15-62 showed that they have a similar phenotype in BL, but a slightly higher Exc495:545 ratio in GL relative to A15-62 (**figure 41**).

### 3. Discussion

Together with the biochemical characterization of MpeW, these preliminary results suggest that the CA4-B genomic island is both *necessary* and *sufficient* for conferring 3c strains the ability to chromatically acclimate. However, more work is needed to confirm this proof-of-concept. First and most importantly, this study currently lacks a negative control, in the form of empty plasmid transformation. This experiment has been performed but there were still no colony at the date of writing this thesis manuscript. Another important missing control is the transformation of the CA4-B genomic island into PT 3a strains (GL specialists). It is also worth noting that recipients and CA4-B donor strains were purposely selected as being phylogenetically close relatives, in order to minimize possible incompatibilities (e.g. different codon usage). It would therefore be interesting to redo the experiment using other strains with a PT 3c background but more distantly related than were strains CC9605 (subclade IIc) and WH8102 (subclade IIIa) from A15-62 (subclade IIc), e.g. using as recipients strains A15-127 (clade WPC1), A15-60 (subclade VIIa) and/or A18-25c (subclade VIIa; Farrant *et al.*, 2016; Mazard *et al.*, 2012).

Interestingly, the comparison of the spectral properties of the 3c transformants and the WT 3dB strain A15-62 showed that GL acclimation was not complete in transformants, with an Exc<sub>495:545</sub> of ~0.95 in GL-grown transformants compared with ~0.75 for the GL-grown A15-62 (**figure 41**). One possible explanation could be an effect of the endogenous *unk10* in recipient PT3c strains. Indeed, we cloned all four genes of the CA4-B genomic island, i.e. *fciA*, *fciB*, *mpeW* and *unk10*. *unk10* is present in both CA4-A and CA4-B genomic islands, and although it is not yet characterized, its overexpression in BL relative to GL in both 3dA and 3dB strains (Sanfilippo *et al.*, 2016; unpublished work by Doré, Ratin and Garczarek) suggest a role in acclimation to BL. However, PT 3c strains also have a copy of *unk10* in their PBS genomic region (see Chapter I), which is likely constitutively expressed. If these two hypotheses are correct, the endogeneous *unk10* would hinder the complete GL acclimation. Alternatively, even if the endogenous *unk10* is effectively regulated by the newly introduced CA4-B genomic island, the two copies of *unk10* in 3c transformants might influence the GL phenotype through a dose

effect. Different expression levels from genomic and plasmid genes might also be at play, e.g. because of plasmid copy number. In RS9916 transformants with plasmids bearing *fciA* and/or *fciB*, these genes were 5-10 fold more expressed than in the WT with no plasmid (Sanfilippo *et al.*, 2016). However, *fciA* and *fciB* are not CA4-regulated (Sanfilippo *et al.*, 2016), and CA4 regulation might counteract the difference between plasmid and genomic expression: plasmid-induced expression could affect GL-phenotype through high Unk10 levels, but it would probably also result into higher MpeW levels under BL, which would reduce  $\text{Exc}_{495:545}$  in BL. As we did not observe any difference in this ratio between WT 3c strains and transformants grown under BL, it is possible that plasmid expression does not affect CA4 in our experimental setup. To test the effect of the endogenous *unk10*, we constructed a plasmid bearing a shortened CA4-B genomic island without *unk10*, but so far we have not been successful in introducing it into any 3c strain. Interestingly, interruption mutants in the C83  $\alpha$ -PEII PEB lyase-isomerase genes *mpeY*<sup>3cdB</sup> (CA4-B) and *mpeZ* (CA4-A) exhibit about the same  $\text{Exc}_{495:545}$  ratio in BL (0.84 and 0.9 for *mpeY*<sup>3cdB</sup>- and *mpeZ*- respectively) as do the CA4-B transformants in GL (0.95), further reinforcing the *unk10* hypothesis. Indeed, in such BL-acclimated mutants, the value of the  $\text{Exc}_{495:545}$  ratio likely results from the absence of lyase-isomerase (compensated by the cognate lyase being expressed at constant level; Sanfilippo *et al.*, *in prep.*) and the CA4-induced high abundance of Unk10, although we cannot definitely rule out other CA4-induced mechanisms (Sanfilippo *et al.*, 2016). Altogether, these results suggest that Unk10, a protein with no homology to any known phycobilin lyase/lyase-isomerase, might play a role in PUB attachment on C140  $\alpha$ PE-II and/or C139  $\alpha$ PE-I, the two other sites besides C83  $\alpha$ -PEII that change chromophores during CA4 (Shukla *et al.*, 2012).

Genomic islands have been identified in the first sequenced genomes of *Synechococcus* by their atypical nucleotide composition (Palenik *et al.*, 2006; Dufresne *et al.*, 2008). These islands have highly variable gene content, and island genes exhibit a phylogeny departing from that of core genes, suggesting that they are inherited by lateral transfer (Palenik *et al.*, 2006; Dufresne *et al.*, 2008). Genes present in the genomic islands of *Synechococcus* and *Prochlorococcus* that have been assigned a putative function are often involved the response to environmental factors, such as nutrient uptake (Martiny *et al.*, 2006, 2009a, 2009b; Berube *et al.*, 2015), light acclimation (Shukla *et al.*, 2012, Sanfilippo *et al.*, 2016), motility (McCarren and Brahamsha, 2005), oxidative stress (Stuart *et al.*, 2013) or biotic interactions such as response to phage infection (Avrani *et al.*, 2011; Fedida and Lindell, 2017) or allelopathy (Paz-Yepes *et al.*, 2013), thus playing an important role in shaping *Synechococcus* ecology. The early results presented here showing that an important function can be conferred by gain of a small genomic island constitute one of the first attempts to experimentally challenge the possibility and effect of

genomic island lateral transfer in marine *Synechococcus*, a crucial step in understanding these important drivers of *Synechococcus* genome plasticity and evolution.

## 4. Materials and methods

### *Strains and growth conditions*

*Synechococcus* strains A15-62, WH8102, CC9605 and BOUM118 were obtained from the RCC (RCC numbers 1089, 539, 753 and 2421, respectively). Information about these strains can be found in **table s3**. Wild-type and mutant strains were maintained in PCR-S11 medium at 22°C using transparent polycarbonate flasks and continuous white light ( $20 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ ). Mutants were maintained with  $75 \mu\text{g} \cdot \text{mL}^{-1}$  spectinomycin. Strains were acclimated for at least two weeks under  $20 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$  blue light (BL) or green light (GL) provided by LED ramps (Luxeon Rebel LED LXML-PB01-0040 and LXML-PM01-0100 for BL and GL, respectively; Alpheus, France) before their spectral properties were measured using a LS-50B spectrofluorimeter (Perkin Elmer, Waltham, MA, USA).

### *Plasmid construction*

All primers and plasmids used in this study are listed in **table s4** and **table s5**, respectively. *Synechococcus* sp. A15-62 CA4-B genomic island was PCR amplified, and the PCR fragment and the pRL153 derivative pJS2-fciA were digested with BamHI and EagI. Digested fragments were gel-purified before assembly, and the construct sequenced.

### *Conjugation*

Conjugation between *Synechococcus* strains and *E. coli* MC1061 containing pTG\_A15-62\_CA4B\_GI and helper plasmids was performed as previously described (Brahamsa, 1996; Shukla *et al.*, 2012).

## 5. Supplemental material

See next page

**Table S3: Strains used in the “FrankenSynechococcus” study.**

| Strain name | RCC number | Clade | Pigment type | Isolation date | Sea/Ocean            | Oceanic Region     | Latitude | Longitude | Isolation depth (m) | Reference                                |
|-------------|------------|-------|--------------|----------------|----------------------|--------------------|----------|-----------|---------------------|--|
| A15-62      | 1089       | IIc   | 3dB          | 04 Oct 2004    | North Atlantic Ocean | -                  | 17°37'N  | 20°57'W   | 15                  | Humily et al., 2013; Mazard et al., 2012 |
| WH8102      | 539        | IIIa  | 3c           | 15 Mar 1981    | North Atlantic Ocean | -                  | 22°30'N  | 65°36'W   | -                   | Six et al., 2007                         |
| CC9605      | 753        | IIc   | 3c           | 1996           | North Pacific Ocean  | California Current | 30°25'N  | 123° 58'W | 51                  | Six et al., 2007                         |
| BOUM118     | 2421       | IIIa  | 3c           | 1 Oct 2008     | Mediterranean Sea    | Eastern basin      | 33°38'N  | 32°38'E   | 5                   | Farrant et al., 2016; This study         |

**Table S4: Primers used in the “FrankenSynechococcus” study.**

| Primer                     | Sequence (5' → 3')                | Use                  | Source     |
|----------------------------|-----------------------------------|----------------------|------------|
| Syn_A15-62_mpeU_887F_BamHI | ATAGGATCCAGATTTGCAGGAGAAACAGG     | Plasmid construction | This study |
| Syn_A15-62_cpeZ_601F_EagI  | AATCGGCCGTGGCACGACTTCAGAATCTA     | Plasmid construction | This study |
| Syn_A15-62_unk10_38R_EagI  | ATACGGCCGAGAACCTGAAGAAAGCGTATG    | Plasmid construction | This study |
| Syn_A15-62_fciA_550F       | AGCCAACCATCAACGAAACC              | Sequencing           | This study |
| Syn_A15-62_fciB_658F       | TCCCCTCTGTCAATTGCTTGA             | Sequencing           | This study |
| Syn_A15-62_mpeW_147R       | TGAAGAACAAGCTCCCAACC              | Sequencing           | This study |
| Syn_A15-62_unk10_199F      | GGAAGACTATGCCCAAGAT               | Sequencing           | This study |
| Syn_A15-62_fciA_873R_BamHI | AATGGATCCGAGACCTGTTGAGGAATGTGGAGT | Sequencing           | This study |
| Syn_A15-62_fciB_43F_BamHI  | AATGGATCCGAAGAGCTTGCAAAACTCGGTCAG | Sequencing           | This study |
| Syn_A15-62_fciA_550F       | AGCCAACCATCAACGAAACC              | Sequencing           | This study |
| Syn_A15-62_fciB_658F       | TCCCCTCTGTCAATTGCTTGA             | Sequencing           | This study |
| Syn_A15-62_mpeW_1041R      | CGGCTCCAATAATGATTCTT              | Sequencing           | This study |
| Syn_A15-62_mpeW_6F         | GCCAAGTTCACAGCGATTG               | Sequencing           | This study |

**Table S5: Plasmids used in the “FrankenSynechococcus” study.**

| Plasmid                    | Description   | Source                  |
|----------------------------|---|-------------------------|
| pJS2                       | Spec <sup>R</sup> derivative of pRL153 with <i>fciA</i> cloned between the BamHI and EagI sites                   | Sanfilippo et al., 2016 |
| pTG_A15-62_CA4B_GI         | pJS2 derivative, with <i>fciA</i> insert replaced by <i>Synechococcus</i> sp. A15-62 CA4-B genomic island         | This study              |
| pTG_A15-62_CA4B_GI_noUnk10 | pJS2 derivative, with <i>fciA</i> insert replaced by partial <i>Synechococcus</i> sp. A15-62 CA4-B genomic island | This study              |



# CONCLUSION & PERSPECTIVES

---

« It would be difficult to find another series of colouring-matters  
of greater beauty.»

H.C. Sorby, 1875

## I. Considerations on the phycobilin lyases function

A part of the work presented here has consisted in the characterization of two related phycobilin lyases (Chapter III): MpeW (present in 3dB strains) and one MpeY allele (MpeY<sup>3cdB</sup>, present in all PT 3dB and 3c strains). We have demonstrated that together with two previously characterized lyases, the MpeY<sup>3dA</sup> allele (Sanfilippo *et al.*, in prep.) and MpeZ (present only in PT 3dA strains, Humily *et al.*, 2013; Shukla *et al.*, 2012), they form a lyase family whose members all act on the same phycobiliprotein attachment site, namely Cys-83  $\alpha$ PE-II, and either solely attach PEB (MpeW, MpeY<sup>3dA</sup> and likely MpeY<sup>3a</sup>) or attach PEB and concomitantly isomerize it into PUB (MpeY<sup>3cdB</sup> and MpeZ). This highlights that genetic microdiversity, in our case allelic variation, can have important phenotypic consequences, which in turn impacts the distribution and ecology of *Synechococcus* (Chapter II).

In recent years, a large number of phycobilin lyases representative of the three known structural clans (T, S/U and E/F, as defined by Bretaudeau *et al.*, 2013; see also the Cyanolyase web site; <http://cyanolyase.genouest.org/>) have been biochemically characterized in a variety of organisms, including freshwater (*Fremyella diplosiphon*, *Nostoc* sp., *Synechococcus* spp.) and marine (*Synechococcus* spp., *Prochlorococcus marinus*) cyanobacteria, as well as some virus-encoded orthologs (see Introduction). Together with comparative physiology and genomics, this has dramatically expanded our understanding of the biosynthetic pathway of phycobiliproteins, enlightening some general features for each of the different phycobilin lyases families but still leaving some gaps concerning the site and phycobilin specificities of the numerous PE-II lyases.

By combining the results obtained in the framework of this PhD (chapter III) with the extensive literature existing on phycobilin lyases including most recent results (Mahmoud *et al.*, 2017; Gasper *et al.*, 2017; Sanfilippo *et al.*, in prep), it is now possible either to know with certainty or at least predict with some confidence the function of all phycobilin lyases present in marine *Synechococcus* spp. (**Table 6**) but also in *Prochlorococcus* spp. (**Table 7**). Indeed, despite the absence of phycobilisomes, *Prochlorococcus* LL ecotypes were shown to synthesize a seemingly functional though divergent PE (called PE-III) and HL ecotypes a  $\beta$ -PE remnant (Hess *et al.*, 1996, 1999). The following paragraphs list the arguments that led me to draw this overall picture on phycobilin lyase functions.

The PE-III present in LL *Prochlorococcus* spp. strains has four chromophore binding sites, one on the  $\alpha$ -subunit ( $\alpha$ -73, equivalent to the consensus position  $\alpha$ -82) and three on the  $\beta$ -subunit ( $\beta$ -82,  $\beta$ -163 and likely  $\beta$ -50,61). Consequently, *Prochlorococcus* LL strains have a (reduced) PBS genomic region containing four (putative) phycobilin lyases, CpeS, CpeT, CpeY/Z and CpeF (Hess *et al.*, 1999). In contrast, the  $\beta$ -PE remnant found in HL ecotypes

possesses one chromophore binding site at Cys-82 and these strains consistently have only one putative phycobilin lyase, CpeS, which was shown to catalyze the attachment of PEB at this site in *P. marinus* MED4 (Wiethaus *et al.*, 2010). In *Fremyella diplosiphon*, CpeS also catalyzes the attachment of PEB to  $\beta$ -82 of PE-I (Biswas *et al.*, 2011). Thus, even if CpeS has not been formally characterized in marine *Synechococcus* nor in LL *Prochlorococcus*, it is most likely that the CpeS orthologs present in these strains also catalyze the attachment of PEB to  $\beta$ -82 of PE-I (**Tables 6 and 7**). Given the broad phycobiliprotein substrate exhibited by CpcS (acting on  $\alpha$ - and  $\beta$ -82 of APC and on  $\beta$ -82 of PC; Zhao *et al.*, 2007), it is also possible that CpeS might also catalyze the attachment of PEB to  $\beta$ -82 of PE-II in marine *Synechococcus* (**Table 6**).

CpcT has been shown to bind PCB to Cys-153 of  $\beta$ -PC in the freshwater *Synechococcus* sp. strain PCC7002 (Shen *et al.*, 2006), and to Cys-155 of  $\beta$ -PEC in *Nostoc* sp. PCC7120 (Zhao *et al.*, 2007b). The genomic context, homology to CpcT, and phenotype of strains possessing it all concordantly suggest that RpcT is acting on the same position in marine *Synechococcus* to add PEB (**Table 6**; Blot *et al.*, 2009, Schluchter *et al.*, 2010). Similarly, CpeT likely attaches PEB to  $\beta$ -165 PE-I in marine *Synechococcus* (**Table 6**) and to  $\beta$ -163 PE-III in LL *Prochlorococcus* (**Table 7**). Thus, members of the T clan appear highly specific of the chromophore-binding site  $\beta$ -153 (consensus site) but act on different phycobiliproteins and/or add different phycobilins.

Finally, several members of the E/F clan have been characterized allowing one to complete the picture. This includes the CpeY/Z heteroduplex in *Fremyella diplosiphon*, which attaches PEB to  $\alpha$ -82 of PE-I (Biswas *et al.*, 2011), and likely has the same site specificity in marine *Synechococcus* and LL *Prochlorococcus* (**Tables 6 and 7**). Similarly, the characterization of the PEB lyase CpcE/F in *Synechococcus* PCC7002 (Zhou *et al.*, 1992; Fairchild *et al.*, 1992) and of the PEB lyase-isomerase RpcG of marine *Synechococcus* (Blot *et al.*, 2009) concordantly suggest that CpcE/F and RpcE/F all act on  $\alpha$ -84 PC, binding PCB and PEB, respectively (**Table 6**). At last, the MpeWYZ family appears to bind either PEB or PUB (after isomerization) specifically on Cys-83  $\alpha$ PE-II at this position (Chapter III; Shukla *et al.*, 2012; Sanfilippo *et al.*, in prep.).

No lyase has been characterized so far for  $\beta$ 50,61 PE-I and PE-II in marine *Synechococcus* and for  $\beta$ 50,61 PE-III in LL *Prochlorococcus* (**Table 7**). However, CpeF is the only (putative) phycobilin lyase left in LL *Prochlorococcus* for which no orthologs has been characterized. This strongly suggests that CpeF is acting on  $\beta$ 50,61 PE-III. Extending this hypothesis to marine *Synechococcus*, we assume that CpeF could be acting on  $\beta$ 50,61 PE-I and maybe PE-II (**Table 7**). As CpeF is only present in PT 2, 3a and 3dA, it is most likely that CpeF adds PEB at these sites (**Table 6**). MpeU is the closest paralog of CpeF, and is present in marine

*Synechococcus* strains that have a high or variable PUB:PEB ratio (3c, 3dA, 3dB). Moreover, the inactivation of *mpeU* in the 3dA strain *Synechococcus* RS9916 showed that this gene is necessary for reaching such high PUB:PEB ratios (Mahmoud *et al.*, 2017). Together with the hypothesized function of CpeF, this advocates for an isomerase function, and I propose that MpeU would be a phycobilin lyase-isomerase attaching PUB at  $\beta$ 50,61 PE-I and perhaps PE-II (Table 7). If these predictions hold true, this would imply that different members of the same structural clan (E/F) exhibit different phycobiliprotein site specificity, opening interesting questions about the determinants and evolution of this specificity.

**Table 6: Marine *Synechococcus* phycobiliprotein chromophorylation sites and cognate phycobilin lyases (or lyases-isomerases).** Italicized lyases and chromophores are predictions. Chromophorylation sites modified during CA4 are indicated in bold. Uncharacterized phycobilin lyases (isomerases) are indicated by a grey background. See also the introduction of this manuscript for details about the different characterized phycobilin-lyases.

|          |               | PC                 |                    |                  | PE-I             |            |                    |     |     | PE-II    |  |            |             |     |
|----------|---------------|--------------------|--------------------|------------------|------------------|------------|--------------------|-----|-----|----------|--|------------|-------------|-----|
|          |               | $\alpha$           |                    | $\beta$          | $\alpha$         |            | $\beta$            |     |     | $\alpha$ |  |            | $\beta$     |     |
|          |               | 84                 | 82                 | 153              | 82               | <b>139</b> | 50,<br>61          | 82  | 165 | 75       | <b>83</b>                              | <b>140</b> | 50,<br>61   | 82  |
| S/U clan | CpcS/U        |                    | PCB <sup>1,2</sup> |                  |                  |            |                    |     |     |          |  |            |             |     |
|          | CpeS/U        |                    |                    |                  |                  |            | PEB <sup>3,4</sup> |     |     |          |  |            | PEB         |     |
| T clan   | CpcT          |                    |                    | PCB <sup>5</sup> |                  |            |                    |     |     |          |  |            |             |     |
|          | <i>RpcT</i>   |                    |                    | <i>PEB?</i>      |                  |            |                    |     |     |          |  |            |             |     |
|          | <i>CpeT</i>   |                    |                    |                  |                  |            |                    | PEB |     |          |  |            |             | PEB |
| E/F clan | CpcE/F        | PCB <sup>6,7</sup> |                    |                  |                  |            |                    |     |     |          |  |            |             |     |
|          | <i>RpcE/F</i> | <i>PEB</i>         |                    |                  |                  |            |                    |     |     |          |  |            |             |     |
|          | RpcG          | PUB <sup>8</sup>   |                    |                  |                  |            |                    |     |     |          |  |            |             |     |
|          | CpeY/Z        |                    |                    |                  | PEB <sup>9</sup> |            |                    |     |     |          |  |            |             |     |
|          | MpeW          |                    |                    |                  |                  |            |                    |     |     |          | PEB <sup>†</sup>                       |            |             |     |
|          | MpeY          |                    |                    |                  |                  |            |                    |     |     |          | PEB <sup>†</sup> ,<br>PUB <sup>‡</sup> |            |             |     |
|          | MpeZ          |                    |                    |                  |                  |            |                    |     |     |          | PUB <sup>10</sup>                      |            |             |     |
|          | CpeF          |                    |                    |                  |                  |            | <i>PEB?</i>        |     |     |          |  |            | <i>PEB?</i> |     |
|          | MpeU          |                    |                    |                  |                  |            | <i>PUB?</i>        |     |     |          |  |            | <i>PUB?</i> |     |

<sup>1</sup> Zhao *et al.*, 2007<sup>2</sup> Saunée *et al.*, 2008<sup>3</sup> Wiethaus *et al.*, 2010<sup>4</sup> Biswas *et al.*, 2011<sup>5</sup> Shen *et al.*, 2006<sup>6</sup> Fairchild *et al.*, 1992<sup>7</sup> Zhou *et al.*, 1992<sup>8</sup> Blot *et al.*, 2009<sup>9</sup> Biswas *et al.*, 2011<sup>10</sup> Shukla *et al.*, 2012<sup>†</sup> Sanfilippo *et al.*, in prep<sup>‡</sup> This work

**Table 7** *Prochlorococcus* phycobiliprotein chromophorylation sites and cognate phycobilin lyases (isomerases). Italicized lyases and chromophores are predictions. Uncharacterized phycobilin lyases (isomerases) are indicated by a grey background. See also the introduction of this manuscript for details about the different characterized phycobilin-lyases.

|          |           | PE-III               |         |                  |                      |
|----------|-----------|----------------------|---------|------------------|----------------------|
|          |           | $\alpha$             | $\beta$ |                  |                      |
|          |           | 73                   | 50, 61  | 82               | 163                  |
| S/U clan | CpeS      |                      |         | PEB <sup>1</sup> |                      |
| T clan   | CpeT      |                      |         |                  | PEB/PUB <sup>2</sup> |
| E/F clan | CpeY/CpeZ | PEB/PUB <sup>3</sup> |         |                  |                      |
|          | CpeF      |                      | PEB/PUB |                  |                      |

<sup>1</sup> Wiethaus *et al.*, 2010

<sup>2</sup> Zhao *et al.*, 2007

<sup>3</sup> Biswas *et al.*, 2011

The only (putative) lyases in marine *Synechococcus* for which no orthologs have been characterized so far are CpeF and CpeU, while MpeU is only partially characterized (phenotype of an inactivation mutant; Mahmoud *et al.*, 2017). The phycobiliproteins sites for which the cognate lyase remains elusive because they do not fit with the characterized site specificities of the three known clans (i.e.,  $\beta$ -155 for clan T;  $\beta$ -82 for clan S/U and  $\alpha$ -82 for clan E/F) are  $\alpha$ -139 and  $\beta$ -50,61 of PE-I, as well as  $\alpha$ -75,  $\alpha$ -140,  $\beta$ -50,61 of PE-II. Even if some of the predictions made above prove to be false, it seems highly unlikely that these three lyases together are sufficient for the attachment of chromophores to these five sites, and we are faced with a lack of candidate lyases given the number of chromophorylation sites. In particular, the chromophorylation of Cys-75  $\alpha$ PE-II appears to be somewhat enigmatic. Indeed, this site is only present on PE-II, and not in PE-I. Yet, the only lyase specific of PE-II-containing *Synechococcus* is MpeY, which we demonstrated to be acting on Cys-83  $\alpha$ PE-II. It is unlikely that CpeF acts on this residue, as the *cpeF* gene is present in PT 2 strains, which lack PE-II, and it also appears unlikely for MpeU to attach chromophore on this site since the *mpeU* gene is only present in a subset of PE-II-containing strains. Similarly, CpeU is probably a facilitator of CpeS, as demonstrated for CpcU/CpcS (Saunée *et al.*, 2008; **Table 6**). The only genes found specifically in PE-II-containing strains are *unk8-unk7* and *unk9*, which are localized in the PE-II sub-region of the PBS rod genomic region (see Chapter I), strongly suggesting that they display a PE-II-related function. I thus suggest that one or both of these genes are involved in the attachment of PEB or PUB at  $\alpha$ -75 PE-II. Moreover, we know that in PT 3a, PE-II bears only one PUB chromophore, located either on  $\beta$ -50,61 or on  $\alpha$ -75 (Six *et al.*, 2007c; Ong and Glazer 1991; Shukla *et al.*, 2012). If the prediction about CpeF function is true, i.e. if CpeF adds PEB at  $\beta$ -50,61, then it appears likely that *unk8-unk7* and/or *unk9* is involved in the attachment of PUB at Cys-75  $\alpha$  PE-II. Similarly, we obtained some very indirect evidence for a role of Unk10 in PUB attachment during CA4 (see Chapter III). As the two sites Cys-139  $\alpha$  PE-I and Cys-140  $\alpha$  PE-II

are modified during CA4, Unk10 could act on one or both of these positions in BL. However, the phycobilin lyase or Unk protein acting on these sites in GL remains elusive. Strikingly, none of these genes bear any similarity to known phycobilin-lyases. These genes could thus encode a totally new class of phycobilin-lyases, or alternatively could interact with enzymes of the phycobiliprotein biosynthesis pathway, e.g. by binding to phycobiliprotein subunits and making one chromophorylation site “look like” another one from the perspective of a phycobilin lyase.

## II. Type IV Chromatic Acclimation

Type IV Chromatic Acclimation has been a central aspect of my work. In addition to showing that MpeW and MpeY<sup>3cdB</sup> are key effectors of CA4-B (see previous paragraph and Chapter III), we demonstrated that both versions of CA4 are widespread and ecologically significant (Chapter II). We outlined the dominance of PT 3dA genotypes (i.e. populations likely able to perform CA4-A) at high latitudes, a finding that was confirmed in the newer *Tara* Polar Circle metagenomes (data not included in the article because they were not released yet at the time of submission), in which they represented 100% of the *Synechococcus* populations, as also independently observed in subpolar waters of the NW Pacific by Xia *et al.*, (2017b). These subpolar environments are characterized by frequent deep mixing of the water column, recurrent blooms, and high seasonal variability in both light intensity and quality at the surface of the ocean. As an example, winter months above 60°N are characterized by the prevalence of indirect irradiance by light backscattering in the atmosphere and little to no direct sunlight irradiance, and the spectrum of light reaching the surface of the ocean in such conditions is dominated by wavelengths in the 450-500 nm range, creating a blue light environment similar to that observed at depth in clear intertropical waters (Hansen *et al.*, 1997; Eilertsen and Degerlund, 2010; Stomp *et al.*, 2007). On the contrary, during the diatom spring blooms observed in polar environments (Smetacek, 1999; Klein *et al.*, 2002; Tremblay *et al.*, 2002, 2006; Mock *et al.*, 2017), the light colour spectrum is likely green-shifted due to the strong absorption of blue wavelengths by xanthophylls and chlorophylls found in these organisms (Guglielmi *et al.*, 2005; Hoepffner and Sathyendranath, 1991). The phenotypic plasticity allowed by CA4 could be an evolutionary response to such a highly variable environment. Additionally, 3dA isolates have the highest number of linker proteins encoded in their genomes (Six *et al.*, 2007c; Chapter I), suggesting larger phycobilisomes and thus more efficient light harvesting, which would be useful in these light-limited environments. The recent 3D structure of whole phycobilisomes in the red algae *Griffithsia pacifica* showed the presence of additional short peripheral rods branching onto the main phycobilisome rods (Zhang *et al.*, 2017). Some of the supplementary linkers in 3dA compared to other PT might be involved in the formation of such structures. It would thus be highly interesting to perform cryo-electron microscopy of phycobilisomes isolated from 3dA and other PT strains to compare their respective structures, and investigate the role of the various linker proteins. Together with time-resolved spectroscopy, this would help better understand the fundamental differences between *Synechococcus* PTs, which probably extend beyond the mere change in spectral properties and could include differences in light absorption efficiency and excitation energy transfer.

An interesting addition to the distribution analysis presented in this thesis would be to analyse the acclimation state of *in situ* populations. This should be technically straightforward with the sequencing of metatranscriptomes from the *Tara* Ocean expedition. By relating the relative transcript abundance of *mpeZ* and *mpeW*, which are respectively more expressed under BL in CA4-A and GL in CA4-B, to the corresponding *mpeY* allele, which is not differentially regulated by light color, one can probably infer the physiological status (i.e., either BL or GL-acclimated) of populations capable of chromatic acclimation. This would allow relating the acclimation status to both abiotic and biotic variables, and gain new insights about the adaptive value of this process in the natural environment.

The recent characterization *fciA* and *fciB*, two genes present in tandem in both CA4 genomic island versions, demonstrated that they are key regulators of the acclimation process (Sanfilippo *et al.*, 2016). However, the sensor(s) that primarily drive(s) CA4 remains elusive. Comparative analyses showed that no gene other than those found in the CA4 island are specific to CA4-able strains. Moreover, the results presented at the end of Chapter III, together with previous work (Humily *et al.*, 2013; Shukla *et al.*, 2012; Sanfilippo *et al.*, 2016) suggest that the CA4 island is both *necessary* and *sufficient* for CA4 in the genomic background tested. As the only genes common to both versions, *fciA* and *fciB* themselves appear as the best candidates for a light-colour sensing function. While the C-terminal domain of these regulators is predicted to bind DNA (Humily *et al.*, 2013), their N-terminus has no homology to known domains and might have such a function. It would be highly interesting to try purifying these proteins from cultures of *Synechococcus* through heterologous expression to investigate their DNA-binding and light-absorption properties, as well as the eventual presence of a ligand. In this regard, the presence of one highly conserved cysteine residue in the N-terminal domain of both FciA and FciB might be indicative of a covalently bound phycobilin. Alternatively, the sensor could be encoded in other parts of the genome, in which case FciA and FciB would only be involved in the regulation pathway of CA4. For example, the highly degenerated phycobiliprotein found in all HL *Prochlorococcus* binds a PEB and has been suggested to serve as a green light sensor (Steglich *et al.*, 2005), whereas the distant phycobiliprotein paralog AplA has been shown to covalently bind a chromophore yet is not involved in photosynthetic light harvesting (Montgomery *et al.*, 2004). Anyway, much work is still necessary before we can fully understand the role and functionality of these two proteins. The role of *fciC*, which is only present in the CA4-A island and strongly up-regulated in BL is also quite intriguing (Humily *et al.*, 2013, Sanfilippo *et al.*, 2016). Homology analyses suggest a regulatory role, but this raises the questions of why CA4-A would necessitate an additional regulator compared to CA4-B, and whether and how FciC interacts with FciA/B. In this regard, the fact that CA4-A strains respond differently to low ( $20 \mu\text{mol photons m}^{-2} \text{ s}^{-1}$ ) and high ( $75 \mu\text{mol photons m}^{-2} \text{ s}^{-1}$ ) levels of BL and GL, whereas CA4-B does not seem

affected by light intensity (Humily *et al.*, 2013) might indicate a more complex regulation in the former.

**Table 8: Genes identified by (Sanfilippo *et al.*, 2016) as presenting different transcript levels under blue- and green-light in the wild-type or between wild-type and *fciA*- or *fciB*- mutants.** Results from microarray experiments and RNA blot analysis done by A. Shukla during his PhD are also reported (Shukla, 2013). Genes of the CA4-A genomic island are highlighted with blue background, while genes located outside this island consistently identified in both studies are highlighted with red one. CK, Cyanorak ([www.sb-roscoff.fr/cyanorak/](http://www.sb-roscoff.fr/cyanorak/)); WT, wild-type; BL, blue light; GL, green light; n/a, not available; ns, non significant.

| Locus tag                    | CK cluster  | CK name      | CK product  | Shukla (2013)           |             | Sanfilippo <i>et al.</i> (2016)<br>RNA Seq |                           |                           |
|------------------------------|-------------|--------------|---|-------------------------|-------------|--|---------------------------|---------------------------|
|                              |             |              |   | $\mu$ array<br>BL vs GL | RNA<br>blot | WT<br>BL vs GL                             | <i>fciA</i> - vs WT<br>BL | <i>fciB</i> - vs WT<br>GL |
| RS9916_38072                 | CK_00000001 | <i>rpoD8</i> | RNA polymerase sigma factor type II                             | 2.92                    | n/a         | 2.1  | ns                        | ns                        |
| RS9916_26829                 | CK_00001475 |              | uncharacterized membrane protein DoxX family                    | 4.6                     | n/a         | 2.5  | ns                        | ns                        |
| RS9916_39536<br>RS9916_40506 | CK_00009137 | <i>fciC</i>  | ribbon-helix-helix domain-containing protein                    | 3.94                    | n/a         | 9.8  | 12.9                      | 8.6                       |
| RS9916_39531                 | CK_00009110 | <i>mpeZ</i>  | phycoerythrobilin:Cys-83 alpha-phycoerythrin II lyase-isomerase | 5.45                    | yes         | 35.3                                       | 27.1                      | 38.5                      |
| RS9916_39551                 | CK_00002279 | <i>unk10</i> | conserved hypothetical gene                                     | 35.45                   | yes         | 40.1                                       | 38.7                      | 44.1                      |
| RS9916_36677                 | CK_00001561 |              | small heat shock protein (HSP20) family protein                 | 2.9                     | yes         | ns   | ns                        | -3.4                      |
| RS9916_27639                 | CK_00001275 |              | conserved hypothetical protein                                  | 2.12                    | n/a         | ns   | ns                        | -2.1                      |
| RS9916_30189                 | CK_00000167 | <i>ahpC</i>  | 2-Cys peroxiredoxin   | ns                      | n/a         | ns   | ns                        | 2.6                       |
| RS9916_31622                 | CK_00001499 |              | ABC-type sugar transport system permease component              | ns                      | n/a         | ns   | ns                        | 2.8                       |
| RS9916_32652                 | CK_00002762 | <i>pilA</i>  | type IV pilin PilA  | 6.93                    | yes         | ns   | ns                        | 5                         |
| RS9916_36712                 | CK_00044664 |              | rubredoxin family protein                                       | 2.42                    | n/a         | ns   | ns                        | 6.9                       |
| RS9916_36722                 | CK_00001187 |              | NADPH-dependent FMN reductase family protein                    | ns                      | n/a         | ns   | ns                        | 8.4                       |
| RS9916_36727                 | CK_00001732 | <i>pirA</i>  | pirin-related protein   | ns                      | n/a         | ns   | ns                        | 10.3                      |

Finally, a yet overlooked question is the possible integration of CA4 with other cellular processes. As an example, the transcript levels of several genes were found to be significantly different between BL and GL, or between wild-type and *fciA*- or *fciB*- mutants, using RNA Seq (Table 8; Sanfilippo *et al.*, 2016). Surprisingly, not all these genes are directly involved in the biosynthesis of phycobilisome, and might constitute false-positives inherent to such studies. However, two of these genes (RS9916\_32652 and RS9916\_36712) had been previously found to have different transcript levels under BL or GL using microarray, and the transcript levels of RS9916\_32652, which encodes the Type IV pilin PilA, increased within 20 min after transfer of

cells grown under GL to BL, and increased to about 8-fold the GL abundance after three hours (**Table 8**; Shukla, 2013). Thus, different experiments using distinct approaches consistently show that this gene is light-colour regulated, suggesting a pleiotropic effect of CA4.

### III. Adaptive value of the different pigment types

Previous studies of the *in situ* distribution of “high PUB” and “low PUB” populations of picocyanobacteria in the marine environment (Olson *et al.*, 1990; Wood *et al.*, 1998) suggested a competitive advantage for “high PUB” lineages in clear blue waters and for “low PUB” in coastal green waters. Although the results we obtained for the distribution of all *Synechococcus* pigment types (Chapter II) were globally consistent with these findings, we also highlighted the unsuspected abundance in wild populations of chromatic acclimators, which were previously overlooked. However, this study only gives us a snapshot of *Synechococcus* populations and does not provide information on the population dynamics or on the extent of fitness differences between PTs. In this context, some insights have been obtained from culture-based experiments with other cyanobacteria (Stomp *et al.*, 2004, 2008), but as mentioned in the introduction, these experiments have been performed on cyanobacteria differing in their ecology, in particular in their salinity preferenda, but also in their morphology, making it difficult to conclude about the sole effect of light colour. Moreover, these studies focused on freshwater/euryhaline cyanobacteria, which only have “green” and “red” pigmentations (corresponding to marine *Synechococcus* PT 1 and PT 2, respectively), while the results presented in chapter II show that it is actually PT 3 that dominates *Synechococcus* populations in open waters of the world ocean. Thus, the central question of the adaptive value of the different *Synechococcus* PTs is still open. To answer it, one could design dedicated studies to quantify the costs associated to CA4 (including additional gene content, production of light sensors, etc.). Such information is crucial for understanding the ecological and evolutionary processes that led to the apparition and maintenance of the different PTs. For example, we hypothesized that the CA4 cost led to the loss of the CA4 regulation machinery in CA4-deficient *Synechococcus* populations, which dominate in ultra-oligotrophic areas of the South Pacific Ocean (chapter II).

Future culture-based experiments aiming at answering the question of the adaptive value of PTs should try to minimize the possible fitness differences arising from other traits than just the pigmentation. Ideally, one should compare strains from the same clade but displaying different PTs. In this regard, clade II is the best candidate for a comparative study, as this clade is the dominant one *in situ* (Farrant *et al.*, 2016; Sohm *et al.*, 2016) and displays the widest range of pigmentation (2, 3a, 3c, 3dB; see dataset 3 of Grébert *et al.*, 2017). Clade I is also a good choice, as PT 2, 3a and 3dA isolates have been reported for this clade. In order to control for possible effects of the genetic background of the strains, at least two strains of each PT/clade combination should be tested. Alternatively, one way to entirely alleviate variations due to differences in the genomic content of strains would be to genetically engineer a chassis strain by replacing the whole PBS genomic region with the different PT versions of it. This would allow comparing all

pigmentations without being limited by the availability of isolates, but would involve challenging genetic manipulation.

The light conditions to be tested should include constant BL, constant GL, different ratios of BL/GL, as well as successive shifts between BL and GL to simulate light color variations observed in the field. The frequency of such shifts seems to be particularly important, and different time-scales should be tested (Stomp *et al.*, 2008; Agostoni *et al.*, 2016). Experiments could compare fitness of individual pure cultures or of co-cultures of different PTs. Such experiments would allow quantifying precisely the performance of the different PTs under highly controlled conditions and relating the fitness differences with measured environmental parameters (light intensity, light spectrum, photosynthetic performance, cell density, growth rate, etc.). This should ultimately bring some fundamental insights about the selective pressure light colour exerts on *Synechococcus* PT evolution, and how this selective pressure might interact with other factors such as light intensity (Six *et al.*, 2007b; Mackey *et al.*, 2017), temperature (see chapter II) or nutrient availability.

#### **IV. Evolution of pigmentation and evolution of *Synechococcus***

An interesting outcome from the comparative genomic analysis presented in chapter I is the very strong genetic linkage between genes and/or genomic regions involved in the synthesis/regulation of PBS, despite them being not physically co-located, as it is the case for the co-occurring PBS rod region and CA4-A island in all 3dA *Synechococcus* strains sequenced so far. Yet, the contrary is not necessarily true, as some strains displaying a PT 3a PBS rod region also possess a CA4-A island (e.g. WH8016, KORDI-49 see chapter I and II). However, none of these strains is able to perform CA4, highlighting the importance of having both a 3dA PBS rod region and a CA4-A island for achieving fully functional CA4. Similarly, all PT 3 but strain SYN20 have an additional linker protein gene (*mpeE*) compared to PT 2. The location of this extra gene is highly variable between strains as it can be found within the PBS rod region or in unrelated locations on the genome, but its phyletic pattern suggests an important functional role in the PBS. This raises the question of the molecular mechanisms maintaining such strong genetic linkages.

We also demonstrated in chapter I that incomplete lineage sorting (ILS) is the most parsimonious explanation for the distribution pattern observed for PTs within clades. One key question arising from this result is how the wide diversity of pigmentation is maintained and transmitted within (and sometimes between) the different marine *Synechococcus* lineages. In a recent review about microbial evolution and speciation (Shapiro and Polz, 2015), the authors identified three different (non-exclusive) explanations to the apparent paradox of the high genetic diversity in cohesive microbial populations: frequency-dependent selection through biotic interactions (sometimes referred to as “kill the winner” hypothesis), fitness trade-offs (the different variants being adaptive in one niche and maladaptive in another one), and nearly neutral variants in terms of fitness. In the latter, which corresponds to ILS, the allelic diversity can be preserved through speciation, and will in fact be much older than the population itself, which is exactly what we suggest for the different *Synechococcus* PTs (Chapter I). Yet, we did not provide evidence for the assumption of “neutral alleles”. However, it should be noted that if one of the PTs had a much higher adaptive value than others (i.e. if cells with this PT had higher fitness), it would likely have driven the others to extinction, and we would not observe this pigment diversity nowadays. The ILS hypothesis thus raises the question of the selective forces and their dynamics for maintaining pigment diversity within the marine *Synechococcus* radiation: are the different PTs globally equivalent in terms of fitness over evolutionary time scales? If we assume that at any time, the PT that best absorbs dominant wavelengths has the highest fitness as suggested previously (Haverkamp *et al.*, 2008b; Stomp *et al.*, 2007), could it be that light conditions in the natural environment are too variable and transient to allow fixation of the “best”

allele? Such changes have been proposed by (Thompson *et al.*, 2005) as an explanation to the high genetic diversity observed in the marine heterotroph *Vibrio splendidus*: although different genotypes might be adaptive in some “microniches”, these differences disappear when averaged over the water column, and are thus effectively neutral. Similarly, variations of the light field over ecologically-relevant timescales could actively maintain *Synechococcus* pigment diversity. Alternatively, the different PTs could also correspond to different fitness trade-offs. For example, best absorbing the dominant light color is likely key under conditions where light is the growth limiting factor (e.g. at the bottom of the euphotic layer), but might be detrimental at stressful high light levels (e.g. in the upper mixed layer) where photoinhibition and photodamages are likely to occur (Six *et al.*, 2004; Mackey *et al.*, 2017). Experiments proposed in the previous paragraph should help one to precise which of these two scenarios is most probable. Finally, the population-scale genetic processes considered here are inherently linked to evolutionary forces acting on individual cells (selection) as well as to molecular mechanisms such as recombination. In this regard, *Synechococcus* PTs could constitute a good model to gain fundamental information about lateral gene transfer and recombination rates within and between *Synechococcus* clades as well as on the selective pressures acting on them, and thus on the balance between these important processes in shaping the population genomics and evolutionary dynamics of this ecologically key phytoplankton.





# REFERENCES

---

## A

- Acuña AM**, Lemaire C, Grondelle R van, Robert B, Stokkum IHM van. (2017). Energy transfer and trapping in *Synechococcus* WH 7803. *Photosynth Res* 1–10.
- Adir N**. (2005). Elucidation of the molecular structures of components of the phycobilisome: reconstructing a giant. *Photosynth Res* **85**: 15–32.
- Adir N**, Dines M, Klartag M, McGregor A, Melamed-Frank M. (2006). Assembly and disassembly of phycobilisomes. In: Microbiology Monographs. *Complex Intracellular Structures in Prokaryotes*. Springer, Berlin, Heidelberg, pp 47–77.
- Agostoni M**, Lucker BF, Smith MAY, Kanazawa A, Blanchard GJ, Kramer DM, *et al.* (2016). Competition-based phenotyping reveals a fitness cost for maintaining phycobilisomes under fluctuating light in the cyanobacterium *Fremyella diplosiphon*. *Algal Res* **15**: 110–119.
- Agusti S**, González-Gordillo JI, Vaqué D, Estrada M, Cerezo MI, Salazar G, *et al.* (2015). Ubiquitous healthy diatoms in the deep sea confirm deep carbon injection by the biological pump. *Nat Commun* **6**: ncomms8608.
- Ahlgren NA**, Rocap G. (2006). Culture isolation and culture-independent clone libraries reveal new marine *Synechococcus* ecotypes with distinctive light and N physiologies. *Appl Environ Microbiol* **72**: 7193–7204.
- Alvey RM**, Biswas A, Schluchter WM, Bryant DA. (2011). Effects of modified phycobilin biosynthesis in the cyanobacterium *Synechococcus* sp. strain PCC 7002. *J Bacteriol* **193**: 1663–1671.
- Apt KE**, Collier JL, Grossman AR. (1995). Evolution of the phycobiliproteins. *J Mol Biol* **248**: 79–96.
- Arnold W**, Oppenheimer JR. (1950). Internal conversion in the photosynthetic mechanism of blue-green algae. *J Gen Physiol* **33**: 423–435.
- Arteni AA**, Ajlani G, Boekema EJ. (2009). Structural organisation of phycobilisomes from *Synechocystis* sp. strain PCC6803 and their interaction with the membrane. *Biochim Biophys Acta* **1787**: 272–279.
- Avrani S**, Wurtzel O, Sharon I, Sorek R, Lindell D. (2011). Genomic island variability facilitates *Prochlorococcus*-virus coexistence. *Nature* **474**: 604–608.
- Azam F**. (1998). Microbial control of oceanic carbon flux: the plot thickens. *Science* **280**: 694–696.
- Azam F**, Fenchel T, Field JG, Gray JS, Meyer-Reil LA, Thingstad F. (1983). The ecological role of water-column microbes in the sea. *Mar Ecol Prog Ser* **10**: 257–263.

## B

- Bandara S**, Ren Z, Lu L, Zeng X, Shin H, Zhao K-H, *et al.* (2017). Photoactivation mechanism of a carotenoid-based photoreceptor. *Proc Natl Acad Sci* **114**: 6286–6291.
- Bao H**, Melnicki MR, Pawlowski EG, Sutter M, Agostoni M, Lechno-Yossef S, *et al.* (2017). Additional families of orange carotenoid proteins in the photoprotective system of cyanobacteria. *Nat Plants* **3**: 17089.
- Becker JW**, Berube PM, Follett CL, Waterbury JB, Chisholm SW, DeLong EF, *et al.* (2014). Closely related phytoplankton species produce similar suites of dissolved organic matter. *Front Microbiol* **5**. e-pub ahead of print, doi: 10.3389/fmicb.2014.00111.
- Béjà O**, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP, *et al.* (2000).

- Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**: 1902–1906.
- Berns DS.** (1967). Immunochemistry of Biliproteins. *Plant Physiol* **42**: 1569–1586.
- Bersanini L,** Battchikova N, Jokel M, Rehman A, Vass I, Allahverdiyeva Y, *et al.* (2014). Flavodiiron protein Flv2/Flv4-related photoprotective mechanism dissipates excitation pressure of PSII in cooperation with phycobilisomes in Cyanobacteria. *Plant Physiol* **164**: 805–818.
- Berube PM,** Biller SJ, Kent AG, Berta-Thompson JW, Roggensack SE, Roache-Johnson KH, *et al.* (2015). Physiology and evolution of nitrate acquisition in *Prochlorococcus*. *ISME J* **9**: 1195–1207.
- Biard T.** (2015). Diversité, biogéographie et écologie des Collodaires (Radiolaires) dans l’océan mondial. phdthesis, Université Pierre et Marie Curie - Paris VI. <https://tel.archives-ouvertes.fr/tel-01298213/document> (Accessed October 26, 2017).
- Bibby TS,** Nield J, Partensky F, Barber J. (2001). Oxyphotobacteria: Antenna ring around photosystem I. *Nature* **413**: 590–590.
- Biggins J,** Bruce D. (1989). Regulation of excitation energy transfer in organisms containing phycobilins. *Photosynth Res* **20**: 1–34.
- Biller SJ,** Berube PM, Lindell D, Chisholm SW. (2014a). *Prochlorococcus*: the structure and function of collective diversity. *Nat Rev Microbiol* **13**: nrmicro3378.
- Biller SJ,** Coe A, Chisholm SW. (2016). Torn apart and reunited: impact of a heterotroph on the transcriptome of *Prochlorococcus*. *ISME J* **10**: 2831–2843.
- Biller SJ,** McDaniel LD, Breitbart M, Rogers E, Paul JH, Chisholm SW. (2017). Membrane vesicles in sea water: heterogeneous DNA content and implications for viral abundance estimates. *ISME J* **11**: 394–404.
- Biller SJ,** Schubotz F, Roggensack SE, Thompson AW, Summons RE, Chisholm SW. (2014b). Bacterial vesicles in marine ecosystems. *Science* **343**: 183–186.
- Biswas A,** Boutaghou MN, Alvey RM, Kronfel CM, Cole RB, Bryant DA, *et al.* (2011). Characterization of the activities of the CpeY, CpeZ, and CpeS bilin lyases in phycoerythrin biosynthesis in *Fremyella diplosiphon* strain UTEX 481. *J Biol Chem* **286**: 35509–35521.
- Biswas A,** Vasquez YM, Dragomani TM, Kronfel ML, Williams SR, Alvey RM, *et al.* (2010). Biosynthesis of cyanobacterial phycobiliproteins in *Escherichia coli*: chromophorylation efficiency and specificity of all bilin lyases from *Synechococcus* sp. strain PCC 7002. *Appl Environ Microbiol* **76**: 2729–2739.
- Blank CE.** (2004). Evolutionary timing of the origins of mesophilic sulphate reduction and oxygenic photosynthesis: a phylogenomic dating approach. *Geobiology* **2**: 1–20.
- Blank CE,** Sánchez-Baracaldo P. (2010). Timing of morphological and ecological innovations in the cyanobacteria - a key to understanding the rise in atmospheric oxygen. *Geobiology* **8**: 1–23.
- Blankenship RE.** (2015). Structural and functional dynamics of photosynthetic antenna complexes. *Proc Natl Acad Sci U S A* **112**: 13751–13752.
- 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.
- Blot N,** Wu X-J, Thomas J-C, Zhang J, Garczarek L, Böhm S, *et al.* (2009).

- Phycourobilin in trichromatic phycocyanin from oceanic cyanobacteria is formed post-translationally by a phycoerythrobilin lyase-isomerase. *J Biol Chem* **284**: 9290–9298.
- Bouman HA**, Ulloa O, Barlow R, Li WKW, Platt T, Zwirgmaier K, *et al.* (2011). Water-column stratification governs the community structure of subtropical marine picophytoplankton. *Environ Microbiol Rep* **3**: 473–482.
- Brahamsha B.** (1996). A genetic manipulation system for oceanic cyanobacteria of the genus *Synechococcus*. *Appl Environ Microbiol* **62**: 1747–1751.
- Bretaudeau A**, Coste F, Humily F, Garczarek L, Le Corguillé G, Six C, *et al.* (2013). CyanoLyase: a database of phycobilin lyase sequences, motifs and functions. *Nucleic Acids Res* **41**: D396–401.
- Buchan A**, LeCleir GR, Gulvik CA, González JM. (2014). Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nat Rev Microbiol* **12**: nrmicro3326.
- Bussell AN**, Kehoe DM. (2013). Control of a four-color sensing photoreceptor by a two-color sensing photoreceptor reveals complex light regulation in cyanobacteria. *Proc Natl Acad Sci U S A* **110**: 12834–12839.
- C**
- Cabello-Yeves PJ**, Haro-Moreno JM, Martin-Cuadrado A-B, Ghai R, Picazo A, Camacho A, *et al.* (2017). Novel *Synechococcus* genomes reconstructed from freshwater reservoirs. *Front Microbiol* **8**. e-pub ahead of print, doi: 10.3389/fmicb.2017.01151.
- Camacho C**, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, *et al.* (2009). BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.
- Campbell D.** (1996). Complementary chromatic adaptation alters photosynthetic strategies in the cyanobacterium *Calothrix*. *Microbiology* **142**: 1255–1263.
- Campbell L**, Iturriaga R. (1988). Identification of *Synechococcus* spp. in the Sargasso Sea by immunofluorescence and fluorescence excitation spectroscopy performed on individual cells. *Limnol Oceanogr* **33**: 1196–1201.
- Campbell L**, Landry MR, Constantinou J, Nolla HA, Brown SL, Liu H, *et al.* (1998). Response of microbial community structure to environmental forcing in the Arabian Sea. *Deep Sea Res Part II Top Stud Oceanogr* **45**: 2301–2325.
- Campbell L**, Vaultot D. (1993). Photosynthetic picoplankton community structure in the subtropical North Pacific Ocean near Hawaii (station ALOHA). *Deep Sea Res Part Oceanogr Res Pap* **40**: 2043–2060.
- Caron DA**, Countway PD, Jones AC, Kim DY, Schnetzer A. (2012). Marine protistan diversity. *Annu Rev Mar Sci* **4**: 467–493.
- Celepli N**, Sundh J, Ekman M, Dupont CL, Yooseph S, Bergman B, *et al.* (2017). Meta-omic analyses of Baltic Sea cyanobacteria: diversity, community structure and salt acclimation. *Environ Microbiol* **19**: 673–686.
- Chandler JW**, Lin Y, Gainer PJ, Post AF, Johnson ZI, Zinser ER. (2016). Variable but persistent coexistence of *Prochlorococcus* ecotypes along temperature gradients in the ocean's surface mixed layer. *Environ Microbiol Rep* **8**: 272–284.
- Chang L**, Liu X, Li Y, Liu C-C, Yang F, Zhao J, *et al.* (2015). Structural organization of an intact phycobilisome

- and its association with photosystem II. *Cell Res* **25**: 726–737.
- Chen F**, Wang K, Kan J, Bachoon DS, Lu J, Lau S, *et al.* (2004). Phylogenetic diversity of *Synechococcus* in the Chesapeake Bay revealed by Ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) large subunit gene (*rbcL*) sequences. *Aquat Microb Ecol* **36**: 153–164.
- Chen PE**, Shapiro BJ. (2015). The advent of genome-wide association studies for bacteria. *Curr Opin Microbiol* **25**: 17–24.
- Chisholm SW**. (2017). *Prochlorococcus*. *Curr Biol* **27**: R447–R448.
- Chisholm SW**, Frankel SL, Goericke R, Olson RJ, Palenik B, Waterbury JB, *et al.* (1992). *Prochlorococcus marinus* nov. gen. nov. sp.: an oxyphototrophic marine prokaryote containing divinyl chlorophyll *a* and *b*. *Arch Microbiol* **157**: 297–300.
- Chisholm SW**, Olson RJ, Yentsch CM. (1988a). Flow cytometry in oceanography: status and prospects. *Eos Trans Am Geophys Union* **69**: 562–572.
- Chisholm SW**, Olson RJ, Zettler ER, Goericke R. (1987). Red-fluorescing ultra-plankton at the bottom of the euphotic zone: Analysis by flow cytometry (abstract). In: *Eos Trans. AGU* Vol. 68. p 1706.
- Chisholm SW**, Olson RJ, Zettler ER, Goericke R, Waterbury JB, Welschmeyer NA. (1988b). A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* **334**: 340–343.
- Choi DH**, Noh JH. (2009). Phylogenetic diversity of *Synechococcus* strains isolated from the East China Sea and the East Sea. *FEMS Microbiol Ecol* **69**: 439–448.
- Choi DH**, Noh JH, An SM, Choi YR, Lee H, Ra K, *et al.* (2016). Spatial distribution of cold-adapted *Synechococcus* during spring in seas adjacent to Korea. *ALGAE ALGAE* **31**: 231–241.
- Christie-Oleza JA**, Armengaud J, Guerin P, Scanlan DJ. (2015a). Functional distinctness in the exoproteomes of marine *Synechococcus*. *Environ Microbiol* **17**: 3781–3794.
- Christie-Oleza JA**, Scanlan DJ, Armengaud J. (2015b). ‘You produce while I clean up’, a strategy revealed by exoproteomics during *Synechococcus-Roseobacter* interactions. *PROTEOMICS* **15**: 3454–3462.
- Christie-Oleza JA**, Sousoni D, Lloyd M, Armengaud J, Scanlan DJ. (2017). Nutrient recycling facilitates long-term stability of marine microbial phototroph-heterotroph interactions. *Nat Microbiol* **2**: nmicrobiol2017100.
- Chukhutsina V**, Bersanini L, Aro E-M, van Amerongen H. (2015). Cyanobacterial light-harvesting phycobilisomes uncouple from photosystem I during dark-to-light transitions. *Sci Rep* **5**. e-pub ahead of print, doi: 10.1038/srep14193.
- Chung C-C**, Gong G-C, Huang C-Y, Lin J-Y, Lin Y-C. (2015). Changes in the *Synechococcus* assemblage composition at the surface of the East China Sea due to flooding of the Changjiang river. *Microb Ecol* **70**: 677–688.
- Cock PJA**, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, *et al.* (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**: 1422–1423.
- Coleman ML**, Chisholm SW. (2007). Code and context: *Prochlorococcus* as a model for cross-scale biology. *Trends Microbiol* **15**: 398–407.
- Coleman ML**, Sullivan MB, Martiny AC, Steglich C, Barry K, DeLong EF, *et al.* (2006). Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311**: 1768–1770.

- Coutinho F**, Tschoeke DA, Thompson F, Thompson C. (2016). Comparative genomics of *Synechococcus* and proposal of the new genus *Parasynechococcus*. *PeerJ* **4**. e-pub ahead of print, doi: 10.7717/peerj.1522.
- D**
- Dammeyer T**, Bagby SC, Sullivan MB, Chisholm SW, Frankenberg-Dinkel N. (2008). Efficient phage-mediated pigment biosynthesis in oceanic cyanobacteria. *Curr Biol* **18**: 442–448.
- De Los Ríos A**, Grube M, Sancho LG, Ascaso C. (2007). Ultrastructural and genetic characteristics of endolithic cyanobacterial biofilms colonizing Antarctic granite rocks. *FEMS Microbiol Ecol* **59**: 386–395.
- Delmont TO**, Kiefl E, Kilinc O, Esen OC, Uysal I, Rappe MS, *et al.* (2017). The global biogeography of amino acid variants within a single SAR11 population is governed by natural selection. *bioRxiv* 170639.
- Deutsch O**, Landan G, Roettger M, Gruenheit N, Kowallik KV, Allen JF, *et al.* (2008). Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor. *Mol Biol Evol* **25**: 748–761.
- Díez B**, Nylander JAA, Ininbergs K, Dupont CL, Allen AE, Yooseph S, *et al.* (2016). Metagenomic analysis of the indian ocean picocyanobacterial community: structure, potential function and evolution. *PLOS ONE* **11**: e0155757.
- Dolganov N**, Grossman AR. (1999). A polypeptide with similarity to phycocyanin  $\alpha$ -subunit phycocyanobilin lyase involved in degradation of phycobilisomes. *J Bacteriol* **181**: 610–617.
- Dong C**, Tang A, Zhao J, Mullineaux CW, Shen G, Bryant DA. (2009). ApcD is necessary for efficient energy transfer from phycobilisomes to photosystem I and helps to prevent photoinhibition in the cyanobacterium *Synechococcus* sp. PCC 7002. *Biochim Biophys Acta* **1787**: 1122–1128.
- Doron S**, Fedida A, Hernández-Prieto MA, Sabehi G, Karunker I, Stazic D, *et al.* (2016). Transcriptome dynamics of a broad host-range cyanophage and its hosts. *ISME J* **10**: 1437–1455.
- Doust AB**, Marai CNJ, Harrop SJ, Wilk KE, Curmi PMG, Scholes GD. (2004). Developing a structure-function model for the Cryptophyte phycoerythrin 545 using ultrahigh resolution crystallography and ultrafast laser spectroscopy. *J Mol Biol* **344**: 135–153.
- Drews G**. (2005). Contributions of Theodor Wilhelm Engelmann on phototaxis, chemotaxis, and photosynthesis. *Photosynth Res* **83**: 25–34.
- Ducret A**, Sidler W, Frank G, Zuber H. (1994). The complete amino acid sequence of R-phycocyanin-I alpha and beta subunits from the red alga *Porphyridium cruentum*. Structural and phylogenetic relationships of the phycocyanins within the phycobiliprotein families. *Eur J Biochem* **221**: 563–580.
- Dufresne A**, Garczarek L, Partensky F. (2005). Accelerated evolution associated with genome reduction in a free-living prokaryote. *Genome Biol* **6**: R14.
- Dufresne A**, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik BP, *et al.* (2008). Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol* **9**: R90.
- Dupont CL**, Johnson DA, Phillippy K, Paulsen IT, Brahamsha B, 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.

**Dvořák P**, Casamatta DA, Pouličková A, Hašler P, Ondřej V, Sanges R. (2014). *Synechococcus*: 3 billion years of global dominance. *Mol Ecol* **23**: 5538–5551.

## E

**Edelman M**, Mattoo AK. (2008). D1-protein dynamics in photosystem II: the lingering enigma. *Photosynth Res* **98**: 609–620.

**Eilertsen HC**, Degerlund M. (2010). Phytoplankton and light during the northern high-latitude winter. *J Plankton Res* **32**: 899–912.

**Eisenberg I**, Caycedo-Soler F, Harris D, Yochelis S, Huelga SF, Plenio MB, *et al.* (2017). Regulating the energy flow in a cyanobacterial light-harvesting antenna complex. *J Phys Chem B* **121**: 1240–1247.

**Engelmann TW**. (1882). Ueber Sauerstoffausscheidung von Pflanzenzellen im Mikrospektrum. *Arch Für Gesamte Physiol Menschen Tiere* **27**: 485–489.

**von Esenbeck N**. (1836). Ueber einen blau-rothen Farbstoff, der sich bei der Zersetzung von *Oscillatorien* bildet. *Ann Pharm* **17**: 75–82.

**Everroad C**, Six C, Partensky F, Thomas J-C, Holtzendorff J, Wood AM. (2006). Biochemical cases of type IV chromatic adaptation in marine *Synechococcus* spp. *J Bacteriol* **188**: 3345–3356.

**Everroad C**, Wood AM. (2012). Phycoerythrin evolution and diversification of spectral phenotype in marine *Synechococcus* and related picocyanobacteria. *Mol Phylogenet Evol* **64**: 381–392.

## F

**Fairchild CD**, Zhao J, Zhou J, Colson SE, Bryant DA, Glazer AN. (1992). Phycocyanin alpha-subunit phycocyanobilin lyase. *Proc Natl Acad Sci U S A* **89**: 7017–7021.

**Falk H**. (1989). The chemistry of linear oligopyrroles and bile pigments. Springer Vienna: Vienna.

**Falkowski PG**, Barber RT, Smetacek V. (1998). Biogeochemical controls and feedbacks on ocean primary production. *Science* **281**: 200–206.

**Farrant GK**, Doré H, Cornejo-Castillo FM, Partensky F, Ratin M, Ostrowski M, *et al.* (2016). Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proc Natl Acad Sci* **113**: E3365–E3374.

**Fedida A**, Lindell D. (2017). Two *Synechococcus* genes, two different effects on cyanophage infection. *Viruses* **9**. e-pub ahead of print, doi: 10.3390/v9060136.

**Field CB**, Behrenfeld MJ, Randerson JT, Falkowski P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**: 237–240.

**Flombaum P**, Gallegos JL, Gordillo RA, Rincón J, Zabala LL, Jiao N, *et al.* (2013). Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc Natl Acad Sci* **110**: 9824–9829.

**Fox GE**, Stackebrandt E, Hespell RB, Gibson J, Maniloff J, Dyer TA, *et al.* (1980). The phylogeny of prokaryotes. *Science* **209**: 457–463.

**Frankenberg N**, Mukougawa K, Kohchi T, Lagarias JC. (2001). Functional genomic analysis of the HY2 family of ferredoxin-dependent bilin reductases from oxygenic photosynthetic organisms. *Plant Cell* **13**: 965–978.

- Frankenberg-Dinkel N.** (2004). Bacterial heme oxygenases. *Antioxid Redox Signal* **6**: 825–834.
- Fridman S,** Flores-Uribe J, Larom S, Alalouf O, Liran O, Yacoby I, *et al.* (2017). A myovirus encoding both photosystem I and II proteins enhances cyclic electron flow in infected *Prochlorococcus* cells. *Nat Microbiol* **2**: 1350.
- Fuller NJ,** Marie D, Partensky F, Vault D, Post AF, Scanlan DJ. (2003). Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the red sea. *Appl Environ Microbiol* **69**: 2430–2443.
- Fuller NJ,** Tarran GA, Yallop M, Orcutt KM, Scanlan DJ. (2006). Molecular analysis of picocyanobacterial community structure along an Arabian Sea transect reveals distinct spatial separation of lineages. *Limnol Oceanogr* **51**: 2515–2526.
- G**
- Gantt E,** Conti SF. (1966). Phycobiliprotein localization in algae. *Brookhaven Symp Biol* **19**: 393–405.
- Gantt E,** Conti SF. (1965). The ultrastructure of *Porphyridium cruentum*. *J Cell Biol* **26**: 365–381.
- Garczarek L,** Dufresne A, Rousvoal S, West NJ, Mazard S, Marie D, *et al.* (2007). High vertical and low horizontal diversity of *Prochlorococcus* ecotypes in the Mediterranean Sea in summer. *FEMS Microbiol Ecol* **60**: 189–206.
- Garczarek L,** Hess WR, Holtzendorff J, van der Staay GWM, Partensky F. (2000). Multiplication of antenna genes as a major adaptation to low light in a marine prokaryote. *Proc Natl Acad Sci U S A* **97**: 4098–4101.
- Gasper R,** Schwach J, Hartmann J, Holtkamp A, Wiethaus J, Riedel N, *et al.* (2017). Auxiliary metabolic genes - distinct features of cyanophage-encoded T-type phycobiliprotein lyase  $\theta$ CpeT. *J Biol Chem* jbc.M116.769703.
- Ge B,** Lin X, Chen Y, Wang X, Chen H, Jiang P, *et al.* (2017). Combinational biosynthesis of dual-functional streptavidin-phycobiliproteins for high-throughput-compatible immunoassay. *Process Biochem* **58**: 306–312.
- Gierga G,** Voss B, Hess WR. (2012). Non-coding RNAs in marine *Synechococcus* and their regulation under environmentally relevant stress conditions. *ISME J* **6**: 1544–1557.
- Glazer AN.** (1989). Light guides. Directional energy transfer in a photosynthetic antenna. *J Biol Chem* **264**: 1–4.
- Glazer AN.** (1994). Phycobiliproteins - a family of valuable, widely used fluorophores. *J Appl Phycol* **6**: 105–112.
- Grébert T,** Doré H, Partensky F, Farrant GK, Emmanuel B, Picheral M, *et al.* (in revision). Light color acclimation: a key process in the global ocean distribution of *Synechococcus* cyanobacteria. *Proc Natl Acad Sci*.
- Grossman AR,** Bhaya D, He Q. (2001). Tracking the light environment by cyanobacteria and the dynamic nature of light harvesting. *J Biol Chem* **276**: 11449–11452.
- Grossman AR,** Schaefer MR, Chiang GG, Collier JL. (1993). The phycobilisome, a light-harvesting complex responsive to environmental conditions. *Microbiol Rev* **57**: 725–749.
- Guglielmi G,** Cohen-Bazire G, Bryant DA. (1981). The structure of *Gloeobacter*

- violaceus* and its phycobilisomes. *Arch Microbiol* **129**: 181–189.
- Guglielmi G**, Lavaud J, Rousseau B, Etienne A-L, Houmard J, Ruban AV. (2005). The light-harvesting antenna of the diatom *Phaeodactylum tricorutum*. *FEBS J* **272**: 4339–4348.
- 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**: 465–470.
- Guindon S**, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**: 307–321.
- Gutu A**, Kehoe DM. (2012). Emerging perspectives on the mechanisms, regulation, and distribution of light color acclimation in cyanobacteria. *Mol Plant* **5**: 1–13.
- Gutu A**, Nesbit AD, Alverson AJ, Palmer JD, Kehoe DM. (2013). Unique role for translation initiation factor 3 in the light color regulation of photosynthetic gene expression. *Proc Natl Acad Sci* **110**: 16253–16258.
- ## H
- Haeckel EHPA**. (1866). *Generelle morphologie der organismen. Allgemeine grundzüge der organischen formen-wissenschaft, mechanisch begründet durch die von Charles Darwin reformirte descendenztheorie*. Berlin, G. Reimer.
- Hansen G-A**, Henriksen K, Eilertsen H-C. (1997). Underwater spectral measurements. In: NATO ASI Series. *Solar Ultraviolet Radiation*. Springer, Berlin, Heidelberg, pp 143–154.
- Harris D**, Tal O, Jallet D, Wilson A, Kirilovsky D, Adir N. (2016). Orange carotenoid protein burrows into the phycobilisome to provide photoprotection. *Proc Natl Acad Sci* **113**: E1655–E1662.
- Haverkamp T**, Acinas SG, Doeleman M, Stomp M, Huisman J, Stal LJ. (2008a). Diversity and phylogeny of Baltic Sea picocyanobacteria inferred from their ITS and phycobiliprotein operons. *Environ Microbiol* **10**: 174–188.
- Haverkamp THA**, Schouten D, Doeleman M, Wollenzien U, Huisman J, Stal LJ. (2008b). Colorful microdiversity of *Synechococcus* strains (picocyanobacteria) isolated from the Baltic Sea. *ISME J* **3**: 397–408.
- He Q**, Dolganov N, Bjorkman O, Grossman AR. (2001). The high light-inducible polypeptides in *Synechocystis* PCC6803. Expression and function in high light. *J Biol Chem* **276**: 306–314.
- Herdman R**, Castenholz RW, Itean I, Waterbury JB, Rippka R. (2001). Subsection I (Formerly Chroococcales Wettstein 1924, emend. Rippka, Deruelles, Waterbury, Herdman and Stanier 1979). In: Vol. 1 The archaea and the deeply branching and phototrophic bacteria. *Bergey's Manual of Systematics of Archaea and Bacteria*. D.R. Boone, R.W. Catsenholtz and G.M. Garrity, pp 493–514.
- Hernández-Prieto MA**, Li Y, Postier BL, Blankenship RE, Chen M. (2017). Far-red light promotes biofilm formation in the cyanobacterium *Acaryochloris marina*. *Environ Microbiol*. e-pub ahead of print, doi: 10.1111/1462-2920.13961.
- Hess WR**, Partensky F, Staay GW van der, Garcia-Fernandez JM, Börner T, Vaultot D. (1996). Coexistence of phycoerythrin and a chlorophyll a/b antenna in a marine prokaryote. *Proc Natl Acad Sci* **93**: 11126–11130.
- Hess WR**, Steglich C, Lichtlé C, Partensky F. (1999). Phycoerythrins of the

- oxyphotobacterium *Prochlorococcus marinus* are associated to the thylakoid membrane and are encoded by a single large gene cluster. *Plant Mol Biol* **40**: 507–521.
- Hirose Y**, Misawa N, Yonekawa C, Nagao N, Watanabe M, Ikeuchi M, *et al.* (2017). Characterization of the genuine type 2 chromatic acclimation in the two *Geminocystis cyanobacteria*. *DNA Res* **24**: 387–396.
- Hirose Y**, Narikawa R, Katayama M, Ikeuchi M. (2010). Cyanobacteriochrome CcaS regulates phycoerythrin accumulation in *Nostoc punctiforme*, a group II chromatic adapter. *Proc Natl Acad Sci* **107**: 8854–8859.
- Hirose Y**, Rockwell N, Nishiyama K, Narikawa R, Ukaji Y, Inomata K, *et al.* (2013). Green/red cyanobacteriochromes regulate complementary chromatic acclimation via a protochromic photocycle. *Proc Natl Acad Sci U S A* **110**. e-pub ahead of print, doi: 10.1073/pnas.1302909110.
- Hirose Y**, Shimada T, Narikawa R, Katayama M, Ikeuchi M. (2008). Cyanobacteriochrome CcaS is the green light receptor that induces the expression of phycobilisome linker protein. *Proc Natl Acad Sci* **105**: 9528–9533.
- Ho M-Y**, Soulier NT, Canniffe DP, Shen G, Bryant DA. (2017). Light regulation of pigment and photosystem biosynthesis in cyanobacteria. *Curr Opin Plant Biol* **37**: 24–33.
- Hoepffner N**, Sathyendranath S. (1991). Effect of pigment composition on absorption properties of phytoplankton. *Mar Ecol-Prog Ser - MAR ECOL-PROGR SER* **73**: 11–23.
- Hoffmann L**. (1989). Algae of terrestrial habitats. *Bot Rev* **55**: 77–105.
- Hoge FE**, Wright CW, Kana TM, Swift RN, Yungel JK. (1998). Spatial variability of oceanic phycoerythrin spectral types derived from airborne laser-induced fluorescence emissions. *Appl Opt* **37**: 4744–4749.
- Huang S**, Wilhelm SW, Harvey HR, Taylor K, Jiao N, Chen F. (2012). Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *ISME J* **6**: 285–297.
- Humily F**. (2013). Étude génomique, métagénomique et physiologique de la diversité pigmentaire chez les cyanobactéries du genre *Synechococcus*. phdthesis, Université Pierre et Marie Curie - Paris VI. <https://tel.archives-ouvertes.fr/tel-00851664/document> (Accessed September 7, 2016).
- Humily F**, Farrant GK, Marie D, Partensky F, Mazard S, Perennou M, *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.
- Humily F**, Partensky F, Six C, Farrant GK, Ratin M, Marie D, *et al.* (2013). A gene island with two possible configurations is involved in chromatic acclimation in marine *Synechococcus*. *PLoS ONE* **8**: e84459.
- Hunter-Cevera KR**, Neubert MG, Olson RJ, Solow AR, Shalapyonok A, Sosik HM. (2016a). Physiological and ecological drivers of early spring blooms of a coastal phytoplankton. *Science* **354**: 326–329.
- Hunter-Cevera KR**, Post AF, Peacock EE, Sosik HM. (2016b). Diversity of *Synechococcus* at the Martha's Vineyard Coastal Observatory: insights from culture isolations, clone libraries, and flow cytometry. *Microb Ecol* **71**: 276–289.
- Hutchins DA**, Witter AE, Butler A, Luther GW. (1999). Competition among marine phytoplankton for different chelated iron species. *Nature* **400**: 858–861.

## J

**Jiang T**, Chai C, Wang J, Zhang L, Cen J, Lu S. (2016). Temporal and spatial variations of abundance of phycocyanin- and phycoerythrin-rich *Synechococcus* in Pearl River Estuary and adjacent coastal area. *J Ocean Univ China* **15**: 897–904.

**Johnson PW**, Sieburth JM. (1979). Chroococcoid cyanobacteria in the sea: a ubiquitous and diverse phototrophic biomass. *Limnol Oceanogr* **24**: 928–935.

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

**Joshua S**, Bailey S, Mann NH, Mullineaux CW. (2005). Involvement of phycobilisome diffusion in energy quenching in cyanobacteria. *Plant Physiol* **138**: 1577–1585.

**Joshua S**, Mullineaux CW. (2004). Phycobilisome diffusion is required for light-state transitions in cyanobacteria. *Plant Physiol* **135**: 2112–2119.

## K

**Kana TM**, Glibert PM. (1987). Effect of irradiances up to 2000  $\mu\text{E m}^{-2} \text{s}^{-1}$  on marine *Synechococcus* WH7803-II. Photosynthetic responses and mechanisms. *Deep Sea Res Part Oceanogr Res Pap* **34**: 497–516.

**Karradt A**, Sobanski J, Mattow J, Lockau W, Baier K. (2008). NblA, a key protein of phycobilisome degradation, interacts with ClpC, a HSP100 chaperone partner of a cyanobacterial Clp protease. *J Biol Chem* **283**: 32394–32403.

**Kashtan N**, Roggensack SE, Berta-Thompson JW, Grinberg M, Stepanauskas R, Chisholm SW. (2017).

Fundamental differences in diversity and genomic population structure between Atlantic and Pacific *Prochlorococcus*. *ISME J* **11**: 1997–2011.

**Kashtan N**, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A, *et al.* (2014). Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* **344**: 416–420.

**Katoh K**, Standley DM. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* **30**: 772–780.

**Ke B**. (2001). Photosynthesis. Springer Science & Business Media.

**Kehoe DM**, Grossman AR. (1996). Similarity of a chromatic adaptation sensor to phytochrome and ethylene receptors. *Science* **273**: 1409–1412.

**Kehoe DM**, Gutu A. (2006). Responding to color: the regulation of complementary chromatic adaptation. *Annu Rev Plant Biol* **57**: 127–150.

**Kerfeld CA**, Melnicki MR, Sutter M, Dominguez-Martin MA. (2017). Structure, function and evolution of the cyanobacterial orange carotenoid protein and its homologs. *New Phytol* **215**: 937–951.

**Kirilovsky D**, Kerfeld CA. (2016). Cyanobacterial photoprotection by the orange carotenoid protein. *Nat Plants* **2**: nplants2016180.

**Klein B**, LeBlanc B, Mei Z-P, Beret R, Michaud J, Mundy C-J, *et al.* (2002). Phytoplankton biomass, production and potential export in the North Water. *Deep Sea Res Part II Top Stud Oceanogr* **49**: 4983–5002.

**Knoll AH**, Nowak MA. (2017). The timetable of evolution. *Sci Adv* **3**: e1603076.

- Kondo K**, Mullineaux CW, Ikeuchi M. (2009). Distinct roles of CpcG1-phycobilisome and CpcG2-phycobilisome in state transitions in a cyanobacterium *Synechocystis* sp. PCC 6803. *Photosynth Res* **99**: 217–225.
- Kondo K**, Ochiai Y, Katayama M, Ikeuchi M. (2007). The membrane-associated CpcG2-phycobilisome in *Synechocystis*: a new photosystem I antenna. *Plant Physiol* **144**: 1200–1210.
- Konstantinidis KT**, Tiedje JM. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A* **102**: 2567–2572.
- Kopp RE**, Kirschvink JL, Hilburn IA, Nash CZ. (2005). The Paleoproterozoic snowball Earth: A climate disaster triggered by the evolution of oxygenic photosynthesis. *Proc Natl Acad Sci U S A* **102**: 11131–11136.
- Korn A**, Ajlani G, Lagoutte B, Gall A, Sétif P. (2009). Ferredoxin:NADP<sup>+</sup> oxidoreductase association with phycocyanin modulates its properties. *J Biol Chem* **284**: 31789–31797.
- Kützing FT**. (1843). *Phycologia generalis: oder, Anatomie, physiologie und systemkunde der tange*. Leipzig: F. A. Brockhaus.
- Kuzin AP**, Su M, Seetharaman J, Forouhar F, Wang D, Janjua H, *et al.* The crystal structure of fatty acid-binding protein-like Ycf58 from *Thermosynechococcus elongatus*. *Be Publ.* e-pub ahead of print, doi: 10.2210/pdb3bdr/pdb.
- Kylin H**. (1912). Über die roten und blauen Farbstoffe der Algen. *Hoppe-Seylers Z Für Physiol Chem* **76**: 396–425.
- Kylin H**. (1910). Über Phykoerythrin und Phykocyan bei *Ceramium rubrum* (Huds.) Ag. *Hoppe-Seylers Z Für Physiol Chem* **69**: 169–239.
- L**
- Lami R**, Cottrell MT, Ras J, Ulloa O, Obernosterer I, Claustre H, *et al.* (2007). High abundances of aerobic anoxygenic photosynthetic bacteria in the south pacific ocean. *Appl Environ Microbiol* **73**: 4198–4205.
- Lantoine F**, Neveux J. (1997). Spatial and seasonal variations in abundance and spectral characteristics of phycoerythrins in the tropical northeastern Atlantic Ocean. *Deep Sea Res Part Oceanogr Res Pap* **44**: 223–246.
- Larsson J**, Celepli N, Ininbergs K, Dupont CL, Yooseph S, Bergman B, *et al.* (2014). Picocyanobacteria containing a novel pigment gene cluster dominate the brackish water Baltic Sea. *ISME J* **8**: 1892–1903.
- Ledermann B**, Béjà O, Frankenberg-Dinkel N. (2016). New biosynthetic pathway for pink pigments from uncultured oceanic viruses. *Environ Microbiol* **18**: 4337–4347.
- Lemberg R**. (1933). Die Phycobiline der Rot-algen. Überführung in Mesobilirubin und Dehydro-mesobilirubin. *Justus Liebigs Ann Chem* **505**: 151–177.
- Leverenz RL**, Sutter M, Wilson A, Gupta S, Thurotte A, Carbon CB de, *et al.* (2015). A 12 Å carotenoid translocation in a photoswitch associated with cyanobacterial photoprotection. *Science* **348**: 1463–1466.
- Li B**, Sher D, Kelly L, Shi Y, Huang K, Knerr PJ, *et al.* (2010). Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria. *Proc Natl Acad Sci* **107**: 10430–10435.
- Li WKW**. (1998). Annual average abundance of heterotrophic bacteria and *Synechococcus* in surface ocean waters. *Limnol Oceanogr* **43**: 1746–1753.

- Li WKW.** (1994). Primary production of prochlorophytes, cyanobacteria, and eucaryotic ultraphytoplankton: Measurements from flow cytometric sorting. *Limnol Oceanogr* **39**: 169–175.
- Lindell D,** Jaffe JD, Johnson ZI, Church GM, Chisholm SW. (2005). Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438**: 86–89.
- Lindell D,** Post AF. (1995). Ultraphytoplankton succession is triggered by deep winter mixing in the Gulf of Aqaba (Eilat), Red Sea. *Limnol Oceanogr* **40**: 1130–1141.
- Lindell D,** Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F, Chisholm SW. (2004). Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc Natl Acad Sci U S A* **101**: 11013–11018.
- Liu H,** Jing H, Wong THC, Chen B. (2014). Co-occurrence of phycoerythrin- and phycoerythrin-rich *Synechococcus* in subtropical estuarine and coastal waters of Hong Kong: PE-rich and PC-rich *Synechococcus* in subtropical coastal waters. *Environ Microbiol Rep* **6**: 90–99.
- Liu H,** Zhang H, Niedzwiedzki DM, Prado M, He G, Gross ML, *et al.* (2013). Phycobilisomes supply excitations to both photosystems in a megacomplex in Cyanobacteria. *Science* **342**: 1104–1107.
- Liu L-N,** Chen X-L, Zhang Y-Z, Zhou B-C. (2005). Characterization, structure and function of linker polypeptides in phycobilisomes of cyanobacteria and red algae: an overview. *Biochim Biophys Acta* **1708**: 133–142.
- Lokstein H,** Steglich C, Hess WR. (1999). Light-harvesting antenna function of phycoerythrin in *Prochlorococcus marinus*. *Biochim Biophys Acta BBA - Bioenerg* **1410**: 97–98.
- Lundell DJ,** Yamanaka G, Glazer AN. (1981). A terminal energy acceptor of the phycobilisome: the 75,000-dalton polypeptide of *Synechococcus* 6301 phycobilisomes--a new biliprotein. *J Cell Biol* **91**: 315–319.
- ## M
- 
- MacColl** (1998). Cyanobacterial phycobilisomes. *J Struct Biol* **124**: 311–334.
- Mackey KRM,** Post AF, McIlvin MR, Saito MA. (2017). Physiological and proteomic characterization of light adaptations in marine *Synechococcus*. *Environ Microbiol* **19**: 2348–2365.
- Mahmoud RM,** Sanfilippo JE, Nguyen AA, Strnat JA, Partensky F, Garczarek L, *et al.* (2017). Adaptation to blue light in marine *Synechococcus* requires MpeU, an enzyme with similarity to phycoerythrobilin lyase isomerases. *Front Microbiol* **8**. e-pub ahead of print, doi: 10.3389/fmicb.2017.00243.
- Malviya S,** Scalco E, Audic S, Vincent F, Veluchamy A, Poulain J, *et al.* (2016). Insights into global diatom distribution and diversity in the world's ocean. *Proc Natl Acad Sci U S A* **113**: E1516–E1525.
- Martiny AC,** Coleman ML, Chisholm SW. (2006). Phosphate acquisition genes in *Prochlorococcus* ecotypes: evidence for genome-wide adaptation. *Proc Natl Acad Sci U S A* **103**: 12552–12557.
- Martiny AC,** Huang Y, Li W. (2009a). Occurrence of phosphate acquisition genes in *Prochlorococcus* cells from different ocean regions. *Environ Microbiol* **11**: 1340–1347.
- Martiny AC,** Kathuria S, Berube PM. (2009b). Widespread metabolic potential for nitrite and nitrate assimilation among *Prochlorococcus* ecotypes. *Proc Natl Acad Sci U S A* **106**: 10787–10792.
- Mary I,** Garczarek L, Tarran GA, Kolowrat C, Terry MJ, Scanlan DJ, *et al.* (2008). Diel rhythmicity in amino acid uptake by

- Prochlorococcus*. *Environ Microbiol* **10**: 2124–2131.
- Matsen FA**, Kodner RB, Armbrust EV. (2010). pplacer: linear time maximum-likelihood and bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* **11**: 538.
- Mazard S**, Ostrowski M, Partensky F, Scanlan DJ. (2012). Multi-locus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*. *Environ Microbiol* **14**: 372–386.
- McCarren J**, Brahamsha B. (2005). Transposon mutagenesis in a marine *Synechococcus* strain: isolation of swimming motility mutants. *J Bacteriol* **187**: 4457–4462.
- 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**. e-pub ahead of print, doi: 10.3389/fmicb.2012.00285.
- Melnicki MR**, Leverenz RL, Sutter M, López-Igual R, Wilson A, Pawlowski EG, *et al.* (2016). Structure, diversity, and evolution of a new family of soluble carotenoid-binding proteins in Cyanobacteria. *Mol Plant* **9**: 1379–1394.
- Miao D**, Ding W-L, Zhao B-Q, Lu L, Xu Q-Z, Scheer H, *et al.* (2016). Adapting photosynthesis to the near-infrared: non-covalent binding of phycocyanobilin provides an extreme spectral red-shift to phycobilisome core-membrane linker from *Synechococcus* sp. PCC7335. *Biochim Biophys Acta BBA - Bioenerg* **1857**: 688–694.
- Milias-Argeitis A**, Rullan M, Aoki SK, Buchmann P, Khammash M. (2016). Automated optogenetic feedback control for precise and robust regulation of gene expression and cell growth. *Nat Commun* **7**: ncomms12546.
- Miller CA**, Leonard HS, Pinsky IG, Turner BM, Williams SR, Harrison L, *et al.* (2008). Biogenesis of phycobiliproteins. III. CpcM is the asparagine methyltransferase for phycobiliprotein beta-subunits in cyanobacteria. *J Biol Chem* **283**: 19293–19300.
- Mock T**, Otilar RP, Strauss J, McMullan M, Paaanen P, Schmutz J, *et al.* (2017). Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. *Nature* **541**: nature20803.
- Montgomery BL**. (2017). Seeing new light: recent insights into the occurrence and regulation of chromatic acclimation in cyanobacteria. *Curr Opin Plant Biol* **37**: 18–23.
- Montgomery BL**, Casey ES, Grossman AR, Kehoe DM. (2004). AplA, a member of a new class of phycobiliproteins lacking a traditional role in photosynthetic light harvesting. *J Bacteriol* **186**: 7420–7428.
- 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. *Mar Ecol-Prog Ser - MAR ECOL-PROGR SER* **116**: 259–275.
- Moore LR**, Rocap G, Chisholm SW. (1998). Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393**: 464–467.
- Morris JJ**, Kirkegaard R, Szul MJ, Johnson ZI, Zinser ER. (2008). Facilitation of robust growth of *Prochlorococcus* colonies and dilute liquid cultures by ‘helper’ heterotrophic bacteria. *Appl Environ Microbiol* **74**: 4530–4534.
- Mullineaux CW**. (2014). Electron transport and light-harvesting switches in cyanobacteria. *Front Plant Sci* **5**. e-pub ahead of print, doi: 10.3389/fpls.2014.00007.

**Mullineaux CW**, Tobin MJ, Jones GR. (1997). Mobility of photosynthetic complexes in thylakoid membranes. *Nature* **390**: 421–424.

**Muramatsu M**, Hihara Y. (2012). Acclimation to high-light conditions in cyanobacteria: from gene expression to physiological responses. *J Plant Res* **125**: 11–39.

## N

**Neveux J**, Lantoine F, Vaultot D, Marie D, Blanchot J. (1999). Phycoerythrins in the southern tropical and equatorial Pacific Ocean: evidence for new cyanobacterial types. *J Geophys Res Oceans* **104**: 3311–3321.

**Not F**, Massana R, Latasa M, Marie D, Colson C, Eikrem W, *et al.* (2005). Late summer community composition and abundance of photosynthetic picoeukaryotes in Norwegian and Barents Seas. *Limnol Oceanogr* **50**: 1677–1686.

## O

**Ochoa de Alda JAG**, Esteban R, Diago ML, Houmard J. (2014). The plastid ancestor originated among one of the major cyanobacterial lineages. *Nat Commun* **5**: ncomms5937.

**Olson EJ**, Hartsough LA, Landry BP, Shroff R, Tabor JJ. (2014). Characterizing bacterial gene circuit dynamics with optically programmed gene expression signals. *Nat Methods* **11**: 449–455.

**Olson RJ**, Chisholm SW, Zettler ER, Armbrust EV. (1990). Pigments, size, and distributions of *Synechococcus* in the North Atlantic and Pacific Oceans. *Limnol Oceanogr* **35**: 45–58.

**Olson RJ**, Frankel SL, Chisholm SW, Shapiro HM. (1983). An inexpensive

flow cytometer for the analysis of fluorescence signals in phytoplankton: Chlorophyll and DNA distributions. *J Exp Mar Biol Ecol* **68**: 129–144.

**Ong LJ**, Glazer AN. (1991). Phycoerythrins of marine unicellular cyanobacteria. I. Bilin types and locations and energy transfer pathways in *Synechococcus* spp. phycoerythrins. *J Biol Chem* **266**: 9515–9527.

**Ong LJ**, Glazer AN. (1987). R-phycoerythrin II, a new phycoerythrin occurring in marine *Synechococcus* species. Identification of the terminal energy acceptor bilin in phycoerythrins. *J Biol Chem* **262**: 6323–6327.

## P

**Paasche E**. (2001). A review of the coccolithophorid *Emiliania huxleyi* (Prymnesiophyceae), with particular reference to growth, coccolith formation, and calcification-photosynthesis interactions. *Phycologia* **40**: 503–529.

**Paerl RW**, Johnson KS, Welsh RM, Worden AZ, Chavez FP, Zehr JP. (2011). Differential distributions of *Synechococcus* subgroups across the California current system. *Front Microbiol* **2**: 59.

**Palenik B**. (2001). Chromatic adaptation in marine *Synechococcus* strains. *Appl Environ Microbiol* **67**: 991–994.

**Palenik B**. (1994). Cyanobacterial community structure as seen from RNA polymerase gene sequence analysis. *Appl Environ Microbiol* **60**: 3212–3219.

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

- Partensky F**, Blanchot J, Vaultot D. (1999a). Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters: a review. *Bull Inst Océan* 457–475.
- Partensky F**, Garczarek L. (2010). *Prochlorococcus*: advantages and limits of minimalism. *Annu Rev Mar Sci* 2: 305–331.
- Partensky F**, Garczarek L. (2003). The photosynthetic apparatus of chlorophyll b- and d-containing oxyphotobacteria. In: Larkum AWD, Douglas SE, Raven JA (eds) *Advances in Photosynthesis and Respiration. Photosynthesis in Algae*. Springer Netherlands, pp 29–62.
- Partensky F**, Hess WR, Vaultot D. (1999b). *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev MMBR* 63: 106–127.
- Partensky F**, Roche JL, Wyman K, Falkowski PG. (1997). The divinyl-chlorophyll a/b-protein complexes of two strains of the oxyphototrophic marine prokaryote *Prochlorococcus* - characterization and response to changes in growth irradiance. *Photosynth Res* 51: 209–222.
- Paulsen ML**, Doré H, Garczarek L, Seuthe L, Müller O, Sandaa R-A, *et al.* (2016). *Synechococcus* in the Atlantic gateway to the Arctic Ocean. *Front Mar Sci* 3. e-pub ahead of print, doi: 10.3389/fmars.2016.00191.
- Paz-Yepes J**, Brahmsha B, Palenik B. (2013). Role of a microcin-C-like biosynthetic gene cluster in allelopathic interactions in marine *Synechococcus*. *Proc Natl Acad Sci U S A* 110: 12030–12035.
- Pedersen D**, Miller SR. (2017). Photosynthetic temperature adaptation during niche diversification of the thermophilic cyanobacterium *Synechococcus* A/B clade. *ISME J* 11: 1053–1057.
- Peng P-P**, Dong L-L, Sun Y-F, Zeng X-L, Ding W-L, Scheer H, *et al.* (2014). The structure of allophycocyanin B from *Synechocystis* PCC 6803 reveals the structural basis for the extreme redshift of the terminal emitter in phycobilisomes. *Acta Crystallogr D Biol Crystallogr* 70: 2558–2569.
- Pittera J**. (2015). Adaptations des cyanobactéries marines du genre *Synechococcus* au gradient latitudinal de température. phdthesis, Université Pierre et Marie Curie. <https://tel.archives-ouvertes.fr/tel-01287967/document> (Accessed October 23, 2017).
- Pittera J**, Humily F, Thorel M, Grulois D, Garczarek L, Six C. (2014). Connecting thermal physiology and latitudinal niche partitioning in marine *Synechococcus*. *ISME J* 8: 1221–1236.
- Pittera J**, Partensky F, Six C. (2016). Adaptive thermostability of light-harvesting complexes in marine picocyanobacteria. *ISME J*. e-pub ahead of print, doi: 10.1038/ismej.2016.102.
- Pizarro SA**, Sauer K. (2001). Spectroscopic study of the light-harvesting protein C-phycocyanin associated with colorless linker peptides. *Photochem Photobiol* 73: 556–563.
- Planet PJ**. (2006). Tree disagreement: Measuring and testing incongruence in phylogenies. *J Biomed Inform* 39: 86–102.

## R

- Rippka R**, Deruelles J, Waterbury JB, Herdman M, Stanier RY. (1979). Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *Microbiology* 111: 1–61.
- Rocap G**, Distel DL, Waterbury JB, Chisholm SW. (2002). Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal

- DNA internal transcribed spacer sequences. *Appl Environ Microbiol* **68**: 1180–1191.
- Rocap G**, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA, *et al.* (2003). Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**: 1042–1047.
- Rossi G**, Chance RR, Silbey R. (1989). Conformational disorder in conjugated polymers. *J Chem Phys* **90**: 7594–7601.
- ## S
- Saer RG**, Blankenship RE. (2017). Light harvesting in phototrophic bacteria: structure and function. *Biochem J* **474**: 2107–2131.
- Sagan L**. (1967). On the origin of mitosing cells. *J Theor Biol* **14**: 225–IN6.
- Saito MA**, Rocap G, Moffett JW. (2005). Production of cobalt binding ligands in a *Synechococcus* feature at the Costa Rica upwelling dome. *Limnol Oceanogr* **50**: 279–290.
- Sánchez-Baracaldo P**. (2015). Origin of marine planktonic cyanobacteria. *Sci Rep* **5**. e-pub ahead of print, doi: 10.1038/srep17418.
- Sánchez-Baracaldo P**, Ridgwell A, Raven JA. (2014). A neoproterozoic transition in the marine nitrogen cycle. *Curr Biol* **24**: 652–657.
- Sanfilippo JE**, Nguyen AA, Karty JA, Shukla A, Schluchter WM, Garczarek L, *et al.* (2016). Self-regulating genomic island encoding tandem regulators confers chromatic acclimation to marine *Synechococcus*. *Proc Natl Acad Sci* **113**: 6077–6082.
- Sato T**, Minagawa S, Kojima E, Okamoto N, Nakamoto H. (2010). HtpG, the prokaryotic homologue of Hsp90, stabilizes a phycobilisome protein in the cyanobacterium *Synechococcus elongatus* PCC 7942. *Mol Microbiol* **76**: 576–589.
- Sauer K**, Scheer H. (1988). Excitation transfer in C-phycocyanin. Förster transfer rate and exciton calculations based on new crystal structure data for C-phycocyanins from *Agmenellum quadruplicatum* and *Mastigocladus laminosus*. *Biochim Biophys Acta BBA - Bioenerg* **936**: 157–170.
- Saunée NA**, Williams SR, Bryant DA, Schluchter WM. (2008). Biogenesis of Phycobiliproteins II. CpcS-I and CpcU comprise the heterodimeric bilin lyase that attaches phycocyanobilin to cys-82 of  $\beta$ -phycocyanin and cys-81 of allophycocyanin subunits in *Synechococcus* sp. PCC 7002. *J Biol Chem* **283**: 7513–7522.
- Scanlan DJ**. (2012). Marine Picocyanobacteria. In: *Ecology of Cyanobacteria II*. Springer, Dordrecht, pp 503–533.
- Scanlan DJ**, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR, *et al.* (2009). Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol Rev* **73**: 249–299.
- Scheer H**, Zhao K-H. (2008). Biliprotein maturation: the chromophore attachment. *Mol Microbiol* **68**: 263–276.
- Schirmer T**, Bode W, Huber R. (1987). Refined three-dimensional structures of two cyanobacterial C-phycocyanins at 2.1 and 2.5 Å resolution. A common principle of phycobilin-protein interaction. *J Mol Biol* **196**: 677–695.
- Schirrmeister BE**, Gugger M, Donoghue PCJ. (2015). Cyanobacteria and the Great Oxidation Event: evidence from genes and fossils. *Palaeontology* **58**: 769–785.

- Schirrmeister BE**, Sanchez-Baracaldo P, Wacey D. (2016). Cyanobacterial evolution during the Precambrian. *Int J Astrobiol* **15**: 187–204.
- Schluchter WM**, Shen G, Alvey RM, Biswas A, Saunée NA, Williams SR, *et al.* (2010). Phycobiliprotein biosynthesis in cyanobacteria: structure and function of enzymes involved in post-translational modification. *Adv Exp Med Biol* **675**: 211–228.
- Shapiro BJ**, Polz MF. (2015). Microbial Speciation. *Cold Spring Harb Perspect Biol* **7**: a018143.
- Shen G**, Leonard HS, Schluchter WM, Bryant DA. (2008a). CpcM posttranslationally methylates asparagine-71/72 of phycobiliprotein beta subunits in *Synechococcus* sp. strain PCC 7002 and *Synechocystis* sp. strain PCC 6803. *J Bacteriol* **190**: 4808–4817.
- Shen G**, Saunée NA, Williams SR, Gallo EF, Schluchter WM, Bryant DA. (2006). Identification and characterization of a new class of bilin lyase. *J Biol Chem* **281**: 17768–17778.
- Shen G**, Schluchter WM, Bryant DA. (2008b). Biogenesis of Phycobiliproteins I. cpcS-I and cpcU mutants of the cyanobacterium *Synechococcus* sp. PCC 7002 define a heterodimeric phycocyanobilin lyase specific for  $\beta$ -phycocyanin and allophycocyanin subunits. *J Biol Chem* **283**: 7503–7512.
- Sherry ND**, Wood AM. (2001). Phycoerythrin-containing picocyanobacteria in the Arabian Sea in february 1995: diel patterns, spatial variability, and growth rates. *Deep Sea Res Part II Top Stud Oceanogr* **48**: 1263–1283.
- Shih PM**, Wu D, Latifi A, Axen SD, Fewer DP, Talla E, *et al.* (2013). Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci* **110**: 1053–1058.
- Shimodaira H.** (2002). An approximately unbiased test of phylogenetic tree selection. *Syst Biol* **51**: 492–508.
- Shukla A.** (2013). Molecular characterization of the mechanisms controlling blue-green light responses in marine *Synechococcus*. Ph.D., Ann Arbor, United States. <https://search.proquest.com/docview/1419346288/abstract/6E4EF372C5974B4CPQ/1> (Accessed November 7, 2017).
- Shukla A**, Biswas A, Blot N, Partensky F, Karty JA, Hammad LA, *et al.* (2012). Phycoerythrin-specific bilin lyase-isomerase controls blue-green chromatic acclimation in marine *Synechococcus*. *Proc Natl Acad Sci* **109**: 20136–20141.
- Sidler W**, Nutt H, Kumpf B, Frank G, Suter F, Brenzel A, *et al.* (1990). The complete amino-acid sequence and the phylogenetic origin of phycocyanin-645 from the cryptophyten alga *Chroomonas* sp. *Biol Chem Hoppe Seyler* **371**: 537–547.
- Sidler WA.** (1994). Phycobilisome and phycobiliprotein structures. In: Advances in Photosynthesis. *The Molecular Biology of Cyanobacteria*. Springer, Dordrecht, pp 139–216.
- Sieburth JM.** (1978). About bacterioplankton. In: *Phytoplankton manual*. Sournia A: UNESCO, Paris, pp 283–287.
- Six C**, Finkel ZV, Irwin AJ, Campbell DA. (2007a). Light variability illuminates niche-partitioning among marine Picocyanobacteria. *PLOS ONE* **2**: e1341.
- Six C**, Joubin L, Partensky F, Holtzendorff J, Garczarek L. (2007b). UV-induced phycobilisome dismantling in the marine picocyanobacterium *Synechococcus* sp. WH8102. *Photosynth Res* **92**: 75–86.
- Six C**, Thomas JC, Brahamsha B, Lemoine Y, Partensky F. (2004). Photophysiology of the marine cyanobacterium *Synechococcus* sp. WH8102, a new

- model organism. *Aquat Microb Ecol* **35**: 17–29.
- Six C**, Thomas J-C, Garczarek L, Ostrowski M, Dufresne A, Blot N, *et al.* (2007c). Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.: a comparative genomics study. *Genome Biol* **8**: R259.
- Six C**, Thomas J-C, Thion L, Lemoine Y, Zal F, Partensky F. (2005). Two novel phycoerythrin-associated linker proteins in the marine cyanobacterium *Synechococcus* sp. strain WH8102. *J Bacteriol* **187**: 1685–1694.
- Smetacek V.** (1999). Diatoms and the ocean carbon cycle. *Protist* **150**: 25–32.
- Sohm JA**, Ahlgren NA, Thomson ZJ, Williams C, Moffett JW, Saito MA, *et al.* (2016). Co-occurring *Synechococcus* ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. *ISME J* **10**: 333–345.
- Sorby HC.** (1875). On the characteristic colouring-matters of the red groups of algae. *J Linn Soc Lond Bot* **15**: 34–40.
- Squires AH**, Moerner WE. (2017). Direct single-molecule measurements of phycocyanobilin photophysics in monomeric C-phycocyanin. *Proc Natl Acad Sci* **114**: 9779–9784.
- Stadnichuk IN**, Krasilnikov PM, Zlenko DV, Freidzon AY, Yanyushin MF, Rubin AB. (2015). Electronic coupling of the phycobilisome with the orange carotenoid protein and fluorescence quenching. *Photosynth Res* **124**: 315–335.
- Stanier RY**, Cohen-Bazire G. (1977). Phototrophic prokaryotes: the Cyanobacteria. *Annu Rev Microbiol* **31**: 225–274.
- Steglich C**, Frankenberg-Dinkel N, Penno S, Hess WR. (2005). A green light-absorbing phycoerythrin is present in the high-light-adapted marine cyanobacterium *Prochlorococcus* sp. MED4. *Environ Microbiol* **7**: 1611–1618.
- Stokes GG.** (1854). Ueber die Metallreflexion an gewissen nichtmetallischen Substanzen. *Ann Phys* **167**: 300–314.
- van Stokkum IHM**, Gwizdala M, Tian L, Snellenburg JJ, van Grondelle R, van Amerongen H, *et al.* (2017). A functional compartmental model of the *Synechocystis* PCC 6803 phycobilisome. *Photosynth Res.* e-pub ahead of print, doi: 10.1007/s11120-017-0424-5.
- Stomp M.** (2008). Colourful coexistence : a new solution to the plankton paradox.
- Stomp M**, van Dijk MA, van Overzee HMJ, Wortel MT, Sigon CAM, Egas M, *et al.* (2008). The timescale of phenotypic plasticity and its impact on competition in fluctuating environments. *Am Nat* **172**: E169–E185.
- Stomp M**, Huisman J, de Jongh F, Veraart AJ, Gerla D, Rijkeboer M, *et al.* (2004). Adaptive divergence in pigment composition promotes phytoplankton biodiversity. *Nature* **432**: 104–107.
- Stomp M**, Huisman J, Stal LJ, Matthijs HCP. (2007). Colorful niches of phototrophic microorganisms shaped by vibrations of the water molecule. *ISME J* **1**: 271–282.
- Stuart RK**, Dupont CL, Johnson DA, Paulsen IT, Palenik B. (2009). Coastal strains of marine *Synechococcus* species exhibit increased tolerance to copper shock and a distinctive transcriptional response relative to those of open-ocean strains. *Appl Environ Microbiol* **75**: 5047–5057.
- Sullivan MB**, Lindell D, Lee JA, Thompson LR, Bielawski JP, Chisholm SW. (2006). Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol* **4**: e234.

- Sunagawa S**, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, *et al.* (2015). Structure and function of the global ocean microbiome. *Science* **348**: 1261359.
- Suttle CA.** (2007). Marine viruses - major players in the global ecosystem. *Nat Rev Microbiol* **5**: nrmicro1750.
- ## T
- Tai V**, Paulsen IT, Phillippy K, Johnson DA, Palenik B. (2009). Whole-genome microarray analyses of *Synechococcus-Vibrio* interactions. *Environ Microbiol* **11**: 2698–2709.
- Tal O**, Trabelcy B, Gerchman Y, Adir N. (2014). Investigation of phycobilisome subunit interaction interfaces by coupled cross-linking and mass spectrometry. *J Biol Chem* **289**: 33084–33097.
- Tamary E**, Kiss V, Nevo R, Adam Z, Bernát G, Rexroth S, *et al.* (2012). Structural and functional alterations of cyanobacterial phycobilisomes induced by high-light stress. *Biochim Biophys Acta BBA - Bioenerg* **1817**: 319–327.
- Tandeau de Marsac N.** (2003). Phycobiliproteins and phycobilisomes: the early observations. *Photosynth Res* **76**: 193–205.
- Tandeau de Marsac N**, Cohen-bazire G. (1977). Molecular composition of cyanobacterial phycobilisomes. *Proc Natl Acad Sci* **74**: 1635–1639.
- Tang K**, Ding W-L, Höppner A, Zhao C, Zhang L, Hontani Y, *et al.* (2015). The terminal phycobilisome emitter, L<sub>CM</sub>: A light-harvesting pigment with a phytochrome chromophore. *Proc Natl Acad Sci* **112**: 15880–15885.
- Tara Oceans Consortium C**, Tara Oceans Expedition P. (2017). Registry of all samples from the Tara Oceans Expedition (2009-2013). e-pub ahead of print, doi: <https://doi.org/10.1594/PANGAEA.875582>.
- Taylor BB**, Taylor MH, Dinter T, Bracher A. (2013). Estimation of relative phycoerythrin concentrations from hyperspectral underwater radiance measurements-A statistical approach. *J Geophys Res Oceans* **118**: 2948–2960.
- Teale FWJ**, Dale RE. (1970). Isolation and spectral characterization of phycobiliproteins. *Biochem J* **116**: 161–169.
- Tetu SG**, Brahamsha B, Johnson DA, Tai V, Phillippy K, Palenik B, *et al.* (2009). Microarray analysis of phosphate regulation in the marine cyanobacterium *Synechococcus* sp. WH8102. *ISME J* **3**: 835–849.
- Tetu SG**, Johnson DA, Varkey D, Phillippy K, Stuart RK, Dupont CL, *et al.* (2013). Impact of DNA damaging agents on genome-wide transcriptional profiles in two marine *Synechococcus* species. *Front Microbiol* **4**: 232.
- Thomas J-C**, Ughy B, Lagoutte B, Ajlani G. (2006). A second isoform of the ferredoxin:NADP oxidoreductase generated by an in-frame initiation of translation. *Proc Natl Acad Sci* **103**: 18368–18373.
- Thompson JR**, Pacocha S, Pharino C, Klepac-Ceraj V, Hunt DE, Benoit J, *et al.* (2005). Genotypic diversity within a natural coastal bacterioplankton population. *Science* **307**: 1311–1313.
- Thompson LR**, Zeng Q, Chisholm SW. (2016). Gene expression patterns during light and dark infection of *Prochlorococcus* by Cyanophage. *PLOS ONE* **11**: e0165375.
- Ting CS**, Rocap G, King J, Chisholm SW. (2002). Cyanobacterial photosynthesis in the oceans: the origins and significance of divergent light-harvesting strategies. *Trends Microbiol* **10**: 134–142.

- Toledo G**, Palenik B, Brahamsha B. (1999). Swimming marine *Synechococcus* strains with widely different photosynthetic pigment ratios form a monophyletic group. *Appl Environ Microbiol* **65**: 5247–5251.
- Tomitani A**, Knoll AH, Cavanaugh CM, Ohno T. (2006). The evolutionary diversification of cyanobacteria: molecular-phylogenetic and paleontological perspectives. *Proc Natl Acad Sci U S A* **103**: 5442–5447.
- Toole CM**, Allnut FCT. (2003). Red, cryptomonad and glaucocystophyte algal phycobiliproteins. In: *Advances in Photosynthesis and Respiration. Photosynthesis in Algae*. Springer, Dordrecht, pp 305–334.
- Tragin M**, Lopes dos Santos A, Christen R, Vault D. (2016). Diversity and ecology of green microalgae in marine systems: an overview based on 18S rRNA gene sequences. *Perspect Phycol* 141–154.
- Tremblay J-E**, Gratton Y, Fauchot J, Price NM. (2002). Climatic and oceanic forcing of new, net, and diatom production in the North Water. *Deep Sea Res Part II Top Stud Oceanogr* **49**: 4927–4946.
- Tremblay J-É**, Michel C, Hobson KA, Gosselin M, Price NM. (2006). Bloom dynamics in early opening waters of the Arctic Ocean. *Limnol Oceanogr* **51**: 900–912.
- Tsiola A**, Pitta P, Fodelianakis S, Pete R, Magiopoulos I, Mara P, *et al.* (2016). Nutrient limitation in surface waters of the oligotrophic eastern Mediterranean Sea: an enrichment microcosm experiment. *Microb Ecol* **71**: 575–588.
- Turner S**, Pryer KM, Miao VP, Palmer JD. (1999). Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *J Eukaryot Microbiol* **46**: 327–338.
- Ueno Y**, Aikawa S, Kondo A, Akimoto S. (2016). Energy transfer in cyanobacteria and red algae: confirmation of spillover in intact megacomplexes of phycobilisome and both photosystems. *J Phys Chem Lett* **7**: 3567–3571.
- Urbach E**, Robertson DL, Chisholm SW. (1992). Multiple evolutionary origins of prochlorophytes within the cyanobacterial radiation. *Nature* **355**: 267–270.
- Vargas C de**, Audic S, Henry N, Decelle J, Mahé F, Logares R, *et al.* (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**: 1261605.
- Varkey D**, Mazard S, Ostrowski M, Tetu SG, Haynes P, Paulsen IT. (2016). Effects of low temperature on tropical and temperate isolates of marine *Synechococcus*. *ISME J* **10**: 1252–1263.
- Vault D**, Eikrem W, Viprey M, Moreau H. (2008). The diversity of small eukaryotic phytoplankton ( $\leq 3 \mu\text{m}$ ) in marine ecosystems. *FEMS Microbiol Rev* **32**: 795–820.
- Venayak N**, Anesiadis N, Cluett WR, Mahadevan R. (2015). Engineering metabolism through dynamic control. *Curr Opin Biotechnol* **34**: 142–152.
- Wang Q**, Moerner WE. (2015). Dissecting pigment architecture of individual photosynthetic antenna complexes in solution. *Proc Natl Acad Sci* **112**: 13880–13885.
- Watanabe M**, Ikeuchi M. (2013). Phycobilisome: architecture of a light-

- harvesting supercomplex. *Photosynth Res* **116**: 265–276.
- Watanabe M**, Semchonok DA, Webber-Birungi MT, Ehira S, Kondo K, Narikawa R, *et al.* (2014). Attachment of phycobilisomes in an antenna-photosystem I supercomplex of cyanobacteria. *Proc Natl Acad Sci* **111**: 2512–2517.
- Waterbury JB**, Rippka R. (1989). Subsection I. Order Chroococcales Wettstein 1924, emend. Rippka *et al.*, 1979. In: Vol. 3. *Bergey's Manual of Systematics of Archaea and Bacteria*. J.T. Staley, M.P. Bryant, N. Pfennig, and J.G. Holt, pp 1728–1746.
- Waterbury JB**, Watson SW, Guillard RRL, Brand LE. (1979). Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. *Nature* **277**: 293–294.
- Waterbury JB**, Watson SW, Valois FW, Franks DG. (1986). Biological and ecological characterization of the marine unicellular cyanobacterium *Synechococcus*. *Can Bull Fish Aquat Sci* **214**: 71–120.
- Whitton BA** (ed). (2012). Ecology of Cyanobacteria II. Springer Netherlands: Dordrecht.
- Wiethaus J**, Busch AWU, Dammeyer T, Frankenberg-Dinkel N. (2010a). Phycobiliproteins in *Prochlorococcus marinus*: biosynthesis of pigments and their assembly into proteins. *Eur J Cell Biol* **89**: 1005–1010.
- Wiethaus J**, Busch AWU, Kock K, Leichert LI, Herrmann C, Frankenberg-Dinkel N. (2010b). CpeS is a lyase specific for attachment of 3Z-PEB to Cys82 of  $\beta$ -phycoerythrin from *Prochlorococcus marinus* MED4. *J Biol Chem* **285**: 37561–37569.
- Wilhelm SW**, Suttle CA. (1999). Viruses and nutrient cycles in the sea: viruses play critical roles in the structure and function of aquatic food webs. *BioScience* **49**: 781–788.
- Wiltbank LB**, Kehoe DM. (2016). Two cyanobacterial photoreceptors regulate photosynthetic light harvesting by sensing teal, green, yellow, and red light. *mBio* **7**: e02130-15.
- Woese CR**. (1987). Bacterial evolution. *Microbiol Rev* **51**: 221–271.
- Wood AM**, Lipsen M, Coble P. (1999). Fluorescence-based characterization of phycoerythrin-containing cyanobacterial communities in the Arabian Sea during the Northeast and early Southwest Monsoon (1994–1995). *Deep Sea Res Part II Top Stud Oceanogr* **46**: 1769–1790.
- Wood AM**, Phinney DA, Yentsch CS. (1998). Water column transparency and the distribution of spectrally distinct forms of phycoerythrin-containing organisms. *Mar Ecol Prog Ser* **162**: 25–31.
- Worden AZ**, Follows MJ, Giovannoni SJ, Wilken S, Zimmerman AE, Keeling PJ. (2015). Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science* **347**: 1257594.
- Wyman M**, Gregory RPF, Carr NG. (1985). Novel role for phycoerythrin in a marine cyanobacterium, *Synechococcus* strain DC2. *Science* **230**: 818–820.
- ## X
- Xia X**, Guo W, Tan S, Liu H. (2017a). *Synechococcus* assemblages across the salinity gradient in a salt wedge estuary. *Front Microbiol* **8**. e-pub ahead of print, doi: 10.3389/fmicb.2017.01254.
- Xia X**, Liu H, Choi D, Noh JH. (2017b). Variation of *Synechococcus* pigment genetic diversity along two turbidity

gradients in the China Seas. *Microb Ecol* 1–12.

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

## Y

**Yang Z**, Nielsen R. (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol* **19**: 908–917.

**Yelton AP**, Acinas SG, Sunagawa S, Bork P, Pedrós-Alió C, Chisholm SW. (2016). Global genetic capacity for mixotrophy in marine picocyanobacteria. *ISME J* **10**: 2946–2957.

**Yona D**, Park MO, Oh SJ, Shin WC. (2014). Distribution of *Synechococcus* and its phycoerythrin pigment in relation to environmental factors in the East Sea, Korea. *Ocean Sci J* **49**: 367–382.

## Z

**Zehr JP**. (2011). Nitrogen fixation by marine cyanobacteria. *Trends Microbiol* **19**: 162–173.

**Zhang J**, Ma J, Liu D, Qin S, Sun S, Zhao J, *et al.* (2017). Structure of phycobilisome from the red alga *Griffithsia pacifica*. *Nature advance online publication*. e-pub ahead of print, doi: 10.1038/nature24278.

**Zhao F**, Qin S. (2006). Evolutionary analysis of phycobiliproteins: implications for their structural and

functional relationships. *J Mol Evol* **63**: 330–340.

**Zhao KH**, Deng MG, Zheng M, Zhou M, Parbel A, Storf M, *et al.* (2000). Novel activity of a phycobiliprotein lyase: both the attachment of phycocyanobilin and the isomerization to phycoviolobin are catalyzed by the proteins PecE and PecF encoded by the phycoerythrocyanin operon. *FEBS Lett* **469**: 9–13.

**Zhao K-H**, Porra RJ, Scheer H. (2012). Phycobiliproteins. In: Handbook of Porphyrin Science Vol. Volume 22. *Handbook of Porphyrin Science*. World Scientific Publishing Company, pp 1–66.

**Zhao K-H**, Su P, Böhm S, Song B, Zhou M, Bubenzer C, *et al.* (2005a). Reconstitution of phycobilisome core-membrane linker, L<sub>CM</sub>, by autocatalytic chromophore binding to ApcE. *Biochim Biophys Acta BBA - Bioenerg* **1706**: 81–87.

**Zhao K-H**, Su P, Li J, Tu J-M, Zhou M, Bubenzer C, *et al.* (2006). Chromophore attachment to phycobiliprotein beta-subunits: phycocyanobilin:cysteine-beta84 phycobiliprotein lyase activity of CpeS-like protein from *Anabaena* sp. PCC7120. *J Biol Chem* **281**: 8573–8581.

**Zhao K-H**, Su P, Tu J-M, Wang X, Liu H, Plöscher M, *et al.* (2007a). Phycobilin:cystein-84 biliprotein lyase, a near-universal lyase for cysteine-84-binding sites in cyanobacterial phycobiliproteins. *Proc Natl Acad Sci U S A* **104**: 14300–14305.

**Zhao K-H**, Wu D, Zhou M, Zhang L, Böhm S, Bubenzer C, *et al.* (2005b). Amino acid residues associated with enzymatic activities of the isomerizing phycoviolobin-lyase PecE/F. *Biochemistry (Mosc)* **44**: 8126–8137.

**Zhao K-H**, Zhang J, Tu J-M, Böhm S, Plöscher M, Eichacker L, *et al.* (2007b). Lyase activities of CpcS- and CpcT-like proteins from *Nostoc* PCC7120 and sequential reconstitution of binding sites of phycoerythrocyanin and phycocyanin

- beta-subunits. *J Biol Chem* **282**: 34093–34103.
- Zhou J**, Gasparich GE, Stirewalt VL, de Lorimier R, Bryant DA. (1992). The *cpcE* and *cpcF* genes of *Synechococcus* sp. PCC 7002. Construction and phenotypic characterization of interposon mutants. *J Biol Chem* **267**: 16138–16145.
- Zhou W**, Ding W-L, Zeng X-L, Dong L-L, Zhao B, Zhou M, *et al.* (2014). Structure and mechanism of the phycobiliprotein lyase CpcT. *J Biol Chem* **289**: 26677–26689.
- Zwirgmaier K**, Heywood JL, Chamberlain K, Woodward EMS, Zubkov MV, Scanlan DJ. (2007). Basin-scale distribution patterns of picocyanobacterial lineages in the Atlantic Ocean. *Environ Microbiol* **9**: 1278–1290.
- Zwirgmaier K**, Jardillier L, Ostrowski M, Mazard S, Garczarek L, Vaultot D, *et al.* (2008). Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ Microbiol* **10**: 147–161.
- (2011). Method of the Year 2010. *Nat Methods* **8**: 1–1.



## **Pigment diversity in marine *Synechococcus* sp.: molecular basis, evolution and ecological role**

Marine *Synechococcus* are the second most abundant photosynthetic organisms on the planet. These picocyanobacteria present a wide diversity of pigmentation, which comes from differences in the composition of their light-harvesting antenna, called phycobilisome, allowing them to efficiently exploit a wide range of spectral niches. Yet, the evolution, ecology and molecular bases of the different *Synechococcus* pigment types are not well understood. By comparing the genomic region gathering most genes involved in the synthesis of phycobilisome rods from 54 sequenced isolates spanning all cultured pigment types and from natural *Synechococcus* populations, I proposed a scenario for the evolution of the different pigment types, and showed that the pigment diversity of marine *Synechococcus* can be traced back to before the diversification of this genus. Then, I developed a bioinformatic pipeline for reliably quantifying all known *Synechococcus* pigment types from metagenomic data. Applying it to the *Tara* Oceans dataset allowed me to describe for the first time their distribution in the global ocean, and revealed that type IV chromatic acclimation, a process by which cells can match their absorption properties to the ambient light colour, is widespread and constitutes the dominant pigmentation in *Synechococcus* populations. It also showed that natural chromatic acclimation mutants prevail in wide oligotrophic areas of the southern Pacific Ocean. Finally, I genetically characterized two members of an enzyme family binding chromophores to phycoerythrin-II, a major component of phycobilisomes. This provided new insights into the molecular bases of the chromatic acclimation process, and revealed the importance of allelic variation for the diversity of pigment types.

Keywords: cyanobacteria, marine microbiology, phycobiliprotein, chromatic acclimation, metagenomics, functional genomics, microbial ecology, evolution

---

## **Diversité pigmentaire des *Synechococcus* marins : bases moléculaires, évolution et importance écologique**

Les *Synechococcus* marins sont les seconds organismes photosynthétiques les plus abondants sur la planète. Ces picocyanobactéries marines présentent une grande diversité pigmentaire du fait de différences dans la composition de leur antenne collectrice de lumière appelée phycobilisome, ce qui leur permet d'utiliser efficacement une grande partie du spectre lumineux. Cependant, l'évolution, l'écologie et les bases moléculaires de cette diversité restent mal comprises. La comparaison de la région génomique regroupant la plupart des gènes impliqués dans la synthèse des bras de phycobilisomes provenant de 54 souches séquencées ainsi que de populations naturelles m'a permis de proposer un scénario évolutif pour l'apparition des différents types pigmentaires, et de montrer que cette diversité pigmentaire précède la diversification des *Synechococcus* marins. Par la suite, j'ai développé une procédure bioinformatique permettant de quantifier de façon fiable l'abondance relative de tous les types pigmentaires connus à partir de données de métagénomique. L'utilisation de cette méthode sur l'ensemble des métagénomomes de *Tara* Oceans m'a permis de décrire pour la première fois leur répartition à l'échelle mondiale, et a révélé que l'acclimatation chromatique de type IV, qui permet aux cellules de modifier leur spectre d'absorption en fonction de la couleur de la lumière, est très répandue et domine les populations naturelles de *Synechococcus*. Cela a aussi montré que des mutants naturels de l'acclimatation chromatique prédominent dans les larges étendues oligotrophes de l'océan Pacifique sud. Enfin, j'ai caractérisé génétiquement deux membres d'une famille d'enzymes liant les chromophores à la phycoérythrine II, un constituant majeur des phycobilisomes. Ces résultats apportent de nouvelles perspectives quant aux bases moléculaires de l'acclimatation chromatique, et ont révélé l'importance des variations alléliques dans la diversité des types pigmentaires.

Mots-clés : cyanobactéries, microbiologie marine, phycobiliprotéine, acclimatation chromatique, métagénomique, génomique fonctionnelle, écologie microbienne, évolution